

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
CENTRO DE POLÍTICAS PÚBLICAS E AVALIAÇÃO DA EDUCAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO PROFISSIONAL EM GESTÃO E
AVALIAÇÃO DA EDUCAÇÃO PÚBLICA

LAENE ASCENSO LUSTOSA

UM ESTUDO SOBRE O PLANO AMOSTRAL DO SAEB

JUIZ DE FORA

2020

LAENE ASCENSO LUSTOSA

UM ESTUDO SOBRE O PLANO AMOSTRAL DO SAEB

Dissertação apresentada ao Programa de Pós-Graduação Profissional em Gestão e Avaliação da Educação Pública, da Universidade Federal de Juiz de Fora, como requisito parcial à obtenção do título de Mestre em Gestão e Avaliação da Educação Pública. Área de concentração: Gestão e Avaliação da Educação Pública.

Orientador: Prof. Dr. Marcel de Toledo Vieira

JUIZ DE FORA

2020

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Lustosa, Laene Ascenso.

Um estudo sobre o plano amostral do Saeb / Laene Ascenso
Lustosa. -- 2020.
164 p.

Orientador: Marcel de Toledo Vieira

Dissertação (mestrado profissional) - Universidade Federal de Juiz de Fora, Faculdade de Educação/CAEd. Programa de Pós-Graduação em Gestão e Avaliação da Educação Pública, 2020.

1. Saeb. 2. Amostragem. 3. Simulação. 4. Avaliação externa. 5. Inep. I. Vieira, Marcel de Toledo, orient. II. Título.

LAENE ASCENSO LUSTOSA

UM ESTUDO SOBRE O PLANO AMOSTRAL DO SAEB

Dissertação apresentada ao Programa de Pós-Graduação Profissional em Gestão e Avaliação da Educação Pública, da Universidade Federal de Juiz de Fora, como requisito parcial à obtenção do título de Mestre em Gestão e Avaliação da Educação Pública. Área de concentração: Gestão e Avaliação da Educação Pública.

Aprovada em 03 de novembro de 2020

BANCA EXAMINADORA



Prof. Dr. Marcel de Toledo Vieira
Universidade Federal de Juiz de Fora



Prof. Dr. Clécio da Silva Ferreira
Universidade Federal de Juiz de Fora



Prof. Dr. Pedro Luis do Nascimento Silva
Escola Nacional de Ciências Estatísticas

AGRADECIMENTOS

Encerro esse ciclo com o coração cheio de gratidão. Reconheço que não realizei essa caminhada sozinha e que o êxito somente foi possível por causa do apoio recebido das pessoas que estão comigo.

Agradeço a Deus por ter me capacitado e fortalecido para alcançar essa vitória, por ter me sustentado e me guiado desde a decisão de fazer um mestrado até a defesa da dissertação.

Meus pais, Felismon e Zileide, minhas irmãs, Zilmone, Liane e Luana, meu cunhado, Geovany, e meus sobrinhos, Micaela e Gabriel, me estimularam durante todo o curso. Obrigada por serem meu alicerce a cada dia e me ajudarem a renovar as minhas forças e minha disposição. Vocês são o que tenho de mais valioso!

A amizade de Luana me fortaleceu durante esses dois anos de curso. Obrigada por seu abraço e suas palavras de conforto e sabedoria nos momentos mais difíceis.

Os preciosos amigos também foram indispensáveis! Algumas amizades nasceram e outras se fortaleceram. Às companheiras de caminhada que se tornaram amigas – Ana Flávia, Anarcisa, Cléia, Flávia, Rita e Roshni – eu agradeço pela parceria nos vários trabalhos em grupo, nas hospedagens por vários dias longe de casa, nos almoços, pizzas, cafés, sorvetes e brigadeiros. A amizade de vocês foi um doce presente que o mestrado me trouxe.

O Jaloto me incentivou desde a decisão de iniciar o curso e me honrou com a leitura atenta de versões preliminares dessa dissertação. A Patrícia me assistiu revisando o texto. Vários amigos queridos contribuíram com valiosas dicas sobre o R. A ajuda de cada um de vocês foi muito importante para a conclusão dessa pesquisa.

Agradeço ao professor e orientador Marcel, por me acompanhar e auxiliar no estudo, me ajudando a realizar um trabalho melhor. Aos professores da Banca Examinadora, Pedro e Clécio, agradeço a disponibilidade e as sugestões trazidas para enriquecer a dissertação. Agradeço ao Leonardo que me acompanhou desde o início sendo a “ASA” que me permitiu voar nesse processo.

O Inep e a UFJF, por meio de seus representantes, criaram a oportunidade desse mestrado. Portanto, agradeço a cada profissional que viabilizou a realização do curso em cada detalhe do processo.

Agradeço aos familiares e amigos que, de uma forma ou de outra, sempre estiveram presentes, tornando os dias mais felizes e as cargas mais leves, facilitando a minha caminhada e me conduzindo ao meu objetivo. A todos vocês, muito obrigada!

RESUMO

A presente dissertação foi desenvolvida no âmbito do Mestrado Profissional em Gestão e Avaliação da Educação Pública (PPGP) do Centro de Políticas Públicas e Avaliação da Educação da Universidade Federal de Juiz de Fora (CAEd/UFJF). O caso de gestão estudado discutiu alguns aspectos do processo de coleta de dados do Saeb, mais especificamente no que diz respeito ao plano amostral e à estimação dos resultados, a fim de responder à seguinte pergunta: diante das diversas mudanças nos elementos que orientam o processo de coleta de dados do Sistema Nacional de Avaliação da Educação Básica (Saeb) desde a sua criação, o modelo adotado ainda se mostra adequado? Portanto, foi estabelecido como objetivo desta dissertação avaliar o plano amostral do Saeb e como se dá a estimação dos seus resultados, a fim de identificar oportunidades de melhorias nas metodologias adotadas. Para alcançar esse objetivo, a dissertação apresentou o histórico de reformas educacionais ocorridas internacionalmente e como esse movimento influenciou o surgimento da avaliação em larga escala no mundo, e sua chegada ao Brasil dando origem ao Saeb. Além disso, foi estudado o histórico de evolução do Saeb e seu plano amostral para entender o cálculo e a expansão dos resultados com base no plano amostral atual, a fim de identificar limitações desse processo. A partir daí, foi possível propor estratégias para minimizar os efeitos das limitações identificadas tornando o Saeb mais adequado à avaliação educacional em larga escala no Brasil. Como hipótese, assumiu-se que, depois de passar por diversas modificações, o plano amostral do Saeb precisaria ser estudado e avaliado para entender elementos que poderiam ser atualizados nesse modelo. Para tanto, a metodologia adotada é a revisão bibliográfica e documental, seguida de análise de dados coletados pelo Saeb 2017 com uso de simulação para medir eficiência do plano amostral atual por meio de comparação de resultados estimados a partir de Amostragem com os resultados obtidos na aplicação censitária. A comparação entre diferentes metodologias de estimação de variância para o cálculo do tamanho da amostra mostrou que a metodologia adotada atualmente é eficiente e adequada, mas que há espaço para a promoção de melhorias no plano amostral. Também se assumiu que o Saeb, como avaliação em larga escala, apresenta características e limitações intrínsecas a esse modelo de avaliação. O conhecimento desse contexto conduziu à identificação de oportunidades de melhoria no modelo da avaliação e à elaboração de um plano de ação pautado na implementação de mudanças que possibilitem uma evolução do Saeb. Assim, foram propostas ações voltadas para o aperfeiçoamento do plano amostral, para a ampliação do ciclo de coleta de dados visando ao aprofundamento das

informações obtidas e para o fornecimento de mais insumos de contextualização dos resultados da avaliação.

Palavras-Chave: Saeb. Amostragem. Simulação. Avaliação externa. Inep.

ABSTRACT

The case studied discuss aspects of Saeb's data collection process, specifically about sampling and estimation, to answer the question: In view of the various changes in the elements that guide the data collection process of the Saeb since its creation, is the adopted model still adequate? This thesis aims to analyze the Saeb's sampling design and how to estimate its results, looking for opportunities to improve the methodologies. To achieve this goal, this thesis presents the history of educational reforms around the world and how it influenced the emergence of large-scale assessment in the world, in Brazil and the creation of Saeb. The thesis also brings the evolution history of Saeb and its sampling design to understand the calculation of results using the current sampling design, trying to find limitations in the process. After that, it was possible to suggest strategies to reduce the effects of these limitations, making Saeb more adequate to large-scale assessment in Brazil. The theoretical reference tells about the origin of large-scale educational assessment, how this movement influenced the Saeb's creation and the evolution of its sampling design through the years. As hypothesis, it was assumed that after many changes the Saeb's sampling design must be studied and analyzed to understand how it can be updated. The employed methodology was bibliographic and document revision, followed by analysis collected data from Saeb 2017, using simulation to evaluate the efficiency of current sampling design, comparing estimated results from Sampling and results from Census. The comparison between different methodologies shows that the current methodology is efficient and adequate, but there are opportunities to improve it. It was also assumed that as a large-scale educational assessment, Saeb has characteristics and limitations natural from this assessment model. The knowledge of this context leads the opportunities of improvement for the assessment model and create an action plan that can make changes and that have a capacity to evolve Saeb. Therefore, the action plan presents some suggestions to improve the sampling design, to enlarge the data collection cycle, with the goal to obtained deepen information through the data collected and given context inputs for the assessment results.

Keywords: Saeb. Sampling. Simulation. External evaluation. Inep.

LISTA DE ILUSTRAÇÕES

Quadro 1 - Evolução de características centrais do Saeb	32
Gráfico 1 - Número de alunos válidos por dependência administrativa.....	66
Gráfico 2 - Dispersão dos valores de θiLP para o Brasil.....	71
Gráfico 3 - Distribuição dos valores de θiLP para o Brasil.....	72
Gráfico 4 - Distribuição dos valores de θiLP por UF	72
Gráfico 5 - Dispersão dos valores de θiMT para o Brasil	73
Gráfico 6 - Distribuição dos valores de θiMT para o Brasil.....	74
Gráfico 7 - Distribuição dos valores de θiMT por UF.....	75
Figura 1 - Ciclo de políticas e programas públicos.....	93
Quadro 2 - Síntese das limitações identificadas no Saeb	94
Quadro 3 - Resumo das propostas para aprimoramento do Saeb	98
Quadro 4 - Ciclo avaliativo proposto para o Saeb.....	104
Gráfico 8 - Dispersão dos valores de θiLP para o Brasil (Simulação ρc).....	123
Gráfico 9 - Distribuição dos valores de θiLP para o Brasil (Simulação ρc).....	123
Gráfico 10 - Distribuição dos valores de θiLP por UF (Simulação ρc).....	124
Gráfico 11 - Dispersão dos valores de θiMT para o Brasil (Simulação ρc)	124
Gráfico 12 - Distribuição dos valores de θiMT para o Brasil (Simulação ρc).....	125
Gráfico 13 - Distribuição dos valores de θiMT por UF (Simulação ρc).....	125
Gráfico 14 - Dispersão dos valores de θiLP para o Brasil (Simulação ρint_2).....	130
Gráfico 15 - Distribuição dos valores de θiLP para o Brasil (Simulação ρint_2)	130
Gráfico 16 - Distribuição dos valores de θiLP por UF (Simulação ρint_2)	131
Gráfico 17 - Dispersão dos valores de θiMT para o Brasil (Simulação ρint_2).....	131
Gráfico 18 - Distribuição dos valores de θiMT para o Brasil (Simulação ρint_2).....	132
Gráfico 19 - Distribuição dos valores de θiMT por UF (Simulação ρint_2).....	132

LISTA DE TABELAS

Tabela 1 -	Tamanho da amostra em número de alunos e turmas.....	63
Tabela 2 -	Quantitativos da amostra esperada.....	64
Tabela 3 -	Comparativo de quantidades entre censo e amostra.....	67
Tabela 4 -	Descrição de notações.....	67
Tabela 5 -	Medidas de desempenho para Língua Portuguesa.....	69
Tabela 6 -	Medidas de desempenho para Matemática.....	70
Tabela 7 -	Cobertura dos intervalos de confiança e Margem de Erro.....	76
Tabela 8 -	Comparativo dos resultados da Simulação <i>ρ_{int}</i> e da Simulação <i>ρ_c</i>	79
Tabela 9 -	Comparativo dos valores de correlação intraclasse.....	80
Tabela 10 -	Comparativo do tamanho da amostra de turmas.....	81
Tabela 11 -	Comparativo do EQM e do RREQM para Língua Portuguesa.....	82
Tabela 12 -	Comparativo do EQM e do RREQM para Matemática.....	83
Tabela 13 -	Comparativo da cobertura e da margem de erro em Língua Portuguesa.....	84
Tabela 14 -	Comparativo da cobertura e da margem de erro em Matemática.....	85
Tabela 15 -	Tamanho da amostra em número de alunos e turmas (Simulação <i>ρ_c</i>).....	120
Tabela 16 -	Comparativo de quantidades entre censo e amostra (Simulação <i>ρ_c</i>).....	121
Tabela 17 -	Medidas de desempenho para Língua Portuguesa (Simulação <i>ρ_c</i>).....	121
Tabela 18 -	Medidas de desempenho para Matemática (Simulação <i>ρ_c</i>).....	122
Tabela 19 -	Cobertura dos intervalos de confiança e Margem de Erro (Simulação <i>ρ_c</i>).....	126
Tabela 20 -	Tamanho da amostra em número de alunos e turmas (Simulação <i>ρ_{int_2}</i>).....	127
Tabela 21 -	Comparativo de quantidades entre censo e amostra (Simulação <i>ρ_{int_2}</i>).....	128
Tabela 22 -	Medidas de desempenho para Língua Portuguesa (Simulação <i>ρ_{int_2}</i>).....	128
Tabela 23 -	Medidas de desempenho para Matemática (Simulação <i>ρ_{int_2}</i>).....	129
Tabela 24 -	Cobertura dos intervalos de confiança e Margem de Erro (Simulação <i>ρ_{int_2}</i>)	

LISTA DE SIGLAS

A2E	Amostragem por Conglomerado em Dois Estágios
AALE	Avaliação de Aprendizagem em Larga Escala
AAS	Amostragem Aleatória Simples
AC	Amostragem por Conglomerados
ADEMP	<i>Aims, Data-generating mechanisms, Estimands, Methods, Performance measures</i>
AE	Amostragem Estratificada
ANA	Avaliação Nacional de Alfabetização
Aneb	Avaliação Nacional da Educação Básica
Anresc	Avaliação Nacional do Rendimento Escolar
AS	Amostragem Sistemática
CEP	Código de Endereçamento Postal
Daeb	Diretoria de Avaliação da Educação Básica
EPA	Efeito do Plano Amostral
EQM	Erro Quadrático Médio
Ideb	Índice de Desenvolvimento da Educação Básica
IEA	Internacional Association for the Evaluation of Educacional Achievement
Inep	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
LDB	Lei de diretrizes e bases da educação nacional
MEC	Ministério da Educação
OCDE	Organização para a Cooperação e Desenvolvimento Econômico
PNE	Plano Nacional de Educação
PPT	Probabilidade Proporcional ao Tamanho
RREQM	Raiz quadrada do erro quadrático médio em termos relativos ao valor do parâmetro
SAD	Sistema de Avaliação de Desempenho
Saeb	Sistema de Avaliação da Educação Básica
Saep	Sistema de Avaliação do Ensino Público de 1º Grau
TCT	Teoria Clássica dos Testes
TRI	Teoria de Resposta ao Item
UF	Unidade da Federação
Var	Variância

SUMÁRIO

1	INTRODUÇÃO.....	12
2	SAEB: A PRINCIPAL AVALIAÇÃO EDUCACIONAL EM LARGA ESCALA DO BRASIL	16
2.1	SURGIMENTO DA AVALIAÇÃO EDUCACIONAL EM LARGA ESCALA	16
2.2	O PAPEL DA UNIÃO NA AVALIAÇÃO DA EDUCAÇÃO BÁSICA BRASILEIRA E A SOBREPOSIÇÃO DOS SISTEMAS DE AVALIAÇÃO	23
2.2.1	O Inep – representante da União na avaliação da educação básica	25
2.2.2	A sobreposição dos sistemas de avaliação no regime de colaboração	27
2.3	O SAEB	30
2.4	HISTÓRICO DO PLANO AMOSTRAL DO SAEB	36
3	UMA ANÁLISE SOBRE O PLANO AMOSTRAL DO SAEB.....	45
3.1	TEORIA DE AMOSTRAGEM	45
3.1.1	O que é Amostragem	47
3.1.2	Tipos de Amostragem	52
3.2	ANÁLISE DO PLANO AMOSTRAL DO SAEB	54
3.3	SIMULAÇÕES DA REALIZAÇÃO AMOSTRAL A PARTIR DA APLICAÇÃO CENSITÁRIA	58
3.3.1	Seleção da amostra.....	59
3.3.2	Estimação dos resultados	64
3.3.3	Análise dos resultados	66
3.3.4	Metodologias alternativas	77
3.3.5	Limitações do plano amostral	87
3.4	OUTRAS LIMITAÇÕES DO SAEB	88
4	PROPOSTAS PARA SUPERAR LIMITAÇÕES DO SAEB.....	95
4.1	PROPOSTA PARA O PLANO AMOSTRAL	98
4.2	PROPOSTA DE APLICAÇÃO EM CICLO DE QUATRO ANOS INTERCALANDO APLICAÇÃO CENSITÁRIA E APLICAÇÃO AMOSTRAL	100
4.3	PROPOSTA DE CÁLCULO DE PESOS PARA OS DADOS CONTEXTUAIS	106
5	CONSIDERAÇÕES FINAIS.....	108
	REFERÊNCIAS	112
	APÊNDICE A – Resultados da simulação realizada utilizando ρc	120
	APÊNDICE B – Resultados da simulação realizada utilizando ρ_{int_2}.....	127

APÊNDICE C – Scripts utilizados para processamento das simulações 134

1 INTRODUÇÃO

Desde o final dos anos 1980, surgiu no Brasil um movimento de apoio à adoção da avaliação em larga escala no contexto educacional. Esse movimento foi incentivado por ações internacionais que despertaram na comunidade acadêmica e nos governantes ligados à educação um novo entendimento. A partir de então, entendeu-se que para a promoção de um processo educacional de qualidade seria necessário conhecer um diagnóstico de como estava a qualidade da educação ofertada (COELHO, 2008). Essa tendência culminou na criação do Sistema Nacional de Avaliação da Educação Básica (Saeb), que fomentou a implementação de sistemas de avaliações locais conduzidos nos contextos estaduais e municipais (COELHO, 2008; MACHADO; ALAVARSE; ARCAS, 2015).

Desde que foi instituído em 1990, o Saeb passou por diversos ajustes a fim de ser aprimorado para o cumprimento do objetivo de realizar um diagnóstico da educação básica brasileira. Os ajustes realizados impactaram a forma de coleta de dados, incluindo adaptações do plano amostral inicial. A situação-problema a ser estudada nesta dissertação é que, apesar das alterações do Saeb, não foram encontrados registros a respeito das análises realizadas quanto à adequação do desenho amostral aos objetivos da avaliação. Diante disso, faz-se necessário estudar a metodologia de Amostragem para verificar a qualidade dos resultados do Saeb.

Medidas que resumem um conjunto de dados, como os resultados do Saeb, carregam consigo alguma perda de informação ou erro (BUSSAB; MORETIN, 2004). Avaliações censitárias podem ser acometidas por erros não amostrais advindos de possíveis fragilidades no processo, imprecisão ou ausência de resposta, erros de edição, codificação e tabulação dos dados. Assim, pesquisas amostrais exigem mais cuidado, pois os possíveis erros não amostrais são acrescidos aos erros amostrais, que são resultantes do fato de apenas um subconjunto da população estar sendo observado (COCHRAN, 1977; GROVES; LYBERG, 2010). Além disso, a forma de coleta de dados impacta diretamente no custo do processo, podendo acarretar gastos desnecessários se esse desenho não estiver alinhado com os objetivos da avaliação (COCHRAN, 1977). Isto posto, o debate a respeito do modelo adequado de coleta de dados, seja censitário ou amostral, é indispensável para a construção de uma avaliação em larga escala coerente e com resultados confiáveis.

A legislação brasileira estabelece que a União, em regime de colaboração com Estados, Municípios e o Distrito Federal, deve monitorar a qualidade da educação ofertada aos alunos brasileiros. Para tanto, o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio

Teixeira (Inep) foi designado como órgão representante da União nessa tarefa e o Saeb foi criado para esse propósito. Estudos apontam que além do Saeb, pelo menos 21 Estados e mais da metade dos Municípios têm, ou pretendem implementar, sistemas próprios de avaliação. A maioria desses sistemas são muito semelhantes ao Saeb tanto em forma quanto em aspectos e público avaliados, criando um cenário de sobreposição entre os diversos sistemas de avaliação da educação básica existentes.

Ao refletir sobre os sistemas de avaliação da educação básica no Brasil, é importante trazer ao debate questões relacionadas à validação estatística da avaliação, ao melhor desenvolvimento de políticas públicas e a uma melhor alocação dos recursos públicos. Tais questões estão associadas à sobreposição dos sistemas de avaliação, ao papel da União na avaliação da educação básica brasileira, à adequação da metodologia de coleta de dados e à comparabilidade dos resultados ao longo do tempo.

Um dificultador na condução do debate, no contexto do Saeb, é a insuficiência dos registros documentais sobre os detalhes da construção do modelo de coleta de dados para esta avaliação ao longo dos anos. A documentação disponibilizada pelo Inep apresenta o plano amostral em termos gerais. São descritos aspectos relevantes do processo, como a divisão em estratos, os critérios para dimensionamento da amostra e os procedimentos de seleção e cálculos amostrais. Contudo, não são apresentadas todas as fórmulas e nem se aprofundam aspectos que justifiquem a escolha das metodologias adotadas (INEP, 2014, 2017, 2018a). Entretanto, ao longo desta dissertação são apresentados fundamentação e conceitos para balizar esse debate.

Um aspecto relevante é a utilização dos resultados do Saeb para o cálculo do Índice de Desenvolvimento da Educação Básica (Ideb), unificando-os com os resultados de fluxo escolar, obtidos a partir dos dados sobre aprovação dos alunos informados no Censo Escolar. A reunião desses dois conceitos resulta em um importante condutor de política pública para promoção da melhoria da educação (INEP, 2020). Portanto, resultados imprecisos do Saeb, bem como desse índice, poderiam acarretar a elaboração de políticas públicas inócuas, caso fossem fundamentadas em resultados que não retratassem de maneira confiável a situação da educação básica no Brasil.

Eu atuo como Pesquisadora-Tecnologista em Informações e Avaliações Educacionais, na área de avaliação da educação básica do Inep, em equipe dedicada ao planejamento amostral do Saeb e aos cálculos dos seus resultados. Os resultados de avaliações da educação básica em larga escala de abrangência nacional oferecem subsídios para a elaboração, o monitoramento e o aprimoramento de políticas com base em evidências, permitindo que os diversos níveis governamentais avaliem a qualidade da educação praticada no país. Portanto, ao trabalhar com

o cálculo destes resultados é indispensável uma fundamentação científica para os procedimentos adotados em cada etapa da avaliação.

A avaliação em larga escala investiga os argumentos e os instrumentos abordados na sua concepção, podendo gerar resultados contestáveis, caso não seja bem construída e bem delimitada (ARELLANO et al., 2012). Um dos aspectos da construção da avaliação que pode embutir erro nos resultados é a forma de coleta dos dados.

Conforme é detalhado nesta dissertação, o Saeb passou por mudanças intensas desde sua criação. A lista das mudanças implementadas passa por alterações no público avaliado, nas áreas de conhecimento testadas, bem como na abrangência da aplicação. Além disso, foram ampliados os objetivos da avaliação.

Assim, a questão central desta investigação é: diante das diversas mudanças nos elementos que orientam o processo de coleta de dados do Saeb desde a sua criação, o modelo adotado ainda se mostra adequado?

Nesse sentido, o objetivo desta dissertação é avaliar o plano amostral do Saeb, bem como a estimação dos seus resultados, a fim de identificar oportunidades de melhorias nas metodologias adotadas. Para tanto, este texto discorre sobre o processo de avaliação da educação básica brasileira e a sobreposição dos sistemas de avaliação existentes. Além disso, avalia a adequação do plano amostral atual e dos métodos de estimação adotados pela comparação dos resultados obtidos por meio de diferentes metodologias. Por fim, propõem-se ajustes no plano amostral e no modelo de coleta de dados do Saeb como alternativas para uma melhor adequação aos objetivos da avaliação e à organização do sistema educacional brasileiro.

A pesquisa feita para cumprir com os objetivos propostos foi desenvolvida por meio de análise documental da legislação que envolve a educação básica brasileira e sua avaliação, arquivos dos microdados das diversas edições do Saeb, bem como relatórios técnicos fornecidos pelo Inep a respeito dos planos amostrais da avaliação. A leitura dos diversos documentos permitiu a identificação das particularidades de cada edição do Saeb e como elas refletiram no plano amostral, possibilitando a reconstrução da memória de como o desenho amostral do Saeb evoluiu ao longo dos anos. Foi realizada pesquisa bibliográfica sobre a avaliação educacional em larga escala, sobre o Saeb e sobre Teoria da Amostragem, além de simulações para comparação de diferentes metodologias de dimensionamento amostral aplicáveis ao Saeb.

Além da introdução e das considerações finais, o texto se organiza em três capítulos de desenvolvimento. O primeiro discorre sobre o processo de avaliação da educação básica brasileira e a sobreposição dos sistemas de avaliação existentes. Apresenta o Saeb como principal avaliação em larga escala do Brasil, surgindo como consequência de uma série de

reformas educacionais, seguindo uma tendência internacional e se tornando modelo para a criação de sistemas de avaliação estaduais e municipais muito semelhantes a ele. Também relata a transformação do Saeb ao longo dos anos e a evolução do plano amostral da avaliação.

O segundo capítulo avalia aspectos do modelo atual de coleta de dados do Saeb a partir da realidade dessa avaliação. Com base nos Microdados do Saeb 2017, é realizado um estudo sobre a metodologia adotada para a seleção da amostra e para o cálculo dos resultados, por meio da realização de simulações que permitiram a comparação de três diferentes metodologias. Para cada método simulado, foram sorteadas 2.000 amostras e, a partir dos resultados obtidos, foi possível verificar que a metodologia atual, apesar de eficiente e confiável, oferece oportunidades de melhoria que podem ser facilmente implementadas, como o aprimoramento do dimensionamento da amostra e melhor delimitação do universo de referência amostral. Para encerrar o capítulo, foi realizado o levantamento de algumas limitações do Saeb, entre as quais se pode destacar a metodologia de estimação das variâncias para dimensionamento da amostra, a periodicidade de coleta dos dados e a pouca contextualização dos resultados.

O último capítulo descreve um plano de ação, incluindo sugestões de melhorias relacionadas ao plano amostral de Saeb e às limitações identificadas no capítulo anterior. O plano de ação pode ser resumido em três propostas: a) aprimoramento do plano amostral, com o objetivo de contornar limitações técnicas relacionadas ao dimensionamento da amostra; b) ampliação do ciclo da avaliação passando de dois para quatro anos e abrangendo de forma direta as seis dimensões da qualidade educacional avaliadas pelo Saeb; e c) cálculo de pesos específicos para os questionários facilitando a elaboração de um sistema de indicadores que permita melhor contextualização dos resultados e melhor avaliação das diversas dimensões da qualidade educacional.

2 SAEB: A PRINCIPAL AVALIAÇÃO EDUCACIONAL EM LARGA ESCALA DO BRASIL

Neste capítulo, são tratados temas que irão alicerçar o debate dos capítulos seguintes. O objetivo deste capítulo é descrever o processo de avaliação da educação básica brasileira, contemplando o papel da União nesse processo e a sobreposição dos sistemas de avaliação existentes, traçando o histórico do surgimento do Saeb e da elaboração do seu plano amostral.

Na primeira seção, será apresentado o histórico do surgimento da avaliação em larga escala no Brasil, abordando as reformas educacionais e o surgimento das avaliações em larga escala em outros países.

Com base na legislação vigente, a segunda seção apresenta o papel da União no regime de colaboração para promoção da qualidade da educação no Brasil. A descrição tratará também do papel dos Estados e Municípios nessa promoção e a sobreposição dos sistemas de avaliação realizados pela União, Estados e Municípios.

O processo de surgimento do Inep e suas atribuições no que diz respeito à realização de avaliações educacionais em larga escala serão apresentados, criando um panorama geral sobre o órgão responsável pela concepção e realização do Saeb. Além disso, será descrita a organização estrutural do Inep, que define a sua Diretoria de Avaliação da Educação Básica (Daeb) como responsável pela avaliação da educação básica.

Em seguida, a seção três deste capítulo conecta as duas primeiras, descrevendo o histórico de surgimento do Saeb; mostra como a influência externa culminou na criação do Saeb e a responsabilidade do Inep, e mais especificamente da Daeb, para sua realização. Além disso, apresenta a evolução do Saeb ao longo dos seus quase 30 anos de existência.

Encerrando o capítulo, será apresentado um panorama sobre a metodologia de coleta de dados adotada para o Saeb e o histórico de evolução do seu plano amostral. Será apresentada a necessidade de avaliação de sua adequação ao modelo de Saeb que se desenhou nas seções anteriores. Para essa construção, foi realizada ampla pesquisa documental, por meio da consulta a diversos documentos técnicos fornecidos pelo Inep e por profissionais que participaram da elaboração dos planos amostrais do Saeb, tais como consultores externos.

2.1 SURGIMENTO DA AVALIAÇÃO EDUCACIONAL EM LARGA ESCALA

Desde a época da 2ª Guerra mundial, a educação na Grã-Bretanha passou por reformas profundas. Primeiramente, instaurou-se um sistema gratuito e aberto a todos, subdividido em

três tipos de escola: *grammar school*, *secondary modern* e escolas técnicas. Os alunos eram encaminhados para uma dessas escolas de acordo com as aptidões identificadas por meio de uma avaliação aplicada às crianças aos 11 anos de idade. Nos anos iniciais de estudo, anteriores ao exame dos 11 anos, os alunos eram distribuídos de acordo com seu nível de desempenho em turmas com currículos diferenciados, em uma prática seletiva chamada de *streaming*. Esse modelo tinha como principal desvantagem a limitação da possibilidade de desenvolvimento acadêmico das crianças (BROOKE, 2012).

A partir de 1964, quando o partido trabalhista assumiu o governo britânico, instalou-se uma nova reforma. Os trabalhistas unificaram a maioria das *grammar school* e a *secondary modern*, propondo um modelo de escola multipropósito e academicamente diversificada, denominada compreensiva. Esse novo modelo eliminou as escolas secundárias e a enturmação dos alunos de acordo com seu nível de desempenho. A maior desvantagem desse modelo era o fato de que, na busca pela igualdade de oportunidades, as metas relacionadas à qualidade do ensino haviam sido estabelecidas em nível correspondente ao daqueles que apresentavam desempenho inferior (BROOKE, 2012).

Quando a Primeira-Ministra Margaret Thatcher, do Partido Conservador, assumiu o governo em 1979, a Grã-Bretanha se viu novamente em meio a uma reforma educacional. Brooke (2012) destaca que, desta vez, a reforma atendia a reivindicações dos Documentos Negros sobre a Educação – publicações que registraram a inconformidade dos conservadores com as opiniões do governo trabalhista a respeito da educação. A reforma tinha como elementos principais devolver aos pais o direito de escolher sobre a educação dos filhos, promover um currículo básico nacional, frear as atividades dos progressistas, recuperar a seletividade, avaliar o desempenho dos alunos ao longo da trajetória escolar e oferecer um sistema de vales-educação (*vouchers*).

Esse movimento de reformas educacionais não se limitou à realidade da Grã-Bretanha e pode ser observado também em outros países. No período entre 1957 e 1968, no contexto da guerra fria e da corrida espacial, os Estados Unidos se viram ameaçados pelo lançamento do Sputnik pela União Soviética. Esta ameaça serviu de gatilho para uma reforma educacional a fim de promover o aumento do seu poderio científico, técnico e industrial, recuperando sua fibra como Nação. Para tanto, foram criados programas de bolsas de estudos e de fortalecimento do ensino de Matemática, Ciências e línguas estrangeiras, voltados para alunos do nível elementar e médio, assim como bolsas de estudos para alunos de pós-graduação. Essa reforma também abriu caminho para a promulgação de leis que redefiniram as relações entre o governo e a comunidade educacional dos Estados Unidos (NSF, 2012a).

Nesse contexto, foi estabelecido nos Estados Unidos o programa de aperfeiçoamento do conteúdo do ensino para contribuir com a modernização do ensino de Matemática, Ciências e Engenharia, nas escolas de nível elementar, médio e superior. A partir desse programa, extraem-se cinco lições sobre o ensino científico: a) necessidade de atenção equilibrada às diferentes áreas de conhecimento; b) necessidade de disponibilização de recursos de qualidade e capacitação aos professores; c) necessidade de colaboração entre os estudiosos das áreas de pesquisa e ensino; d) necessidade de investimentos substanciais para o desenvolvimento de programas educacionais; e e) necessidade de entendimento da continuidade deste processo (NSF, 2012b).

Em 1959, no Brasil, circulou o manifesto ao povo e ao governo, documento redigido por Fernando de Azevedo, publicado por revistas educacionais e órgãos da imprensa à época da tramitação da Lei nº 4.024 no Congresso Nacional. Esse documento sugeria que a Nação é uma “realidade moral” e que a educação não poderia desconhecer nenhum dos aspectos dessa realidade, mas deveria se colocar a serviço dos interesses coletivos do povo e da cultura nacional. Portanto, a educação deveria ser reestruturada para contribuir para o progresso científico e técnico, para o trabalho produtivo e o desenvolvimento econômico. Isso porque a era tecnológica marcaria o fim do processo de educação com foco na adaptação e submissão do indivíduo ao seu ambiente. Essa ruptura iniciou um processo de ensino focado na evolução do homem e do seu universo (MANIFESTO AO POVO E AO GOVERNO apud BROOKE, 2012).

Entre 1969 e 1972, o Brasil, juntamente com outros quatro países latino-americanos, a saber, Colômbia, El Salvador, Peru e Venezuela, promoveu uma reforma do ensino médio. Esta reforma teve a finalidade de direcionar a maior parte dos alunos para cursos técnico-profissionalizantes, como estratégia para o desenvolvimento social e econômico. No entanto, o Brasil, diferentemente dos outros países, se manteve com ensino médio majoritariamente acadêmico (BROOKE, 2012).

Apesar de não se ter a motivação recebida pelos Estados Unidos por conta do lançamento do Sputnik, o Brasil entendeu que o espírito científico deveria ser estimulado nos jovens alunos e com isso vivenciou uma sucessão de reformas. O objetivo almejado era solucionar os problemas educacionais e promover uma igualdade de oportunidades educacionais.

Nos anos 70 e 80 vários países da América Latina recorreram à descentralização administrativa e financeira dos sistemas educacionais com a intenção de reduzir os gastos do governo central com a educação. Esse modelo de reforma foi o predominante na região durante aquele período. Entretanto, o Brasil não precisou aderir a esse movimento, pois a maior parte

dos seus sistemas de ensino básico já estava sob responsabilidade de Estados e Municípios (CARNOY; CASTRO, 2012).

Ao mesmo tempo, discutia-se a nova Constituição Federal Brasileira, promulgada em 1988, trazendo em seu texto uma associação entre a avaliação e a qualidade da educação. O texto aprovado apresenta em seu art. 206 a “garantia de padrão de qualidade” entre os princípios fundamentais do ensino e, no art. 209, institui a possibilidade de avaliação de qualidade do ensino ministrado por instituições privadas pelo setor público. Mais adiante, no art. 214, o texto constitucional estabelece o Plano Nacional de Educação (PNE), o qual, além de articular o sistema nacional de educação, deve definir diretrizes, objetivos, metas e estratégias que, por meio de ações integradas dos poderes públicos das diferentes esferas federativas, conduzam à “melhoria da qualidade do ensino” (BRASIL, 1988).

Durante a década de 1990, espalharam-se por diferentes regiões do mundo múltiplas e simultâneas mudanças no cenário educacional. As reformas desse período evidenciaram a percepção dos governos de que a qualidade da educação ofertada é elemento fundamental para a promoção do desenvolvimento e da modernização. Na América Latina, percebem-se semelhanças notáveis entre as reformas vivenciadas por vários países naquele período, como relatou Brooke (2012, p. 326):

Na maioria dos países, são especificados os objetivos gerais de descentralizar a gestão, melhorar a qualidade, equidade e eficiência dos sistemas, dar maior autonomia e também de cobrar maior responsabilidade da escola, investir mais e melhor na formação do professor e conectar a escola com as demandas da sociedade.

Além da similaridade entre os objetivos dos modelos propostos, a implantação de sistemas de avaliação padronizada se destacou como elemento de uniformidade entre os modelos, sendo adotada por treze países da América Latina no período de 1990 a 1998 (BROOKE, 2012).

No Brasil, em 1996, a Lei nº 9.934/96, Lei de diretrizes e bases da educação nacional (LDB), foi aprovada como documento que regulamenta a educação brasileira. Como afirmaram Dambros e Mussio (2014, p. 2):

As reformas no âmbito educacional ganharam força a partir da crise do Estado de Bem-estar, quando, em todo o mundo, ocorreu um reordenamento das políticas sociais, calcadas no contexto de descentralização do Estado e nos critérios de eficiência e qualidade, complementadas com uma paulatina “transferência” das responsabilidades estatais para a comunidade.

Todo este movimento de reforma educacional ocorre com o objetivo de promover a qualificação do capital humano nacional que, apesar da dificuldade de mensuração, é a maior riqueza de uma nação e por esta razão precisa ser preservado e valorizado (SHULTZ, 2012).

A melhor forma de valorizar o capital humano de uma nação é investindo na formação das pessoas. Com isso, surgiu a Revolução do Capital Humano, que defendia que o investimento em pessoas por meio da educação e do treinamento traria resultados compensadores em termos de desenvolvimento. A expectativa era de que o acesso à educação traria maior igualdade de status e de renda (SOBEL, 2012).

Essa busca pela valorização do capital humano e por uma forma de garantir a qualidade da educação deu origem às avaliações educacionais em larga escala, que começaram a ganhar importância internacionalmente a partir dos anos 1950. Destaca-se que, até o fim dos anos 1980, o usual era a realização de avaliações amostrais com divulgação de resultados agregados. São chamados de resultados agregados aqueles calculados considerando agrupamentos de alunos nos diferentes níveis previstos na concepção da avaliação.

Coelho (2008) relata que no final dos anos 1960 surgiram internacionalmente fatores que influenciaram a implantação da avaliação da educação básica no Brasil, como a aprovação do *Primary and Secondary Education Act*, em 1965, nos Estados Unidos. A criação da *International Association for the Evaluation of Educational Achievement* (IEA), com a finalidade de realizar estudos internacionais de avaliação educacional, também influenciou esse movimento. Coelho (2008) ainda cita a contribuição dos indicadores internacionais de qualidade da educação desenvolvidos pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE) para impulsionar a avaliação educacional no cenário internacional.

Somados a essa influência, foram desenvolvidos estudos acerca de desigualdades sociais e educacionais, que apontavam para a ineficiência da escola em minimizar os efeitos das diferenças econômicas, sociais e culturais nas famílias dos estudantes. Tais estudos reforçaram a necessidade de medir a qualidade do ensino oferecido para que a escola pudesse se transformar e alterar a realidade verificada nas pesquisas (COELHO, 2008).

Freitas (2005) constatou em sua pesquisa que o uso da avaliação em larga escala como instrumento de regulação da educação básica se iniciou em decorrência de uma crise no Estado desenvolvimentista. Naquele momento, buscava-se a recomposição do poder político, simbólico e operacional de regulação pelo Estado Central e de restrições à sua atuação na área social. Essa crise apontava para a necessidade de reformas do Estado e modernização do País. Nesse contexto, a avaliação mostrou-se como estratégia útil para a gestão a fim de se alinhar às demandas da área social.

Nesse período, estava sendo estudada a efetividade da escola com a intenção de identificar as características comuns às escolas que se destacavam. Eram consideradas escolas efetivas aquelas que levavam seus estudantes a apresentarem desempenho diferenciado em relação aos outros alunos inseridos no mesmo nível socioeconômico desfavorecido e de minorias culturais. A proposta de escola efetiva começou a ser difundida no Brasil nos anos 1980 discutindo que o destaque dado aos aspectos tecnológicos da escola restringiu sua efetividade a um espaço instrucional. Outra consequência apontada foi a redução da cultura a padrões de excelência escolar, como currículo e oportunidade de aprendizagem, sem considerar os aspectos culturais externos à escola (COELHO, 2008).

Vianna (2014) explica que a avaliação em larga escala representa uma visão externa, portanto, o uso dos seus resultados deve ser feito com cautela. É importante identificar quais análises traduzem de maneira confiável a intensidade da influência da escola no desempenho educacional dos estudantes avaliados e até que ponto esses resultados podem ser encarados como medida de efetividade de cada escola individualmente.

Seguindo esse raciocínio, Silva (2013) defende que os resultados da avaliação em larga escala não podem ser usados como instrumento classificatório e controlador, desconsiderando as especificidades de cada escola. Silva (2013) ainda explica que o papel da avaliação em larga escala é fornecer subsídios para o debate do panorama educacional, verificando a aprendizagem dos alunos em relação à proposta curricular. Entretanto, é comum que se usem as avaliações internas como testes preparatórios para as avaliações externas, distorcendo a função de ambos os instrumentos.

Arellano et al. (2012, p.28) apresentam o Sistema de Avaliação de Desempenho (SAD) como um “projeto de reforma e de ação administrativa voltadas para legitimação do governo ora vigente”, descrevendo-o como uma ferramenta a ser desenvolvida e interpretada internamente pelas organizações, segundo suas diferentes finalidades. Além disso, um SAD precisa contar com indicadores, fundamentados nos resultados que se pretende obter, que permitam comparação do desempenho ao longo do tempo. Assim, para ser um instrumento efetivo, a avaliação demanda um ciclo de evolução constante.

Procurar por uma fórmula única que se apresente como receita para elaboração de um SAD que se converta em um mapa fácil e completo para o diagnóstico de uma organização é um erro bastante comum. Contudo, espera-se que um SAD contribua nos processos de tomada de decisão; no entendimento da organização, gerando consensos internos; no monitoramento do uso de recursos públicos e no alcance de resultados; e no diagnóstico organizacional para resolução de problemas (ARELLANO et al., 2012).

Os sistemas de avaliação educacional em larga escala, implementados nos sistemas de educação no Brasil podem ser classificados como SADs, pois têm como objetivo principal oferecer um diagnóstico que possa direcionar a tomada de decisões para o fomento da educação. Sendo assim, eles se constituem como ferramentas de prestação de contas das ações governamentais, cujos resultados devem ser sustentados por uma dinâmica de aprendizagem organizacional.

As políticas de *accountability* propõem que professores e diretores possam mudar o seu comportamento ou sua metodologia de trabalho, a partir dos resultados dos indicadores, a fim de proporcionar melhores condições de aprendizado aos alunos. Esse tema é polêmico em decorrência da confusão entre responsabilização e culpa. Entretanto Fernandes e Grémaud (2009, p.7) afirmam que “acreditar que os responsáveis pela educação podem alterar procedimentos e, assim, melhorar o ensino não significa dizer que sejam culpados por seus alunos não estarem aprendendo mais”.

Um desafio enfrentado nesse contexto é o entendimento das relações causais estabelecidas entre o cenário avaliado e os resultados obtidos na avaliação. Arellano et al. (2012) apontam que é extremamente complicado mostrar que a cadeia causal realmente produz resultados que sejam explicáveis. Diante disso, temos que é igualmente difícil que uma avaliação de desempenho identifique as causas e efeitos relacionados ao resultado obtido.

Arellano et al. (2012) afirmam ainda que os indicadores são aproximações numéricas do que se quer avaliar, e que os sistemas de avaliação de desempenho são um conjunto desse tipo de aproximações. Assim, os sistemas de avaliação tentam explicitar os marcos conceituais, instrumentos e razões que se propõem a explicar de maneira simplificada como uma realidade se move. Em vista disso, é imprescindível reconhecer as limitações desse processo, entre as quais Arellano et al. (2012, p. 23) destacam “que não necessariamente ele servirá para avaliar de forma neutra os resultados de um programa ou organização”. Apesar disso, avalia a força dos argumentos e instrumentos adotados por meio de aproximações válidas e confiáveis. Portanto, conforme Arellano et al. (2012), os resultados produzidos pela avaliação serão considerados confiáveis ou duvidosos a depender da força dessas aproximações.

2.2 O PAPEL DA UNIÃO NA AVALIAÇÃO DA EDUCAÇÃO BÁSICA BRASILEIRA E A SOBREPOSIÇÃO DOS SISTEMAS DE AVALIAÇÃO

No contexto do surgimento da avaliação educacional em larga escala e de sua chegada ao Brasil, se inicia o debate a respeito do papel da União, das Unidades da Federação (UFs) e dos Municípios, no regime de colaboração para gestão e avaliação educacional.

No Brasil, a avaliação em larga escala tem se firmado como política pública, com o propósito de garantir o direito à educação. Sua consolidação ocorre a partir dos anos 1990 em decorrência de uma série de reformas educacionais provocadas por ideais de modernização e desenvolvimento. Desde então, a avaliação em larga escala tem ocupado lugar de destaque no debate educacional brasileiro e a busca pela melhoria da qualidade da educação está ligada à implementação de sistemas de avaliação das redes de ensino.

A Constituição Federal de 1988, em seu art. 206, sintetiza os princípios que o processo educacional deve seguir, dentre eles, a garantia de padrão de qualidade. Dessa forma, apesar de não explicitar a avaliação educacional, infere que a avaliação se faz necessária para verificação da qualidade.

O art. 211 da Constituição Federal apresenta a divisão de responsabilidades em relação à gestão educacional brasileira e estabelece que “a União, os Estados, o Distrito Federal e os Municípios organizarão em regime de colaboração seus sistemas de ensino” (BRASIL, 1988). O regime de colaboração é detalhado nos parágrafos que seguem o caput do referido artigo:

§ 1º A União organizará o sistema federal de ensino e o dos Territórios, financiará as instituições de ensino públicas federais e exercerá, em matéria educacional, função redistributiva e supletiva, de forma a garantir equalização de oportunidades educacionais e padrão mínimo de qualidade do ensino mediante assistência técnica e financeira aos Estados, ao Distrito Federal e aos Municípios.

§ 2º Os Municípios atuarão prioritariamente no ensino fundamental e na educação infantil.

§ 3º Os Estados e o Distrito Federal atuarão prioritariamente no ensino fundamental e médio.

§ 4º Na organização de seus sistemas de ensino, a União, os Estados, o Distrito Federal e os Municípios definirão formas de colaboração, de modo a assegurar a universalização do ensino obrigatório.

§ 5º A educação básica pública atenderá prioritariamente ao ensino regular. (BRASIL, 1988, p.124)

Assim, temos que a União tem o papel de administrar a rede de ensino do país a fim de que haja ensino de qualidade ofertado de maneira equalizada. Aos Estados, Distrito Federal e

Municípios cabe a priorização do ensino fundamental de maneira compartilhada e sob a responsabilidade dos Estados e Distrito Federal está a gestão prioritária do ensino médio.

A LDB reforça o instituído pela Constituição Federal ampliando a lista de princípios que norteiam o processo educacional. A lei incumbe à União a tarefa de “assegurar processo nacional de avaliação do rendimento escolar no ensino fundamental, médio e superior, em colaboração com os sistemas de ensino, objetivando a definição de prioridades e a melhoria da qualidade do ensino” (BRASIL, 1996, p. 10). Nesse momento, fica clara a exigência de um sistema de avaliação de desempenho educacional. Além disso, estabelece-se como responsabilidade da União a elaboração do Plano Nacional de Educação (PNE), também em regime de colaboração com Estados, Distrito Federal e Municípios.

O PNE, publicado em 2014, transformou o que eram princípios em metas a serem alcançadas. Para tanto, criou parâmetros para produção de indicadores a respeito da educação básica brasileira, a partir dos dados do Sistema Nacional de Avaliação da Educação Básica, mais uma vez realizado em regime de colaboração (BRASIL, 2014).

Art. 11 O Sistema Nacional de Avaliação da Educação Básica, coordenado pela União, em colaboração com os Estados, o Distrito Federal e os Municípios, constituirá fonte de informação para a avaliação da qualidade da educação básica e para a orientação das políticas públicas desse nível de ensino (BRASIL, 2014, p.2).

O art.11 do PNE ainda estabelece que cabem ao Inep a elaboração e o cálculo do Ideb e de indicadores de rendimento escolar e de avaliação institucional. Esses indicadores devem ser estimados por etapa, estabelecimento de ensino, rede escolar, Unidade da Federação (UF) e, em nível agregado nacional, a partir de informações coletadas pelo Sistema Nacional de Avaliação da Educação Básica. No que diz respeito à avaliação de desempenho dos estudantes, ela poderá ser realizada diretamente pela União ou em regime de colaboração com os Estados, o Distrito Federal e os Municípios (BRASIL, 2014).

O PNE vem complementar o texto constitucional, detalhando que a União deverá fornecer referências de qualidade da educação, por meio de indicadores calculados com base em resultados de avaliações da educação básica. Registra-se ainda o papel da União nesse processo, designando o Inep como responsável pela elaboração e cálculo desses indicadores. Entretanto não limita à União o papel de realizar as avaliações, nem delimita a periodicidade de coleta dos dados e fornecimento dos indicadores.

2.2.1 O Inep – representante da União na avaliação da educação básica

O Inep é uma autarquia federal vinculada ao Ministério da Educação (MEC) cuja missão é “subsidiar a formulação de políticas educacionais dos diferentes níveis de governo com intuito de contribuir para o desenvolvimento econômico e social do país” (INEP, 2015a). Para tanto, o Inep atua em avaliações, exames e indicadores da educação básica e da educação superior; ações internacionais; estatísticas educacionais; biblioteca e arquivos da educação; e publicações.

Em 1937, o Instituto Nacional de Pedagogia foi criado por meio da Lei nº 378, como instituição de educação escolar, integrante ao Ministério da Educação e Saúde. O Instituto era “destinado a realizar pesquisas sobre os problemas do ensino, nos seus diferentes aspectos” (BRASIL, 1937). Entretanto, não começou a funcionar imediatamente.

No ano seguinte, as atividades foram iniciadas com a publicação do Decreto-Lei nº 580, de 30 de julho de 1938, “que dispõe sobre a organização do Instituto Nacional de Estudos Pedagógicos” (BRASIL, 1938). Foi estabelecido que o Instituto Nacional de Pedagogia passasse a ser denominado Instituto Nacional de Estudos Pedagógicos, nome que originou a sigla Inep. O seu objetivo era funcionar como “centro de estudos de todas as questões educacionais relacionadas com os trabalhos do Ministério da Educação e Saúde” (BRASIL, 1938).

Anos mais tarde, o Decreto nº 71.407, de 20 de novembro de 1972, transformou o Instituto Nacional de Estudos Pedagógicos em Instituto Nacional de Estudos e Pesquisas Educacionais. A partir de então, passou a ter “a finalidade de, como órgão central de direção superior, exercer todas as atividades necessárias ao estímulo, coordenação, realização e difusão da pesquisa educacional no País” (BRASIL, 1972). O Inep continuou vinculado ao Ministério da Educação e Cultura, porém com autonomia administrativa e financeira, previstas no artigo 172 do Decreto-lei número 200, de 25 de fevereiro de 1967.

O Inep foi transformado em autarquia federal vinculada ao Ministério da Educação e do Desporto, pela Lei nº 9.448, em 1997, que ampliou suas finalidades, inclusive aquelas relacionadas ao planejamento, orientação e coordenação dos sistemas de avaliação educacional.

Em 2001, o Inep tem seu nome alterado pela Lei nº 10.269, de 29 de agosto de 2001, passando a denominar-se Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira.

Desde a sua criação, o Inep tem entre suas atribuições a responsabilidade de conduzir o processo de avaliação educacional, a fim de subsidiar o processo de formulação de políticas

públicas, para favorecer o cumprimento do princípio constitucional referente à garantia da qualidade da educação. Atualmente o regimento interno do Inep reforça essa responsabilidade, explicitando-a entre as suas finalidades, a saber:

- I. subsidiar o poder público no monitoramento e na avaliação do Sistema Nacional de Educação;
- II. subsidiar o planejamento de políticas para a garantia do direito à educação de qualidade para todos e para cada um;
- [...]
- VI. planejar, gerir, monitorar e coordenar as atividades necessárias à operação logística das avaliações realizadas pela Autarquia;
- [...]
- IX. apoiar os estados, o Distrito Federal e os municípios no desenvolvimento de sistemas de avaliação educacional, em articulação com o sistema nacional de avaliação e seus respectivos sistemas de educação (BRASIL, 2017b, p.49)

A estrutura organizacional do Inep divide-se em órgãos de assistência imediata ao presidente do Inep, órgãos seccionais, e órgãos específicos singulares. No contexto da avaliação da educação básica, destaca-se a Diretoria de Avaliação da Educação Básica (Daeb), que tem as seguintes competências:

- I. definir e propor parâmetros, critérios e mecanismos de realização das avaliações da educação básica;
- II. promover, em articulação com os sistemas estaduais e municipais de ensino, a realização das avaliações da educação básica;
- III. definir e propor parâmetros, critérios e mecanismos para o estabelecimento de processos de certificação de competências;
- VI. apoiar os estados, o Distrito Federal e os municípios no desenvolvimento de projetos e sistemas de avaliação da educação básica;
- V. promover a realização de avaliações comparadas, em articulação com instituições nacionais e organismos internacionais;
- VI. liderar e acompanhar a elaboração de documentos normativos referentes a exames e avaliações da educação básica (BRASIL, 2017b, p.57).

As competências descritas evidenciam a Daeb como órgão responsável pela condução dos processos relacionados à avaliação da educação básica do Brasil.

A Daeb organiza-se em três coordenações gerais, sendo uma delas voltada para a realização de exames para certificação; outra, especificamente ligada ao sistema nacional de avaliação da educação básica; e a terceira, responsável pelas medidas da avaliação. Assim, no contexto da avaliação da educação básica, estão envolvidas a Coordenação-Geral do Sistema Nacional de Avaliação da Educação Básica e a Coordenação Geral de Instrumentos e Medidas. A primeira é responsável pela concepção das avaliações, disseminação dos resultados e

elaboração dos instrumentos e, a segunda, pelo cálculo dos resultados, elaboração dos planos amostrais e suporte às atividades das demais coordenações dessa diretoria. Entre as avaliações realizadas pelo Inep está o Saeb, desenvolvido sob a coordenação da Daeb e apresentado na seção 2.3.

2.2.2 A sobreposição dos sistemas de avaliação no regime de colaboração

A responsabilidade da União em relação à realização de avaliações educacionais está claramente expressa na legislação, bem como a atribuição do Inep, como representante da União na condução desse processo. Entretanto, as legislações estudadas não definem o papel dos Estados, Distrito Federal e Municípios no regime de colaboração com a União, no contexto da avaliação da educação.

Apesar de não estar regulamentada na legislação a forma de participação dos Estados e Municípios no Sistema Nacional de Avaliação da Educação Básica, vários deles estabeleceram, a partir do Saeb, sistemas de avaliação próprios a fim de avaliar o desempenho de seus alunos e a qualidade do ensino ofertado por suas escolas.

Nesse contexto, desenvolveram-se diversas ações voltadas para prestação de contas e responsabilização, incluindo a bonificação dos profissionais das instituições de ensino que obtivessem melhores resultados, tanto nas avaliações locais, quanto no Saeb.

Apesar de as avaliações educacionais estarem historicamente associadas à avaliação de programas federais, o uso dos resultados das avaliações educacionais estaduais brasileiras para este fim é pouco observado. Isso significa que geralmente os resultados das avaliações estaduais não estão relacionados à avaliação de programas ou políticas educacionais. Entretanto, a utilização desses resultados para o monitoramento genérico da melhoria da educação é crescente, como pode se verificar por meio do número de indicadores criados por Estados e Municípios, incentivados pela criação do Ideb em nível nacional (BROOKE; CUNHA, 2011).

Estudo desenvolvido por Brooke e Cunha (2011) mostrou que o desenvolvimento das avaliações estaduais tem caminhado no sentido de fornecer melhores instrumentos de disseminação dos resultados com o intuito de se aproximar da comunidade escolar. Entretanto, mesmo com a evolução desses instrumentos, como os boletins pedagógicos, os professores fazem a crítica de que os resultados das avaliações não são passíveis de ensejar ações de imediato.

Apesar de todo o esforço das secretarias de educação, os resultados desse tipo de avaliação enfrentam dificuldades para alcançar a comunidade escolar de maneira adequada. As

vantagens das avaliações em larga escala são reconhecidas. Entretanto, existem diversas críticas relacionadas aos efeitos negativos das avaliações, geralmente associados ao mau uso dos resultados que impõem pressões a professores e gestores, que acabam tomando decisões que prejudicam o planejamento pedagógico. Alguns efeitos observados são o afunilamento curricular, exclusão de alunos com piores desempenhos, a competição entre instituições e redes de ensino em função da formação de rankings e o desgaste dos envolvidos nesse processo (BAUER; AVELARSE; OLIVEIRA, 2015).

Brooke e Cunha (2011) relatam problemas apresentados em Estados onde se implementam políticas de incentivos salariais a partir dos resultados das avaliações, por exemplo: a exclusão dos alunos com desempenho mais baixo no momento de aplicação dos testes a fim de aumentar a média da escola, a manipulação de outras regras do processo, a concentração de atenção e dedicação apenas às etapas avaliadas, deixando de lado as demais séries. Além disso, os autores enfatizam a dificuldade de estabelecer um conjunto de regras que seja percebido como justo e legítimo, além da falta de evidências de que o sistema de bonificação realmente suscite melhorias na qualidade da educação ou aumente a motivação do professor.

Estudo realizado por Bauer, Avelarse e Oliveira (2015) aponta que, pelo menos, 19 estados brasileiros desenvolvem iniciativas de avaliações educacionais em larga escala alcançando 67% dos municípios brasileiros. Além disso, o estudo aponta que 37% dos Municípios têm iniciativas próprias de avaliação educacional e outros 21% pretendem implantar algum modelo de avaliação própria. Outro levantamento, realizado em 2016, revelou que o número de Estados brasileiros que possuíam iniciativa própria de avaliação externa subiu para 21 (BAUER, 2019).

Esse movimento tem sido estimulado pelo reconhecimento da importância dos resultados para a elaboração de políticas públicas. Os resultados das avaliações oferecem informações de qualidade para o planejamento, tomada de decisões, incentivo à melhoria dos resultados e fomento do comprometimento dos professores com o aprendizado dos alunos. Porém, o processo não é trivial e exige o cumprimento de condições específicas para que esses efeitos se manifestem, como evidenciam Bauer, Avelarse e Oliveira (2015, p.1379).

Diante do que se poderia denominar de visão ingênua das virtudes das avaliações em larga escala, inclusive porque constatamos características que deveriam ser suprimidas, consideramos que essas avaliações têm potencialidades para produzir avanços no conhecimento do universo educacional. Entendemos que podem, garantidas determinadas condições e adequada utilização, estabelecer pontos de apoio para políticas de melhoria dos sistemas educacionais no sentido da garantia do direito à educação a todos.

Machado, Alavarse e Arcas (2015) relatam que a motivação para o crescente número de avaliações estaduais poderia ser resumida: a) no prazo necessário para divulgação dos resultados do Saeb, considerado longo; b) no caráter amostral da avaliação que limitava a aplicação dos resultados em relação a necessidades específicas das redes de ensino e escolas; e c) na periodicidade de realização bianual, considerada espaçada para os gestores. Diante disso, a expectativa era de que a criação da Prova Brasil mitigasse esse crescimento, entretanto o que se observou foi o reforço do movimento.

Portanto, pode-se observar certa sobreposição entre as avaliações realizadas em âmbito nacional e aquelas realizadas na esfera estadual ou municipal, conforme comentam Sousa e Oliveira (2010, p.812):

Quando o governo federal, unilateralmente, toma a iniciativa de realizar uma testagem censitária, a Prova Brasil, apresentam-se dois tipos de problemas. De um lado, os estados passam a questionar-se acerca da validade de manutenção de sistemas próprios, com seus altos custos e complexidade logística, para realizar a mesma função que o governo federal já realiza. De outro, se estabelece uma frequência muito alta de aplicação de provas, o que inviabiliza, até mesmo, o tempo necessário para que seus resultados sejam absorvidos, analisados e transformados em ações por parte das redes de ensino.

Observa-se ainda que os sistemas estaduais de avaliação da educação coletam informações por meio de testes padronizados e questionários contextuais. Os testes avaliam predominantemente as áreas de conhecimento de Língua Portuguesa e Matemática e são aplicados aos estudantes da educação básica (ensino fundamental e ensino médio). Os questionários são respondidos por gestores, professores, alunos e seus responsáveis. Os dados coletados são utilizados para a geração de índices de qualidade da educação (MACHADO; ALAVARSE; ARCAS, 2015). Esse modelo muito se assemelha ao Saeb e ao Ideb, reforçando a ideia de sobreposição dos sistemas.

Outro fator destacado por Machado, Alavarse e Arcas (2015) é que as avaliações estaduais têm uma possibilidade maior de oferecer resultados que dialoguem de maneira mais próxima com os sistemas educacionais e as unidades escolares. A natureza das relações políticas que se estabelecem entre eles favorece o uso dos resultados como instrumento de gestão.

Diante desse cenário, compreender o Saeb, modelo de avaliação da educação básica implementado pelo Inep, ajuda na compreensão dos sistemas estaduais e municipais que replicam esse modelo. Independentemente da abrangência, todas essas avaliações estão dedicadas à tarefa de avaliar para garantir a qualidade educacional.

2.3 O SAEB

No Brasil, os primeiros passos para a implementação de uma avaliação educacional em larga escala ocorreram entre 1987 e 1990, com a proposição do Sistema de Avaliação do Ensino Público de 1º Grau – Saep. O modelo testava o rendimento dos alunos como medida de desempenho das escolas e sistemas. Essa avaliação desenvolveu-se em meio a amplos debates sobre a função seletiva e excludente da educação brasileira, apesar das pesquisas que alertavam para as consequências da avaliação na educação e na sociedade (COELHO, 2008).

Criado em 1990 em substituição ao Saep, o Sistema de Avaliação da Educação Básica (Saeb) ganhou maior importância e visibilidade em 1995 com a reestruturação do Inep. Nesse momento, o Saeb passou a fazer parte de um sistema de avaliação e informação da educação composto de censos e testes de desempenho em todos os níveis de ensino. Os resultados do Saeb coletados ao longo dos anos reforçaram as polêmicas quanto à efetividade da escola, pois mostravam a insuficiência nos índices de desempenho relacionados às habilidades e competências dos alunos nas áreas de conhecimento avaliadas (COELHO, 2008).

Análises dos resultados do Saeb a partir da edição de 2001 e da série histórica de 1995 a 2003 levantaram polêmicas, pois revelaram insuficiência nos níveis de aprendizagem dos alunos avaliados. Tais resultados levantaram questionamentos não apenas em relação às políticas educacionais, mas também suscitaram dúvidas a respeito da qualidade do ensino ofertada. Porém, uma vez que se observou nesse período um aumento no volume de matrículas, esse fato foi levantado como explicação para os resultados insatisfatórios, contestando a hipótese de que os alunos não estavam recebendo um bom ensino (COELHO, 2008).

Apesar das polêmicas despertadas, Vianna (2014) classifica o Saeb como o melhor e mais bem delineado projeto do Ministério da Educação. Sendo a educação básica o fundamento da formação cidadã e o alicerce das demais etapas educacionais, o autor defende que os responsáveis pela avaliação deveriam se preocupar com a sua validade de conteúdo e sua validade consequential. Além disso, Vianna (2014, p. 201) declara que as avaliações externas, como o Saeb, são recomendáveis em função de seu caráter isento.

As avaliações externas [...] são recomendáveis, na medida em que representam um trabalho não comprometido com a administração educacional e as políticas que a orientam. São avaliações que traduzem uma visão de fora e supostamente isenta em relação a possíveis idiosincrasias próprias dos sistemas educacionais.

Silva (2013) relata que em 2005 foi criada a Prova Brasil em decorrência de críticas sobre a eficiência da avaliação, que indicavam que o Saeb não era útil para formulação de

políticas públicas, porque seus resultados não alcançavam escolas e Municípios. Entretanto, a expectativa de que a aproximação das escolas favorecesse a qualidade do ensino não foi atendida. Pelo contrário, não foi possível observar melhorias relevantes nos resultados (SILVA, 2013). Além disso, pesquisa conduzida por Placco, Almeida e Souza (2011) revelou que mais de vinte anos depois da implementação do Saeb, os coordenadores pedagógicos parecem não compreender os resultados da avaliação e não sabem como utilizá-lo mesmo após a criação da Prova Brasil e do Ideb.

O Saeb foi uma das primeiras ações brasileiras voltadas para entender o nível de aprendizagem dos alunos e hoje se tornou um amplo instrumento de avaliação externa do Brasil e um dos mais sofisticados da América Latina. Ele permite identificar que o cenário educacional brasileiro tem sido insuficiente na função de promover educação de qualidade e com equidade (ARAÚJO; LUZIO, 2005). Essa identificação fomentou o desenvolvimento de programas e políticas voltados à melhoria dos indicadores educacionais fornecidos pelo Saeb. Araújo e Luzio (2005, p.20) ainda trazem ao debate que “os números são esclarecedores e apontam que é preciso introduzir mudanças significativas no modelo de gestão da Educação, bem como na prática cotidiana e no processo pedagógico nas escolas brasileiras.”

Reforçando o movimento de realização de avaliações educacionais externas, ou seja, aquelas realizadas por instituições alheias à escola, surgiram a partir dos anos 90 as avaliações de sistemas estaduais de ensino no Brasil, juntamente com o Saeb. A maioria desses sistemas adota aplicação amostral, por se mostrar mais econômica, mais fácil logisticamente e com resultados igualmente confiáveis aos das avaliações censitárias (VIANNA, 2014).

O Saeb foi instituído com o propósito de avaliar as redes de ensino brasileiras a fim de fornecer insumos para: a) o entendimento dos problemas enfrentados pelo sistema educacional; e b) a elaboração de políticas públicas que promovam a equidade e a melhoria da qualidade da educação básica.

Inicialmente, o Saeb era constituído de uma única avaliação amostral, que avaliava as redes de ensino pública e privada, oferecendo resultados agregados por UFs, Regiões e para o Brasil. Este modelo seguiu até o ano de 2003, pois a partir de 2005 o Saeb passou a ser composto por duas avaliações: a Avaliação Nacional da Educação Básica (Aneb), que manteve o modelo do Saeb com avaliação amostral, e a Avaliação Nacional do Rendimento Escolar (Anresc), também conhecida como Prova Brasil. Esta passou a avaliar de forma censitária as instituições públicas de educação básica, na série de encerramento de cada ciclo do ensino fundamental, oferecendo resultados para as instituições avaliadas.

Em 2013, foi incluída no Saeb a Avaliação Nacional de Alfabetização (ANA) com a finalidade de aferir os níveis de alfabetização e letramento em Língua Portuguesa (leitura e escrita) e Matemática. Em 2017, o ensino médio passou a integrar a Prova Brasil.

Até a edição mais recente, considerando o momento de início da escrita desta dissertação, aplicada em 2017, o Saeb passou por diversos ajustes a fim de se aprimorar no cumprimento do objetivo de realizar um diagnóstico da educação básica brasileira. Ao longo dos anos, foram modificadas não apenas as etapas avaliadas, mas a forma de coleta dos dados e a abrangência dos resultados gerados. Entretanto não foram encontrados registros sobre as análises de adequação do processo de coleta dos dados e, conseqüentemente, sobre a forma de cálculo dos resultados. O Quadro 1 resume as principais mudanças do Saeb, desde sua criação.

Quadro 1 - Evolução de características centrais do Saeb

EDIÇÃO	DISCIPLINAS AVALIADAS	POPULAÇÃO ALVO	SÉRIES AVALIADAS	TIPO DE APLICAÇÃO
1990	Língua Portuguesa, Matemática, Ciências Naturais e Redação	Escolas públicas	1 ^a , 3 ^a , 5 ^a e 7 ^a séries do EF	Amostral
1993	Língua Portuguesa, Matemática, Ciências Naturais e Redação	Escolas públicas	1 ^a , 3 ^a , 5 ^a e 7 ^a séries do EF	Amostral
1995	Língua Portuguesa (leitura) e Matemática	Escolas públicas e Escolas particulares	4 ^a e 8 ^a séries do EF, 3 ^a série do EM	Amostral
1997	Língua Portuguesa, Matemática, Ciências Naturais, Física, Química e Biologia	Escolas públicas e Escolas particulares	4 ^a e 8 ^a séries do EF, 3 ^a série do EM	Amostral
1999	Língua Portuguesa, Matemática, Ciências Naturais, Física, Química, Biologia, História e Geografia	Escolas públicas e Escolas particulares	4 ^a e 8 ^a séries do EF, 3 ^a série do EM	Amostral
2001	Língua Portuguesa, Matemática	Escolas públicas e Escolas particulares	4 ^a e 8 ^a séries do EF, 3 ^a série do EM	Amostral
2003	Língua Portuguesa, Matemática	Escolas públicas e Escolas particulares	4 ^a e 8 ^a séries do EF, 3 ^a série do EM	Amostral

2005	Língua Portuguesa, Matemática	Escolas públicas e Escolas particulares	4ªsérie/5º ano, 8ªsérie/9º ano do EF, 3ª série do EM	Parte censitária e parte amostral
2007	Língua Portuguesa, Matemática	Escolas públicas e Escolas particulares	4ªsérie/5º ano, 8ªsérie/9º ano do EF, 3ª série do EM	Parte censitária e parte amostral
2009	Língua Portuguesa, Matemática	Escolas públicas e Escolas particulares	4ªsérie/5º ano, 8ªsérie/9º ano do EF, 3ª série do EM	Parte censitária e parte amostral
2011	Língua Portuguesa, Matemática	Escolas públicas e Escolas particulares	5º ano/4ªsérie, 9º ano/ 8ªsérie do EF, 3ª e 4ª série do EM	Parte censitária e parte amostral
2013	Língua Portuguesa, Matemática	Escolas públicas e Escolas particulares	5º ano/4ªsérie, 9º ano/ 8ªsérie do EF, 3ª e 4ª série do EM	Parte censitária e parte amostral
2015	Língua Portuguesa, Matemática	Escolas públicas e Escolas particulares	5º ano e 9ºano do EF, 3ª e 4ª série do EM	Parte censitária e parte amostral
2017	Língua Portuguesa, Matemática	Escolas públicas e Escolas particulares	5º ano e 9ºano do EF, 3ª e 4ª série do EM (tradicional e integrado)	Parte censitária e parte amostral
2019	Língua Portuguesa, Matemática, Ciências Humanas e Ciências da Natureza	Escolas públicas e Escolas particulares	2º ano, 5º ano e 9ºano do EF, 3ª e 4ª série do EM (tradicional e integrado)	Parte censitária e parte amostral

Fonte: Elaborado pela autora

Entre as alterações que o Saeb experimentou, podemos destacar a mudança expressiva na metodologia adotada para a coleta dos dados. Nas edições de 1990 e 1993 os dados eram coletados de maneira amostral apenas entre as escolas públicas que ofereciam a 1ª, 3ª, 5ª e 7ª séries do ensino fundamental, para avaliar a aprendizagem dos alunos nos componentes curriculares de Língua Portuguesa, Matemática, Ciências Naturais e Redação (INEP, 2019d).

Para a edição de 1995, as fontes consultadas não são unânimes. O histórico do Saeb apresentado no portal do Inep informa que essa edição seguiu o mesmo formato das duas

primeiras aplicações. Os demais documentos consultados registram a avaliação da 4^a e 8^a séries do ensino fundamental e a 3^a série do ensino médio em Língua Portuguesa (Leitura) e Matemática, envolvendo escolas públicas e particulares. Por serem mais completas, foram consideradas informações dos documentos relacionados ao planejamento amostral (ANDRADE; SILVA; BUSSAB, 2001; FLETCHER, 2006; INEP, 2006a, 2019d; SILVA; BUSSAB; ANDRADE, 2003b).

A edição de 1995 tem a particularidade de ter sido a primeira em que se adotou a Teoria de Resposta ao Item (TRI) para a construção de testes e análise dos resultados. Essa medida permitiu que os resultados das diferentes edições passassem a ser comparáveis a partir de então.

Nas primeiras edições, os resultados eram analisados apenas sob a ótica da Teoria Clássica dos Testes (TCT), em que a proficiência do aluno é função do seu número de acertos. Essa metodologia ajuda a “compreender as dificuldades dos participantes em relação aos itens do teste” (INEP, 2019c, p.45), mas dificulta a comparação dos resultados de alunos submetidos a testes aplicados em momentos diferentes (INEP, 2019c).

O que a TRI propõe são modelos para descrever “características do indivíduo que não podem ser observadas diretamente” (ANDRADE; TAVARES; VALLE, 2000, p. 3), chamadas de traço latente. Assim, a TRI busca modelar a probabilidade de um item ser acertado por um indivíduo de acordo com a medida de seu traço latente, oferecendo uma aproximação para a medida que não pode ser obtida de maneira direta (ANDRADE; TAVARES; VALLE, 2000). No caso do Saeb, o traço latente é a proficiência do aluno em cada área de conhecimento avaliada. A fim de viabilizar a comparabilidade entre as edições estabeleceu-se uma escala comum que tem como média o valor de 250 pontos e desvio padrão de 50 pontos: escala (250,50) (INEP, 2019c). Os resultados são originalmente calculados em escala (0,1). A definição da escala foi fixada com base nos resultados da edição de 1997, quando se arbitrou que a média da 8^a série daquela edição valeria 250 e o desvio padrão 50. Dessa edição em diante, os resultados obtidos na escala (0,1) são equalizados na escala (250, 50) fixada anteriormente (KLEIN, 2009).

A coleta dos dados a partir de 1997 confirmou as mudanças encontradas para a edição de 1995, envolvendo a Amostragem de escolas públicas e particulares. Também confirmou a alteração das etapas de ensino, avaliando as 4^a e 8^a séries do ensino fundamental e a 3^a série do ensino médio. Além disso, passou a considerar os componentes curriculares de Língua Portuguesa, Matemática e Ciências Naturais para o ensino fundamental e Língua Portuguesa, Matemática, Física, Química e Biologia para o ensino médio (INEP, 2006b). Em 1999 foram incluídas avaliações em História e Geografia e, a partir de 2001, houve a redução das disciplinas

avaliadas, passando a incluir apenas Língua Portuguesa e Matemática (ANDRADE; SILVA; BUSSAB, 2001; INEP, 2006c).

Em 2005, o Saeb foi reestruturado pela Portaria Ministerial nº 931/2005, aproximando-se do modelo atual e passando a ser composto por duas avaliações: Avaliação Nacional da Educação Básica (Aneb) e Avaliação Nacional do Rendimento Escolar (Anresc), conhecida como Prova Brasil. A Prova Brasil avalia de forma censitária a aprendizagem de Língua Portuguesa e Matemática na 4ª série/5º ano e 8ª série/9º ano do ensino fundamental de escolas públicas com pelo menos 30 alunos matriculados na etapa avaliada, gerando resultados por escola e município. A Aneb avalia a aprendizagem de Língua Portuguesa e Matemática na 4ª série/5º ano e 8ª série/9º ano do ensino fundamental e na 3ª série do ensino médio em escolas com no mínimo 10 estudantes matriculados, das redes públicas e privadas, gerando resultados para UF, região e Brasil. Para tanto, são selecionadas amostras complementares para cobrir a parte da Aneb que não foi avaliada pela Prova Brasil (BRASIL, 2005a; BRASIL, 2005b; BUSSAB; MIAZAKI; RABELLO, 2005).

Esse modelo vem sendo aplicado até hoje com pequenas alterações: na edição de 2009, o mínimo de 30 alunos da Prova Brasil foi reduzido para 20 alunos e na edição de 2017 para 10 alunos. Além disso, a partir da edição de 2011, a 4ª série do ensino médio passou a ser considerada como etapa de conclusão do ensino médio, juntamente com a 3ª série. Em 2017, a avaliação das 3ª e 4ª séries do Ensino médio passou a integrar a Prova Brasil, tendo aplicação censitária para as escolas públicas. Também foi incluído no universo da avaliação o ensino médio integrado em escolas públicas (ANDRADE, 2011a; BRASIL, 2009; BRASIL, 2017a; INEP, 2018a).

Na edição de 2019, foram incluídas alterações relevantes no que diz respeito à delimitação da avaliação. Para a definição do universo, foram acrescentados, com aplicação amostral, o 2º ano do ensino fundamental e a 3ª e 4ª séries do ensino médio integrado de escolas particulares. Além disso, houve a ampliação do conteúdo avaliado passando a englobar as áreas de Ciências Humanas e Ciências da Natureza, para uma amostra de alunos do 9º ano do ensino fundamental (BRASIL, 2019).

As alterações relatadas impactaram diretamente o plano amostral considerado para a coleta dos dados, especialmente as mudanças de população alvo e do tipo de aplicação. Além dessas adaptações, foram realizados ajustes na abrangência da avaliação e em pontos específicos do desenho amostral. Diante disso, faz-se necessário um detalhamento da evolução do plano amostral do Saeb, que será apresentado na próxima seção.

2.4 HISTÓRICO DO PLANO AMOSTRAL DO SAEB

Para compor um histórico da evolução do plano amostral do Saeb, foi realizada análise documental, a partir de arquivos fornecidos pelo Inep, e documentação complementar, fornecida por profissionais que em algum momento participaram da elaboração do plano amostral do Saeb. Para as edições de 1990 e 1993 não foram encontrados documentos descritivos do plano amostral. Em relação aos anos de 1995 a 2011, os microdados¹ publicados no portal do Inep fornecem no arquivo *leia-me* um capítulo destinado à descrição do plano amostral aplicado na edição a que se refere cada um dos arquivos. A partir de 2013, os microdados passaram a incluir um documento separado a respeito deste assunto, denominado Relatório de Amostragem. Além desses arquivos públicos, foram fornecidos diretamente pelo Inep, por meio de demanda no Serviço de Informação ao Cidadão alguns relatórios mais detalhados do processo de construção do planejamento amostral. A complementação documental obtida diretamente com profissionais envolvidos no processo de planejamento da Amostragem do Saeb consistiu em relatórios referentes aos anos de 1999 e 2001.

Desde a criação do Saeb, foram realizadas alterações na metodologia de coleta dos dados. Entretanto, não foram encontrados registros sobre as análises realizadas para a fundamentação das modificações nos métodos de Amostragem adotados, a fim de se adequarem aos novos objetivos da avaliação. No entanto, permaneceu inalterada a realidade de que uma pesquisa do porte do Saeb exige uma amostra complexa. Assim, a amostra do Saeb se desdobra em amostras independentes, referentes a cada série avaliada. Cada uma dessas amostras é estratificada e conglomerada, com seleção em múltiplas etapas. Esses conceitos serão explicados posteriormente na seção 3.1.2.

Em 1995 o plano amostral visava gerar resultados agregados nacionais, por UF e por nove regiões metropolitanas para cada uma das séries e disciplinas avaliadas. Para isso, foi selecionada uma amostra em três etapas: na primeira etapa foram sorteados alguns Municípios; na segunda etapa foram sorteadas algumas escolas dentro dos Municípios já selecionados; por último, dentro das escolas selecionadas eram sorteadas algumas de suas turmas (FLETCHER, 2006).

¹ Microdados são arquivos que disponibilizam os dados coletados em uma pesquisa ou levantamento, em seu menor nível de desagregação. Os microdados do Inep estão disponíveis em seu próprio portal, bem como no portal de dados abertos do Governo Federal e incluem, além dos arquivos de dados, diversos materiais explicativos. Para leitura das bases, o usuário precisa utilizar *softwares* específicos em função do seu grande volume.

A seleção dos Municípios deu-se com Probabilidades Proporcionais ao Tamanho e tendo como referência um número mínimo de alunos esperado. Caso esse número não fosse atingido era formado um bloco composto por pequenos Municípios semelhantes que, juntos, apresentassem o número de matrículas necessário. Para a seleção das escolas dentro dos Municípios já escolhidos, foi realizado o ordenamento das escolas de acordo com o número de matrículas. Nesse momento, foi novamente adotado o procedimento de agrupamento de unidades, caso a escola selecionada tivesse menos de 6 alunos matriculados na série avaliada. Outro critério adotado foi que cada escola fosse avaliada em apenas uma das séries, fazendo parte de apenas uma das amostras do Saeb 1995. Por fim, foi selecionada em média uma turma em cada escola. Porém, como o procedimento de seleção levou em consideração a quantidade de alunos em cada turma, poderiam ser selecionadas todas as turmas ou nenhuma das turmas, retirando a escola da amostra. Ao término do processo de seleção, a amostra total do Saeb 1995 foi composta por aproximadamente 45.000 alunos das quatro séries avaliadas (FLETCHER, 2006).

Na edição de 1997, foram excluídas escolas federais, escolas rurais da Região Norte, escolas profissionalizantes e turmas multisseriadas. Para a seleção da amostra, a população de cada UF foi dividida em subgrupos, chamados de estratos, compostos pela combinação de dependência administrativa (estadual, municipal, particular), localização (urbana ou rural) e área (capital ou interior). Deixaram de ser consideradas as regiões metropolitanas para estratificação da população. O procedimento de seleção continuou sendo realizado em etapas, mas o número de etapas variou dependendo do estrato: três etapas para os estratos de escolas urbanas do interior, duas etapas para as escolas urbanas das capitais e rurais do interior e uma etapa para as escolas rurais das capitais. Sendo assim, para as escolas urbanas do interior as etapas de seleção foram município, escola e turma; para as escolas localizadas em áreas urbanas das capitais, selecionaram-se escolas e em seguida as turmas. Para as escolas rurais, considerou-se que a maioria delas possuía apenas uma turma. Portanto, essa etapa foi suprimida, restando a seleção do município e da escola para as rurais do interior e apenas a seleção da escola para as rurais das capitais. O método de seleção com Probabilidades Proporcionais ao Tamanho continuou sendo utilizado. Outro critério modificado foi que nessa edição uma mesma escola poderia fazer parte da amostra de mais de uma série.

Os tamanhos de amostra calculados foram ajustados, por meio do aumento no número de turmas, para minimizar as perdas do processo de aplicação da avaliação. Este procedimento ainda é adotado nas edições mais recentes do Saeb, com o acréscimo de 25% da quantidade de

turmas para o 5º e o 9º anos do ensino fundamental e de 50% para o ensino médio (ANDRADE, 2011b; INEP, 2006b; 2013b; 2017; 2018a).

Ao término do processo de seleção, a amostra do Saeb 1997 foi composta por aproximadamente 167.000 alunos das três séries avaliadas (INEP, 2006b).

Em 1999 foram consideradas as seguintes exclusões: a) as escolas federais, por serem em número muito reduzido; b) escolas rurais de todo o Brasil para a amostra da 8ª série do EF e 3ª série do EM; e escolas rurais do Sul, Sudeste, DF, MT, GO para a amostra da 4ª série do EF, em função de seus resultados não impactarem de forma significativa os resultados nacionais; e c) turmas multisseriadas, pela dificuldade de aplicação dos testes (BUSSAB et al., 1999; INEP, 2006c).

Ainda na edição de 1999, o processo de seleção da amostra foi simplificado e unificado, seguindo o mesmo procedimento para todos os estratos. Foi adotada a seleção em duas etapas: primeiramente as escolas e, em seguida, as turmas. Destaca-se que, para escolas com uma ou duas turmas, foi sorteada uma turma. Para escolas com três turmas ou mais, o sorteio contemplou duas turmas de cada escola. Os alunos participaram respondendo a questões de uma das disciplinas. Assim, do ponto de vista da Amostragem, para cada disciplina a seleção foi realizada em três etapas, acrescentando a etapa de seleção dos alunos. Para garantir o espalhamento geográfico da amostra, os estratos foram organizados em zonas de Amostragem segundo o Código de Endereçamento Postas (CEP) das escolas, e em cada zona de Amostragem foram selecionadas duas escolas (BUSSAB, *et al.*, 1999; INEP, 2006c). A organização dos estratos em zonas de Amostragem para seleção das escolas continuou vigente até o Saeb 2017.

O cálculo do tamanho da amostra do Saeb 1999 se deu em três passos: a) para cada estrato de interesse, a partir de um número fixo de alunos, calculou-se o número de turmas necessário, com base no número médio de alunos por turma; b) o número de turmas a ser sorteado foi distribuído em cada estrato de interesse entre dois subgrupos referentes ao porte da escola (pequeno se tiver até duas turmas ou grande se tiver três ou mais turmas) proporcionalmente ao número de turmas existente em cada subgrupo; c) com base na quantidade de turmas selecionadas em cada subgrupo, foi calculado o número de escolas a serem selecionadas, arredondando o número de escolas para múltiplos de dois a fim de facilitar a seleção de escolas que foi realizada aos pares: duas escolas em cada estrato de Amostragem (INEP, 2006c). Esse processo ainda é utilizado, porém, a partir de 2011, o cálculo adotou como referência inicial um número de alunos estimado com base na variabilidade da população estudada.

Para seleção das escolas da amostra do Saeb de 1999, o método de Amostragem com Probabilidades Proporcionais ao Tamanho continuou a ser utilizado. Porém, foi adotado o caso particular da Amostragem Sequencial de Poisson, procurando “favorecer” a sobreposição de escolas selecionadas para a avaliação nas diversas séries consideradas. Ao término do processo de seleção, a amostra total do Saeb 1999 foi composta por aproximadamente 360.500 alunos das três séries avaliadas (INEP, 2006c).

O plano amostral do Saeb 2001 foi elaborado pela mesma equipe do Saeb 1999, o que resultou em propostas bastante semelhantes. A edição de 2001 manteve praticamente inalterados os procedimentos de seleção e de cálculo do tamanho da amostra, bem como as séries avaliadas. Foi incluído na Amostragem Sequencial de Poisson um esquema de rotação parcial controlada² da amostra de escolas com a repetição de parâmetros da edição anterior, permitindo que fossem incluídas na amostra de 2001 parte das escolas selecionadas na amostra da edição anterior. Além disso, os alunos foram avaliados somente em Língua Portuguesa e Matemática, com a participação de metade dos alunos de cada turma em cada uma das duas disciplinas. O tamanho inicial da amostra em número de alunos foi calculado considerando 300 alunos avaliados em cada disciplina e estrato de interesse para cada série. As exclusões permaneceram as mesmas do Saeb 1999 (ANDRADE; SILVA; BUSSAB, 2001; INEP, 2006d). Ao término do processo de seleção, a amostra do Saeb 2001 foi composta por aproximadamente 370.000 alunos das três séries avaliadas (INEP, 2006d).

Para 2003, mais uma vez a equipe técnica responsável pela elaboração do plano amostral foi mantida. Assim, o procedimento de Amostragem continuou semelhante ao de 2001, mas passando a incluir as escolas federais e as escolas rurais do 4º ano com pelo menos 10 alunos. Nessa edição voltou a consideração das regiões metropolitanas para a estratificação da população. Ao término do processo de seleção, a amostra do Saeb 2003 foi composta por aproximadamente 275.000 alunos das três séries avaliadas (INEP, 2006e; SILVA; BUSSAB; ANDRADE, 2003a; 2003b).

A edição de 2005 contou com uma alteração da equipe técnica, mas manteve um dos profissionais envolvidos nos trabalhos de 2003. Dessa forma, o plano amostral dessa edição

² Ao justificar a adoção do esquema de rotação parcial controlada, Andrade, Silva e Bussab (2001) explicam que “ao fixar os números aleatórios atribuídos a cada escola ao longo do tempo, abre-se também a possibilidade de controlar a rotação da amostra de escolas em Saeb’s consecutivos. Fizemos uso deste recurso para buscar manter na amostra do SAEB01 aproximadamente a metade das escolas pesquisadas no Saeb99, caso fosse repetido exatamente o mesmo plano amostral e tamanhos de amostra. [...] a possibilidade de manter na amostra algumas escolas já pesquisadas no Saeb99, abre possibilidades interessantes para análises longitudinais, impossíveis sem essa sobreposição” (ANDRADE; SILVA; BUSSAB, 2001, p. 20).

manteve-se semelhante ao de 2003, com a diferença que a estratificação voltou a desconsiderar as regiões metropolitanas dividindo as escolas em dois grupos, segundo a área (capital ou interior). Outra alteração foi em relação às exclusões; também foram excluídos os alunos matriculados em escolas exclusivamente de educação especial, naquelas localizadas em áreas indígenas e quilombolas e, ainda, naquelas com menos de 10 alunos. Ao término do processo de seleção, a amostra do Saeb 2005 foi composta por aproximadamente 262.000 alunos das três séries avaliadas (BUSSAB; MIAZAKI; RABELO, 2005a; 2005b; INEP, 2008).

O plano amostral do Saeb 2007 foi elaborado por equipe diferente daquela que trabalhou nos anos de 1999 a 2005, mas a nova equipe manteve os padrões adotados nas edições anteriores. As alterações identificadas foram a inclusão dos alunos matriculados no ensino fundamental de 9 anos, assim foram avaliados alunos da 4ª série / 5º ano e 8ª série / 9º ano, além dos alunos da 3ª série do ensino médio. Ao término do processo de seleção, a amostra do Saeb 2007 foi composta por aproximadamente 304.500 alunos de todas as séries avaliadas (DUARTE; BATISTA; ALVES, 2007a; 2007b).

Um dos consultores participantes do planejamento amostral no período de 1999 a 2005 voltou a compor a equipe técnica responsável pela elaboração do plano amostral do Saeb em 2009. Nesse momento, a estratificação por dependência administrativa passou a contar com dois grandes grupos (pública ou privada) agrupando as escolas federais, estaduais e municipais em um único estrato. Além disso, o esquema de rotação parcial controlada da amostra de escolas foi interrompido em função da indisponibilidade de informações suficientes sobre os parâmetros considerados em 2007. Ao término do processo de seleção, a amostra total do Saeb 2009 foi composta por aproximadamente 148.500 alunos de todas as séries avaliadas (BUSSAB; NISHIMURA, 2009).

Com mais uma alteração da equipe técnica, o planejamento amostral da edição de 2011 acompanhou os planos adotados desde 2003. Mas, pela primeira vez, contou com um plano amostral que levava em consideração a aplicação censitária da Prova Brasil, criada em 2005. Assim, a amostra do Saeb 2011 foi “formada por escolas particulares do 5º ano e 9º ano do Ensino Fundamental com 10 ou mais alunos em turmas regulares, e por escolas públicas e particulares da 3ª série do Ensino Médio com 10 ou mais alunos em turmas regulares” (INEP, 2013, p. 7). Também foram incluídas amostras das escolas públicas de 5º ano e 9º ano, com 10 a 19 alunos pois não estavam contempladas na aplicação da Prova Brasil, que alcançava apenas escolas com pelo menos 20 alunos. A estratificação por dependência administrativa incluiu um estrato específico para as escolas federais que passaram a ser avaliadas em caráter censitário e foi retomada a avaliação das escolas rurais. Outra diferença dessa edição é que os alunos foram

submetidos às provas das duas disciplinas, levando a seleção da amostra a ser realizada efetivamente em duas etapas (escolas e turmas). Foram mantidos os procedimentos de seleção da amostra utilizados anteriormente, adaptados conforme os novos limites da população alvo (ANDRADE, 2011a; 2011b; INEP, 2013).

O cálculo do tamanho da amostra foi alterado, pois as referências para esse cálculo passaram a ser as estimativas da média e da variância das proficiências calculadas na edição anterior do Saeb, e não mais um número fixo. Também foi considerado um erro amostral máximo de 10 pontos absolutos na escala do Saeb. Os procedimentos de divisão dos estratos em zonas de Amostragem e de seleção das escolas por Amostragem Sequencial de Poisson foram mantidos. Ao término do processo de seleção, a amostra total do Saeb 2011 foi composta por aproximadamente 143.000 alunos de todas as séries avaliadas (ANDRADE, 2011a; 2011b; INEP, 2013).

Os documentos referentes ao planejamento amostral do Saeb 2013 não indicam a equipe técnica envolvida, mas os procedimentos adotados nesse planejamento seguiram o padrão das edições anteriores, com poucas modificações. O esquema de rotação parcial controlada das escolas foi retomado, e o erro amostral máximo esperado passou a ser 4 pontos absolutos na escala do Saeb. Foi incluído no planejamento amostral um estudo experimental para o ensino médio integrado e para avaliação de Ciências no 9º ano (8ª série) do ensino fundamental e na 3ª série (4ª série) do ensino médio. Entretanto os documentos não explicitam os objetivos do estudo. A seleção da amostra das escolas públicas de 5º e 9º anos com 10 a 19 alunos foi diferenciada por se tratar de complementação aos dados coletados na Prova Brasil. Nesse caso, foram sorteadas 4 escolas urbanas e 6 escolas rurais de cada UF. Ao término do processo de seleção, a amostra total do Saeb 2013 foi composta por aproximadamente 201.000 alunos de todas as séries avaliadas (INEP; 2013a; 2013b; 2014).

A partir de 2015, o planejamento amostral do Saeb passou a ser realizado por equipe composta por servidores do Inep. A equipe contou com o apoio de uma Comissão de Assessoramento em Amostragem, composta por especialistas com destacada atuação nos processos de Amostragem e/ou análises em avaliações educacionais em larga escala. O planejamento amostral de 2015 seguiu os mesmos procedimentos de 2013, com a única diferença de que não foi mantido o estudo das escolas com ensino médio integrado e de Ciências. Ao término do processo de seleção, a amostra total do Saeb 2015 foi composta por aproximadamente 170.000 alunos de todas as séries avaliadas (INEP, 2017).

Em 2017, a avaliação do ensino médio foi incluída na Prova Brasil, assim como as escolas públicas de 5º e 9º anos com 10 a 19 alunos. Assim, a amostra dessa edição considerou

apenas as escolas particulares com pelo menos 10 alunos matriculados, ou seja, as escolas com menos de 10 alunos não foram avaliadas. O ensino médio integrado voltou a ser avaliado pelo Saeb. Porém, foi incluído somente na Prova Brasil e não foi contemplado pela amostra. Além disso, a Portaria Inep nº 447/2017 permitiu a avaliação por adesão para escolas do ensino médio e previu a geração de resultados individuais para as escolas participantes dessa etapa. Em decorrência disso, para o ensino médio não foi realizado o sorteio de turmas, tendo sido avaliados todos os alunos das escolas sorteadas. Ao término do processo de seleção, a amostra total do Saeb 2017 foi composta por aproximadamente 118.000 alunos de todas as séries avaliadas (BRASIL, 2017a; INEP, 2018a; 2018b).

O Saeb 2019, regido pela Portaria Inep nº366/2019, previu quatro amostras diferentes:

II - uma amostra de escolas privadas localizadas em zonas urbanas e rurais que possuam 10 (dez) ou mais estudantes matriculados em turmas de 5º ano e 9º ano do Ensino Fundamental e de 3ª e 4ª série do Ensino Médio (tradicional e integrado), distribuídas nas vinte e sete Unidades da Federação;

III - rurais que possuam 10 (dez) ou mais estudantes matriculados em turmas 9º ano do Ensino Fundamental, distribuídas nas vinte e sete Unidades da Federação, para aplicação dos instrumentos descritos no inciso V do art. 11 da presente Portaria

IV - uma amostra de escolas públicas e privadas localizadas em zonas urbanas e rurais que possuam 10 (dez) ou mais estudantes matriculados em turmas de 2º ano do Ensino Fundamental, distribuídas nas vinte e sete Unidades da Federação, para aplicação exclusiva dos instrumentos previstos no inciso VI do art. 11.

V - uma amostra de instituições públicas ou conveniadas com o setor público, localizadas em zonas urbanas e rurais que possuam turmas de creche ou pré-escola da etapa da Educação Infantil, para aplicação exclusiva dos instrumentos previstos nos incisos I, II e III, do art. 11, em caráter de estudo-piloto (BRASIL, 2019, p.2)

Para viabilizar essa nova realidade, fizeram-se necessários ajustes no plano amostral do Saeb. A nota técnica nº 10/2019/CGIM/DAEB explica que em 2019 todos os alunos responderam a testes de Língua Portuguesa, Matemática e, no 9º ano, metade dos alunos da amostra participarão da avaliação de Ciências Humanas e a outra metade de Ciências da Natureza. O erro máximo esperado continuou sendo de 4 pontos absolutos na escala do Saeb, assim como foram mantidos os procedimentos de seleção e cálculo do tamanho da amostra. O ensino médio integrado foi incluído na amostra (INEP, 2019a).

Para a educação infantil, como estudo piloto, não havia previsão de divulgação de resultados, mas o plano amostral tinha por objetivo coletar dados para aprimoramento e

refinamento do procedimento proposto e elaboração de indicadores para avaliação desta etapa da Educação Básica (INEP, 2019a).

Para as demais etapas, houve previsão de que a amostra, juntamente com os dados coletados de forma censitária, permitisse a geração de resultados para 2º, 5º e 9º anos do ensino fundamental, e 3ª e 4ª série do ensino médio conjuntamente como etapa de conclusão do ensino médio, por UF, Região e Brasil. Assim, ficariam garantidos resultados confiáveis para os estratos de interesse que consideraram a dependência administrativa (federal, estadual, municipal ou privada), área (capital ou interior) e localização (urbana ou rural). As regras se aplicaram tanto para Língua Portuguesa e Matemática, quanto para Ciências Humanas e Ciências da Natureza (exclusivamente para o 9º ano). Para manter a comparabilidade dos resultados do ensino médio com as edições anteriores do Saeb, os resultados do ensino médio integrado seriam tratados separadamente dos dados do ensino médio tradicional, avaliado até 2017. Mas também poderiam ser gerados resultados conjuntos para as duas modalidades de ensino médio (INEP, 2019a).

No processo de construção desse histórico, observou-se que a documentação desse processo é incompleta, restringindo-se basicamente a resumos do plano amostral incluídos no documento “Leia-me” constante dos microdados da avaliação. Para as edições de 1999, 2001, 2003, 2007, 2011 e 2013, pode-se contar com relatórios técnicos descritivos mais completos. Para as demais edições não foi encontrada documentação completa e explicativa sobre a metodologia adotada. Em função da rotatividade dos profissionais envolvidos, as alterações nos planos amostrais nas primeiras edições do Saeb eram mais intensas. No período de 1999 até 2005, essa equipe permaneceu mais estável, estabelecendo um modelo de trabalho replicado desde então, com pequenos ajustes a cada edição.

Outro aspecto que impacta a mutabilidade do plano amostral é a instabilidade da própria avaliação que vem incorporando mudanças decorrentes de ajustes políticos e pedagógicos, a cada edição. No decorrer dos anos foram alteradas as etapas de ensino e disciplinas avaliadas, bem como a abrangência dos resultados. Desde 2005, as mudanças foram mais sutis, o que refletiu também na redução das alterações do plano amostral.

A partir da edição de 2015, o processo de Amostragem do Saeb, que antes era realizado exclusivamente por consultores externos, passou a ser internalizado pela equipe de pesquisadores do Inep. Nas edições seguintes, o Saeb também passou por ajustes, mais fortemente nas edições de 2017 e 2019 com a alteração da população avaliada. Em 2017 foi incluído o ensino médio na Prova Brasil. Foram acrescentadas em 2019 a avaliação do ensino médio integrado, a avaliação da alfabetização aplicada de maneira amostral ao 2º ano do ensino

fundamental e a avaliação amostral das áreas de Ciências Humanas e Ciências da Natureza para os alunos do 9º ano do ensino fundamental.

Mesmo com a ampliação da Prova Brasil e o aumento do volume de dados coletados de forma censitária, a coleta amostral continua sendo necessária. Como será explicado mais à frente, os resultados obtidos em uma pesquisa não são válidos para uma população que não foi representada na coleta de dados, seja de forma censitária ou amostral. Assim, como a coleta censitária abrange apenas escolas públicas, a supressão da amostra deixaria a população das escolas privadas descoberta pela avaliação.

Diante de todas as mudanças realizadas ao longo dos quase 30 anos de existência do Saeb, faz-se necessária avaliação sobre a adequação de alguns aspectos da metodologia adotada para a coleta dos dados para a realização da avaliação.

A elaboração do plano amostral tem se dado com base em portarias que estabelecem os critérios considerados em cada edição do Saeb. A partir daí, avaliam-se quais alterações são necessárias no plano amostral, tomando-se o cuidado de manter as referências principais que permitam a comparabilidade dos resultados em relação aos anos anteriores. Também são realizados estudos e discussões para auxiliar na tomada de decisão em relação aos critérios a serem definidos em portaria.

3 UMA ANÁLISE SOBRE O PLANO AMOSTRAL DO SAEB

O capítulo anterior apresentou o histórico de surgimento da avaliação educacional em larga escala e sua chegada ao Brasil. A partir daí, mostrou que, seguindo a tendência internacional, foram desenvolvidos sistemas de avaliação educacional com a finalidade de verificar a qualidade da educação ofertada pelas redes de ensino.

Identificou-se que o sistema nacional, representado pelo Saeb, e os sistemas estaduais se sobrepõem em função da sua semelhança. No entanto, apesar de ambos os modelos fornecerem informações no contexto de políticas públicas, eles se diferenciam no que diz respeito ao uso dos seus resultados. O Saeb é mais usado para a avaliação de políticas já implementadas, enquanto os sistemas estaduais são mais requeridos como ferramentas de gestão, tendo seus resultados vinculados a políticas específicas de bonificação e responsabilização de forma mais frequente.

Independentemente do uso dos resultados, uma avaliação em larga escala precisa garantir a qualidade e a confiabilidade dos seus resultados. Para isso, a adequação da metodologia considerada para a coleta de dados, mais especificamente no caso das avaliações amostrais, é indispensável.

A realização de uma avaliação em larga escala não implica uma aplicação necessariamente censitária, ou seja, a todos os elementos do universo pesquisado. Dependendo dos objetivos da avaliação, uma aplicação amostral é suficiente e pode trazer vantagens e facilidades ao processo, garantindo qualidade próxima àquela observada em uma aplicação censitária (BOLFARINE; BUSSAB, 2005).

Neste capítulo serão apresentados alguns conceitos básicos da Teoria da Amostragem e como esses conceitos se aplicam no contexto do Saeb. Além disso, serão realizadas análises do plano amostral adotado no Saeb hoje. Para tanto, serão feitas análises dos resultados obtidos e do plano amostral adotado, bem como comparação entre resultados obtidos com a aplicação censitária e resultados obtidos a partir de seleção amostral. Por fim, serão discutidos outros aspectos relacionados à periodicidade da avaliação e ao uso dos resultados que podem representar limitações para a avaliação.

3.1 TEORIA DE AMOSTRAGEM

Um levantamento pode-se referir a uma pesquisa realizada por recenseamento (ou censo) ou por Amostragem. O que diferencia esses dois tipos de levantamento é a quantidade

de elementos estudados, pois no primeiro tipo todos os elementos são estudados enquanto no segundo apenas parte deles (BOLFARINE; BUSSAB, 2005).

Bolfarine e Bussab (2005) explicam que faz parte do senso comum a crença de que apenas um recenseamento pode fornecer a “verdade” sobre uma determinada população. De fato, se forem aplicados nas mesmas condições, os dados resultantes de um levantamento censitário são mais precisos que aqueles provenientes de uma amostra. No entanto, esse cenário pode se inverter quando existem limitações financeiras ou logísticas, tornando um levantamento amostral mais fidedigno que o censo.

Os autores exemplificam o caso de uma pesquisa sobre o estado de saúde da população de uma cidade em que se propõem dois modelos de coleta de dados. Pode-se optar pela aplicação de questionários a todos os cidadãos, ou pela realização de exames clínicos e laboratoriais, com a participação de médicos e outros profissionais especializados, em uma amostra. Neste caso, a opção por uma amostra provavelmente fornecerá resultados com mais qualidade que a aplicação censitária (BOLFARINE; BUSSAB, 2005).

Diante disso, Bolfarine e Bussab (2005) recomendam que se adote o recenseamento quando a população é pequena, os erros amostrais são muito grandes, há baixo custo envolvido ou as consequências relacionadas a uma decisão errada sejam muito graves. Em contrapartida, o levantamento amostral deve ser preferido quando a população é grande e/ou quando o custo para a obtenção das informações é alto. Em todo caso, deve prevalecer o bom senso na escolha do tipo de levantamento a ser adotado.

A Amostragem é base de construção de conhecimento, tanto no contexto cotidiano quanto em pesquisas científicas. Os processos de Amostragem são muito variados e se aplicam de modo igualmente variado. Seja na opinião formada por um turista que passa 10 dias em uma cidade e se torna apto a dar recomendações a respeito do lugar, seja em um estudo científico desenvolvido durante anos para o entendimento de um aspecto específico da formação cultural da mesma cidade. Os dois processos exemplificados são diferentes em profundidade e propósito, entretanto, ambos se referem à análise de um todo a partir da observação de um recorte, ou seja, ambos podem ser resultados de processos amostrais (COCHRAN, 1977). Assim, nas próximas páginas serão apresentados fundamentos sobre a Amostragem para que seja possível compreender as principais características de um levantamento desse tipo.

3.1.1 O que é Amostragem

Para iniciar a discussão sobre Amostragem, é importante entender o seu conceito. Para tanto, pode-se recorrer ao dicionário que define Amostragem como um conjunto de métodos para a “seleção de amostra para ser examinada como representante de um todo” (FERREIRA, 2004, p.118). Para complementar o entendimento, vale também alinhar o conceito de amostra, que é definida no contexto da estatística como “parte de uma população selecionada para pesquisa de características dela” (FERREIRA, 2004, p.118).

Assim, pode-se resumir que Amostragem é o processo de determinação de um subgrupo da população a ser examinado, a fim de fornecer informações que representem a população da qual faz parte.

O uso da Amostragem muitas vezes é encarado com desconfiança especialmente por aqueles que, possivelmente por não terem conhecimento técnico e por estarem acostumados a lidar com o resultado de levantamentos censitários, apresentam resistência à adoção de um novo modelo de coleta de dados. Entretanto, esse posicionamento pode ser modificado quando as vantagens do uso da Amostragem são conhecidas (COCHRAN, 1977).

Cochran (1977) listou as quatro principais vantagens da Amostragem em comparação ao censo:

- **Custo reduzido:** Permite que sejam coletados dados seguros consultando apenas parte da população. Essa redução na quantidade de unidades pesquisadas faz com que os custos envolvidos sejam inferiores aos de uma coleta censitária.
- **Maior agilidade:** Além da economicidade, a redução do número de unidades pesquisadas proporciona a redução no tempo de coleta e análise dos dados. A evolução tecnológica observada desde a época em que Cochran escreveu seu livro favoreceu significativa redução de tempo no processamento de grande volume de dados. Entretanto, o tempo de coleta dos dados ainda é maior proporcionalmente ao volume de dados a serem coletados.
- **Maior abrangência:** As pesquisas amostrais permitem a realização de estudos mais abrangentes e com maior flexibilidade em relação aos temas que serão pesquisados. A vantagem que fica evidente em alguns tipos de levantamento de dados, em que a aplicação censitária se torna inviável por causa da necessidade de pessoal altamente treinado e equipamentos especializados ou com disponibilidade limitada. Nesses casos, a amostra se apresenta como única opção para a realização da pesquisa. Entretanto, quando a pesquisa requer precisão de

resultados para muitas subdivisões da população, ou para domínios muito pequenos, a amostra pode ser tão grande que não se justifique, sendo mais adequada a coleta censitária.

- **Maior acurácia:** Outra vantagem decorrente da redução no volume de trabalho é a possibilidade de melhor capacitação das equipes de coleta e supervisão mais próxima no trabalho de campo. Este cenário favorece a obtenção de resultados mais precisos do que em uma coleta censitária.

As pesquisas amostrais podem apresentar variados níveis de complexidade, e questões que são triviais em uma situação podem ser verdadeiros desafios em outro contexto. Assim, para orientar a elaboração de um desenho amostral, Cochran (1977) listou onze passos, que serão apresentados a seguir.

- **Objetivos da pesquisa:** Para que uma pesquisa amostral complexa apresente resultados consistentes para a tomada de decisão é importante que os objetivos da pesquisa estejam definidos com clareza. Objetivos vagos ou mal definidos dificultam o planejamento da pesquisa e facilitam a tomada de decisões em desacordo com os objetivos.
- **População a ser amostrada:** De maneira geral, população se refere ao agrupamento do qual se seleciona a amostra. A população a ser amostrada (de referência) deve coincidir com a população sobre a qual se quer obter informações (população alvo), pois os resultados obtidos de uma amostra se aplicam somente à população de referência. Particularidades do processo podem fazer com que a população de referência seja mais restrita que a população alvo. Quando isso acontece, a extensão desses resultados à população alvo dependem de outras fontes de informação que possam complementar o entendimento das diferenças entre a população alvo e de referência.
- **Dados a serem coletados:** É primordial que os dados essenciais ao objetivo da pesquisa sejam incluídos no processo e que seja evitada a coleta de dados irrelevantes. Instrumentos de pesquisa longos demais prejudicam a qualidade dos dados tanto para as questões importantes quanto para aquelas que são dispensáveis.
- **Grau de precisão desejado:** Em pesquisas amostrais, os resultados carregam consigo alguma incerteza decorrente de serem obtidos a partir de consulta a apenas parte da população e de erros de medida. A adoção de amostras maiores,

de desenhos amostrais mais eficientes e de melhores instrumentos de medição favorecem a redução da incerteza dos resultados, mas podem exigir maior tempo e mais recursos financeiros. Diante disso, é indispensável a definição de qual o nível de erro tolerável para que as estimativas sejam suficientes para a tomada de boas decisões. Para tanto, as equipes que utilizarão os resultados deverão fazer essa definição procurando um ponto de equilíbrio entre precisão desejada e recursos disponíveis.

- **Métodos de medição:** A construção dos formulários de coleta consome grande parte do trabalho de preparação de uma pesquisa. Esse processo se inicia com a visualização da estrutura de tabelas que serão utilizadas para a análise dos dados. A partir daí é necessário que se faça a escolha do instrumento de medição e da estratégia de abordagem da população. Em questionários simples, as perguntas podem ser pré-codificadas, facilitando a inserção dos dados em sistemas para a realização das análises.
- **Cadastro:** Para seleção da amostra, a população é organizada em partes chamadas de unidades amostrais ou, simplesmente, unidades. A lista destas unidades é chamada de cadastro. A construção e manutenção dessas listas é um dos problemas práticos mais comuns em Amostragem, pois quanto mais específica é a população mais difícil é o delineamento do cadastro. As unidades devem cobrir toda a população sem que haja sobreposição, de forma que cada elemento da população pertença a uma e somente uma unidade amostral. Assim, em alguns casos a definição das unidades é óbvia, mas em outros é resultado de uma escolha.
- **Seleção da amostra:** Existe uma grande variedade de métodos pelos quais uma amostra pode ser selecionada. Para a escolha do método a ser adotado, devem ser comparados o tempo e os custos envolvidos. Para cada método considerado, podem ser calculadas estimativas aproximadas de tamanho da amostra, que podem ser calculadas a partir do conhecimento do grau de precisão desejado.
- **Pré-teste:** O teste dos instrumentos e metodologias de trabalho de campo em escala menor é importante para revelar problemas que seriam graves em uma escala maior de aplicação. Essa etapa geralmente fornece insumos para o aprimoramento dos instrumentos, favorecendo a obtenção de melhores resultados para a pesquisa.

- **Organização do trabalho de campo:** A orientação da equipe de campo a respeito dos objetivos da pesquisa e dos métodos de medição adotados, bem como a supervisão adequada do trabalho podem evitar diversos problemas de gerenciamento, comuns em grandes pesquisas. A verificação da qualidade dos dados durante o processo de coleta e o planejamento sobre como lidar com a não resposta e a falha na coleta da informação para alguma unidade amostral também são indispensáveis para que o trabalho de campo seja bem sucedido.
- **Crítica e análise dos dados:** O primeiro passo aqui é a decisão de como tratar as respostas com falhas de registro e dados claramente errados, considerando inclusive a possibilidade de adotar métodos de imputação. Em seguida, são processados os cálculos das estimativas, aplicando os métodos de estimação mais adequados a cada conjunto de dados. Por último, os resultados são apresentados, preferencialmente acompanhados de medidas de precisão para as estimativas.
- **Informações obtidas para futuras pesquisas:** Toda amostra realizada fornece informações sobre médias, desvios-padrão, natureza da variabilidade das principais medidas e sobre os custos envolvidos na coleta dos dados, que podem ser usadas como referência para um próximo planejamento amostral. Esse processo é importante porque quanto mais informações estão disponíveis a respeito da população, mais fácil é elaborar um plano amostral que fornecerá estimativas acuradas. Além disso, dificilmente uma Amostragem complexa se realiza exatamente como foi planejada, portanto, cada experiência torna o amostrista atento mais capaz se preparar para que erros de execução não se repitam em pesquisas futuras.

Além disso, é importante a compreensão da diferença entre população alvo e população de referência, bem como conhecer sua delimitação, pois este é o critério inicial para o delineamento da amostra. População alvo diz respeito ao grupo sobre o qual se gostaria de obter informações por meio de uma pesquisa. Já a população de referência delimita o grupo que será efetivamente coberto pela pesquisa.

Quando se trata de estatísticas oficiais, os levantamentos amostrais precisam ser cuidadosos e “refletir os mais altos padrões de procedimento e análise, fundamentados numa base metodológica bastante sólida” (FLETCHER, 2006, p.139). Isso é importante, pois os resultados divulgados devem ser objeto de confiança do público. Fletcher (2006) relata que nos

primeiros 30 anos do século XX era comum a preocupação com a qualidade das estatísticas oficiais, envolvendo o que se chamava à época de “método representativo”. Nesse contexto, uma pergunta recorrente era: como selecionar amostras que representassem satisfatoriamente a população estudada e permitissem que as conclusões fossem válidas? Esse questionamento fomentou grande desenvolvimento da teoria de Amostragem.

O conhecimento e o uso hábil da estrutura populacional podem redundar em apreciáveis benefícios de precisão em determinadas circunstâncias. Essas considerações tiveram a maior importância no estabelecimento de padrões de prática nos levantamentos por Amostragem, hoje universalmente endossados. Esse ponto é de especial relevância no contexto das estatísticas oficiais (FLETCHER, 2006, p.139).

O propósito da Teoria da Amostragem é tornar a estimação mais eficiente, tentando desenvolver métodos de seleção da amostra e de cálculo das estimativas, com o menor custo possível e fornecendo estimativas suficientes para os objetivos definidos. Por isso, o foco deve estar voltado para esse objetivo mesmo nas etapas do processo em que esta teoria desempenha um papel menos marcante, tais como a organização do trabalho de campo e a definição da população, dos dados coletados e dos instrumentos utilizados (COCHRAN, 1977).

Uma grande parte da teoria da Amostragem gira em torno de encontrar fórmulas para os cálculos de médias e variâncias. Essa construção somente é garantida por meio da adoção de métodos probabilísticos de Amostragem, definidos como os procedimentos que permitem que seja calculada a distribuição de probabilidades das estimativas geradas caso o procedimento amostral fosse aplicado repetidas vezes à mesma população. Além disso, esse procedimento permite que sejam calculadas as probabilidades de seleção das diferentes amostras possíveis e que se calculem estimativas a partir dos dados de uma amostra específica (COCHRAN, 1977).

Cochran (1977) alerta que amostras não probabilísticas podem até fornecer resultados úteis, mas que são de difícil verificação. O autor aponta que o único modo de verificar quão bons são estes resultados é realizando comparações com os resultados de uma amostra probabilística ou de uma situação em que são conhecidos os resultados da população como um todo. Mesmo que um método não probabilístico apresente bons resultados em uma comparação, não há garantias de que a mesma confiabilidade será observada em outras circunstâncias.

Diante disso, nos próximos tópicos deste texto sempre que for usado o termo Amostragem, será referente à Amostragem probabilística.

3.1.2 Tipos de Amostragem

Um dos aspectos que caracterizam a Amostragem probabilística é a forma de seleção dos elementos componentes da amostra, em que cada elemento da população pode ser selecionado de acordo com uma probabilidade conhecida (ou calculável) e pré-definida. Existem diversas técnicas de Amostragem que variam de acordo com a metodologia de seleção da amostra. Os tipos de planejamento amostral mais comuns serão brevemente apresentados a seguir.

Amostragem Aleatória Simples (AAS): “é o método mais simples e mais importante para a seleção de uma amostra” (BOLFARINE; BUSSAB, 2005, p. 61). Consiste em um sorteio simples, em que cada elemento tem a mesma probabilidade de ser selecionado. Além disso, cada possível amostra tem também a mesma probabilidade de ser escolhida. Na AAS a seleção das unidades amostrais pode ocorrer com ou sem reposição, ou seja, um elemento já selecionado pode ser novamente sorteado ou cada elemento pode ser selecionado apenas uma vez para compor a amostra. O principal pré-requisito para o uso da AAS é a existência de um sistema de referência (cadastro) completo – uma relação que descreve todos os elementos da população. A ausência desse cadastro completo muitas vezes se torna impeditiva para a aplicação desse modelo mais simples de Amostragem. Outra característica da AAS é seu frequente uso na composição de planos amostrais complexos realizados em múltiplos estágios, por isso a AAS é tão importante (BOLFARINE; BUSSAB, 2005).

Amostragem Sistemática (AS): esse método de Amostragem também exige uma listagem dos elementos da população, pois é operacionalizado a partir dessa lista que é ordenada conforme um aspecto importante do plano amostral. Após o ordenamento é efetuada a seleção aleatória do primeiro elemento da amostra entre os k primeiros elementos do sistema de referência. Os seguintes são selecionados sistematicamente a cada k -ésimo elemento. As vantagens desse modelo são a facilidade de implementação e a garantia de espalhamento da amostra pois a população é estratificada em grupos de k elementos e apenas um elemento é selecionado em cada grupo (COCHRAN, 1977; BOLFARINE; BUSSAB, 2005).

Amostragem Estratificada (AE): para a utilização do plano de Amostragem estratificada a população precisa ser dividida em subpopulações, chamadas de estratos. É indispensável que os estratos sejam mutuamente excludentes e exaustivos, ou seja, que não se sobreponham e que juntos cubram toda a população. Assim, cada elemento da população faz parte de um, e somente um, estrato. Então seleciona-se uma amostra independente em cada estrato, que conjuntamente compõem uma amostra que representa a população. Se para seleção

dessas amostras for utilizada a AAS, então denomina-se esse plano de Amostragem Estratificada Simples (COCHRAN, 1977).

A AE é uma técnica comum por várias razões, entre as quais pode-se destacar: a) se forem necessários dados com precisão conhecida para determinados subgrupos da população, é recomendável tratar cada subgrupo como uma população específica; b) os estratos podem ser definidos por conveniência administrativa; c) os problemas de Amostragem podem apresentar grandes diferenças de acordo com a realidade de cada parte da população e a estratificação permite que o planejamento seja adaptado às particularidades de cada uma delas; d) a estratificação pode proporcionar ganhos de precisão nas estimativas de características da população como um todo.

Os ganhos de precisão ocorrem porque uma população heterogênea pode ser subdividida em grupos internamente homogêneos. Dessa forma, as medidas dos elementos da amostra de cada estrato variam pouco resultando em estimativas mais precisas para cada estrato e a combinação dessas várias estimativas gera estimativas populacionais também com maior precisão (COCHRAN, 1977).

Amostragem por Conglomerados (AC): também se refere a um processo em que a população é subdividida em grupos, dessa vez chamados de conglomerados. A partir dessa divisão, alguns conglomerados são selecionados para compor a amostra e, quando a Amostragem é realizada em um estágio, todos os seus elementos são observados. Geralmente esse processo é menos eficiente que a AAS ou a AE, porém tende a ser mais econômico. A AC é recomendada quando o sistema de referência está incompleto ou desatualizado e a sua atualização é um dificultador, e quando a população pode ser dividida em muitos subgrupos. O processo de AC se torna mais eficiente quando os conglomerados são mais heterogêneos internamente e mais homogêneos entre si, de forma que cada conglomerado seja uma microrrepresentação do universo estudado (BOLFARINE; BUSSAB, 2005).

Amostragem com Probabilidade Proporcional ao Tamanho (PPT): é o procedimento em que a escolha das unidades amostrais se dá com probabilidade de seleção proporcional a uma medida de tamanho, que poderia ser ao número de elementos que as compõem. Assim, a probabilidade de que cada unidade seja incluída na amostra é desigual. Para tanto, é necessário que o número de elementos de cada unidade amostral seja conhecido (COCHRAN, 1977).

Amostragem por Conglomerado em Dois Estágios (A2E): este método de planejamento amostral se dá quando uma amostra é selecionada em duas etapas. A população é subdividida como na AC e seleciona-se uma amostra de conglomerados em um primeiro

estágio, em seguida o segundo estágio consiste em selecionar uma amostra dentro de cada conglomerado já selecionado. Para cada estágio do procedimento pode se adotar uma técnica de amostragem diferente podendo ser uma das previamente apresentadas nessa seção. A A2E apresenta uma grande variedade de aplicações e sua principal vantagem é agregar flexibilidade ao processo (COCHRAN,1977; BOLFARINE; BUSSAB, 2005).

A partir do entendimento das principais características de uma pesquisa por Amostragem e dos principais métodos de Amostragem, bem como do histórico de evolução do plano amostral de Saeb, trazidos nessa seção, é possível realizar uma avaliação da sua adequação, identificando suas vantagens e desvantagens.

3.2 ANÁLISE DO PLANO AMOSTRAL DO SAEB

Consolidando as informações apresentadas até este ponto, pode-se construir um retrato mais completo do atual plano amostral do Saeb. Tendo em vista que o Saeb 2019 ainda estava em andamento quando da escrita desta dissertação, considera-se como metodologia de amostragem atual aquela adotada para o Saeb 2017.

Para compreender a metodologia de coleta de dados de uma pesquisa é importante conhecer os seus objetivos, visto que todo o planejamento deve ter como alvo o alcance dos objetivos propostos. Segundo o documento de referência do Saeb, esse sistema tem por objetivos no âmbito da Educação Básica:

- (i) avaliar a qualidade, a equidade e a eficiência da educação praticada no país em seus diversos níveis governamentais;
- (ii) produzir indicadores educacionais para o Brasil, suas regiões e Unidades da Federação e, quando possível, para os municípios e as instituições escolares, tendo em vista a manutenção da comparabilidade dos dados, permitindo, assim, o incremento das séries históricas;
- (iii) subsidiar a elaboração, o monitoramento e o aprimoramento de políticas públicas baseadas em evidências, com vistas ao desenvolvimento social e econômico do Brasil; e
- (vi) desenvolver competência técnica e científica na área de avaliação educacional, ativando o intercâmbio entre instituições educacionais de ensino e pesquisa (INEP, 2018d, p.6).

Sendo assim, os dados coletados precisam fornecer informações suficientes para o atendimento de uma missão de abrangência diversificada e imprecisa, que faz mais referência à forma de uso dos resultados do que em relação ao que se pretende medir. O público-alvo e a população de referência da avaliação são descritos pela portaria Inep nº 447/2017, que estabelece as diretrizes para o Saeb 2017. A população alvo abrange os alunos de escolas

públicas e privadas, matriculados no 5º ou 9º anos do ensino fundamental ou na 3ª ou 4ª séries do ensino médio (INEP, 2017a). Enquanto a população de referência engloba os alunos de escolas públicas e privadas, matriculados no 5º ou 9º anos do ensino fundamental ou na 3ª ou 4ª séries do ensino médio, em instituições com pelo menos 10 matrículas na série avaliada. São excluídos alunos das turmas de Educação Especial Exclusiva, multisseriadas, de correção de fluxo, de Educação de Jovens e Adultos ou de ensino normal/magistério, assim como aqueles de escolas indígenas que não ministram o ensino em Língua Portuguesa (INEP, 2017a).

As escolas pertencentes à população de referência são identificadas com base nas informações declaradas no Censo Escolar, adotado como cadastro para a realização do Saeb (INEP, 2017a). Em decorrência da aplicação censitária para parte da população de referência, a saber, as escolas públicas, a amostra refere-se apenas às escolas particulares que atendam à delimitação da população de referência. Vale destacar que os resultados gerados pela avaliação somente podem ser expandidos para a população de referência. Diante disso, os resultados obtidos, seja pela aplicação amostral, seja pela censitária, somente se aplicam às instituições que se enquadram nessa delimitação, não sendo diretamente válidos para as escolas incompatíveis com esse perfil.

As disciplinas avaliadas são Língua Portuguesa e Matemática e todos os alunos da amostra respondem a questões de ambas as disciplinas (INEP, 2018a).

O tipo de plano amostral do Saeb é de grande complexidade, combinando quase todos as técnicas apresentadas na seção anterior. Trata-se de um desenho amostral em duas etapas, estratificado, por conglomerados com probabilidades proporcionais a medidas de tamanho. Na primeira etapa, as combinações de escola/série – unidades primárias de amostragem deste plano amostral – são distribuídas em estratos e selecionadas com probabilidades proporcionais ao número de turmas aptas a participarem da avaliação, utilizando a Amostragem Sequencial de Poisson. Na segunda etapa, uma ou duas turmas de cada escola são sorteadas por Amostragem Aleatória Simples, e no dia da aplicação todos os alunos presentes nas turmas sorteadas participam da avaliação. Esse processo se repete para cada uma das séries incluídas na população de referência. Particularmente para o ensino médio a etapa de seleção da turma foi retirada, pois nesta edição havia a intenção de fornecer resultado a todas as escolas participantes e o sorteio de turmas inviabilizaria esse resultado (INEP, 2018a).

Para cada série avaliada, a estratificação da população se dá pela combinação de três aspectos da caracterização das escolas. Uma vez que a dependência administrativa foi limitada apenas às escolas privadas em função da aplicação censitária para as escolas públicas, os elementos de estratificação são: UF (26 Estados + Distrito federal), localização (urbana ou

rural) e área (capital ou interior). A combinação desses critérios resulta em um total de 324 possíveis estratos, sendo 108 de cada série. Entretanto, ao se realizar a distribuição das escolas nesses estratos observa-se que alguns não apresentam escolas que façam parte da população de referência. Dessa forma, a população de referência para o Saeb 2017, construída com base no Censo Escolar 2016, conta com 225 estratos viáveis (INEP, 2018a).

A definição do tamanho da amostra se deu com base no cadastro de escolas, turmas e alunos registrados no Censo Escolar do ano anterior. Esta escolha decorre dos prazos necessários para o preenchimento pela escola dos dados referentes ao Censo Escolar do ano de aplicação da avaliação e os prazos necessários para a operacionalização dos procedimentos relacionados ao dimensionamento da amostra. Tal prática é útil e válida pois não são esperadas mudanças importantes nas características das escolas de um ano para o outro (INEP, 2018a).

Além do Censo Escolar, para identificar a variabilidade da população, utiliza-se as médias e variâncias observadas na Prova Brasil anterior. No caso de 2017 foram usados os resultados da Prova Brasil 2015. Tendo em vista que o Saeb oferece resultados precisos para Língua Portuguesa e Matemática, o tamanho da amostra foi calculado considerando as informações a respeito do desempenho dos alunos em cada uma das áreas de conhecimento. Então foi tomado como tamanho de amostra final aquele que prevê aplicação ao maior número de turmas (INEP, 2018a).

Outro detalhe relatado nos documentos analisados é que o tamanho da amostra é calculado considerando a população de escolas públicas e privadas. Entretanto, como a aplicação nas escolas públicas se dá em caráter censitário, aplica-se a avaliação por Amostragem a uma quantidade de turmas que representa 40% do tamanho de amostra calculado (INEP, 2018a). Inicialmente, a amostra abrangia escolas públicas e privadas, proporcionalmente ao tamanho da população de cada rede de ensino, com o cuidado de que a diferença do tamanho da amostra entre cada uma delas não fosse maior que 20%. Assim, na maioria das UFs prevalecia a divisão da amostra em 60% para a rede pública e 40% para a rede privada. Com a substituição da amostra de escolas públicas pela aplicação censitária, o procedimento foi mantido, porém utilizando-se apenas os 40% referentes à rede privada (INEP, 2013b).

No cenário em que os resultados se referem à população de escolas públicas e privadas, calcula-se o tamanho da amostra independentemente da dependência administrativa das escolas. Para conseguir melhor distribuição das escolas realiza-se a alocação do número de turmas proporcionalmente ao tamanho da população de cada dependência administrativa. Para garantir que a diferença do tamanho da amostra por dependência administrativa não se diferenciasse em mais de 20 pontos percentuais, adotou-se a medida de 40% para a dependência

administrativa com população menor e 60% para a população maior. Como a aplicação para as escolas públicas é censitária, o tamanho da amostra de escolas privadas foi estabelecido como sendo 40% do tamanho da amostra calculado considerando ambas as dependências administrativas (ANDRADE, 2011b).

O tamanho da amostra é calculado em número de turmas para cada UF e em seguida essa quantidade é distribuída entre os estratos de interesse, combinados com a divisão das escolas por porte; a essa combinação dá-se o nome de estrato de tamanho. São classificadas como pequenas as escolas com uma ou duas turmas e como grandes as escolas com três turmas ou mais. A partir dessa distribuição, considerando o número de turmas a ser sorteado em cada escola, chega-se ao tamanho da amostra em número de escolas (INEP, 2018a).

Para a seleção das escolas é aplicada a Amostragem com Probabilidades Proporcionais ao Tamanho através do algoritmo da Amostragem Sequencial de Poisson, descrito em documento publicado pelo Inep (2018a), que pode ser organizado em quatro passos:

- 1) Atribui-se a cada escola do cadastro um número aleatório com base na distribuição uniforme no intervalo (0;1) de forma que o processo de atribuição dos números aleatórios seja independente para cada escola;
- 2) Calcula-se uma medida de tamanho relativo com base no número de turmas da escola (*número de turmas da escola/número de turmas existentes*). Essa medida pode ser chamada de probabilidade de seleção em um sorteio único e a soma dessas medidas deve ser igual a 1;
- 3) Calcula-se para cada escola uma variável aleatória transformada, que é o resultado da divisão do número aleatório atribuído inicialmente pela medida de tamanho relativo da escola;
- 4) Faz-se uma nova subdivisão da lista de escolas em função do seu CEP, agrupando-as em zonas, em que a combinação do estrato de interesse, o porte e a zona da escola é denominada estrato de Amostragem;
- 5) Ordena-se o cadastro de escolas, segundo as variáveis aleatórias transformadas em ordem crescente e inclui-se na amostra as primeiras escolas da lista ordenada, ou seja, aquelas que possuem menores variáveis aleatórias transformadas. A quantidade de escolas a serem incluídas na amostra é definida de acordo com o tamanho calculado para o estrato.

Para a amostra do Saeb, duas escolas são selecionadas em cada um dos estratos de Amostragem. Esse procedimento acarreta um ajuste no tamanho da amostra a fim de que o número de escolas selecionadas seja múltiplo de 2.

3.3 SIMULAÇÕES DA REALIZAÇÃO AMOSTRAL A PARTIR DA APLICAÇÃO CENSITÁRIA

Para verificar a eficiência do desenho amostral do Saeb, foi realizado um estudo por meio de simulações estocásticas considerando a metodologia atual. Os resultados obtidos a partir de amostragem foram comparados com os resultados obtidos a partir dos dados coletados de forma censitária.

Para que essa comparação seja possível foi considerada a aplicação do plano amostral para o universo das escolas públicas, público-alvo da Prova Brasil, que na prática é aplicada de forma censitária. O estudo foi realizado para o 5º ano do ensino fundamental. Este recorte foi escolhido por ser o que passou por menos mudanças nas últimas edições do Saeb e apresentar a menor taxa de não resposta. Em 2017, o ensino médio passou a integrar a Prova Brasil, com a inclusão do ensino médio integrado e a participação de escolas particulares por adesão e retornou 29,39% das provas sem preenchimento. Para o 9º ano a taxa de não resposta foi de 22,84% das provas, enquanto para o 5º ano a taxa foi de 16,40% (INEP, 2018c).

Para a simulação foram utilizados como referência para o dimensionamento da amostra os dados da Prova Brasil de 2017, que estão disponíveis nos microdados publicados pelo Inep. Para processamento dos cálculos e manipulação das bases de dados foi utilizado o ambiente de programação R e o *software SAS Studio* por meio do *SAS OnDemand for Academics*.

Para o planejamento deste estudo de simulação, foi utilizada a abordagem estruturada ADEMP, delineada por Morris, White e Crowther (2019). Este acrônimo vem da combinação das etapas do processo: Objetivos, Mecanismos de geração de dados, Estimativas, Métodos e Medidas de desempenho - em inglês: *Aims, Data-generating mechanisms, Estimands, Methods, Performance measures*.

A simulação realizada foi estruturada conforme detalhado a seguir:

Objetivos: os objetivos de um estudo de simulação geralmente são definidos levando em consideração a verificação de propriedades desejáveis para o estimador de um parâmetro de interesse, sob o plano amostral estudado. No presente caso, o objetivo é investigar a ocorrência de possível viés e eficiência do estimador, sob o plano amostral adotado para a seleção da amostra do Saeb.

Mecanismos de Geração de Dados: os dados podem ser gerados pela produção de desenhos paramétricos a partir de uma distribuição conhecida ou da repetição da seleção da amostra obtida de um conjunto de dados especificado. Neste estudo, foi adotada a repetição da seleção da amostra. O plano amostral do Saeb foi aplicado, com a utilização de diferentes

números aleatórios a cada repetição da Amostragem Sequencial de Poisson. Após a seleção das escolas, o sorteio das turmas foi realizado por AAS a cada rodada da simulação.

Estimativas: a maioria dos estudos de simulação avaliam ou comparam métodos de estimação de quantidades populacionais. As estimativas produzidas neste estudo são de médias de proficiência no Brasil e por UF, para Língua Portuguesa e Matemática.

Métodos: este é um termo genérico que se refere às técnicas ou procedimentos adotados, como regras de decisão, por exemplo. Aqui, o método adotado no estudo de simulação terá como base a comparação dos resultados gerados por meio da Amostragem com aqueles gerados a partir da aplicação censitária.

Medida de Desempenho: descreve uma quantidade numérica usada para avaliar o desempenho de um método e depende dos objetivos e metas do estudo. Para avaliação de um estimador (no caso deste estudo, a média), as medidas de desempenho mais naturais são o viés e medidas de precisão, tais como o erro padrão, erro quadrático médio e o intervalo de confiança.

Além da estruturação da abordagem, um aspecto importante no planejamento de um estudo de simulação é a definição do número de repetições. A análise realizada por Morris, White e Crowther (2019) mostrou que são comuns estudos de simulação com número de repetições $d = 500$ e $d = 1.000$. Para este estudo, o número de repetições foi definido em $d = 2.000$, principalmente em função de limites computacionais, e considerando que este é um número suficientemente grande e superior ao comumente adotado.

Morris, White e Crowther (2019, p. 2097) lembram que “nenhum estudo de simulação é definitivo e novos métodos ou refinamento dos métodos são inevitáveis”. Assim, as conclusões do presente estudo poderão ser, futuramente, aprimoradas.

3.3.1 Seleção da amostra

A seleção da amostra se inicia com o dimensionamento do seu tamanho. Seja B o erro máximo desejado (precisão esperada) e Y a variável estudada, considerando-se uma AAS com reposição, o tamanho da amostra (n_0) é calculado a partir da fórmula apresentada a seguir (BOLFARINE; BUSSAB, 2005),

$$n_0 = \frac{\sigma^2}{D}, \text{ em que } D = \frac{B^2}{Z^{\alpha/2}{}^2}$$

σ^2 é a variância de Y na população e z_α é um valor obtido a partir da tabela da distribuição Normal Padrão e representa o nível de confiança. Para 95% de confiança, $z_{\alpha/2} = 1,96$. Tanto B quanto α podem ser escolhidos pelo pesquisador. Para o Saeb o erro máximo desejado $B = 4$ pontos na escala (250, 50). Como geralmente σ^2 é desconhecido, ele pode ser estimado a partir de pesquisas realizadas no passado para a mesma população alvo ou de pesquisas piloto.

Para planos amostrais mais complexos, as equações para determinação dos tamanhos amostrais (n) são aprimoradas a partir do tamanho de amostra calculado para AAS com a consideração do Efeito do Plano Amostral (EPA) de forma que $n = n_0 * EPA$.

Para Amostragem por conglomerados, o EPA é dado por (BOLFARINE e BUSSAB, 2005)

$$EPA \cong 1 + (\bar{M} - 1)\rho_{int},$$

em que \bar{M} = tamanho médio dos conglomerados e ρ_{int} = Coeficiente de correlação intraclasse (ou intraconglomerado). É importante destacar que o ρ_{int} é “uma medida de homogeneidade ou similaridade entre os elementos pertencentes a determinados grupos” (SILVA; MOURA, 1990). Portanto, quanto menor o coeficiente de correlação intraclasse, mais heterogêneos são os conglomerados e mais eficiente é o uso da AC. Assim,

$$\rho_{int} = \frac{\sigma_{te}^2 + \sigma_e^2}{\sigma_a^2 + \sigma_{te}^2 + \sigma_e^2},$$

em que σ_{te}^2 = variância entre as turmas dentro da escola, σ_e^2 = variância entre as escolas dentro da UF e σ_a^2 = variância entre os alunos. De tal forma que o numerador representa uma variância parcial proveniente da influência da escola e da turma e o denominador representa a variância total que inclui também os efeitos referentes aos alunos. Dessa forma, a correlação intraclasse representa a proporção da variância total que pode ser atribuída à influência da escola e da turma em que o aluno está inserido (SKRONDAL; RABE-HESKETH, 2004).

Silva e Moura (1990) relatam que o coeficiente de correlação intraclasse surgiu no contexto de estudos relacionados à avaliação de similaridade fraternal, passando pela descoberta de que a variância entre médias de diferentes classes pode ser separada, ou isolada da variância total e chegando à conclusão de que o ρ_{int} pode ser expresso como uma razão de duas variâncias. Apesar de ter surgido em meio a estudos de similaridade, o uso do coeficiente de correlação intraclasse se expandiu para diversos outros contextos, ganhando grande relevância para o planejamento de experimentos e para a Amostragem (SILVA; MOURA, 1990). No contexto de conglomeração, o coeficiente de correlação intraclasse (ρ_{int}) representa a uniformidade da informação estudada dentro dos conglomerados de forma que quanto menor for o coeficiente, mais diversificado é aquele grupo (SILVA; MOURA, 1990). Essa medida

ajuda a entender quão adequada é a AC para uma população, pois quanto mais diversificados os conglomerados, melhores são os resultados obtidos com esse método.

Como medida de variabilidade da população estudada para o cálculo da correlação intraclasse, são utilizadas variâncias estimadas considerando um modelo hierárquico de 3 níveis, nulo (ou seja, sem covariáveis). A estimação dos coeficientes se dá pelo método de máxima verossimilhança restrita, adequado à estimação de traços latentes em que se espera um viés negativo na estimação das variâncias, pois prevê um fator de correção do viés esperado (SKRONDAL e RABE-HESKETH, 2004). Também são utilizadas as variâncias observadas entre os alunos de cada UF.

Desde a edição de 1995 do Saeb, foi reconhecido o potencial dos modelos hierárquicos lineares para modelar a variância em amostras complexas como a do Saeb (FLETCHER, 2006). Os modelos hierárquicos, também chamados de multiníveis, “foram desenvolvidos para analisar dados que possuem uma estrutura hierárquica, eles levam em conta a variabilidade dos dados existente dentro de cada nível hierárquico e entre os níveis hierárquicos” (PINHEIRO, 2005, p.1). São aplicados em diversas áreas de conhecimento, entre elas a educação, em que o agrupamento natural dos alunos em turmas e escolas formam os níveis de análise.

Pinheiro (2005) explica que a análise multinível permite verificar simultaneamente o efeito que as variáveis do nível indivíduo e do nível grupo exercem sobre a variável que está sendo estudada. Por exemplo, no contexto educacional, um modelo linear hierárquico pretende identificar que influência as competências individuais do aluno e as características da escola exercem no rendimento obtido pelo aluno.

A estimação dos parâmetros do modelo hierárquico é feita pelo método de máxima verossimilhança restrita. Segundo o dicionário, verossimilhança é a qualidade daquilo que se assemelha com o verdadeiro (FERREIRA, 2004). Assim, uma amostra verossímil é aquela que fornece informações suficientes para o cálculo de estimativas que mais se aproximem do valor real de um parâmetro populacional desconhecido (BUSSAB; MORETIN, 2004).

O princípio da verossimilhança afirma que para melhor representação de um parâmetro populacional desconhecido, o ideal é escolher o valor que torna mais provável a seleção de uma amostra específica já observada (BUSSAB; MORETIN, 2004). O método de máxima verossimilhança é bastante popular para a estimação de parâmetros na estatística, pois decorre desse princípio, adotando como estimativas dos parâmetros os valores que maximizam as chances de uma determinada amostra ser selecionada. Assim, a função de verossimilhança fornece um bom estimador para o verdadeiro valor do parâmetro estudado (BUSSAB; MORETIN, 2004; CASELLA; BERGER, 2016).

O cálculo das estimativas por máxima verossimilhança se utiliza de um processo iterativo iniciado com valores fornecidos para os parâmetros estimados que são aprimorados conforme os cálculos se repetem. As iterações prosseguem até que seja identificada convergência desses valores, finalizando o processo quando se verifica diferença próxima de zero entre os dois últimos valores observados (PINHEIRO, 2005).

Concluídos os cálculos das variâncias e dos coeficientes de correlação intraclasse, todos os elementos para o cálculo do tamanho da amostra estão disponíveis. Assim, dando prosseguimento ao processo após o cálculo do tamanho da amostra, por UF, foi efetuada a distribuição dessa amostra entre os estratos, considerando uma alocação proporcional ao tamanho de cada estrato. Para o recorte utilizado na simulação, foram identificados 216 possíveis estratos, formados a partir da combinação dos elementos que os constituem – 26 UFs e o Distrito Federal, duas dependências administrativas (estadual e municipal), duas localizações (urbana, rural) e duas áreas (capital, interior). Excluindo os estratos que não apresentaram nenhuma escola, o universo de referência foi distribuído em 171 estratos.

Para o cálculo do tamanho da amostra em número de alunos por UF foi utilizada a fórmula $n = n_0 * EPA$, apresentada no início desta seção. Esse cálculo foi repetido considerando a variabilidade dos resultados das duas áreas de conhecimento avaliadas e adotou-se como n o maior valor encontrado.

Em seguida o valor de n foi dividido pelo número médio de alunos por turma de cada UF, que representa o tamanho médio dos conglomerados (\bar{M}). Dessa forma, o tamanho da amostra em número de turmas é dado por $n_t = (n/\bar{M})$. Este valor é arredondado para cima, isto é, para o menor número inteiro maior ou igual ao valor inicialmente calculado. O processo original do Saeb prevê o acréscimo de uma folga para cobrir eventuais perdas da amostra, mas como na simulação realizada não haveria perdas esse acréscimo foi suprimido. Como a população de referência desta simulação é o Saeb, e inclui apenas escolas estaduais e municipais, 100% do tamanho da amostra calculado foi utilizado. As quantidades calculadas para a obtenção do tamanho da amostra estão apresentadas na Tabela 1.

Tabela 1 - Tamanho da amostra em número de alunos e turmas

UF	n_{LP}	n_{MT}	n	\bar{M}	n_t
BR	68.392	69.719	73.308	22	3.324
RO	2.265	2.399	2.399	23	105
AC	3.026	3.455	3.455	24	143
AM	5.285	4.990	5.285	25	214
RR	2.723	3.208	3.208	22	147
PA	2.630	2.049	2.630	22	120
AP	2.016	1.527	2.016	21	94
TO	3.388	3.449	3.449	22	159
MA	2.222	1.730	2.222	21	105
PI	3.621	3.880	3.880	21	186
CE	3.285	4.800	4.800	23	212
RN	2.537	1.962	2.537	20	125
PB	2.212	1.946	2.212	19	114
PE	2.539	2.511	2.539	23	111
AL	3.717	3.888	3.888	24	160
SE	2.097	1.667	2.097	21	100
BA	2.349	1.960	2.349	20	117
MG	2.880	2.895	2.895	22	135
ES	1.721	1.742	1.742	21	83
RJ	2.229	2.356	2.356	24	97
SP	1.994	2.258	2.258	24	94
PR	2.000	2.556	2.556	22	116
SC	1.846	2.162	2.162	21	103
RS	1.835	1.916	1.916	18	106
MS	2.439	2.617	2.617	24	111
MT	1.907	1.921	1.921	22	90
GO	2.166	2.456	2.456	23	106
DF	1.464	1.419	1.464	21	71

Fonte: Elaborado pela autora

Seguindo o procedimento de cálculo do tamanho da amostra do Saeb, o tamanho da amostra em número de turmas foi distribuído entre os estratos de tamanho proporcionalmente ao número de turmas existente em cada um deles. Vale lembrar que se dá o nome de estrato de tamanho à combinação do estrato de interesse com o porte. Assim, o estrato de tamanho é formado por UF, dependência administrativa, localização, área e porte.

Após a alocação do número de turmas entre os estratos, calcula-se o número de escolas, levando-se em consideração que para escolas grandes são selecionadas duas turmas e para escolas pequenas seleciona-se uma turma. Assim, o tamanho da amostra em número de escolas é calculado por $n_e = n_{tp} + (n_{tg}/2)$, onde n_{tp} = número de turmas de escolas pequenas e n_{tg} = número de turmas de escolas grandes.

Em decorrência do processo de alocação e eventuais arredondamentos, o tamanho da amostra em número de turmas pode sofrer uma pequena variação. Também o número de alunos previstos inicialmente é atualizado em função desse ajuste. O número final de alunos previstos é dado pelo número de turmas multiplicado pelo número médio de alunos por turma. A Tabela 2 mostra os quantitativos finais esperados para a amostra.

Tabela 2 - Quantitativos da amostra esperada

Dependência Administrativa	Alunos	Turmas	Escolas
Estadual	24.051	1.089	846
Municipal	49.779	2.246	1.700
Total	73.830	3.335	2.546

Fonte: Elaborado pela autora

Após o cálculo do tamanho da amostra, para cada repetição da simulação foi aplicado o processo de Amostragem sequencial de Poisson, definido na sessão 3.2, a partir de diferentes números aleatórios gerados a cada rodada.

3.3.2 Estimação dos resultados

Uma das primeiras questões levantadas quando são analisados dados provenientes de amostras finitas é como considerar o desenho amostral nessa análise. A estimação das medidas de interesse é um passo importante do processo, havendo na literatura da Amostragem diferentes formas de se chegar a essas medidas. Apesar dessa diversidade metodológica, na prática prevalece o uso dos estimadores ponderados em que um peso amostral específico é associado a cada unidade, especialmente quando se trata de planos amostrais complexos que incluem estratificação (BINDER; ROBERTS, 2003; PESSOA; SILVA, 1998). A metodologia adotada para a análise dos dados do Saeb, ou seja, para a estimação dos resultados, inicia-se com o cálculo dos pesos amostrais para expansão dos resultados obtidos por Amostragem para a população de referência (INEP, 2018a).

Um método muito usado para ponderação dos elementos de uma amostra probabilística, como é o caso do Saeb, é o estimador não viciado de Horvitz-Thompson, também conhecido como estimador π -ponderado, que admite como peso amostral o valor inverso da probabilidade de seleção (π_i) (CHAMBERS, 2003; PESSOA; SILVA, 1998). Além do peso usual, pode-se

aplicar um esquema de calibração³ a fim de tratar os casos de não resposta, considerando que a perda é aleatória e fazendo com que as estimativas amostrais ponderadas coincidam com parâmetros populacionais conhecidos (LITTLE, 2003).

Trazendo o processo de ponderação para a realidade do Saeb, tem-se que cada aluno que respondeu às provas recebe um peso. O valor do peso de cada aluno é resultado do inverso da sua probabilidade de compor a amostra multiplicado por fatores de calibração, que visam também ajustar para perdas naturais do processo de coleta de dados. A metodologia adotada para o Saeb resulta em pesos para os alunos cuja soma é igual ao número de alunos do universo de referência (INEP, 2018a). Neste estudo de simulação, o cálculo dos pesos para cada amostra selecionada foi realizado seguindo o processo descrito no Relatório de Amostragem da Aneb 2017, capítulo 3, seção 2 (INEP, 2018a).

Para o cálculo das medidas descritivas de interesse, como totais, médias e variâncias, com os dados obtidos por meio de uma amostra complexa, é aconselhado o uso dos pesos a fim de compensar os efeitos do plano amostral. Ignorar a etapa de ponderação pode resultar em estimativas viciadas, tornando-as pouco confiáveis (PESSOA; SILVA, 1998).

Para a avaliação da qualidade das estimativas produzidas a partir de uma pesquisa amostral, é fundamental a etapa de estimação da precisão das estimativas. Para tanto, recomenda-se a adoção de estimadores de erro padrão que possam ser calculados por pacotes computacionais especializados. Esses estimadores são uma aproximação obtida levando em consideração os detalhes do plano amostral e o peso de cada unidade da amostra. Essa aproximação é suficiente para as análises necessárias e pode ser facilmente implementada, evitando uma programação sob medida (PESSOA; SILVA, 1998).

Para o cálculo das estimativas das médias de proficiência e dos erros padrão obtidos a partir de cada amostra selecionada para este estudo, foi utilizada a versão 3.36 do pacote Survey do Ambiente de programação R, publicada em abril de 2019. Intitulado como *Analysis of Complex Survey Samples*, que foi desenvolvido por Thomas Lumley para analisar dados provenientes de amostras complexas. Para que isso seja possível, o pacote atende a dois propósitos principais: a) vincular as características do plano amostral aos dados coletados, para que os ajustes necessários à análise sejam realizados de forma confiável e automática; b) disponibilizar estimativas válidas para as variâncias das estatísticas calculadas utilizando este pacote (LUMLEY, 2004).

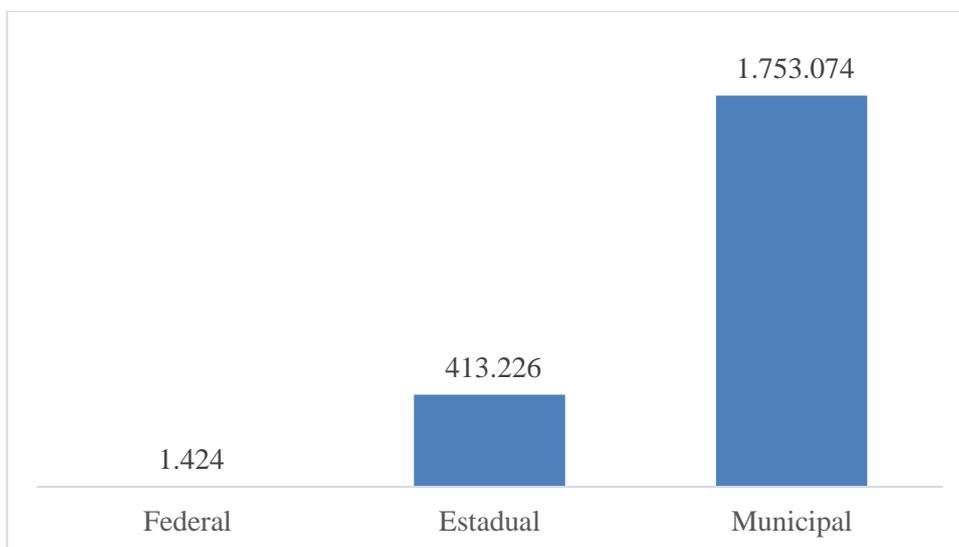
³ O esquema de calibração dá-se pela multiplicação de fatores ao peso usual, sob forma de transformações lineares, a fim de que a soma dos pesos obtidos após a calibração coincida com as contagens populacionais, obtidas do Censo Escolar.

3.3.3 Análise dos resultados

A simulação baseou-se na repetição da seleção da amostra, resultando em 2.000 amostras diferentes, selecionadas a partir da metodologia utilizada para a Amostragem do Saeb. A principal diferença entre as amostras selecionadas e a amostra do Saeb é o público-alvo, uma vez que para esta simulação foi considerado o universo de escolas estaduais e municipais, enquanto o Saeb também abrange as escolas particulares e as federais. As escolas particulares foram excluídas da simulação por não haver resultado de aplicação censitária para comparação. A rede federal, apesar de contar com resultados censitários, não apresenta número de matrículas grande o suficiente para justificar a seleção amostral, conforme demonstra o Gráfico 1 e, por essa razão, não foi incluída na simulação.

Assim, nesta seção, quando se falar em Saeb 2017 como referência para este estudo, está sendo considerado apenas o recorte dos alunos válidos para o cálculo dos resultados das escolas estaduais e municipais. Para o aluno ser considerado válido, é requerida a consistência com o Censo Escolar ($IN_SITUACAO_CENSO = 1$) e a proficiência individual calculada ($IN_PROFICIENCIA = 1$). Como universo de referência para o cálculo do tamanho da amostra e dos pesos para expansão dos resultados, consideraram-se as quantidades obtidas a partir dos microdados do Saeb 2017. Da mesma forma, as escolas participantes do Saeb 2017 constituíram o cadastro para o sorteio das escolas e turmas. Assim, para realização da simulação foram utilizados dados contidos nas bases TS_ALUNO_5EF (INEP, 2018c) e Análise - Tabela da lista das escolas – Detalhado (INEP, 2019b).

Gráfico 1 - Número de alunos válidos por dependência administrativa



Fonte: Elaborado pela autora.

A Tabela 3 mostra o comparativo das quantidades de escolas, turmas e alunos, envolvidos na aplicação do Saeb 2017, independente de sua validade, para as redes estadual e municipal. Mostra também o tamanho esperado para a amostra e a média das quantidades observadas nas amostras desta simulação.

Tabela 3 - Comparativo de quantidades entre censo e amostra

Quantidade	Saeb	Amostra Esperada	Média realizada
Alunos	2.594.335	73.830	73.855
Turmas	98.688	3.335	3.335
Escolas	48.165	2.546	2.546

Fonte: Elaborado pela autora.

Destaca-se que a diferença entre o número de alunos esperado e o realizado se dá em função da variação na quantidade de alunos em cada turma. Foram selecionados escolas, alunos e turmas, considerando os elementos válidos constantes nos microdados do Saeb 2017, que sejam de escolas estaduais ou municipais (ID_DEPENDENCIA_ADM in (2,3)). Relembrando que os critérios de validade da participação do aluno são a consistência com o Censo Escolar (IN_SITUACAO_CENSO = 1) e a proficiência calculada (IN_PROFICIENCIA = 1).

Observa-se ainda que as quantidades amostrais são muito inferiores às quantidades envolvidas na aplicação do Saeb, representando aproximadamente 2,8% do número de alunos, 3,4% do número de turmas e 5,3% do número de escolas.

Para maior clareza nos conceitos que serão trabalhados a seguir, a Tabela 4 apresenta algumas notações adotadas nos próximos parágrafos.

Tabela 4 - Descrição de notações

Símbolo	Descrição
θ	Parâmetro média da proficiência
θ_{LP}	Parâmetro média da proficiência em Língua Portuguesa
θ_{MT}	Parâmetro média da proficiência em Matemática
d	Número de repetições considerado
$i = 1, 2, \dots, d$	Indexador das repetições na simulação
$\hat{\theta}$	Estimador de θ
$\hat{\theta}_i$	Estimador de θ aplicado à i-ésima repetição

$\bar{\theta}$	A média de $\hat{\theta}_i$ entre as repetições (valor esperado estimado via simulação)
----------------	---

Fonte: Elaborado pela autora

Para avaliação das metodologias testadas, foram adotados como medidas de desempenho o viés do estimador, a sua variância (Var), e o seu erro quadrático médio (EQM), nesta dissertação, estimados via simulação. O viés estatístico é uma característica de uma técnica estatística ou de seus resultados, por meio da qual o valor esperado dos resultados difere do verdadeiro parâmetro quantitativo subjacente sendo estimado. O viés mede quanto o valor esperado do estimador $\hat{\theta}$ é diferente de θ , em média, e representa uma medida de acurácia. A variância mede o quanto cada realização do estimador $\hat{\theta}$ se aproximou do seu valor esperado $\bar{\theta}$, representando uma medida de precisão. O EQM é a soma do quadrado do viés com a variância de $\hat{\theta}$, agregando medidas de acurácia e precisão em um único indicador. Quanto menores forem estes valores, podemos considerar que melhor é o desempenho do estimador utilizado (MORRIS; WHITE; CROWTHER, 2019; BUSSAB; MORETIN, 2004; CASELLA; BERGER, 2010).

Foi ainda calculada a raiz quadrada do erro quadrático médio em termos relativos ao valor do parâmetro (RREQM). Ele mede o quanto o EQM representa em porcentagem do valor do parâmetro após neutralizado o efeito quadrático embutido no EQM. Dessa forma, quanto menor esta medida, maior a qualidade dos resultados.

Vale destacar que os valores adotados como referência do valor real (parâmetro) da proficiência média para a população estudada é a média aritmética das proficiências dos estudantes de escolas estaduais e municipais apresentados nos microdados do Saeb 2017. Essa medida foi adotada por estar se considerando o referido universo dos microdados como representação da população estudada.

Para o cálculo das medidas apresentadas nas Tabelas 5 e 6 foram utilizadas as fórmulas a seguir (MORRIS; WHITE; CROWTHER, 2019; BUSSAB; MORETIN, 2004):

$$\bar{\theta} = \frac{1}{d} \sum_{i=1}^d \hat{\theta}_i \quad \text{Viés}(\hat{\theta}) = \frac{1}{d} \sum_{i=1}^d \hat{\theta}_i - \theta \quad \text{Var}(\hat{\theta}) = \frac{1}{d} \sum_{i=1}^d (\hat{\theta}_i - \bar{\theta})^2$$

$$\text{EQM}(\hat{\theta}) = \frac{1}{d} \sum_{i=1}^d (\hat{\theta}_i - \theta)^2 = \text{Viés}(\hat{\theta})^2 + \text{Var}(\hat{\theta}) \quad \text{RREQM}(\hat{\theta}) = \frac{\sqrt{\text{EQM}(\hat{\theta})}}{\theta} 100$$

Pode-se observar nas Tabelas 5 e 6 que os valores estimados para a média se aproximaram da média obtida nos microdados, adotada como parâmetro para a proficiência média de ambas as disciplinas. O viés calculado não foi exatamente nulo como desejado (MORRIS; WHITE; CROWTHER, 2019), mas se aproximou bastante de zero, variando entre -0,29 e 0,51.

Em relação ao EQM, pode-se afirmar que a simulação realizada demonstrou bom desempenho em acurácia e precisão, uma vez que os valores observados são pequenos em relação ao valor da média. Os valores do RREQM corroboram esse entendimento apresentando valores sempre muito próximos de 1% do valor do parâmetro.

Tabela 5 - Medidas de desempenho para Língua Portuguesa

UF	θ_{LP}	$\bar{\theta}_{LP}$	Viés ($\hat{\theta}_{LP}$)	Var ($\hat{\theta}_{LP}$)	EQM ($\hat{\theta}_{LP}$)	RREQM ($\hat{\theta}_{LP}$)
BR	210,03	210,01	-0,03	0,45	0,45	0,32%
RO	210,47	210,57	0,10	4,45	4,46	1,00%
AC	214,62	214,59	-0,03	1,86	1,86	0,64%
AM	200,34	200,26	-0,08	2,17	2,17	0,74%
RR	200,92	201,42	0,51	0,84	1,10	0,52%
PA	185,04	185,02	-0,03	6,40	6,40	1,37%
AP	181,92	181,63	-0,29	3,75	3,83	1,08%
TO	200,97	200,96	0,00	3,06	3,06	0,87%
MA	178,59	178,61	0,02	7,03	7,03	1,48%
PI	194,56	194,56	0,00	2,93	2,93	0,88%
CE	218,04	218,07	0,03	3,94	3,94	0,91%
RN	185,93	185,89	-0,04	4,19	4,19	1,10%
PB	190,04	190,00	-0,04	5,40	5,40	1,22%
PE	193,17	193,08	-0,08	5,59	5,60	1,22%
AL	189,82	189,82	0,00	5,76	5,76	1,26%
SE	182,48	182,41	-0,08	4,07	4,07	1,11%
BA	191,24	191,25	0,00	5,71	5,71	1,25%
MG	223,61	223,57	-0,04	4,96	4,96	1,00%
ES	214,52	214,43	-0,09	5,96	5,97	1,14%
RJ	209,93	209,99	0,06	4,78	4,78	1,04%
SP	226,19	226,13	-0,06	5,68	5,68	1,05%
PR	225,86	225,89	0,03	4,10	4,11	0,90%
SC	225,32	225,31	-0,01	4,91	4,91	0,98%
RS	214,50	214,34	-0,16	4,79	4,81	1,02%
MS	213,26	213,23	-0,03	3,45	3,45	0,87%
MT	205,99	205,94	-0,06	5,29	5,30	1,12%
GO	216,33	216,41	0,07	7,08	7,09	1,23%
DF	220,39	220,44	0,04	5,18	5,18	1,03%

Fonte: Elaborado pela autora

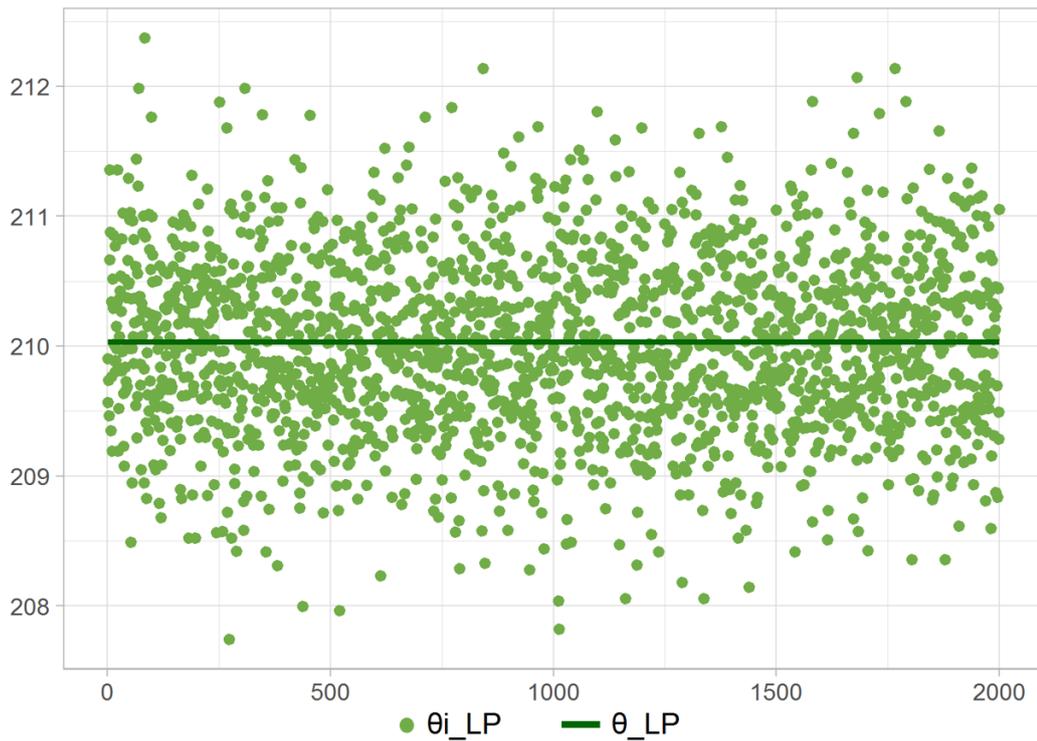
Tabela 6 - Medidas de desempenho para Matemática

UF	θ_{MT}	$\bar{\theta}_{MT}$	Viés ($\hat{\theta}_{MT}$)	Var ($\hat{\theta}_{MT}$)	EQM ($\hat{\theta}_{MT}$)	RREQM ($\hat{\theta}_{MT}$)
BR	219,50	219,47	-0,04	0,50	0,50	0,32%
RO	220,57	220,64	0,07	4,99	5,00	1,01%
AC	226,70	226,66	-0,03	2,18	2,18	0,65%
AM	209,52	209,41	-0,11	2,35	2,37	0,73%
RR	215,73	216,20	0,47	0,90	1,12	0,49%
PA	192,82	192,83	0,00	5,42	5,42	1,21%
AP	188,88	188,63	-0,25	3,22	3,28	0,96%
TO	211,16	211,14	-0,02	3,15	3,15	0,84%
MA	186,58	186,54	-0,03	5,60	5,60	1,27%
PI	204,76	204,77	0,01	2,90	2,90	0,83%
CE	226,11	226,11	0,00	5,58	5,58	1,04%
RN	193,59	193,61	0,02	3,32	3,32	0,94%
PB	199,97	199,91	-0,06	5,02	5,02	1,12%
PE	202,69	202,60	-0,09	5,82	5,83	1,19%
AL	201,69	201,71	0,02	6,61	6,61	1,27%
SE	193,42	193,34	-0,08	3,35	3,36	0,95%
BA	199,52	199,52	0,01	5,07	5,07	1,13%
MG	231,76	231,71	-0,05	5,12	5,12	0,98%
ES	223,56	223,50	-0,06	6,03	6,03	1,10%
RJ	217,86	217,87	0,02	4,91	4,91	1,02%
SP	237,83	237,72	-0,11	6,89	6,90	1,10%
PR	239,14	239,17	0,03	5,29	5,29	0,96%
SC	234,34	234,32	-0,02	5,89	5,89	1,04%
RS	223,48	223,36	-0,12	5,02	5,03	1,00%
MS	220,88	220,85	-0,03	4,05	4,05	0,91%
MT	214,34	214,30	-0,04	5,43	5,44	1,09%
GO	222,79	222,88	0,09	7,68	7,69	1,24%
DF	229,20	229,25	0,04	4,93	4,93	0,97%

Fonte: Elaborado pela autora

Os Gráficos 2 e 5 mostram a dispersão das médias de proficiência em Língua Portuguesa e em Matemática, para o Brasil, a partir das 2.000 amostras selecionadas. Os Gráficos 3, 4, 6 e 7 mostram os histogramas da distribuição das médias de proficiência em Língua Portuguesa e Matemática, para o Brasil e para cada uma das UFs.

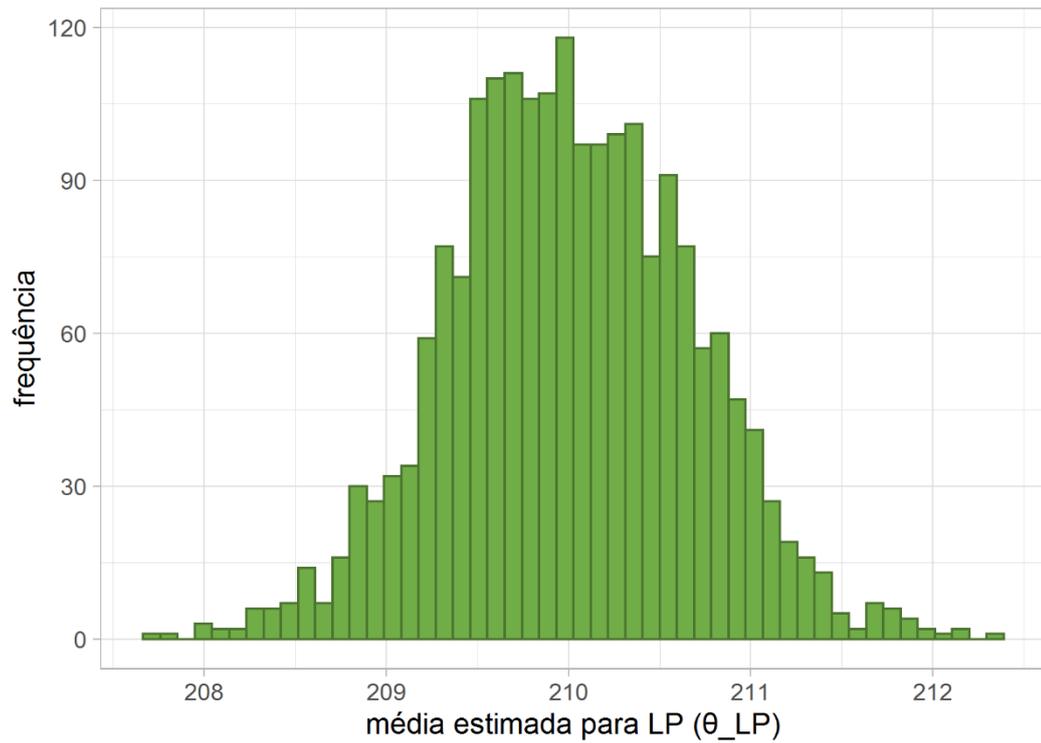
O Gráfico 2 mostra a distribuição dos valores de proficiência média em Língua Portuguesa dos estudantes de 5º ano do ensino fundamental, das redes de ensino estadual e municipal, estimados a partir das 2.000 amostras sorteadas na simulação. A linha horizontal presente no Gráfico representa o valor de referência para a proficiência média real calculado a partir dos resultados dos alunos participantes do Saeb 2017.

Gráfico 2 - Dispersão dos valores de $\hat{\theta}_{iLP}$ para o Brasil

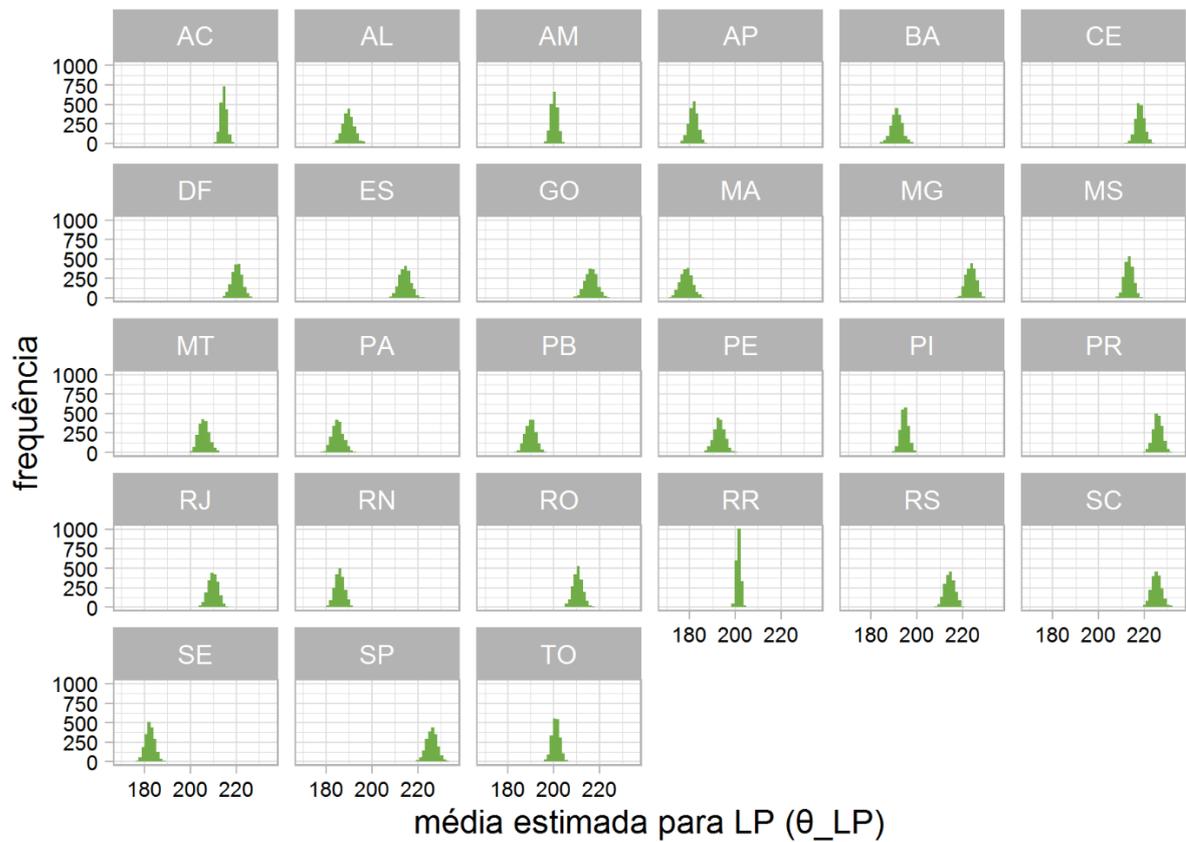
Fonte: Elaborado pela autora

Os pontos distribuem-se de forma equilibrada em torno da linha horizontal, apresentando sempre valores próximos ao valor de referência. Essa distribuição mostra-se como uma representação visual que complementa o entendimento dos dados resumidos nas Tabelas 5 e 6, corroborando a afirmação de que os dados amostrais oferecem aproximações confiáveis para a proficiência média em Língua Portuguesa.

O Gráfico 3 mostra que as 2.000 proficiências médias de Língua Portuguesa dos alunos do Brasil, estimadas a partir das amostras selecionadas na simulação, se distribuem conforme o esperado, apresentando gráfico semelhante ao da distribuição normal.

Gráfico 3 - Distribuição dos valores de $\hat{\theta}_{ILP}$ para o Brasil

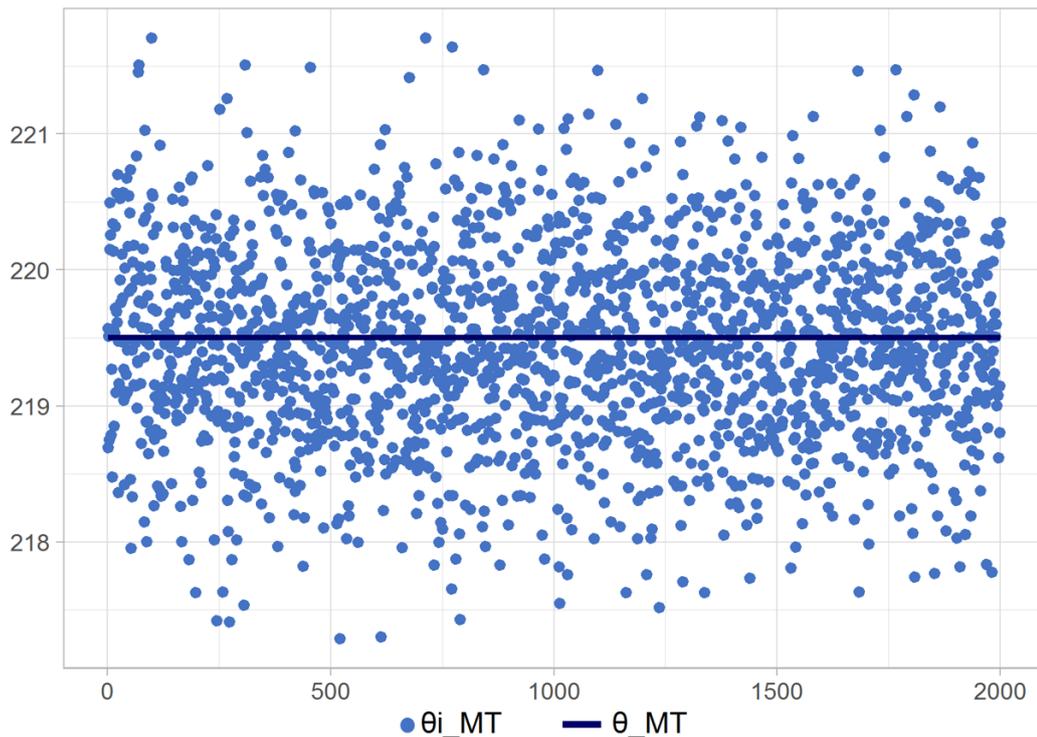
Fonte: Elaborado pela autora

Gráfico 4 - Distribuição dos valores de $\hat{\theta}_{ILP}$ por UF

Fonte: Elaborado pela autora

O Gráfico 5, de maneira equivalente ao Gráfico 2, mostra a distribuição dos valores de proficiência média em Matemática dos estudantes de 5º ano do ensino fundamental, das redes de ensino estadual e municipal, estimados a partir das 2.000 amostras sorteadas na simulação. A linha horizontal presente no Gráfico representa o valor de referência para a proficiência média real calculado a partir dos resultados dos alunos participantes do Saeb 2017.

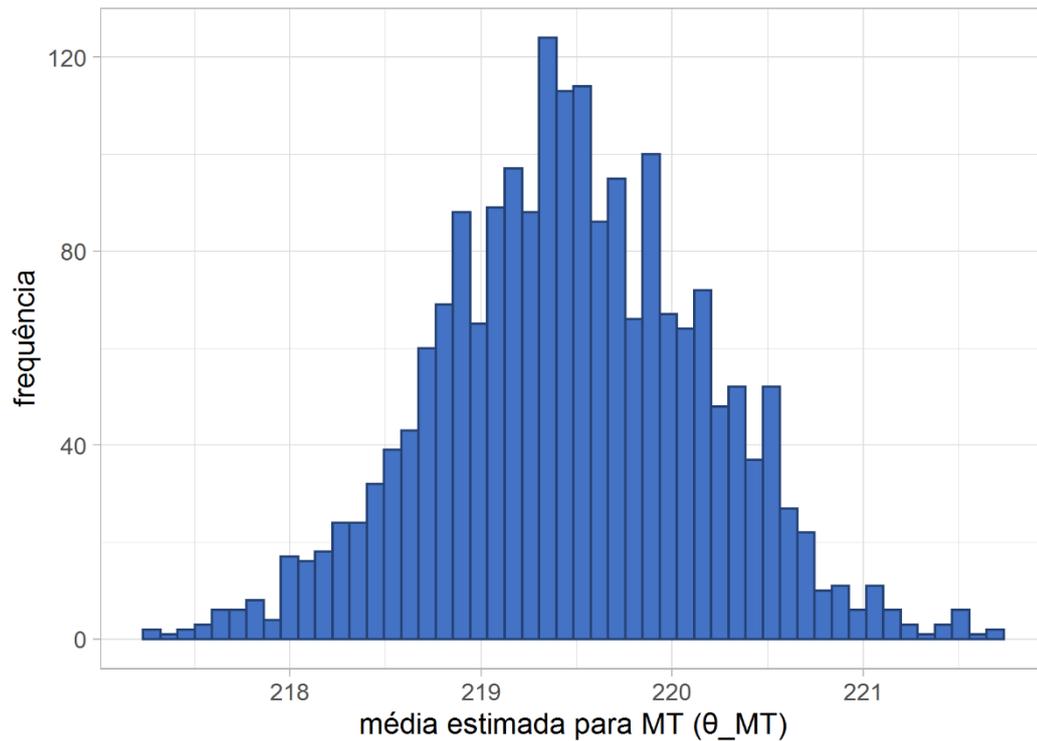
Gráfico 5 - Dispersão dos valores de $\hat{\theta}_{iMT}$ para o Brasil



Fonte: Elaborado pela autora

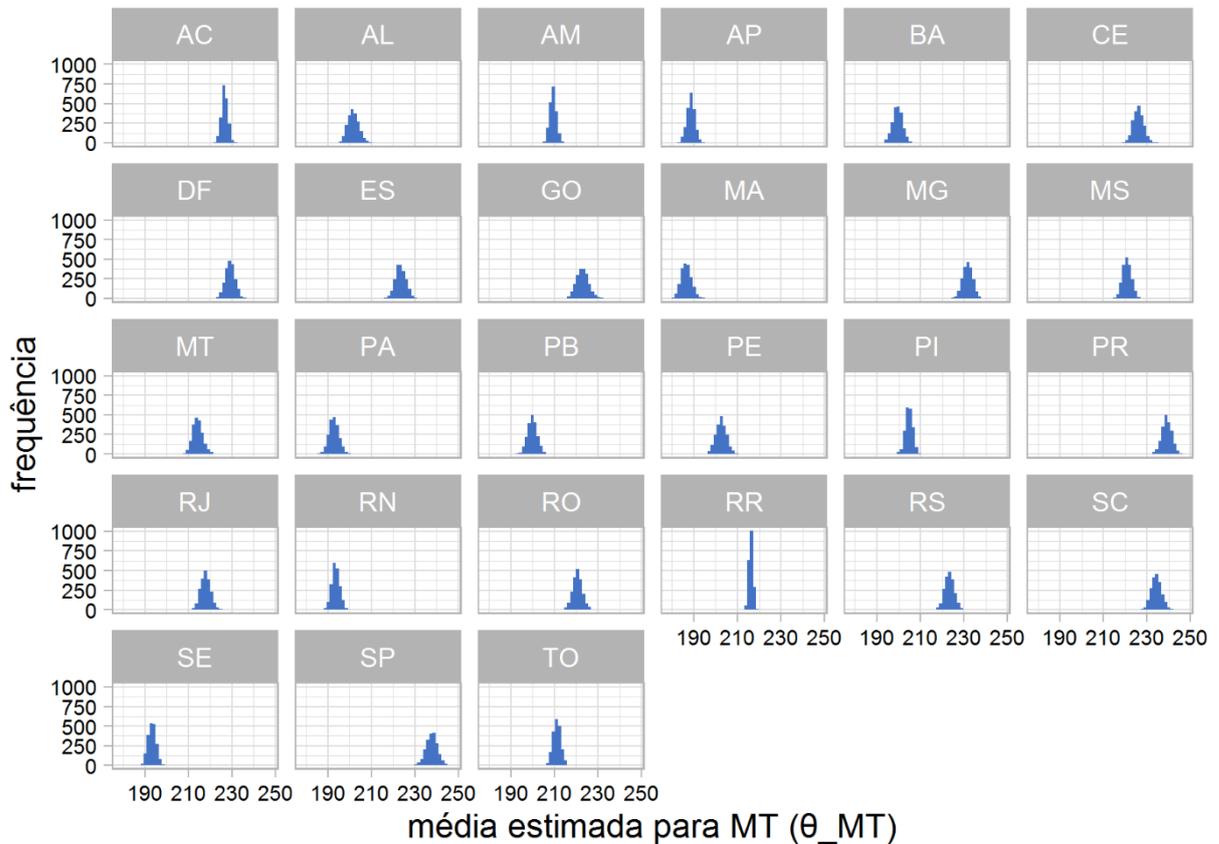
Assim como os resultados para Língua Portuguesa, os resultados para Matemática se distribuem de forma equilibrada em torno da linha horizontal, apresentando sempre valores próximos ao valor de referência. Tal constatação reforça que dados amostrais oferecem aproximações confiáveis para a proficiência média em Matemática da mesma forma que para Língua Portuguesa.

As proficiências médias em Matemática dos alunos brasileiros, estimadas a partir das amostras selecionadas na simulação, também se distribuem conforme o esperado, em formato semelhante ao da distribuição normal, conforme se vê no Gráfico 6.

Gráfico 6 - Distribuição dos valores de $\hat{\theta}_{iMT}$ para o Brasil

Fonte: Elaborado pela autora

O que é observado para o Brasil no Gráfico 6, também se aplica às UFs, de acordo com o que se apresenta no Gráfico 7. Assim, conclui-se que proficiências médias em Matemática dos alunos de cada UF, estimadas a partir das amostras selecionadas na simulação, se distribuem conforme o esperado. No contexto dos resultados de Matemática, Roraima (RR) mais uma vez de destaca em relação à concentração dos dados indicando que a variabilidade da proficiência dos alunos é menor.

Gráfico 7 - Distribuição dos valores de $\hat{\theta}_{iMT}$ por UF

Fonte: Elaborado pela autora

Nos gráficos de dispersão (Gráficos 2 e 5) é possível perceber a distribuição da nuvem de pontos em torno do valor de $\hat{\theta}$, decorrente do viés próximo de zero. Já os histogramas mostram como as médias calculadas a partir de cada uma das 2.000 amostras se distribuem da forma esperada, assemelhando-se ao gráfico da distribuição normal. Tais observações são válidas tanto para Língua Portuguesa quanto para Matemática, fundamentando a conclusão de que os resultados obtidos por meio de uma amostra são suficientes e confiáveis para estimar os resultados populacionais.

Outra medida que pode ser estudada para avaliar a qualidade da medida obtida por simulação obtida é a cobertura do intervalo de confiança. Esta é também uma importante medida do ponto de vista da estatística frequentista e indica a probabilidade de o intervalo de confiança conter o valor de θ (MORRIS; WHITE; CROWTHER, 2019).

Para obter essa medida, calculou-se o intervalo de confiança para a média Brasil e a para as médias por UF, para Língua Portuguesa e Matemática em cada uma das 2.000 repetições.

Cada intervalo de confiança foi calculado a partir da fórmula a seguir, onde EP_i = erro padrão estimado na i -ésima repetição e $\alpha = 5\%$ de tal forma que $Z_{\frac{\alpha}{2}} = 1,96$.

$$IC_{(1-\alpha)\%} = (\hat{\theta}_i - Z_{\frac{\alpha}{2}}(EP_i) ; \hat{\theta}_i + Z_{\frac{\alpha}{2}}(EP_i))$$

Em seguida, a cobertura foi obtida a partir do percentual de intervalos de confiança calculados que contém o valor de θ . A Tabela 7 mostra os resultados de cobertura dos intervalos de confiança e a margem de erro média (ME) observados para esta simulação. Chama-se margem de erro o termo $Z_{\frac{\alpha}{2}}(EP_i)$ usado na construção do intervalo de confiança.

Tabela 7 - Cobertura dos intervalos de confiança e Margem de Erro

UF	Cobertura LP	Cobertura MT	$ME(\hat{\theta}_{LP})$	$ME(\hat{\theta}_{MT})$
BR	96,2%	95,7%	1,42	1,49
RO	97,0%	97,2%	4,73	4,98
AC	99,5%	99,5%	4,07	4,34
AM	98,0%	97,9%	3,48	3,59
RR	99,7%	99,9%	3,38	3,41
PA	95,3%	95,2%	5,16	4,79
AP	96,2%	96,0%	4,32	4,03
TO	98,2%	98,2%	4,37	4,34
MA	94,3%	95,0%	5,31	4,79
PI	97,1%	96,7%	3,79	3,78
CE	95,4%	95,4%	4,15	5,02
RN	96,6%	97,1%	4,41	3,94
PB	95,9%	95,8%	4,83	4,69
PE	94,8%	94,7%	4,86	4,92
AL	96,4%	95,9%	5,19	5,56
SE	95,2%	95,6%	4,17	3,83
BA	94,2%	94,5%	4,82	4,51
MG	96,7%	96,5%	4,62	4,72
ES	95,0%	95,1%	4,97	4,99
RJ	96,0%	96,1%	4,53	4,64
SP	94,1%	94,0%	4,71	5,13
PR	95,5%	96,6%	4,20	4,85
SC	96,0%	96,0%	4,64	5,08
RS	95,6%	95,7%	4,41	4,55
MS	96,6%	96,4%	4,05	4,35
MT	95,7%	95,7%	4,75	4,82
GO	94,9%	95,6%	5,40	5,80
DF	95,7%	97,5%	4,91	5,00

Fonte: Elaborado pela autora

Destaca-se que a cobertura observada chegou aos 95% esperados para a maioria das UFs, bem como para o resultado nacional. Em média, a margem de erro ficou próxima ao erro máximo esperado de 4 pontos, variando entre 1,42 e 5,80, mas extrapolou o valor máximo esperado em 24 UFs para Língua Portuguesa e 22 UFs para Matemática.

3.3.4 Metodologias alternativas

Apesar dos bons resultados verificados na simulação que replica a metodologia amostral atual do Saeb, é possível identificar algumas fragilidades. A estimação das variâncias por meio do modelo hierárquico de três níveis exige informações nos níveis de aluno, turma e escola. Em função de a amostra coletar dados de apenas algumas turmas em cada escola, não há disponibilidade de dados suficientes a respeito das turmas das escolas privadas. Para contornar essa dificuldade, assumiu-se que a variabilidade entre os alunos das escolas privadas é equivalente à dos alunos das escolas públicas. Decisão semelhante foi tomada em relação aos alunos do ensino médio, quando essa etapa era avaliada de forma amostral para alunos das redes pública e particular: assumia-se que a variabilidade da proficiência dos alunos do ensino médio era equivalente à variabilidade dos alunos do 9º ano do ensino fundamental.

Para evitar esse pressuposto, o tamanho da amostra pode ser definido por meio de outras metodologias, visto que a literatura apresenta diversas formas de estimação do coeficiente de correlação intraclasse usado para o cálculo do EPA. No caso de conglomerados selecionados com probabilidades proporcionais ao tamanho, como é o exemplo do Saeb, Silva e Moura (1990) recomendam que a correlação intraclasse seja calculada com a seguinte fórmula.

$$\rho_c = \frac{\sigma_e^2 - \frac{1}{N} \sum_{i=1}^M \frac{N_i}{N_i - 1} \sigma_i^2}{\sigma^2} = 1 - \frac{\frac{1}{N} \sum_{i=1}^M N_i \sigma_i^2 (1 + \frac{1}{N_i - 1})}{\sigma^2},$$

em que, σ_e^2 é a variância entre conglomerados, representada pela variância entre as escolas da UF; σ_i^2 é a variância da característica estudada dentro do i -ésimo conglomerado, representada pela variância das proficiências dos alunos dentro de cada escola; σ^2 é a variância na população, representada pela variância das proficiências dos alunos da UF; N é o número de unidades da população, representado pelo número de alunos na UF; M é o número de conglomerados em que a população foi dividida, representado pelo número de escolas por UF; e N_i é o número de unidades no i -ésimo conglomerado, representado pelo número de alunos em determinadas escola.

Mediante o desconhecimento das variâncias populacionais, ρ_c pode ser estimado por

$$\hat{\rho}_c = \frac{s_e^2 - \frac{1}{N} \sum_{i=1}^M s_i^2}{s_e^2 + \frac{1}{N} \sum_{i=1}^M (N_i - 1) s_i^2},$$

em que s_e^2 é a variância amostral entre os conglomerados e s_i^2 é a variância amostral dentro do i -ésimo conglomerado.

A fórmula para o cálculo do EPA recomendada por Silva e Moura (1990) é a mesma já adotada para o Saeb,

$$EPA = 1 + (\bar{M} - 1)\rho_{int},$$

em que \bar{M} é o número médio de unidades selecionadas em cada conglomerado, representado pelo número médio de alunos por turma e ρ_{int} pode ser substituído por ρ_c ou $\hat{\rho}_c$.

A fórmula do EPA foi desenhada considerando que o \bar{M} seria semelhante para todos os estratos. Entretanto, no caso do Saeb, esse número pode variar bastante em função do porte da escola, uma vez que em escolas menores se seleciona apenas uma turma, enquanto nas maiores duas turmas são sorteadas. Diante disso, o \bar{M} das escolas grandes é aproximadamente o dobro do das pequenas, tornando o efeito da estratificação mais intenso nos estratos das escolas de porte grande. Essa particularidade não foi considerada neste estudo, pois não houve tempo hábil para desenvolvimento de uma fórmula para o cálculo do EPA considerando essa especificidade da estratificação do Saeb, mas esse assunto pode ter tratado em estudos futuros.

Para verificar se o tamanho da amostra gerado com a correlação intraclasse recomendada por Silva e Moura (1990) promoveria melhores resultados, a simulação foi repetida considerando a fórmula de $\hat{\rho}_c$. Entretanto, os resultados não apresentaram diferenças relevantes, como sintetizado na Tabela 8, que mostra os valores do EQM e da margem de erro obtidos na primeira simulação, usando ρ_{int} , e na segunda simulação, usando $\hat{\rho}_c$ no cálculo do tamanho da amostra. Os resultados completos obtidos nessa simulação estão apresentados no Apêndice A. Para facilitar a diferenciação, essa simulação será chamada de Simulação_ $\hat{\rho}_c$, enquanto a anterior será chamada de Simulação_ ρ_{int} .

Outra opção para estimar as variâncias a partir dos dados amostrais é adotar o modelo hierárquico, como no modelo atual, porém em dois níveis, que exigiria apenas dados sobre os alunos e as escolas. Para testar se tal proposta é válida foi realizado o cálculo das variâncias a partir das 2.000 amostras selecionadas na Simulação_ ρ_{int} e estimados os valores esperados do coeficiente de correlação intraclasse, do EPA e do tamanho amostral decorrente deles. Os resultados obtidos por essa metodologia serão identificados como $\rho_{int.2}$. Os resultados completos dessa simulação estão apresentados no Apêndice B.

Tabela 8 - Comparativo dos resultados da Simulação ρ_{int} e da Simulação $\hat{\rho}_c$.

UF	EQM ($\hat{\theta}_{LP}$)		EQM ($\hat{\theta}_{MT}$)		ME ($\hat{\theta}_{LP}$)		ME ($\hat{\theta}_{MT}$)	
	ρ_{int}	$\hat{\rho}_c$	ρ_{int}	$\hat{\rho}_c$	ρ_{int}	$\hat{\rho}_c$	ρ_{int}	$\hat{\rho}_c$
BR	0,45	0,52	0,50	0,56	1,42	1,50	1,49	1,58
RO	4,46	4,47	5,00	4,94	4,73	4,73	4,98	4,98
AC	1,86	1,88	2,18	2,22	4,07	4,11	4,34	4,39
AM	2,17	2,29	2,37	2,52	3,48	3,48	3,59	3,59
RR	1,10	1,11	1,12	1,11	3,38	3,39	3,41	3,42
PA	6,40	6,50	5,42	5,41	5,16	5,18	4,79	4,80
AP	3,83	3,22	3,28	2,70	4,32	4,12	4,03	3,83
TO	3,06	3,18	3,15	3,21	4,37	4,45	4,34	4,43
MA	7,03	7,39	5,60	6,03	5,31	5,54	4,79	5,00
PI	2,93	3,31	2,90	3,29	3,79	4,13	3,78	4,18
CE	3,94	3,99	5,58	5,76	4,15	4,18	5,02	5,05
RN	4,19	5,01	3,32	4,14	4,41	4,71	3,94	4,22
PB	5,40	6,02	5,02	5,81	4,83	5,05	4,69	4,91
PE	5,60	5,90	5,83	6,03	4,86	4,98	4,92	5,04
AL	5,76	6,73	6,61	7,50	5,19	5,46	5,56	5,85
SE	4,07	4,22	3,36	3,48	4,17	4,21	3,83	3,88
BA	5,71	6,38	5,07	5,57	4,82	4,97	4,51	4,66
MG	4,96	6,36	5,12	6,66	4,62	5,19	4,72	5,30
ES	5,97	6,59	6,03	6,71	4,97	5,10	4,99	5,12
RJ	4,78	5,39	4,91	5,70	4,53	4,75	4,64	4,87
SP	5,68	6,23	6,90	7,45	4,71	4,97	5,13	5,43
PR	4,11	4,20	5,29	5,35	4,20	4,20	4,85	4,85
SC	4,91	4,85	5,89	5,89	4,64	4,66	5,08	5,09
RS	4,81	4,82	5,03	5,12	4,41	4,39	4,55	4,54
MS	3,45	3,67	4,05	4,05	4,05	4,05	4,35	4,35
MT	5,30	5,26	5,44	5,36	4,75	4,75	4,82	4,81
GO	7,09	7,23	7,69	7,86	5,40	5,40	5,80	5,81
DF	5,18	4,94	4,93	5,02	4,91	5,04	5,00	5,13

Fonte: Elaborado pela autora

A Tabela 9 mostra que a estimação das variâncias a partir dos dados amostrais (ρ_{int_2}) oferece resultados próximos àqueles obtidos a partir do universo de referência (ρ_{int}) bem como daqueles obtidos sem o uso dos modelos hierárquicos ($\hat{\rho}_c$). Isso fica evidente pela proximidade dos valores calculados para a correlação intraclasse.

Tabela 9 - Comparativo dos valores de correlação intraclasse

UF	Língua Portuguesa			Matemática		
	ρ_{int}	$\hat{\rho}_c$	ρ_{int_2}	ρ_{int}	$\hat{\rho}_c$	ρ_{int_2}
RO	0,16	0,16	0,16	0,19	0,19	0,19
AC	0,19	0,19	0,19	0,23	0,23	0,22
AM	0,35	0,34	0,32	0,36	0,35	0,34
RR	0,21	0,21	0,20	0,25	0,24	0,24
PA	0,22	0,22	0,20	0,21	0,21	0,19
AP	0,17	0,18	0,15	0,17	0,17	0,14
TO	0,25	0,24	0,24	0,27	0,26	0,28
MA	0,19	0,18	0,17	0,19	0,18	0,16
PI	0,29	0,26	0,29	0,33	0,31	0,33
CE	0,20	0,19	0,22	0,31	0,30	0,34
RN	0,19	0,17	0,17	0,19	0,17	0,18
PB	0,18	0,17	0,17	0,19	0,18	0,18
PE	0,17	0,16	0,21	0,20	0,19	0,27
AL	0,23	0,21	0,18	0,27	0,25	0,21
SE	0,18	0,17	0,17	0,17	0,17	0,17
BA	0,19	0,18	0,22	0,20	0,19	0,25
MG	0,19	0,14	0,16	0,23	0,17	0,20
ES	0,12	0,11	0,11	0,15	0,14	0,14
RJ	0,15	0,14	0,14	0,20	0,18	0,18
SP	0,11	0,10	0,09	0,15	0,13	0,12
PR	0,16	0,15	0,17	0,22	0,21	0,24
SC	0,13	0,13	0,12	0,18	0,18	0,16
RS	0,16	0,16	0,16	0,20	0,20	0,20
MS	0,19	0,19	0,17	0,23	0,23	0,20
MT	0,12	0,12	0,12	0,15	0,15	0,14
GO	0,14	0,14	0,14	0,19	0,19	0,19
DF	0,11	0,10	0,10	0,13	0,11	0,11

Fonte: Elaborada pela autora

O fato de os valores obtidos para os coeficientes de correlação intraclasse serem próximos indica que os métodos de estimação testados fornecem resultados similares. Assim, a estimação da variabilidade pelo modelo hierárquico de 2 níveis pode ser adotada de maneira tão confiável quanto pelos outros métodos testados. O tamanho da amostra calculado a partir do ρ_{int_2} também é muito próximo àqueles encontrados a partir do ρ_{int} no modelo hierárquico de três níveis e do $\hat{\rho}_c$ sem a adoção do modelo hierárquico, como mostra a Tabela 10.

Tabela 10 - Comparativo do tamanho da amostra de turmas

UF	N_t	$n_{t\rho_{int}}$	$n_{t\hat{\rho}_c}$	$\bar{n}_{t\rho_{int}2}$
BR	97.782	3.324	3.186	3.316
RO	1.043	105	102	105
AC	517	143	140	138
AM	2.407	214	212	207
RR	418	147	145	145
PA	5.452	120	118	112
AP	561	94	99	90
TO	967	159	152	166
MA	4.389	105	99	100
PI	1.828	186	176	192
CE	4.488	212	208	231
RN	1.739	125	110	116
PB	2.012	114	107	113
PE	4.108	111	105	157
AL	1.681	160	150	128
SE	1.194	100	96	98
BA	7.314	117	110	132
MG	10.135	135	110	125
ES	2.006	83	80	83
RJ	5.336	97	90	91
SP	18.072	94	85	85
PR	5.455	116	113	127
SC	3.545	103	102	98
RS	5.587	106	106	107
MS	1.360	111	111	102
MT	1.824	90	90	88
GO	3.012	106	105	112
DF	1.332	71	65	68

Fonte: Elaborada pela autora

Partindo desse novo tamanho amostral, a simulação foi repetida, com o sorteio de mais 2.000 amostras e o cálculo dos resultados obtidos por meio delas. Foram auferidos valores muito próximos aos anteriores. As tabelas apresentadas a seguir sintetizam as comparações realizadas entre as três metodologias de estimação de variâncias para definição do tamanho da amostra. As Tabelas 11 e 12 mostram o comparativo entre os resultados de EQM e RREQM para Língua Portuguesa e Matemática.

Tabela 11 - Comparativo do EQM e do RREQM para Língua Portuguesa

UF	EQM ($\hat{\theta}_{LP}$)			RREQM ($\hat{\theta}_{LP}$)		
	ρ_{int}	$\hat{\rho}_c$	ρ_{int_2}	ρ_{int}	$\hat{\rho}_c$	ρ_{int_2}
BR	0,45	0,52	0,47	0,32%	0,34%	0,33%
RO	4,46	4,47	4,39	1,00%	1,00%	1,00%
AC	1,86	1,88	1,91	0,64%	0,64%	0,64%
AM	2,17	2,29	2,16	0,74%	0,76%	0,73%
RR	1,10	1,11	1,09	0,52%	0,52%	0,52%
PA	6,40	6,50	6,66	1,37%	1,38%	1,39%
AP	3,83	3,22	4,12	1,08%	0,99%	1,12%
TO	3,06	3,18	2,78	0,87%	0,89%	0,83%
MA	7,03	7,39	7,65	1,48%	1,52%	1,55%
PI	2,93	3,31	2,73	0,88%	0,93%	0,85%
CE	3,94	3,99	3,20	0,91%	0,92%	0,82%
RN	4,19	5,01	4,64	1,10%	1,20%	1,16%
PB	5,40	6,02	5,67	1,22%	1,29%	1,25%
PE	5,60	5,90	3,69	1,22%	1,26%	0,99%
AL	5,76	6,73	8,61	1,26%	1,37%	1,55%
SE	4,07	4,22	4,29	1,11%	1,13%	1,13%
BA	5,71	6,38	4,83	1,25%	1,32%	1,15%
MG	4,96	6,36	5,43	1,00%	1,13%	1,04%
ES	5,97	6,59	5,98	1,14%	1,20%	1,14%
RJ	4,78	5,39	5,25	1,04%	1,11%	1,09%
SP	5,68	6,23	6,36	1,05%	1,10%	1,11%
PR	4,11	4,20	3,69	0,90%	0,91%	0,85%
SC	4,91	4,85	4,90	0,98%	0,98%	0,98%
RS	4,81	4,82	4,92	1,02%	1,02%	1,03%
MS	3,45	3,67	4,05	0,87%	0,90%	0,94%
MT	5,30	5,26	5,36	1,12%	1,11%	1,12%
GO	7,09	7,23	6,13	1,23%	1,24%	1,14%
DF	5,18	4,94	5,08	1,03%	1,01%	1,02%

Fonte: Elaborada pela autora

Tabela 12 - Comparativo do EQM e do RREQM para Matemática

UF	EQM ($\hat{\theta}_{MT}$)			RREQM ($\hat{\theta}_{MT}$)		
	ρ_{int}	$\hat{\rho}_c$	ρ_{int_2}	ρ_{int}	$\hat{\rho}_c$	ρ_{int_2}
BR	0,50	0,56	0,52	0,32%	0,34%	0,33%
RO	5,00	4,94	4,95	1,01%	1,01%	1,01%
AC	2,18	2,22	2,28	0,65%	0,66%	0,67%
AM	2,37	2,52	2,44	0,73%	0,76%	0,75%
RR	1,12	1,11	1,14	0,49%	0,49%	0,49%
PA	5,42	5,41	5,65	1,21%	1,21%	1,23%
AP	3,28	2,70	3,61	0,96%	0,87%	1,01%
TO	3,15	3,21	2,64	0,84%	0,85%	0,77%
MA	5,60	6,03	6,27	1,27%	1,32%	1,34%
PI	2,90	3,29	2,64	0,83%	0,89%	0,79%
CE	5,58	5,76	4,49	1,04%	1,06%	0,94%
RN	3,32	4,14	3,73	0,94%	1,05%	1,00%
PB	5,02	5,81	5,27	1,12%	1,20%	1,15%
PE	5,83	6,03	3,80	1,19%	1,21%	0,96%
AL	6,61	7,50	10,28	1,27%	1,36%	1,59%
SE	3,36	3,48	3,50	0,95%	0,96%	0,97%
BA	5,07	5,57	4,31	1,13%	1,18%	1,04%
MG	5,12	6,66	5,76	0,98%	1,11%	1,04%
ES	6,03	6,71	6,01	1,10%	1,16%	1,10%
RJ	4,91	5,70	5,43	1,02%	1,10%	1,07%
SP	6,90	7,45	7,47	1,10%	1,15%	1,15%
PR	5,29	5,35	4,81	0,96%	0,97%	0,92%
SC	5,89	5,89	5,88	1,04%	1,04%	1,03%
RS	5,03	5,12	5,13	1,00%	1,01%	1,01%
MS	4,05	4,05	4,82	0,91%	0,91%	0,99%
MT	5,44	5,36	5,36	1,09%	1,08%	1,08%
GO	7,69	7,86	6,84	1,24%	1,26%	1,17%
DF	4,93	5,02	5,04	0,97%	0,98%	0,98%

Fonte: Elaborada pela autora

Adicionalmente foram comparadas as medidas de cobertura de intervalos de confiança e margem de erro verificadas nas três simulações realizadas. Mais uma vez os resultados apresentaram diferenças pequenas, como mostram as Tabelas 13 e 14.

Tabela 13 - Comparativo da cobertura e da margem de erro em Língua Portuguesa

UF	Cobertura LP			$ME(\hat{\theta}_{LP})$		
	ρ_{int}	$\hat{\rho}_c$	ρ_{int_2}	ρ_{int}	$\hat{\rho}_c$	ρ_{int_2}
BR	96,2%	96,2%	96,5%	1,42	1,50	1,47
RO	97,0%	97,3%	97,4%	4,73	4,73	4,73
AC	99,5%	99,6%	99,5%	4,07	4,11	4,14
AM	98,0%	97,8%	98,0%	3,48	3,48	3,60
RR	99,7%	99,8%	99,8%	3,38	3,39	3,38
PA	95,3%	95,3%	94,9%	5,16	5,18	5,24
AP	96,2%	97,0%	96,4%	4,32	4,12	4,45
TO	98,2%	98,3%	98,7%	4,37	4,45	4,27
MA	94,3%	95,0%	94,5%	5,31	5,54	5,53
PI	97,1%	97,5%	97,4%	3,79	4,13	3,74
CE	95,4%	95,8%	97,0%	4,15	4,18	3,89
RN	96,6%	96,5%	96,2%	4,41	4,71	4,62
PB	95,9%	95,6%	95,0%	4,83	5,05	4,85
PE	94,8%	95,4%	95,8%	4,86	4,98	3,94
AL	96,4%	95,4%	94,8%	5,19	5,46	6,01
SE	95,2%	95,5%	95,0%	4,17	4,21	4,23
BA	94,2%	94,5%	95,5%	4,82	4,97	4,49
MG	96,7%	95,5%	95,9%	4,62	5,19	4,82
ES	95,0%	94,6%	95,9%	4,97	5,10	4,97
RJ	96,0%	95,7%	95,6%	4,53	4,75	4,70
SP	94,1%	94,6%	93,9%	4,71	4,97	5,00
PR	95,5%	96,1%	95,8%	4,20	4,20	3,99
SC	96,0%	96,3%	96,0%	4,64	4,66	4,68
RS	95,6%	95,2%	94,9%	4,41	4,39	4,39
MS	96,6%	96,9%	96,7%	4,05	4,05	4,26
MT	95,7%	96,0%	95,6%	4,75	4,75	4,79
GO	94,9%	94,8%	95,6%	5,40	5,40	5,10
DF	95,7%	97,3%	96,3%	4,91	5,04	4,90

Fonte: Elaborada pela autora

Tabela 14 - Comparativo da cobertura e da margem de erro em Matemática

UF	Cobertura MT			$ME(\hat{\theta}_{MT})$		
	ρ_{int}	$\hat{\rho}_c$	ρ_{int_2}	ρ_{int}	$\hat{\rho}_c$	ρ_{int_2}
BR	95,7%	95,8%	96,2%	1,49	1,58	1,55
RO	97,2%	96,8%	97,2%	4,98	4,98	4,98
AC	99,5%	99,3%	99,3%	4,34	4,39	4,42
AM	97,9%	97,3%	97,9%	3,59	3,59	3,72
RR	99,9%	99,9%	99,8%	3,41	3,42	3,41
PA	95,2%	95,0%	94,8%	4,79	4,80	4,85
AP	96,0%	97,6%	96,3%	4,03	3,83	4,15
TO	98,2%	98,3%	98,8%	4,34	4,43	4,25
MA	95,0%	94,9%	94,6%	4,79	5,00	5,00
PI	96,7%	97,6%	97,1%	3,78	4,18	3,74
CE	95,4%	96,0%	96,8%	5,02	5,05	4,70
RN	97,1%	95,5%	95,9%	3,94	4,22	4,13
PB	95,8%	95,1%	95,1%	4,69	4,91	4,71
PE	94,7%	95,7%	95,4%	4,92	5,04	3,98
AL	95,9%	94,8%	93,4%	5,56	5,85	6,41
SE	95,6%	95,4%	95,8%	3,83	3,88	3,89
BA	94,5%	95,0%	94,7%	4,51	4,66	4,21
MG	96,5%	95,0%	95,6%	4,72	5,30	4,93
ES	95,1%	94,8%	95,9%	4,99	5,12	5,00
RJ	96,1%	95,2%	95,9%	4,64	4,87	4,82
SP	94,0%	94,6%	94,6%	5,13	5,43	5,45
PR	96,6%	96,6%	96,4%	4,85	4,85	4,62
SC	96,0%	95,9%	96,0%	5,08	5,09	5,12
RS	95,7%	95,3%	95,2%	4,55	4,54	4,54
MS	96,4%	96,0%	95,4%	4,35	4,35	4,59
MT	95,7%	95,8%	95,7%	4,82	4,81	4,86
GO	95,6%	95,1%	95,8%	5,80	5,81	5,47
DF	97,5%	97,9%	97,1%	5,00	5,13	4,99

Fonte: Elaborada pela autora

Para interpretação dos resultados, recomenda-se que as medidas de desempenho sejam analisadas conjuntamente, podendo ser adotadas como referência na comparação de diferentes metodologias para identificação daquela que se revela mais adequada. O EQM, por integrar medidas de viés e variância em um único indicador, apresenta-se como uma medida importante, especialmente quando os métodos comparados não apresentam viés. A cobertura do intervalo de confiança é uma característica chave para indicar o bom resultado de um estimador (MORRIS; WHITE; CROWTHER, 2019). Além disso, o RREQM facilita a análise transformando o EQM em uma medida relativa do erro em relação ao parâmetro. A Margem de

erro complementa o entendimento da cobertura por permitir a compreensão da amplitude do intervalo de confiança.

Nas simulações realizadas, todos os resultados observados foram bastante próximos. Os três métodos de estimação da correlação intraclasse usada para dimensionamento da amostra apresentam um viés muito próximo de zero e, com isso, o EQM ganha importância no processo de comparação entre as metodologias testadas. Apesar de não apresentarem grandes diferenças, tanto os valores do EQM quanto do RREQM e da margem de erro foram em média um pouco maiores na Simulação $\hat{\rho}_c$. Usar os parâmetros do modelo hierárquico para dimensionamento da amostra pode ser questionável quando há a possibilidade de estimação das variâncias por métodos mais diretos, como o testado na metodologia recomendada por Silva e Moura (1990). Entretanto, a proximidade dos resultados observados indica que a estimação das variâncias por meio do modelo hierárquico se mostra adequada no contexto do Saeb. O Apêndice C apresenta os *scripts* utilizados para processamento das simulações.

Pessoa e Silva (1998) corroboram esse entendimento ao explicar que a organização hierárquica é intrínseca a algumas populações, citando como exemplo os sistemas educacionais, em que o agrupamento natural dos alunos em turmas e escolas formam os níveis de análise. Assim, a adoção dos modelos hierárquicos para descrever tais estruturas é motivada por essa característica natural da população, independentemente do procedimento amostral envolvido na obtenção dos dados.

Diante das análises realizadas, pode-se perceber que os resultados obtidos por Amostragem se aproximam do valor verdadeiro da proficiência média, tanto para Língua Portuguesa quanto para Matemática. A análise para a agregação máxima (Brasil) e para as agregações por UF revelam que o método de seleção da amostra do Saeb oferece resultados válidos e confiáveis em termos de aproximação do valor da média obtida por meio da avaliação censitária. Algumas medidas de desempenho não cumprem plenamente o recomendado pela teoria amostral, especialmente a margem de erro, que extrapolou o valor inicialmente estabelecido para a maioria das UFs. Entretanto, mesmo essa medida está próxima do esperado e confirma a validade dos resultados. Diante disso, constata-se que apesar dos bons resultados obtidos por meio da metodologia amostral adotada atualmente no Saeb, algumas melhorias podem ser incorporadas ao processo a fim de minimizar suas limitações, como veremos a seguir.

3.3.5 Limitações do plano amostral

O entendimento do plano amostral do Saeb permite a identificação de possíveis limitações da avaliação associadas a essa etapa do processo, que são decorrentes de diversos fatores.

Parte das limitações identificadas são de ordem operacional e impossibilitam que seja vivenciado um cenário que atenda a todas as especificações da teoria da Amostragem. Entre elas, destaca-se a construção do cadastro e a delimitação do universo de referência para a seleção da amostra e a expansão dos resultados. O cadastro utilizado no Saeb é construído a partir dos dados coletados no Censo Escolar. Entretanto, os prazos necessários para a coleta dos dados das escolas para o ano de aplicação da avaliação e os procedimentos necessários à preparação da aplicação impedem que a mesma referência (cadastro) seja adotada para todas as etapas do processo.

Nesse contexto, os dados do Censo Escolar são adotados em três momentos: a) para o cálculo do tamanho da amostra e seleção das escolas, usa-se o Censo Escolar do ano anterior; b) para a seleção das turmas e identificação dos alunos que participarão da aplicação da avaliação, a referência é Censo Escolar do ano da aplicação em sua versão preliminar; c) para a expansão dos resultados, são utilizadas as quantidades apuradas a partir do Censo Escolar do ano da aplicação em sua versão final. Essa diversidade de referências agrega dificuldade ao processo; contudo, se reconhece que esse procedimento aprimora a identificação da população de referência em relação à identificação realizada a partir do cadastro construído com base em um único momento do Censo Escolar. Diante disso, são adotadas medidas corretivas para minimizar o impacto dessa realidade como, por exemplo, a consistência dos dados provenientes da aplicação do Saeb com a situação atualizada em relação à versão final do Censo.

Outra limitação operacional, relacionada à delimitação do universo de referência, é a exclusão de parte dos alunos da população alvo. Ao desenhar o plano amostral para a coleta dos dados de uma avaliação de aprendizagem em larga escala (AALE) é comum a exclusão de crianças matriculadas em escolas de difícil acesso ou alunos com alguma particularidade que os diferencia dos demais como a língua ou alguma deficiência (UNESCO, 2019). Isso ocorre também no Saeb, onde são excluídas escolas muito pequenas (com menos de 10 alunos), turmas multisseriadas ou de educação especial exclusiva e alunos indígenas que não têm conhecimento da Língua Portuguesa. A UNESCO (2019) destaca que essa delimitação impede que os resultados obtidos pela avaliação expliquem a realidade educacional dos grupos excluídos. Por outro lado, é importante reconhecer que para alcançar todos os perfis de alunos existentes

seriam necessárias tantas adaptações aos instrumentos e aos procedimentos de aplicação que se tornaria inviável a comparabilidade dos resultados entre os diferentes grupos. Esse cenário também poderia oferecer vantagens comparativas para os grupos atendidos pelas adaptações incentivando a declaração de falsas necessidades especiais a fim de se obter melhoria nos resultados (UNESCO, 2019).

Tanto a constituição do cadastro da população de referência quanto a sua delimitação, são limitações de difícil resolução, mas a partir do conhecimento de sua existência, são adotadas medidas que minimizem seus efeitos.

Há ainda limitações decorrentes de aspectos técnicos. Entre eles, pode-se destacar alguns pontos apresentados na seção que trata da descrição do plano amostral do Saeb. O uso de dados oriundos dos resultados das escolas públicas (Prova Brasil), cuja aplicação é censitária, como referência de variabilidade da proficiência dos alunos matriculados em escolas privadas é um deles. Ressalta-se que tal procedimento não tem como base nenhuma hipótese e se tornou uma prática ao longo das sucessivas edições do SAEB.

Outro ponto que pode ser repensado é a metodologia de cálculo do tamanho da amostra que adota como universo de referência da amostra o universo de referência do Saeb como um todo.

O cálculo do tamanho da amostra utiliza como parâmetros iniciais a variância da proficiência e o número médio de alunos por turma. No caso do Saeb, são adotadas como referências as variâncias referentes às proficiências dos alunos das escolas públicas. Porém o número médio de alunos por turma considera o Saeb como um todo, incluindo escolas públicas e particulares. Como a seleção amostral diz respeito apenas a escolas particulares, após o cálculo do tamanho da amostra, aplica-se o percentual de 40% do tamanho da amostra calculado para obtenção do tamanho da amostra de escolas particulares. Esse procedimento agrega complexidade ao processo, partindo de pressuposições que não foram necessariamente testadas.

3.4 OUTRAS LIMITAÇÕES DO SAEB

O Saeb é uma AALE e, conseqüentemente, está suscetível a diversas limitações inerentes a esse modelo avaliativo. A UNESCO (2019) alerta sobre a importância de se reconhecer as limitações das AALEs para que seja possível aproveitar melhor as oportunidades oferecidas pelos resultados da avaliação. As limitações das AALEs destacadas pela UNESCO (2019) são agrupadas em duas grandes categorias: limitações da elaboração e limitações no uso.

Quando se trata das limitações da elaboração, se fala de valorizar além do que se pode medir, enquanto as limitações no uso dizem respeito ao uso para além dos dados como um fim em si.

Na elaboração de uma AALE é importante a diferenciação entre educação e aprendizagem. A educação é plural e serve a múltiplos propósitos, se ligando intimamente a diversos aspectos do desenvolvimento humano pessoal e coletivo. Enquanto “a aprendizagem é um processo complexo de aquisição ou desenvolvimento de conhecimentos, habilidades e atitudes, com base em valores subjacentes” (UNESCO, 2019, p.37). Diante de toda a complexidade que envolve a educação e a aprendizagem, as AALEs restringem-se a poucas áreas de aprendizagem que são facilmente mensuráveis, deixando de abordar uma série de elementos da aprendizagem indispensáveis para o cumprimento dos múltiplos propósitos da educação. Essa característica não é essencialmente problemática, entretanto, em um cenário de dependência excessiva da avaliação, existe o risco de valorizar aquilo que é medido, em vez de incluir na medição o que é valorizado. Além disso, existe o risco de que a aprendizagem seja reduzida aos aspectos que podem ser medidos (UNESCO, 2019).

O conceito de qualidade da educação é tão multidimensional quanto a própria educação, portanto os resultados de aprendizagem, mencionados pela UNESCO (2019) como sinônimo de resultados de desempenho, não podem ser analisados de forma isolada ou adotados como medida única de qualidade da educação. A não contextualização dos indicadores resultará em conclusões tendenciosas, portanto as AALEs têm sua eficácia potencializada quando associadas a outros fatores dentro e fora da escola. Apesar disso, abordagens equilibradas e multidimensionais tem se tornado mais raras na mesma medida em que as AALEs têm ganhado importância sendo consideradas como medida suficiente da qualidade da educação (UNESCO, 2019).

Sobre o uso dos dados provenientes de avaliações, a UNESCO (2019) adverte que o debate a respeito desse tema é imprescindível, uma vez que grande parte das limitações das AALEs estão relacionadas ao uso dos resultados, evidenciando a necessidade de clareza sobre a visão geral dos propósitos e usos previstos para esse instrumento.

A UNESCO (2019, p. 54) entende que “os resultados não falam por si, e a formulação de políticas com base em evidência não decorre *naturalmente* dos dados brutos”. Assim, corrobora que proposição de políticas ou práticas educacionais precisam ser embasadas em evidências decorrentes da análise de um conjunto de resultados e seu contexto, envolvendo julgamento e reflexão

Como os dados das avaliações são, por definição, parciais e imperfeitos, é provável que uma atenção desequilibrada às AALEs ou a falha em integrá-las a outros componentes do sistema dificulte sua capacidade de contribuir para um avanço significativo em direção a uma maior qualidade e equidade da educação (UNESCO, 2019, p.52).

Os processos externos de avaliação de aprendizagem geralmente se encerram com a publicação de relatórios consolidando uma série de conclusões e resultados, que logo são deixados de lado e esquecidos. Contudo, apesar da reconhecida importância das AALEs para a obtenção de informações sobre múltiplas dimensões dos sistemas educacionais, sua eficácia somente será atingida se os resultados forem usados para gerar transformação (UNESCO, 2019).

Para tanto, a validação e adequada divulgação dos resultados das avaliações incrementam seu valor político e reforçam o senso de apropriação. Com isso, os governantes e educadores podem compreender melhor como usar os resultados divulgados no sentido de promover a transformação necessária (UNESCO, 2019).

Por outro lado, a adoção de múltiplos canais de divulgação dos resultados, intensa repercussão na mídia e formação de rankings pode gerar uma excessiva atenção às AALEs, criando um cenário de utilização dos dados das avaliações. Tal cenário pode conduzir ao entendimento equivocado de que o mais importante é a melhoria dos números, em vez de desencadear processos de mudanças estruturais substanciais (UNESCO, 2019).

O uso abundante dos resultados das avaliações não é inerentemente problemático, porém, quando o foco se concentra somente em pontuações ou classificações em detrimento da compreensão dos fatores que os influenciam, as medidas propostas podem se tornar enviesadas (UNESCO, 2019).

Além dos cenários de subutilização e sobreutilização dos resultados, a UNESCO (2019) também alerta sobre os riscos decorrentes da associação das avaliações a políticas de responsabilização. Mais uma vez, o uso dos resultados pode direcionar o foco para as pontuações obtidas, gerando estímulos para a adoção de práticas não recomendáveis, mas que acarretem notas melhores. Podem ser citadas como exemplo dessas práticas a seleção de alunos, a redução do currículo aos conteúdos incluídos nas provas e a concentração nos alunos com maiores chances de responder corretamente às questões. Tais medidas podem garantir melhores pontuações, no entanto podem desorientar o sistema de monitoramento das redes de ensino impedindo a identificação das reais necessidades de melhorias, ao mesmo tempo que prejudicam a qualidade e a equidade da educação geral (UNESCO, 2019).

Com base em estudos realizados por diversos autores, Calderón (2017) corrobora que no Brasil, apesar de haver exemplos contrários, é comum o uso dos resultados das avaliações educacionais para viabilizar práticas pouco recomendáveis. Entre os exemplos citados, pode-se listar distribuição de recursos, bonificação de professores, redução do currículo, tomada de decisão sobre a continuidade de estudo dos alunos e adoção de atividades que preparem os alunos para que tenham bom desempenho nas avaliações. Nesse contexto, Calderón (2017) pontua que, a despeito do consenso a respeito da relevância das avaliações em larga escala, “as críticas aos usos das avaliações externas geralmente são ácidas e acentuadas” (p. 38).

Diante deste paradoxo, Calderón (2017) defende que a importância das avaliações decorre da crescente busca pela conquista dos direitos sociais, que incluem o direito à educação. Esse movimento é incentivado por organismos internacionais como a UNESCO que estabelece aprendizagem como parâmetro do cumprimento do direito à educação. Assim, as avaliações educacionais destacam-se por permitirem a obtenção de diagnósticos que orientem a formulação e o monitoramento de políticas públicas, a partir de informações organizadas e comparáveis sobre diversos aspectos dos sistemas educacionais (CALDERÓN, 2017)

Levantamento realizado por Bauer (2019) confirma que, apesar do grande volume de dados produzidos pelos diversos sistemas de avaliação educacional no Brasil, os seus resultados são pouco utilizados para a gestão educacional ou escolar. Bauer (2019) identificou que estudos ratificam o paradoxo apresentado por Calderón (2017), revelando que mesmo que os gestores reconheçam a importância desses resultados, sua influência na prática da gestão é tênue ou inexistente. Por outro lado, Brooke e Cunha (2011) defendem que houve uma intensificação do uso dos resultados das avaliações como instrumentos de gestão, mesmo que ainda estejam presentes os usos menos recomendados desses resultados.

Outro aspecto evidenciado por diversos estudos analisados por Bauer (2019) é a dificuldade dos gestores em entender adequadamente as informações produzidas pelo Saeb, comprometendo a sua utilização. Assim, em um contexto de expansão do Saeb e dos sistemas de avaliação nos Estados e Municípios, fica claro que o volume de informação produzida é bastante para exercer influência no cenário da educação brasileira. Entretanto não há tempo ou orientação suficiente para o entendimento das informações e seu uso adequado. Essa situação pode ser exemplificada pelos dados coletados por meio dos questionários contextuais. Apesar de os dados serem coletados em todas as edições e incluídos nos microdados, as análises realizadas a partir deles são menos frequentes. O único indicador desenvolvido pelo Inep a partir desses dados é o Indicador de Nível Socioeconômico, que não se refere exatamente aos elementos do processo de ensino e aprendizagem, mas à condição social predominante entre os

alunos. Um possível dificultador para a elaboração de indicadores é o sistema de pesos calculados para a ponderação dos resultados do Saeb. Hoje esse sistema fornece pesos voltados apenas para o cálculo das proficiências, limitando o estudo dos resultados contextuais do Saeb. No entanto, a análise mais aprofundada dos dados coletados por meio dos questionários poderia conduzir ao desenvolvimento de indicadores que, conjuntamente, auxiliem o delineamento das condições em que a promoção do ensino ocorre no Brasil. O produto dessas análises seriam ferramentas valiosas para orientar os gestores na proposição de políticas voltadas à melhoria da qualidade da educação brasileira (INEP, 2017?).

Retomando os objetivos do Saeb, pode-se lembrar que a avaliação está voltada para o fomento de políticas públicas. Souza (2006) explica que quando políticas públicas são implementadas, elas são submetidas a sistemas de monitoramento e avaliação. Isso se dá porque a política pública pode ser entendida com um ciclo deliberativo, constituído de fases que se complementam formando um processo dinâmico e de aprendizagem, tendo a avaliação como um dos estágios desse ciclo.

Não há um modelo único do ciclo da política pública que estabeleça de maneira definitiva quantos e quais estágios o compõem. Raeder (2014) alerta que as etapas podem se misturar e alternar sua sequência e, portanto, o mais importante é compreender que as políticas públicas se organizam em estágios com características específicas.

Apesar da falta de consenso a respeito de quais seriam as fases do ciclo da política pública, a avaliação está sempre presente como um importante elemento para sua retroalimentação, pois encerra o ciclo da política pública fornecendo subsídios para o início de um novo ciclo (SOUZA, 2006; RAEDER, 2014; JANNUZZI, 2011). Um exemplo dos estágios do ciclo foi sintetizado por Jannuzzi (2011) na Figura 1.

Relacionando ao contexto da educação, Kellaghan, Greaney e Murray (2011) destacam que avaliações realizadas em um curto intervalo de tempo não acarretam mudanças mais rápidas no cenário avaliado.

Uma consulta aos resultados das avaliações nacionais que vêm sendo realizadas há várias décadas indica que grandes mudanças nos aproveitamentos dos alunos em um sistema educacional não ocorrem em um curto período, mesmo quando foram feitos esforços para solucionar os problemas identificados em uma avaliação. Dada essa situação, um intervalo de quatro a cinco anos parece sensato (KELLAGHAN, GREANEY; MURRAY, 2011, p. 92).

Figura 1 - Ciclo de políticas e programas públicos



Fonte: Jannuzzi (2011)

Diante disso, entendendo que avaliação é uma das etapas do ciclo de política pública, é necessário que haja entre duas avaliações tempo para implementação de alguma intervenção e para que essa intervenção gere resultados perceptíveis em um novo momento avaliativo. Além disso, o Saeb é utilizado não apenas para o acompanhamento de políticas já vigentes, mas para a proposição de novas políticas públicas que visem corrigir os problemas revelados pelos resultados da avaliação. Portanto, repetir uma avaliação sem que tenha havido o reprocessamento dos demais estágios do ciclo da política pública, configura-se como um procedimento injustificado.

Conclui-se que os estudos e análises realizados permitiram a identificação de algumas limitações do Saeb, que podem ser categorizadas em função de sua origem e estão sintetizadas no Quadro 2 apresentado a seguir.

Quadro 2 - Síntese das limitações identificadas no Saeb

CATEGORIA	LIMITAÇÃO IDENTIFICADA
Operacional	Diversidade de referências para composição do cadastro.
	Delimitação do universo de referência.
Técnica	Variância das proficiências dos alunos das escolas públicas usada como referência de variabilidade da proficiência dos alunos das escolas particulares.
	Universo de referência da amostra confundido com universo de referência da avaliação.
	Cálculo do tamanho da amostra considerando número médio de alunos por turma de escolas públicas e particulares.
Política	Coleta dos dados em periodicidade incompatível com os objetivos de fomentar políticas públicas. Incompatibilidade com ciclo da política pública.
Intrínseca	Pouca contextualização, favorecendo o uso inadequado dos resultados.

Fonte: Elaborado pela autora

Partindo do entendimento do modelo atual do Saeb e da identificação de suas limitações, a próxima etapa da pesquisa propõe um Plano de Ação que visa à implementação de melhorias ao Saeb. As melhorias propostas abrangem especificamente as limitações do planejamento amostral e as demais limitações apresentadas, respeitando as particularidades da avaliação, com o objetivo de sugerir um caminho factível para a promoção de aperfeiçoamentos no modelo atual.

4 PROPOSTAS PARA SUPERAR LIMITAÇÕES DO SAEB

Os capítulos anteriores foram dedicados a uma discussão sobre o contexto em que surgiu o Saeb e sua evolução ao longo dos anos, além da apresentação de conceitos básicos de Amostragem e do detalhamento do plano amostral adotado para o Saeb.

Apoiada nessa discussão, foi realizada uma simulação com o propósito de comparar os resultados obtidos por censo com aqueles obtidos por Amostragem, ampliando o debate que permitiu a identificação de limitações que podem ser classificadas como operacionais, técnicas, políticas e intrínsecas.

As limitações operacionais identificadas são decorrentes da impossibilidade de mudança de fatores que não são diretamente ligados à avaliação, como o processo de coleta dos dados do Censo Escolar da educação básica ou a logística da aplicação. No entanto, em função do entendimento de que o Inep já toma as medidas necessárias para minimizar o impacto destas limitações no resultado da avaliação, as propostas elaboradas se abstiveram em relação às limitações operacionais.

A delimitação do cadastro é uma importante etapa da elaboração de uma pesquisa amostral, entretanto é comum que esta etapa represente um dos principais problemas práticos na operacionalização de uma seleção amostral (COCHRAN, 1977). No caso do Saeb, a dificuldade na constituição do cadastro é minimizada pela qualidade e disponibilidade dos dados do Censo Escolar da Educação Básica. Partindo dessa asserção, a limitação do Saeb associada ao cadastro diz respeito ao tempo necessário para a atualização dos dados do Censo Escolar e a realização da seleção da amostra. Para que seja possível a preparação da aplicação no prazo exigido, a seleção das escolas da amostra precisa ser realizada com base no Censo Escolar do ano anterior. Com o propósito de contornar os efeitos negativos dessa característica do processo, os dados de turmas e alunos elegíveis para participar da avaliação são atualizados pela versão preliminar do Censo Escolar referente ao ano de aplicação. Por fim, os dados são novamente atualizados, com base na versão final do Censo Escolar do ano de aplicação, a fim de identificar os dados coletados que estão vinculados a alunos que deixaram de pertencer ao universo de referência do Saeb após as correções no cadastro (INEP, 2018a). O Inep (2018a) explica que essa alternativa apresenta como vantagem o ajuste dos dados à versão mais atual do universo de referência, facilitando a interpretação dos resultados da avaliação.

Outra limitação operacional comum às avaliações em larga escala está relacionada à delimitação do universo de referência da avaliação. Mesmo de posse de um cadastro de alta qualidade, a adaptação dos instrumentos e da operação logística a fim de que a avaliação alcance

todos os alunos tornaria sua realização inviável (UNESCO, 2019). No caso do Saeb, são realizadas exclusões indispensáveis para viabilização da avaliação, porém mantendo a máxima abrangência possível. Pode-se exemplificar o caso das comunidades indígenas: a tradução dos instrumentos para todas suas línguas aumentaria significativamente a complexidade do processo avaliativo, tornando sua execução improvável. Contudo, com o intuito de se obter uma aproximação da realidade indígena, as escolas com esse perfil que ministram o ensino em Língua Portuguesa participam da avaliação. Essas escolas concentram 94,5% dos alunos indígenas declarados no Censo Escolar (INEP, 2018a, 2018e). Da mesma forma, as demais exclusões são realizadas de forma que o percentual de alunos não cobertos pela avaliação seja o menor possível. No Saeb 2017, esse número foi inferior a 8% do total de alunos do 5º ano do ensino fundamental, a 3% dos alunos do 9º ano do ensino fundamental e a 1% dos alunos da 3ª/4ª séries do ensino médio (INEP, 2018a).

Em relação aos aspectos técnicos, foram investigados os impactos de se calcular o tamanho da amostra considerando diretamente a população que será avaliada. A fórmula utilizada para calcular o tamanho da amostra leva em consideração o número médio de alunos por turma e a estrutura de variação das variáveis de interesse na população.

A metodologia atual de cálculo do tamanho da amostra do Saeb, explicada na seção 3.2, assume que a variabilidade dos alunos das escolas públicas é semelhante à dos alunos das escolas privadas. A adoção dessa referência é decorrente da indisponibilidade de dados para o cálculo das variâncias pelo modelo hierárquico de três níveis para o público avaliado por Amostragem. Usando o número médio de alunos por turma e a estrutura de variação referentes ao universo do Saeb como um todo, calcula-se o tamanho da amostra considerando que ela abrangeria todo o Saeb. Como a amostra cobre apenas escolas particulares, uma vez que as escolas públicas têm aplicação censitária, adota-se a pressuposição de que a proporção de 40% da amostra total é suficiente para avaliação das escolas particulares.

Esse procedimento reflete a estratificação da amostra com alocação proporcional, etapa do processo que distribui o tamanho da amostra entre os estratos de interesse, proporcionalmente ao tamanho da população de cada estrato. Quando a população é subdividida entre os estratos de escolas públicas e escolas privadas, geralmente a proporção dos alunos das escolas públicas é maior. Para evitar uma concentração muito grande da amostra em uma única rede de ensino (pública ou privada), adotava-se 60% para a rede com maior número de alunos e 40% para a rede de ensino com menor quantidade de alunos. Esse procedimento foi utilizado enquanto havia amostra para as duas redes de ensino, especialmente para a avaliação do ensino médio que permaneceu completamente amostral até a edição de 2015. Conforme a aplicação da

avaliação passou a ser censitária para os alunos das escolas públicas, o cálculo se manteve inalterado, desconsiderando os 60% da amostra que seriam destinados a esse público e permanecendo a alocação de 40% da amostra para escolas particulares (INEP, 2017).

Com o propósito de adequar o tamanho da amostra, utilizando as referências específicas da população amostrada, foi estudada a possibilidade de estimação da variabilidade da população a partir de outras metodologias que permitam o uso dos dados coletados por Amostragem. Essa proposta torna desnecessário assumir que a variabilidade referente às escolas públicas é equivalente à das escolas privadas. Além disso, permite o cálculo do tamanho da amostra diretamente para a rede privada, sem a necessidade de aplicar o percentual de 40% do tamanho total calculado.

No que diz respeito às limitações políticas, o plano de ação aborda aspectos relacionados ao ciclo de políticas públicas, que apresenta a avaliação como uma de suas etapas. Diante disso, propõe-se uma periodicidade de coleta dos dados referentes ao Saeb, compatível com o ciclo de políticas públicas a fim de que esta avaliação atenda de maneira mais próxima sua função de subsidiar políticas públicas educacionais.

Em relação às limitações intrínsecas às AALEs, aplica-se mais diretamente ao Saeb a realidade de uso inadequado dos resultados da avaliação. Parte dessa limitação é decorrente da pouca contextualização dos resultados, apesar das estratégias de divulgação adotadas pelo Inep como a publicação de relatórios, do painel educacional e do boletim da escola. Com o objetivo de melhorar a contextualização dos resultados da avaliação e minimizar a possibilidade de uso inadequado dos resultados, a proposta incluiu um método de cálculo de pesos válidos para os dados coletados pelos questionários contextuais. Esta medida pretende facilitar o cálculo de novos indicadores contextuais que auxiliem no correto entendimento das informações obtidas a partir do Saeb.

O Quadro 3 apresenta uma síntese dos elementos apresentados na proposta desse plano de ação.

Quadro 3 - Resumo das propostas para aprimoramento do Saeb

CATEGORIA	LIMITAÇÃO IDENTIFICADA	PROPOSTA
Operacional	Diversidade de referências para composição do cadastro	Não serão apresentadas propostas para esta categoria, pois se entende que o Inep adota medidas adequadas para minimizar o efeito dessas limitações.
	Delimitação do universo de referência	
Técnica	Variância das proficiências dos alunos das escolas públicas usada como referência de variabilidade das proficiências dos alunos das escolas particulares.	Estimação de variâncias usando os resultados amostrais.
	Universo de referência da amostra confundido com universo de referência da avaliação. Cálculo do tamanho da amostra considerando número médio de alunos por turma de escolas públicas e particulares.	Diferenciar o universo de referência da amostra, calculando o tamanho da amostra a partir do número médio de alunos por turma na rede privada.
Política	Coleta dos dados em periodicidade incompatível com os objetivos de fomentar políticas públicas. Incompatibilidade com ciclo da política pública.	Aplicações censitária e amostral intercaladas e fim de alinhar a divulgação dos resultados ao ciclo de política pública, permitindo que haja mais tempo para o entendimento dos resultados e implementação de melhorias.
Intrínseca	Pouca contextualização favorecendo o uso inadequado dos resultados.	Cálculo de peso para os questionários contextuais, facilitando o cálculo de indicadores que favoreçam o uso adequado dos dados.

Fonte: Elaborado pela autora

As propostas deste plano de ação, sintetizadas no Quadro 3, têm o objetivo de apresentar caminhos para superar as limitações identificadas no Saeb ou minimizar os seus efeitos, fortalecendo o processo de avaliação da educação básica.

4.1 PROPOSTA PARA O PLANO AMOSTRAL

Atualmente o plano amostral do Saeb aplica-se apenas às escolas particulares. Em função da indisponibilidade de dados suficientes para o cálculo de variâncias segundo o modelo hierárquico de três níveis para os alunos das escolas privadas, assumiu-se que a variabilidade

entre os alunos das escolas privadas é equivalente à dos alunos das escolas públicas. Decisão semelhante foi tomada em relação aos alunos do ensino médio. Quando essa etapa era avaliada de forma amostral para alunos das redes pública e particular, assumia-se que a variabilidade da proficiência dos alunos do ensino médio era equivalente à variabilidade dos alunos do 9º ano do ensino fundamental. Outra característica do plano amostral atual do Saeb é a utilização de parâmetros do universo de referência do Saeb como um todo, incluindo escolas públicas e particulares, para o cálculo do tamanho da amostra, que atualmente abrange apenas escolas particulares.

Como alternativa para contornar essas limitações, foram testadas duas metodologias para estimação das variâncias usadas no cálculo do tamanho da amostra. As simulações realizadas mostraram que o tamanho de amostra calculado a partir de parâmetros específicos da população de referência é suficiente para a obtenção de resultados confiáveis. Adicionalmente, na seção 3.3.4 ficou demonstrado que há metodologias eficientes para a estimação das variâncias que não exigem coleta censitária dos dados. Além disso, o Censo Escolar fornece informação suficiente para o cálculo do número médio de alunos por turma do universo de referência para a amostra. Conclui-se, portanto, que não há necessidade da utilização de parâmetros referentes a todo o universo do Saeb, quando a avaliação amostral se aplicar apenas às escolas particulares. Pelo contrário, é possível diferenciar o universo de referência do Saeb do universo de referência específico da amostra.

Tendo como referências os resultados observados nas simulações realizadas, a proposta é o plano amostral do Saeb passe a adotar para o cálculo do tamanho da amostra as variâncias estimadas por meio do modelo hierárquico de dois níveis. Assim, o tamanho da amostra seria calculado por meio das fórmulas apresentadas na seção 3.3.1:

$$n_0 = \frac{\sigma^2}{D}, \text{ em que } D = \frac{B^2}{z_{\alpha/2}^2} \quad n = n_0 * EPA \quad EPA \cong 1 + (\bar{M} - 1)\rho_{int_2},$$

mas com um ajuste na fórmula para o cálculo no coeficiente de correlação intraclasse que levaria em consideração os valores de σ_e^2 = variância entre as escolas dentro da UF e σ_a^2 = variância entre os alunos, conforme apresentado a seguir:

$$\rho_{int_2} = \frac{\sigma_e^2}{\sigma_a^2 + \sigma_e^2},$$

Essa proposta tem como principal justificativa a manutenção uso dos modelos hierárquicos que já vem sendo utilizados no Saeb nos últimos anos. Além disso, exigiria menos alterações nos processos que a equipe técnica já está capacitada para executar. Visto que a diferença observada entre os resultados das metodologias testadas é pequena, pode-se concluir

que os três métodos são equivalentes em termos de resultados oferecidos, mas apresentam particularidades que podem orientar a escolha do melhor modelo de acordo com as características do estudo que um pesquisador pretenda realizar.

No caso do Saeb, os modelos multiníveis podem ser mais indicados em função da organização natural dos dados em categorias hierarquizadas (PESSOA; SILVA, 1998). Além disso, os resultados obtidos nas simulações que utilizaram esse método foram ligeiramente melhores do que aqueles obtidos por outra metodologia. A escolha do modelo de dois níveis em detrimento do modelo de três níveis utilizado atualmente se dá em função da possibilidade de estimar a variância a partir dos dados amostrais. Esse aspecto é importante porque elimina a necessidade de calcular o tamanho da amostra a partir de variâncias que não são comprovadamente equivalentes às variâncias da população estudada.

A implementação dessas medidas no cálculo do tamanho da amostra para o Saeb simplifica o procedimento de cálculo, evitando a adoção de hipóteses não confirmadas sobre as características da população, continuando a oferecer estimativas precisas e confiáveis. A mudança não traz prejuízo à comparabilidade dos resultados obtidos em edições anteriores e nem acarreta custos adicionais ao processo.

4.2 PROPOSTA DE APLICAÇÃO EM CICLO DE QUATRO ANOS INTERCALANDO APLICAÇÃO CENSITÁRIA E APLICAÇÃO AMOSTRAL

O ciclo de modificação da educação é de longo prazo, com poucos efeitos perceptíveis em pequenos intervalos de tempo (KELLAGHAN; GREANEY; MURRAY, 2011). Além disso, a avaliação educacional dificilmente será suficiente para a compreensão de todos os fatores que determinam o sucesso ou o fracasso de uma política pública voltada para a educação (UNESCO, 2019). Entretanto, pode ser utilizada como uma bússola para o direcionamento geral das ações executadas em cada ciclo governamental e a manutenção ou modificação das ações iniciadas em um governo anterior.

Durante o período de desenvolvimento desta pesquisa, o MEC publicou a portaria nº 458, de 5 de maio de 2020, que “institui normas complementares necessárias ao cumprimento da Política Nacional de Avaliação da Educação Básica” (BRASIL, 2020, p.1). Esse documento determina a realização anual do Saeb, em caráter censitário e amplia a população alvo, passando a incluir alunos de todas as etapas de ensino.

Os aspectos definidos pelo MEC para o futuro do Saeb trazem em seu bojo cinco objetivos amplos, previstos no art. 6º da portaria:

Art. 6º São objetivos do Saeb:

I - construir uma cultura avaliativa, ao oferecer à sociedade, de forma transparente, informações sobre o processo de ensino-aprendizagem em cada escola, comparáveis em nível nacional, anualmente e com resultados em tempo hábil, para permitir intervenções pedagógicas de professores e demais integrantes da comunidade escolar;

II - produzir indicadores educacionais para o Brasil, suas regiões, unidades da federação, municípios e instituições escolares, tendo em vista a manutenção da comparabilidade dos dados, permitindo, assim, o incremento de séries históricas;

III - avaliar a qualidade, a equidade e a eficiência da educação praticada no País em seus diversos níveis governamentais;

IV - subsidiar a elaboração, o monitoramento e o aprimoramento de políticas públicas em educação baseadas em evidências, com vistas ao desenvolvimento social e econômico do Brasil; e

V - desenvolver competência técnica e científica na área de avaliação educacional, por meio de intercâmbio entre instituições de ensino e pesquisa, bem como de servidores do Inep, docentes e gestores da educação de todos os entes envolvidos (BRASIL, 2020, p.2).

Além disso, o documento prevê que os resultados do Saeb devem possibilitar diferentes formas de uso em um curto espaço de tempo, dificultando a absorção de toda a informação disponibilizada, favorecendo um cenário de uso indevido dos resultados. Percebe-se que o esperado é que a avaliação forneça não apenas um diagnóstico do cenário educacional brasileiro, mas que chegue a ser usada como ferramenta de seleção para o acesso à educação superior e aos benefícios de políticas públicas específicas. Tais aspectos estão previstos no seguinte artigo da portaria:

Art. 9º Os resultados do Saeb deverão possibilitar:

I - a produção de indicadores educacionais para o Brasil, suas regiões, unidades da federação, municípios e instituições escolares;

II - o monitoramento e a avaliação da eficiência, da eficácia e da efetividade da aplicação dos recursos públicos alocados aos programas e projetos da educação básica;

III - a disponibilidade de informações relevantes para subsidiar a elaboração, o monitoramento e o aprimoramento de políticas públicas;

IV - a implementação gradual de aperfeiçoamentos e inovações nos processos de avaliação e exames, mantendo a comparabilidade entre as distintas pesquisas e a manutenção de séries históricas;

V - a difusão, com transparência e de fácil acesso, para toda a sociedade, de diagnósticos e pesquisas sobre o sistema de educação básica, em níveis de desagregação distintos, até a instituição escolar, comparáveis em nível nacional;

VI - sua utilização como mecanismo único, alternativo ou complementar para acesso à educação superior, especialmente a ofertada pelas instituições federais de educação superior;

VII - a avaliação anual da educação básica, abrangendo, quando couber, todas as áreas de formação em todos os anos letivos;

VIII - a realização de devolutivas pedagógicas tempestivas; e

IX - o acesso a programas governamentais de financiamento ou apoio ao estudante da educação superior (BRASIL, 2020, p.2).

O MEC (BRASIL, 2020) ainda reforçou que o Saeb é composto por dimensões que se complementam a fim de fornecer um panorama geral da educação brasileira e identificar os fatores que influenciam para a composição de um ambiente educacional adequado.

Art. 7º O Saeb tem como referência as seguintes dimensões de qualidade para a avaliação da educação básica:

I - atendimento escolar;

II - ensino e aprendizagem;

III - investimento;

IV - profissionais da educação;

V - gestão;

VI - equidade; e

VII - cidadania, direitos humanos e valores (BRASIL, 2020, p.2).

Assim, diante do que foi estudado a respeito do uso dos resultados das avaliações, do ciclo de política pública e da suficiência e confiabilidade dos resultados obtidos por amostragem, a proposta aqui apresentada se mostra como alternativa ao projeto da Portaria nº 458/2020. No lugar da realização de avaliações anuais focando essencialmente na dimensão de ensino e aprendizagem, propõe-se a realização da avaliação subdividida em ciclos de 4 anos, com aplicações focadas em cada dimensão a ser avaliada. Ao invés de aplicações anuais censitárias para todas as etapas de ensino, recomenda-se a intercalação de coletas amostrais e censitárias alcançando os momentos de conclusão de cada etapa da trajetória de formação do aluno. Esse modelo proporcionará resultados mais ricos e completos que os atuais, com qualidade igual ou superior e em um processo menos oneroso. Ademais, proporcionará aos gestores tempo para compreender os resultados da avaliação, bem como para elaborar e implementar planos de intervenção voltados a melhoria de cada dimensão da qualidade educacional.

Em alternativa à tentativa de avaliação anual e simultânea de todas as dimensões de qualidade que fazem parte do Saeb, a avaliação pode ser realizada de forma fracionada abordando de forma mais profunda cada uma das dimensões. A dimensão de ensino e aprendizagem será avaliada por meio de testes para acompanhamento da evolução das proficiências médias, aplicados a cada dois anos aos alunos das etapas de conclusão de cada ciclo educacional, intercalando aplicação amostral e censitária. Esta proposta se baseia no entendimento de que o efeito das políticas públicas não é percebido de forma imediata, mas em intervalos mais longos. Dessa forma, a avaliação da dimensão de ensino e aprendizagem de

forma mais espaçada permite o acompanhamento da evolução dos efeitos das políticas implementadas, com custos bem inferiores aos custos de realização da avaliação censitária e anual. O maior espaçamento dos resultados de proficiência para escolas e Municípios não prejudica o atendimento dos objetivos do Saeb, além de abrir espaço para que outros elementos da qualidade educacional sejam avaliados, complementando o discernimento sobre o quadro educacional brasileiro.

Kellaghan, Greaney e Murray (2011) compararam avaliações nacionais baseadas no censo e baseadas em amostras e relataram suas vantagens. A avaliação amostral tem potencial para fornecer informações para o sistema educacional em geral, setores do sistema e fatores associados ao aproveitamento dos alunos. A avaliação censitária tem potencial para fornecer informações para escolas, turmas ou professores e alunos individualmente, além daquelas já fornecidas pela avaliação amostral. Apesar de a avaliação amostral possibilitar menos opções de desagregação dos dados, apresenta muitas vantagens tais como: a) o processo ser menos oneroso; b) oferecer informações sobre a qualidade do aprendizado dos alunos e o funcionamento do sistema educacional de forma fidedigna, sem necessidade de alcançar cada aluno individualmente; c) não identificar as escolas participantes evitando os efeitos negativos decorrentes da aplicação de sanções vinculadas aos resultados das avaliações censitárias; e d) poder ser aplicada com mais frequência, viabilizando a realização de uma sequência de avaliações voltadas a temas em discussão no momento atual (KELLAGHAN; GREANEY; MURRAY, 2011).

As vantagens listadas por Kellaghan, Greaney e Murray (2011) manifestam-se na proposta deste plano de ação, pois as dimensões que compõem o Saeb poderão ser avaliadas em pesquisas amostrais, abordando aspectos específicos relacionados às políticas públicas em implementação. Os questionários a serem aplicados poderão levantar informações mais aprofundadas a respeito dos elementos de cada uma das dimensões, sem a sobrecarga trazida ao respondente que é submetido a um instrumento muito longo. Com o estudo focado em apenas algumas das dimensões em cada ciclo de aplicação, é possível obter maior qualidade na resposta recebida.

Assim, a aplicação do Saeb ocorreria anualmente, mas abordando elementos diferentes em cada momento de coleta de dados. Nos anos ímpares, a avaliação focaria nos testes de desempenho e na identificação do contexto geral dos alunos e dos profissionais de educação. O foco estaria nas dimensões de ensino e aprendizagem, profissionais de educação e equidade, intercalando a coleta censitária e amostral a fim de que sejam fornecidas informações do contexto macro a cada dois anos e no contexto micro a cada quatro anos.

Complementarmente, o Inep poderá disponibilizar instrumentos de autoavaliação, em formato semelhante ao da Provinha Brasil⁴, para que as instituições de ensino tenham a possibilidade de avaliar o desempenho geral dos seus alunos durante cada exercício ou quando jugarem necessário.

Nos anos pares, a avaliação seria contextual voltada para o entendimento das demais dimensões: atendimento escolar, investimento, gestão e cidadania, direitos humanos e valores, com questionários aplicados aos gestores e profissionais de educação. Acompanhando o modelo intercalado, sugerido para os anos ímpares, a proposta aqui é intercalar a avaliação das dimensões investimento e gestão em um momento e, no seguinte, a avaliação de atendimento escolar e cidadania, direitos humanos e valores. Esse modelo intercalado permite a obtenção de informações mais precisas e completas, pois um modelo em que todas as dimensões são avaliadas simultaneamente, resulta em instrumentos demasiadamente longos ou a superficialidade dos dados coletados sobre cada tema. A dinâmica proposta é apresentada no Quadro 4.

Quadro 4 - Ciclo avaliativo proposto para o Saeb.

Ano do ciclo do Saeb	Dimensões alcançadas	Tipo de Aplicação
Ano 1	Ensino e aprendizagem Equidade Profissionais de educação	Censitária
Ano 2	Investimento Gestão	Amostral
Ano 3	Ensino e aprendizagem Equidade Profissionais de educação	Amostral
Ano 4	Atendimento escolar Cidadania, direitos humanos e valores	Amostral

Fonte: Elaborado pela autora

A adoção desse modelo de avaliação em ciclo de quatro anos traria grande economicidade em comparação com o modelo proposto pela Portaria/MEC nº 458/2020. O

⁴ A Provinha Brasil é uma avaliação diagnóstica que visa municiar professores e gestores educacionais com dados sobre o desenvolvimento de habilidades pelos seus alunos. Mais informações sobre a Provinha Brasil podem ser obtidas no portal do Inep: <http://portal.inep.gov.br/web/guest/provinha-brasil> (INEP, 2015b).

processo de aplicação torna-se menos oneroso, entretanto com a garantia de atendimento dos objetivos da avaliação, preservando a confiabilidade dos resultados e possibilitando maior aprofundamento. Tal modelo também favoreceria o uso adequado dos dados e o melhor entendimento da educação brasileira e dos fatores que influenciam diretamente na construção de um sistema educacional de qualidade. Essa possibilidade decorre do intervalo mais espaçado entre duas avaliações do mesmo elemento, permitindo que haja um período maior para se dedicar à análise de cada aspecto avaliado. Com isso, as equipes dedicadas a esses estudos poderiam se aprofundar mais no entendimento de cada fator, produzindo análises mais detalhadas sobre cada um deles, sem a expectativa de que seja explicado todo cenário educacional em uma única análise.

Para viabilizar a implementação desta proposta uma nova portaria precisaria ser publicada pelo MEC a fim de ajustar alguns pontos estabelecidos pela portaria vigente. Além disso, também precisaria ser revisto o sistema de metas do Ideb, especificamente no âmbito do município e da instituição escolar, pois os resultados de desempenho somente seriam fornecidos nesse contexto a cada quatro anos. Em contrapartida, os instrumentos de autoavaliação disponibilizados pelo Inep permitiriam que cada escola pudesse acompanhar anualmente a evolução individual do seu resultado para o Ideb, da mesma forma que os Municípios poderiam individualmente mobilizar sua rede de ensino para o acompanhamento deste indicador a partir dos instrumentos de autoavaliação. A adaptação do sistema de metas do Ideb seria simplesmente em relação aos prazos de divulgação dos resultados oficiais para Municípios e instituições escolares. A autoavaliação tende a reduzir o efeito do ranqueamento das escolas e desestimular a adoção de medidas que resultem em maiores resultados na avaliação de desempenho sem promover a real melhoria da qualidade da educação praticada. Essa tendência deriva da redução da divulgação de resultados para todas as escolas e que possam ser usados para fins comparativos entre as instituições; porém, anualmente cada entidade interessada poderia medir o seu próprio resultado, com vistas à promoção da melhoria das práticas educacionais adotadas.

Outro ajuste necessário seria a atualização dos questionários contextuais, incluindo a elaboração de novos instrumentos voltados para cada tema especificamente, pois os questionários atuais abordam as dimensões da qualidade da educação de maneira superficial. Diante disto, os dados coletados pelos questionários atuais oferecem apenas um vislumbre do que estudos mais acurados podem revelar.

Considerando que o Ideb tem metas definidas até o ano de 2021, a proposta é que a implementação do modelo sugerido seja iniciada a partir de 2023 a fim de que não haja

interrupção do acompanhamento das metas já estabelecidas. O prazo de 2023 também é importante para que haja tempo suficiente para que as equipes do Inep se organizem no sentido de adaptar os processos e instrumentos para implementação do novo modelo.

4.3 PROPOSTA DE CÁLCULO DE PESOS PARA OS DADOS CONTEXTUAIS

Ao longo desta dissertação evidenciou-se que a coleta de dados por meio amostral seria suficiente para a obtenção de resultados fidedignos sobre a população estudada. Também se compreendeu que a contextualização dos resultados das avaliações de desempenho é indispensável para o entendimento do cenário educacional. Diante disso, as propostas apresentadas nas seções 4.1 e 4.2 dedicaram-se ao aprimoramento dos procedimentos amostrais e à construção de um modelo que colete dados mais completos e aprofundados das sete dimensões que compõem o Saeb.

A última proposta trazida neste plano de ação delineia-se no sentido de complementar as propostas anteriores e de oferecer alternativa para contornar o uso inadequado dos resultados. A proposta visa favorecer a contextualização e a compreensão de que os resultados de desempenho dos alunos não são suficientes para estabelecer um diagnóstico completo.

Existem diversos motivos para a subutilização dos resultados das avaliações educacionais; entre eles, pode-se destacar a falha em comunicar as conclusões da avaliação e a falha em elaborar ações adequadas a partir das conclusões (KELLAGHAN; GREANEY; MURRAY, 2011). Entretanto, para atenuar essas motivações é necessária a compreensão completa dos resultados da avaliação. Nessa circunstância, a contextualização dos resultados de desempenho ganha importância, pois auxilia no entendimento de cada dimensão da qualidade educacional, complementando a informação trazida pelas proficiências. O entendimento do panorama completo, por meio de bons indicadores, permite a comunicação mais clara das conclusões e, conseqüentemente, favorece a elaboração de ações adequadas ao avanço da qualidade da educação.

Para que seja possível a construção de bons indicadores relacionados a cada uma das dimensões avaliadas, é necessário que os dados coletados por amostragem permitam a correta expansão dos resultados para toda a população avaliada. Esse processo somente é possível por meio da constituição de pesos adequados para cada nível de análise. Atualmente o Saeb trabalha apenas com a ponderação dos resultados baseados nos dados de proficiência no nível dos alunos. Tal medida tem impacto minimizado pela coleta dos dados censitários para a rede de ensino pública, mas fragiliza a geração de resultados contextuais no âmbito da rede particular.

No modelo proposto em que os dados seriam coletados majoritariamente de forma amostral, um novo conjunto específico de pesos para os questionários contextuais revela-se indispensável. Para tanto, a cada aplicação precisariam ser calculados os pesos referentes ao instrumento relativo a cada dimensão estudada.

O procedimento de cálculo dos pesos é o mesmo adotado atualmente, sendo atualizadas as referências relacionadas à probabilidade de seleção de cada elemento e os critérios de validação do instrumento respondido.

Adotando como referência inicial os indicadores que sintetizam os resultados medidos pela avaliação, seriam considerados aptas a receber um peso aquelas unidades que tiverem resposta preenchida pelo menos para os itens mais diretamente relacionados ao cálculo dos indicadores no instrumento. A partir da identificação dos instrumentos que retornaram do processo de aplicação com preenchimento aceitável, seria processado o cálculo de acordo com a unidade amostral responsável pela resposta do instrumento. Dessa forma, os instrumentos respondidos pelos alunos teriam o processo de cálculo como se dá o cálculo do peso associado às proficiências. Os pesos para os instrumentos associados aos gestores e profissionais de educação precisariam passar por pequenos ajustes, entretanto seguindo a mesma metodologia, descrita nos relatórios de Amostragem do Saeb.

Para exemplificar a aplicação desta proposta, poder-se-ia calcular pesos para os questionários coletados no Saeb 2017, considerando as amostras selecionadas na simulação realizada e comparar os resultados amostrais com o censitário, semelhantemente ao que foi feito para proficiência. Contudo o Indicador de Nível Socioeconômico, único indicador calculado atualmente pelo Inep a partir dos questionários contextuais do Saeb, não tem divulgação de resultados por aluno que permita uma simulação desse cálculo.

Apesar da impossibilidade em se exemplificar a proposta neste momento, pode-se esperar que as conclusões de validade dos resultados amostrais aplicam-se aos dados contextuais de forma semelhante que para as proficiências, visto que a metodologia da amostragem considerada se aplica tanto às médias quanto para proporções.

5 CONSIDERAÇÕES FINAIS

Esta dissertação desenvolveu-se em torno do cenário da avaliação educacional em larga escala, tendo o Saeb como elemento fulcral. O objetivo primordial foi compreender a metodologia de amostragem adotada no processo de coleta de dados do Saeb, e identificar oportunidades de melhoria por meio da comparação entre os resultados amostrais e a realidade da população avaliada medida por meio de dados censitários. Além disso, identificaram-se limitações do Saeb que dificultam o cumprimento dos seus propósitos.

Com o intuito de propor melhorias para dirimir ou minorar os efeitos das limitações encontradas, foi feita pesquisa documental sobre do histórico de surgimento das avaliações educacionais em larga escala e do Saeb, analisou-se o plano amostral do Saeb e sua evolução, testou-se a confiabilidade dos resultados obtidos por meio de Amostragem e discutiram-se as características, limitações e potencialidades das avaliações educacionais em larga escala. Para apresentar esta pesquisa, a dissertação foi organizada em três capítulos de desenvolvimento.

O primeiro capítulo de desenvolvimento (Capítulo 2) apresentou o problema estudado e caracterizou o Saeb como principal avaliação educacional brasileira. Abordou, ainda, o histórico de surgimento da avaliação educacional em larga escala, sua chegada ao Brasil e o caminho percorrido até a criação do Saeb. Explicou o papel da União na realização das avaliações brasileiras e a sobreposição dos sistemas de avaliação e descreveu a evolução do Saeb e do seu plano amostral ao longo dos anos. A principal dificuldade encontrada para realização desse levantamento foi a quantidade reduzida de documentos que registrem de forma detalhada a evolução do Saeb e dos seus planos amostrais. Entretanto, a leitura de diversos documentos e o conhecimento prático do processo permitiu que lacunas encontradas no estudo de um documento isolado fossem preenchidas.

Em seguida, foi construído um referencial teórico para o debate, o Capítulo 3 apresentou conceitos de Amostragem que permitem a compreensão dos aspectos teóricos e sua aplicação no contexto do Saeb com o detalhamento do seu plano amostral. Além disso, foi realizada uma simulação comparando os resultados obtidos na aplicação censitária do Saeb com aqueles estimados a partir de coletas amostrais que seguiram o desenho amostral atual do Saeb, ratificando que a metodologia amostral do Saeb é adequada, apesar de haver oportunidades de melhoria facilmente implementáveis. Concluindo a fundamentação teórica, foram apresentados elementos inerentes à avaliação educacional e à avaliação como elemento do ciclo de política pública. Com base nas informações levantadas, foi possível identificar limitações do Saeb e

categorizá-las em função de sua origem nos seguintes grupos: Limitações operacionais; Limitações técnicas, Limitações políticas e Limitações intrínsecas.

A partir daí, foi elaborado no Capítulo 4 um plano de ação composto de proposições para aprimoramento do Saeb, que possam solucionar ou contornar os efeitos das limitações identificadas em cada uma das categorias. Vale destacar que as limitações operacionais não foram incluídas no plano de ação, pois o Inep adota medidas suficientes no que diz respeito às limitações identificadas. No que tange às demais categorias, as medidas sugeridas envolvem a) ajustes na metodologia de cálculo do tamanho da amostra; b) ampliação do ciclo de aplicação da avaliação promovendo maior aprofundamento para a coleta de dados sobre as sete dimensões que o Saeb abarca; e c) cálculo de conjunto de pesos específicos para os dados contextuais, fortalecendo os resultados associados às dimensões que não se referem ao ensino e aprendizagem, facilitando o entendimento e uso dos resultados. As medidas que integram o plano de ação têm a finalidade comum de incrementar o conjunto de resultados oferecidos pelo Saeb, favorecendo a satisfação dos objetivos estabelecidos para a avaliação. Assim, o conteúdo apresentado nos três capítulos de desenvolvimento permitiu que a pergunta inicialmente proposta para esta dissertação fosse respondida, da mesma forma que foram cumpridos os objetivos estabelecidos.

A implementação do plano de ação proposto pode ser efetuada de maneira integral ou em partes, tendo como maior desafio a adoção do ciclo de aplicação em quatro anos, pois depende de decisões políticas que acontecem em âmbito distanciado do corpo técnico. Esse desafio foi majorado pela recente publicação da Portaria MEC nº 458/2020, que propõe uma evolução para o Saeb em direção oposta à sugerida no plano de ação. As mudanças infundem a ideia de que a ampliação da aplicação, com periodicidade mais frequente, maior abrangência e foco mais intenso nos resultados de desempenho acarretaria impactos diretos no entendimento do cenário educacional brasileiro e nas políticas públicas educacionais.

O ajuste da metodologia de cálculo do tamanho da amostra depende apenas de decisão da equipe técnica. Não prejudica a comparabilidade da série histórica, nem altera os custos do processo, uma vez que os resultados dos estudos realizados indicaram que a alteração da metodologia não traria muitos impactos de alterações no tamanho da amostra. Tampouco depende da realização dos demais ajustes propostos, portanto pode ser implementado a partir da próxima edição do Saeb. O cálculo do conjunto de pesos também independe das outras melhorias propostas, entretanto, por ser uma atividade nova depende de adaptações administrativas como inclusão no cronograma de atividades e possível reforço da equipe técnica. Apesar disso, o cálculo não exige aumento de custo, além daqueles relacionados ao

custo de pessoal em decorrência do aumento do tempo dedicado às atividades de cálculo de resultados do Saeb. Para melhor aproveitamento desse novo conjunto de pesos, poderiam ser desenvolvidos indicadores específicos para mensurar aspectos das sete dimensões de qualidade da Educação Básica, o que talvez exija alterações nos questionários contextuais.

O desenvolvimento da dissertação enfrentou dificuldades relacionadas à verificação do plano amostral quanto a sua aplicabilidade às sete dimensões do Saeb, atendo-se apenas à dimensão de ensino e aprendizagem, mensurada por meio das proficiências. O desenvolvimento da análise a respeito desse aspecto do Saeb foi refreado pela indisponibilidade de informação sobre os indicadores usados para mensuração das demais dimensões avaliadas, uma vez que não foram encontrados documentos que descrevam esses indicadores ou tornem públicos os seus resultados. Entretanto, apesar das dificuldades enfrentadas, o caminho percorrido para desenvolvimento dessa dissertação trouxe grande aprendizado, aprimorando o entendimento dos procedimentos atualmente adotados no meu dia a dia profissional. Com isso, será possível não apenas melhorar a capacitação de toda a equipe no que diz respeito ao plano amostral do Saeb, mas também abre portas para a melhoria do processo e para a realização de novos estudos relacionados a esse tema.

Diante disso, para estudos futuros, recomenda-se que seja ajustado o cálculo dos pesos para tratamento da não resposta dos questionários, viabilizando a melhoria no cálculo de indicadores educacionais, a fim de detalhar e aprofundar a afirmação de que o plano amostral proposto atende tão bem aos dados contextuais quanto aos dados de proficiência. Indica-se ainda que sejam desenvolvidos estudos que tenham como objetivo a construção de indicadores adequados à mensuração de cada uma das dimensões avaliadas pelo Saeb, bem como a elaboração de instrumentos voltados para a coleta desses dados específicos. Cabe ressaltar que dados bastante relevantes são coletados pelo Censo Escolar e que alguns indicadores são compilados a partir desses dados. Portanto, estudos futuros podem propor uma complementação desse sistema de indicadores e sua associação a cada dimensão avaliada. Outro importante estudo que ainda pode ser desenvolvido é a respeito do ajuste da estimação do EPA, levando em consideração que os tamanhos de amostras sejam diferentes de acordo com as características dos estratos, como é o caso do Saeb, em que o tamanho da amostra varia consideravelmente dependendo do porte da escola.

Esta dissertação não se propõe a esgotar o debate sobre as possibilidades de melhoria do Saeb ou das possíveis metodologias de Amostragem aplicáveis a esse processo. O propósito foi de mostrar que, apesar das diversas oportunidades de melhoria, a metodologia atual é

confiável e que a Amostragem permite o aprofundamento da avaliação, sem a necessidade de ampliação da coleta censitária.

REFERÊNCIAS

ANDRADE, Dalton Francisco. Memorando Técnico Projeto 1 – **Definição do universo da pesquisa**. Para: MEC/Inep. Projeto: Plano amostral Saeb 2011, [s.l.], Jul. 2011a.

ANDRADE, Dalton Francisco. Memorando Técnico **Plano amostral para o Saeb 2011**. Para: MEC/Inep. Projeto: Plano amostral Saeb 2011, [s.l.], Ago. 2011b.

ANDRADE, D.F.; SILVA, P. L. N.; BUSSAB, W.O. Memorando Técnico nº 01/01 – **Plano Amostral para o Saeb01**. Para: MEC/Inep. Projeto: MEC-Inep-Saeb01, [s.l.], Abr. 2001.

ANDRADE, Dalton Francisco; TAVARES, Heliton Ribeiro; VALLE, Raquel da Cunha. **Teoria de Resposta ao Item**: conceitos e aplicações. São Paulo: Associação Brasileira de Estatística, 2000.

ARAÚJO, Carlos Henrique; LUZIO, Nildo. **Avaliação da Educação Básica**: em busca da qualidade e equidade no Brasil. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, Brasília, 2005. Disponível em <<http://portal.inep.gov.br/documents/186968/484421/Avalia%C3%A7%C3%A3o+da+educa%C3%A7%C3%A3o+b%C3%A1sica+em+busca+da+qualidade+e+eq%C3%BCidade+no+Brasil/c7af7a6c-3cc6-473d-a7b5-ace3f3e499ee?version=1.1>>. Acesso em 13 mai. 2019.

ARELLANO, David Gault; LEPORE, Walter; ZAMUDIO, Emilio; BLANCO, Felipe. **Sistemas de evaluación del desempeño para organizaciones públicas**: ¿Como construirlos efectivamente? México, D.F.: Centro de Investigación y Docencia Económicas, 2012.

BAUER, Adriana; ALAVARSE, Ocimar Munhoz; OLIVEIRA, Romualdo Portela. Avaliações em larga escala: uma sistematização do debate. **Educação e Pesquisa**, v. 41, nº esp., p. 1367-1384, dez. 2015, Disponível em: <<http://www.scielo.br/pdf/ep/v41nspe/1517-9702-ep-41-spe-1367.pdf>>. Acesso em: 11 set. 2019.

BAUER, Adriana. Avaliação de Redes de Ensino e Gestão Educacional: aportes teóricos. **Educação & Realidade**, Porto Alegre, v. 44, n. 1, 2019. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S2175-62362019000100603&lng=en&nrm=iso>. Acesso em: 10 abr. 2020.

BINDER, David A.; ROBERTS, Georgia R. Design-based and model-based methods for estimating model parameters. In: CHAMBERS, R. L.; SKINNER, C. J. (Org.). **Analysis of survey data**. West Sussex: John Wiley & Sons, 2003.

BOLFARINE, Heleno, BUSSAB, Wilton O. **Elementos de Amostragem**. São Paulo: Blucher, 2005.

BRASIL. **Lei nº 378**, de 13 de janeiro de 1937. Dá nova organização ao Ministério da Educação e Saúde Pública. Diário Oficial da União, Rio de Janeiro, 15 jan. 1937.

_____. **Decreto-Lei nº 580**, de 30 de julho de 1938. Dispõe sobre a organização do Instituto Nacional de Estudos Pedagógicos. Diário Oficial da União, Rio de Janeiro, 30 jul. 1938.

_____. **Decreto nº 71.407**, de 20 de novembro de 1972. Dispõe sobre o Instituto Nacional de Estudos e Pesquisas Educacionais, do Ministério da Educação e Cultura, e dá outras providências. Diário Oficial da União, Rio de Janeiro, 22 nov. 1972.

_____. **Constituição da República Federativa do Brasil**. Brasília: Senado Federal, 1988.

_____. **Lei nº 9.394**, de 20 de dezembro de 1996. Estabelece as diretrizes e bases da educação nacional. Diário Oficial da União, Brasília, 23 dez 1996.

_____. **Lei nº 9.448**, de 14 de março de 1997. Transforma o Instituto Nacional de Estudos e Pesquisas Educacionais - Inep em Autarquia Federal, e dá outras providências. Diário Oficial da União, Brasília, 15 mar 1997.

_____. **Lei nº 10.269**, de 29 de agosto de 2001. Dá nova denominação ao Instituto Nacional de Estudos e Pesquisas Educacionais - Inep. Diário Oficial da União, Brasília, 30 ago. 2001.

_____. **Lei nº 13.005/2014**, de 25 de julho de 2014. Aprova o Plano Nacional de Educação - PNE e dá outras providências. Brasília: Diário Oficial da União, 26 jun. 2014.

_____. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep. **Portaria nº 69**, de 4 de maio de 2005. Diário Oficial da União, Brasília, DF, 5 de maio de 2005a.

_____. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep. **Portaria nº 87**, de 7 de maio de 2009. Diário Oficial da União, Brasília, DF, 8 de maio de 2009.

_____. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep. **Portaria nº 447**, de 24 de maio de 2017. Estabelece diretrizes para o planejamento e a operacionalização do Sistema de Avaliação da Educação Básica (Saeb) no ano de 2017. Diário Oficial da União, Brasília, 25 maio 2017a.

_____. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep. **Portaria nº 986**, de 21 de dezembro de 2017. Aprova o Regimento Interno do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - Inep. Diário Oficial da União, Brasília, 22 dez. 2017b.

_____. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep. **Portaria nº 366**, de 29 de abril de 2019. Estabelece diretrizes para o planejamento e a operacionalização do Sistema de Avaliação da Educação Básica (Saeb) no ano de 2019. Diário Oficial da União, Brasília, 2 maio 2019.

_____. Ministério da Educação. **Portaria nº 931**, de 21 de março de 2005. Diário Oficial da União, Brasília, DF, 22 mar. 2005b.

_____. Ministério da Educação. **Portaria nº 458**, de 05 de maio de 2020. Diário Oficial da União, Brasília, 6 maio 2020.

BROOKE, Nigel. Crise cultural. In: BROOKE, N. (Org.). **Marcos históricos na reforma da educação**. Belo Horizonte, MG: Fino Traço, 2012.

- BROOKE, Nigel; CUNHA, Maria Amália A. A avaliação externa como instrumento da gestão educacional nos estados. **Estudos & Pesquisas Educacionais**, n. 2, nov. 2011. Disponível em: https://abrilfundacaovictorcivita.files.wordpress.com/2018/04/estudos_e_pesquisas_educacionais_vol_2.pdf. Acesso em: 15 set. 2019.
- BUSSAB, W.O.; ANDRADE, D.F.; SILVA, P. L. N.; FREITAS, M. P. S. Memorando Técnico nº 02/99 – **Definição do Plano Amostral**. Para: MEC/Inep. Projeto: Plano Amostral Saeb-99, [s.l.], out. 1999.
- BUSSAB, Wilton de Oliveira; MIAZAKI, Édina Shisue; RABELLO, Maria Cecília Ramalho. Memorando Técnico nº 01 – **Definição do universo da pesquisa**. Para: MEC/Inep. Projeto: Plano Amostral Saeb2005, [s.l.], ago. 2005a.
- _____. Memorando Técnico nº 03/03 – **Plano amostral para o Saeb2005**. Para: MEC/Inep. Projeto: Plano Amostral Saeb2005, [s.l.], set. 2005b.
- BUSSAB, Wilton O.; MORETIN, Pedro A. **Estatística Básica**. São Paulo: Saraiva, 2004.
- BUSSAB, W.O.; NISHIMURA, R., Memorando Técnico nº 03/03 – **Plano amostral para o Saeb2009**. Para: MEC/Inep. Projeto: Plano Amostral Saeb2009, [s.l.], out. 2009.
- CALDERÓN, Adolfo Ignacio. Usos e apropriações das avaliações em larga escala: tensões e desafios. In: QUIOSSA, Amanda Sangy (Org.). **Série Diálogos e Proposições: planos de ação para a Rede Estadual de Educação de Minas Gerais**, v. II. Juiz de Fora: Projeto CAEd-FADEPE/JF, 2017. Disponível em: <http://www.mestrado.caedufjf.net/wp-content/uploads/2018/01/DI%C3%81LOGOS-E-PROPOSI%C3%87%C3%95ES-vol-II-WEB.pdf>. Acesso em: 10 abr. 2020.
- CARNOY, Martin; CASTRO, Claudio de Moura. A melhoria da educação na América Latina: e agora, para onde vamos? In: BROOKE, Nigel (Org.). **Marcos históricos na reforma da educação**. Belo Horizonte: Fino Traço, 2012.
- CASELLA, George; BERGER, Roger L. **Inferência Estatística**. São Paulo: Cengage Learning, 2016.
- CHAMBERS, R. L. Introduction to part A. In: CHAMBERS, R. L.; SKINNER, C. J. (Org.). **Analysis of survey data**. West Sussex: John Wiley & Sons, 2003.
- COCHRAN, William Gemmill. **Sampling Techniques**. New York: John Wiley & Sons, 1977.
- COELHO, Maria Inês de Matos. Vinte anos de avaliação da educação básica no Brasil: aprendizagens e desafios. **Ensaio: Avaliação e Políticas Públicas em Educação**, Rio de Janeiro, v. 16, n. 59, p. 229-258, abr./jun. 2008. Disponível em <http://www.scielo.br/pdf/ensaio/v16n59/v16n59a05.pdf>. Acesso em: 29 maio 2019.
- DAMBROS, Marlei; MUSSIO, Bruna Roniza. Política educacional brasileira: a reforma dos anos 90 e suas implicações. **Anais do X Anped Sul**. Florianópolis: 2014.

DUARTE, Denise; BATISTA, José Rodrigues; ALVES, Reginaldo Wemersom. Relatório 01: **Definição do universo da pesquisa**. Plano amostral Saeb 2007. Herkenhof & Prates Tecnologia e Desenvolvimento, [s.l.], ago. 2007a.

_____. Relatório 02: **Definição do plano amostral**. Plano amostral Saeb 2007. Herkenhof & Prates Tecnologia e Desenvolvimento, [s.l.], ago. 2007b.

FERNANDES, Reynaldo; GREMAUD, Amaury Patrick. Qualidade da educação: avaliação, indicadores e metas. In: **Educação básica no Brasil: construindo o país do futuro**, 2009. Disponível em: https://www.cps.fgv.br/ibrecps/rede/seminario/reynaldo_paper.pdf. Acesso em: 12 abr. 2019.

FERREIRA, Aurélio Buarque de Holanda. **Miniaurélio**: O minidicionário da Língua Portuguesa. Curitiba: Posigraf, 2004.

FLETCHER, Philip R. **Relatório Técnico sobre a Amostra**. Microdados do Saeb 1995. Brasília: Inep, 2006. Disponível em: <http://portal.inep.gov.br/web/guest/microdados>>. Acesso em: 08 abr. 2019.

FREITAS, Dirce Nei Teixeira. A avaliação da educação básica no Brasil: dimensão normativa, pedagógica e educativa In: **Anais da 28ª Reunião Anual da Anped**. Caxambu, 2005. Disponível em: <http://www.anped.org.br/biblioteca/item/avaliacao-da-educacao-basica-no-brasil-dimensao-normativa-pedagogica-e-educativa>. Acesso em: 30 maio 2019.

GROVES, Robert M.; LYBERG, Lars. Total survey error – Past, present and future. **Public Opinion Quarterly**, v. 74, n. 5, p. 849-879, 2010. Disponível em: <https://bit.ly/2Q8b8M4>. Acesso em: 11 mar. 2020.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Microdados do Saeb 1995**. Brasília: Inep, 2006a. Disponível em: <http://portal.inep.gov.br/web/guest/microdados>. Acesso em: 1 mar. 2019.

_____. **Microdados do Saeb 1997**. Brasília: Inep, 2006b. Disponível em: <http://portal.inep.gov.br/web/guest/microdados>. Acesso em: 7 mar. 2019.

_____. **Microdados do Saeb 1999**. Brasília: Inep, 2006c. Disponível em: <http://portal.inep.gov.br/web/guest/microdados>. Acesso em: 1 mar. 2019.

_____. **Microdados do Saeb 2001**. Brasília: Inep, 2006d. Disponível em: <http://portal.inep.gov.br/web/guest/microdados>. Acesso em: 1 mar. 2019.

_____. **Microdados do Saeb 2003**. Brasília: Inep, 2006e. Disponível em: <http://portal.inep.gov.br/web/guest/microdados>. Acesso em: 11 mar. 2019.

_____. **Microdados do Saeb 2005**. Brasília: Inep, 2008. Disponível em: <http://portal.inep.gov.br/web/guest/microdados>. Acesso em: 14 mar. 2019.

_____. **Microdados do Saeb 2011**. Brasília: Inep, 2013. Disponível em: <http://portal.inep.gov.br/web/guest/microdados>. Acesso em: 14 mar. 2019.

_____. **Documento técnico que contemple a descrição do universo a ser pesquisado na Aneb 2013**. Produto 1. Inep, 2013a.

_____. **Documento técnico contendo o detalhamento do plano amostral da Aneb 2013.** Produto 2. Inep, 2013b.

_____. **Relatório da Amostragem da Aneb 2013.** Microdados da Aneb e da Anresc 2013. Brasília: Inep, 2014. Disponível em: <http://portal.inep.gov.br/web/guest/microdados>. Acesso em: 24 out. 2019.

_____. Conheça o Inep. **Portal do Inep**, 2015a. Disponível em: <http://portal.inep.gov.br/web/guest/conheca-o-inep>. Acesso em: 23 ago. 2019.

_____. Provinha Brasil. **Portal do Inep**, 2015b. Disponível em: <http://portal.inep.gov.br/web/guest/provinha-brasil>. Acesso em: 19 jun. 2020.

_____. **Relatório da Amostragem da Aneb 2015** .Microdados da Aneb e da Anresc 2015. Brasília: Inep, 2017. Disponível em: <http://portal.inep.gov.br/web/guest/microdados>. Acesso em: 28 mar. 2019.

_____. **Nota Técnica.** Indicador de Nível Socioeconômico das Escolas de Educação Básica (Inse). Brasília: Inep, [2017?]. Disponível em: http://download.inep.gov.br/informacoes_estatisticas/indicadores_educacionais/2015/nota_tecnica/nota_tecnica_inep_inse_2015.pdf. Acesso em: 27 ago. 2020.

_____. **Relatório da Amostragem da Aneb 2017.** Microdados da Aneb e da Anresc 2017. Brasília: Inep, 2018a. Disponível em: <http://portal.inep.gov.br/web/guest/microdados>. Acesso em: 26 mar. 2019.

_____. **Nota Técnica.** Detalhamento da população e resultados do Saeb 2017. Microdados da Aneb e da Anresc 2017. Brasília: Inep, 2018b. Disponível em: <http://portal.inep.gov.br/web/guest/microdados>. Acesso em: 26 mar. 2019.

_____. **Microdados do Saeb 2017.** Brasília: Inep, 2018c. Disponível em: <http://portal.inep.gov.br/web/guest/microdados>. Acesso em: 26 mar. 2019.

_____. **Sistema de Avaliação da Educação Básica: documentos de referência versão 1.0.** Brasília: Inep, 2018d. Disponível em: http://download.inep.gov.br/educacao_basica/saeb/2018/documentos/saeb_documentos_de_referencia-versao_1.0.pdf. Acesso em: 30 out. 2019.

_____. **Microdados do Censo da Educação Básica 2017.** Brasília: Inep, 2018e. Disponível em: <http://portal.inep.gov.br/web/guest/microdados>. Acesso em: 18 ago. 2020.

_____. **Nota Técnica nº 10/2019/CGIM/DAEB.** Detalhamento da população e resultados do Saeb 2019. Inep, 2019a. Disponível em: http://download.inep.gov.br/educacao_basica/saeb/2019/documentos/Nota_tecnica_Detalhamento_da_populacao_e_resultados_do_Saeb_2019.pdf. Acesso em: 24 out. 2019.

_____. Catálogo de escolas. **Portal do Inep**, 2019b. Disponível em <http://portal.inep.gov.br/web/guest/dados/catalogo-de-escolas>. Acesso em: 29 abr. 2020.

_____. **Relatório SAEB.** Brasília: INEP, 2019c. Disponível em <http://portal.inep.gov.br/documents/186968/484421/RELAT%C3%93RIO+SAEB+2017/fe63936-8002-43b6-b741-4ac9ff39338f?version=1.0>. Acesso em 6 de jul. 2020.

_____. Histórico. **Portal do Inep**, 2019d. Disponível em, <http://portal.inep.gov.br/web/guest/educacao-basica/saeb/historico>. Acesso em: 15 ago. 2020.

_____. Ideb. **Portal do Inep**, 2020. Disponível em: <http://portal.inep.gov.br/ideb>. Acesso em: 1 mar. 2020.

JANNUZZI, Paulo de Martino. Avaliação de Programas Sociais no Brasil: Repensando Práticas e Metodologias das Pesquisas Avaliativas. **Planejamento e Políticas Públicas**, n. 36, p.251-275, jan./jul. 2011. Disponível em: <http://www.ipea.gov.br/ppp/index.php/PPP/article/view/228/212>. Acesso em: 19 dez. 2019.

KELLAGHAN, Thomas; GREANEY, Vicent; MURRAY, T. Scott. **O uso dos resultados da avaliação de desempenho educacional**. Rio de Janeiro: Elsevier, 2011.

KLEIN, Ruben. Utilização da Teoria de Resposta ao Item no Sistema Nacional de Avaliação da Educação Básica (SAEB). **Revista Meta: Avaliação**, v. 1, n. 2, p. 125-140, set. 2009. Disponível em: <http://revistas.cesgranrio.org.br/index.php/metaavaliacao/article/view/38>. Acesso em: 6 jul. 2020.

LITTLE, Roderick J. Bayesian methods for unit and item nonresponse. In: CHAMBERS, R. L.; SKINNER, C. J. (Org.). **Analysis of survey data**. West Sussex: John Wiley & Sons, 2003.

LUMLEY, Thomas. Analysis of complex survey samples. **Journal of Statistical Software**, v. 9, issue 6, abr. 2004. Disponível em: <https://www.jstatsoft.org/article/view/v009i08>. Acesso em: 11 dez. 2019.

MACHADO, Cristiane; ALAVARSE, Ocimar Munhoz; ARCAS, Paulo Henrique. Sistemas estaduais de avaliação: interfaces com qualidade e gestão da educação. **Revista Brasileira de Política e Administração da Educação**, v. 31, n. 3, p. 667-680, set./dez. 2015. Disponível em: <https://seer.ufrgs.br/rbpae/article/download/63800/37029>. Acesso em: 7 out. 2019.

MANIFESTO AO POVO E AO GOVERNO. Manifesto de educadores democratas em defesa do ensino público (1959); Mais uma vez convocados. In: BROOKE, Nigel (Org.). **Marcos históricos na reforma da educação**. Belo Horizonte: Fino Traço, 2012.

MORRIS, Tim P.; WHITE, Ian R.; CROWTHER, Michael J. Using simulation studies to evaluate statistical methods. **Statistics in Medicine**, v. 38, p. 2074-2102, jan. 2019. Disponível em: <https://doi.org/10.1002/sim.8086>. Acesso em: 8 nov. 2019.

NSF, Fundação Nacional de Ciências. Do Sputnik à era de ouro, 1958-1968. In: BROOKE, Nigel (Org.). **Marcos históricos na reforma da educação**. Belo Horizonte: Fino Traço, 2012a.

NSF, Fundação Nacional de Ciências. O programa de aperfeiçoamento do conteúdo de ensino. In: BROOKE, Nigel (Org.). **Marcos históricos na reforma da educação**. Belo Horizonte: Fino Traço, 2012b.

PESSOA, Djalma Galvão Carneiro; SILVA, Pedro Luis do Nascimento. **Análise de dados amostrais complexos**. SINAPE, 1998.

PLACCO, Vera Maria Nigro de Souza; ALMEIDA, Laurinda Ramalho; SOUZA, Vera Lucia Trevisan. O coordenador pedagógico (CP) e a formação de professores: intenções, tensões e contradições. **Estudos & Pesquisas Educacionais**, n. 2, p. 227- 287, nov. 2011. Disponível em:

https://abrilfundacaovictorcivita.files.wordpress.com/2018/04/estudos_e_pesquisas_educacionais_vol_2.pdf. Acesso em: 24 mar. 2020.

PINHEIRO, Sandra Maria Conceição. **Modelo Linear Hierárquico: um método alternativo para análise de desempenho escolar**. Dissertação de Mestrado. Universidade Federal de Pernambuco. Recife, 2005. Disponível em

https://repositorio.ufpe.br/bitstream/123456789/6476/1/arquivo7241_1.pdf. Acesso em: 6 nov. 2019.

RAEDER, Savio. Ciclo de políticas: uma abordagem integradora dos modelos para análise de políticas públicas. **Perspectivas em Políticas Públicas**, v. VII, n. 13, p. 121-146, jan./jun. 2014. Disponível em;

<http://intranet.uemg.br/comunicacao/arquivos/PubLocal8P20150224110637.pdf>. Acesso em: 19 dez. 2019.

SILVA, Josiane Cristina da Costa. **Um estudo sobre a política e o material de divulgação de resultados da Prova Brasil**. Dissertação de Mestrado Profissional. Universidade Federal de Juiz de Fora. Juiz de Fora, 2013.

SILVA, Pedro Luis do Nascimento; MOURA, Fernando Antônio da Silva. **Efeito de Conglomeração da malha setorial do censo demográfico de 1980**. Fundação Instituto Brasileiro de Geografia e Estatística, 1990.

SILVA, P. L. N.; BUSSAB, W.O.; ANDRADE, D.F. Memorando Técnico – Fundação Cesgranrio nº 01/03 –**Definição do universo a ser investigado**. Para: MEC/Inep. Projeto: MEC-Inep-Saeb03, [s.l.], maio, 2003a.

_____. Memorando Técnico – Fundação Cesgranrio nº 03/03 –**Plano Amostral para o Saeb03**. Para: MEC/Inep. Projeto: MEC-Inep-Saeb03, [s.l.], jun. 2003b.

SOBEL, Irvin. A revolução do capital humano no desenvolvimento econômico: sua história e status atual. In: BROOKE, Nigel (Org.). **Marcos históricos na reforma da educação**. Belo Horizonte: Fino Traço, 2012.

SOUZA, Celina. Políticas públicas: uma revisão da literatura. **Sociologias**, n. 16, p. 20-45, dez. 2006. Disponível em: <http://www.scielo.br/pdf/soc/n16/a03n16.pdf>. Acesso em: 10 abr. 2020.

SHULTZ, Theodore W. Investimento em capital humano. In: BROOKE, Nigel (Org.). **Marcos históricos na reforma da educação**. Belo Horizonte: Fino Traço, 2012.

SKRONDAL, Anders; RABE-HESKETH, Sophia. **Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models**. Boca Raton: Chapman & Hall/CRC, 2004.

SOUZA, Sandra Maria Zákia Lian; OLIVEIRA, Romualdo Portela. Sistemas estaduais de avaliação: uso dos resultados, implicações e tendências. **Cadernos de Pesquisa**, v. 40, n. 141,

p. 793-822, set./dez. 2010. Disponível em:
<http://www.scielo.br/pdf/cp/v40n141/v40n141a07.pdf>. Acesso em: 7 out. 2019.

UNESCO. **A promessa das avaliações de aprendizagem em larga escala**: reconhecer os limites para desbloquear oportunidades. Paris, 2019. Disponível em:
<https://unesdoc.unesco.org/ark:/48223/pf0000372615>. Acesso em: 28 mar. 2020.

VIANNA, Heraldo Marelim. Avaliações nacionais em larga escala: análises e propostas. **Estudos em Avaliação Educacional**, v. 25, n. 60, p. 196-232, n. esp., dez. 2014. Disponível em: <http://publicacoes.fcc.org.br/ojs/index.php/eae/article/view/3336/2952>. Acesso em: 12 abr. 2019.

APÊNDICE A – Resultados da simulação realizada utilizando $\hat{\rho}_c$

Para verificar se o tamanho da amostra gerado com a correlação intraclasse recomendada por Silva e Moura (1990) promoveria melhores resultados que o modelo atual adotado pelo Saeb, foi realizada simulação considerando a fórmula de $\hat{\rho}_c$ para o dimensionamento amostral.

Tabela 15 - Tamanho da amostra em número de alunos e turmas (Simulação $\hat{\rho}_c$)

UF	n_{LP}	n_{MT}	n	\bar{M}	n_t
BR	65.408	66.907	70.256	22	3.186
RO	2.190	2.337	2.337	23	102
AC	2.980	3.379	3.379	24	140
AM	5.224	4.882	5.224	25	212
RR	2.708	3.153	3.153	22	145
PA	2.585	2.014	2.585	22	118
AP	2.122	1.585	2.122	21	99
TO	3.283	3.268	3.283	22	152
MA	2.090	1.651	2.090	21	99
PI	3.342	3.672	3.672	21	176
CE	3.201	4.697	4.697	23	208
RN	2.232	1.754	2.232	20	110
PB	2.077	1.864	2.077	19	107
PE	2.382	2.413	2.413	23	105
AL	3.442	3.643	3.643	24	150
SE	2.010	1.635	2.010	21	96
BA	2.221	1.902	2.221	20	110
MG	2.247	2.348	2.348	22	110
ES	1.644	1.686	1.686	21	80
RJ	2.043	2.173	2.173	24	90
SP	1.782	2.033	2.033	24	85
PR	1.963	2.490	2.490	22	113
SC	1.838	2.135	2.135	21	102
RS	1.841	1.918	1.918	18	106
MS	2.505	2.614	2.614	24	111
MT	1.924	1.934	1.934	22	90
GO	2.184	2.441	2.441	23	105
DF	1.347	1.286	1.347	21	65

Fonte: Elaborado pela autora

Tabela 16 - Comparativo de quantidades entre censo e amostra (Simulação $\hat{\rho}_c$)

Quantidade	Saeb	Amostra Esperada	Média realizada
Alunos	2.594.335	70.974	71.023
Turmas	98.688	3.205	3.205
Escolas	48.165	2.446	2.446

Fonte: Elaborado pela autora.

Tabela 17 - Medidas de desempenho para Língua Portuguesa (Simulação $\hat{\rho}_c$)

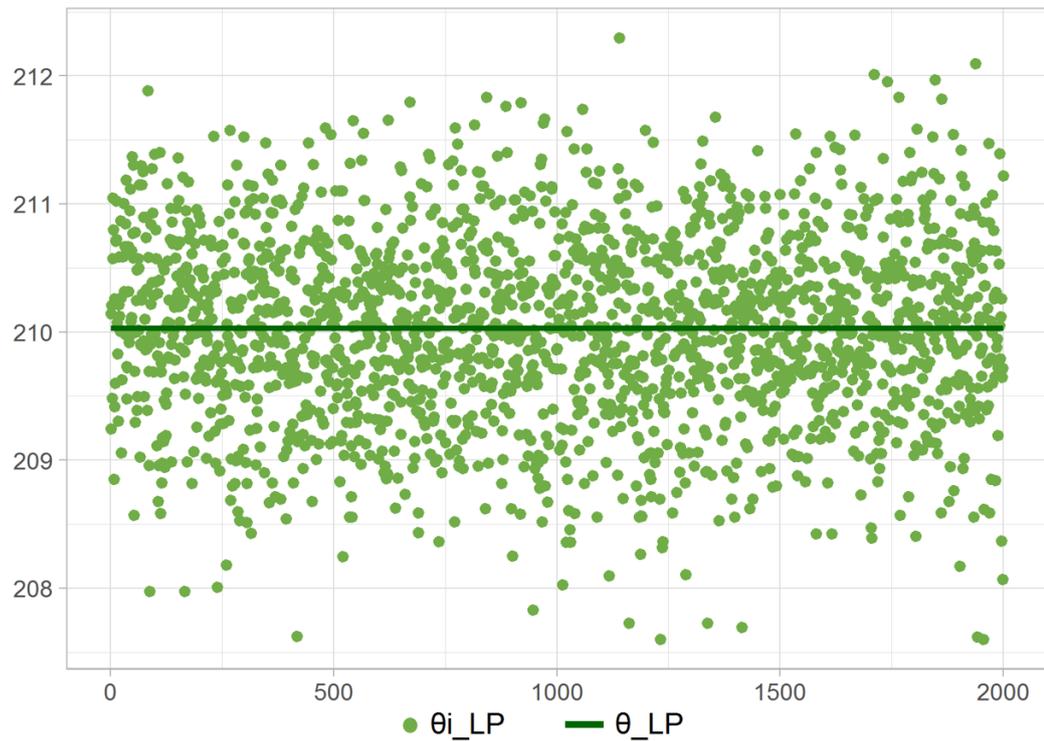
UF	θ_{LP}	$\bar{\theta}_{LP}$	Viés ($\hat{\theta}_{LP}$)	Var ($\hat{\theta}_{LP}$)	EQM ($\hat{\theta}_{LP}$)	RREQM ($\hat{\theta}_{LP}$)
BR	210,032	210,028	-0,004	0,52	0,52	0,34%
RO	210,473	210,577	0,104	4,46	4,47	1,00%
AC	214,617	214,587	-0,031	1,87	1,88	0,64%
AM	200,344	200,252	-0,093	2,28	2,29	0,76%
RR	200,919	201,385	0,466	0,89	1,11	0,52%
PA	185,042	184,978	-0,064	6,49	6,50	1,38%
AP	181,922	181,652	-0,270	3,15	3,22	0,99%
TO	200,965	201,001	0,036	3,18	3,18	0,89%
MA	178,594	178,579	-0,014	7,39	7,39	1,52%
PI	194,560	194,527	-0,033	3,31	3,31	0,93%
CE	218,037	218,055	0,019	3,99	3,99	0,92%
RN	185,933	185,995	0,063	5,01	5,01	1,20%
PB	190,039	189,961	-0,078	6,02	6,02	1,29%
PE	193,166	193,118	-0,048	5,90	5,90	1,26%
AL	189,818	189,861	0,043	6,73	6,73	1,37%
SE	182,485	182,395	-0,090	4,21	4,22	1,13%
BA	191,245	191,264	0,020	6,38	6,38	1,32%
MG	223,612	223,629	0,017	6,36	6,36	1,13%
ES	214,521	214,398	-0,123	6,57	6,59	1,20%
RJ	209,935	210,026	0,092	5,38	5,39	1,11%
SP	226,192	226,208	0,016	6,23	6,23	1,10%
PR	225,862	225,875	0,014	4,20	4,20	0,91%
SC	225,318	225,261	-0,057	4,84	4,85	0,98%
RS	214,500	214,356	-0,145	4,80	4,82	1,02%
MS	213,262	213,257	-0,006	3,67	3,67	0,90%
MT	205,995	205,934	-0,061	5,26	5,26	1,11%
GO	216,335	216,425	0,091	7,22	7,23	1,24%
DF	220,393	220,397	0,00	4,94	4,94	1,01%

Fonte: Elaborado pela autora

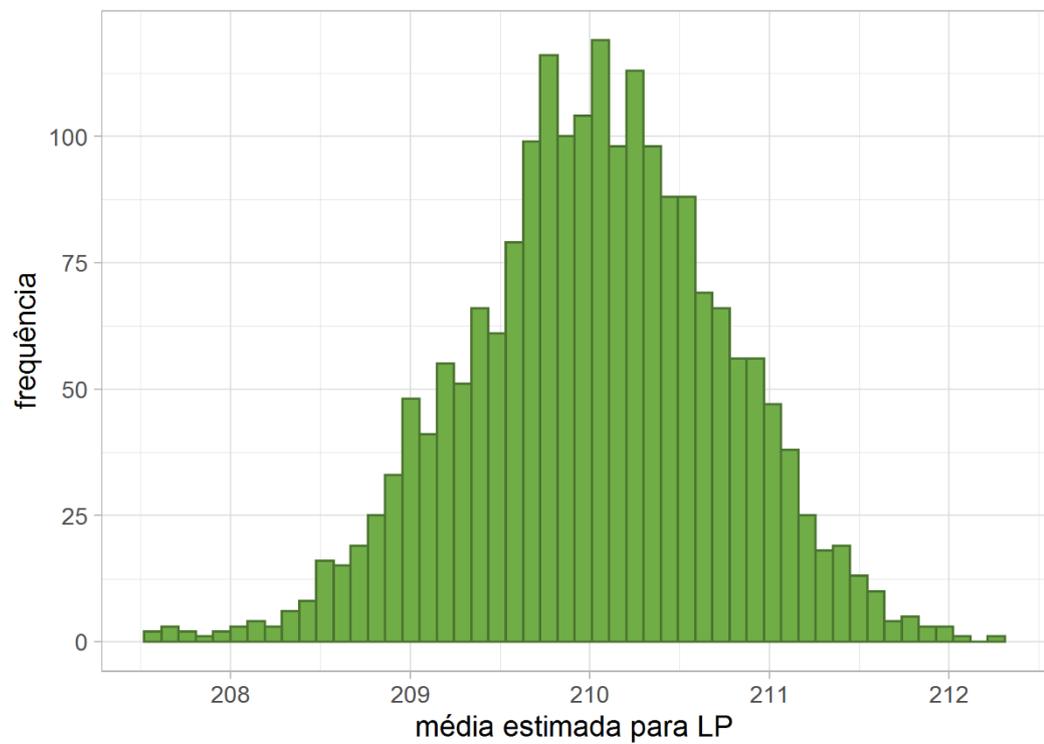
Tabela 18 - Medidas de desempenho para Matemática (Simulação $\hat{\rho}_c$)

UF	θ_{MT}	$\bar{\theta}_{MT}$	Viés ($\hat{\theta}_{MT}$)	Var ($\hat{\theta}_{MT}$)	EQM ($\hat{\theta}_{MT}$)	RREQM ($\hat{\theta}_{MT}$)
BR	219,503	219,492	-0,011	0,56	0,56	0,34%
RO	220,570	220,661	0,090	4,93	4,94	1,01%
AC	226,695	226,668	-0,027	2,22	2,22	0,66%
AM	209,516	209,410	-0,106	2,51	2,52	0,76%
RR	215,726	216,206	0,479	0,88	1,11	0,49%
PA	192,823	192,800	-0,022	5,40	5,41	1,21%
AP	188,880	188,655	-0,225	2,65	2,70	0,87%
TO	211,160	211,159	-0,001	3,21	3,21	0,85%
MA	186,575	186,519	-0,057	6,03	6,03	1,32%
PI	204,761	204,727	-0,034	3,29	3,29	0,89%
CE	226,109	226,116	0,007	5,76	5,76	1,06%
RN	193,595	193,660	0,066	4,13	4,14	1,05%
PB	199,972	199,859	-0,113	5,79	5,81	1,20%
PE	202,692	202,640	-0,053	6,02	6,03	1,21%
AL	201,687	201,713	0,026	7,50	7,50	1,36%
SE	193,417	193,346	-0,071	3,48	3,48	0,96%
BA	199,520	199,506	-0,013	5,57	5,57	1,18%
MG	231,758	231,805	0,046	6,66	6,66	1,11%
ES	223,558	223,470	-0,088	6,70	6,71	1,16%
RJ	217,858	217,876	0,019	5,70	5,70	1,10%
SP	237,826	237,815	-0,012	7,45	7,45	1,15%
PR	239,139	239,151	0,012	5,35	5,35	0,97%
SC	234,345	234,289	-0,056	5,89	5,89	1,04%
RS	223,483	223,388	-0,095	5,11	5,12	1,01%
MS	220,881	220,867	-0,014	4,05	4,05	0,91%
MT	214,341	214,291	-0,051	5,36	5,36	1,08%
GO	222,791	222,928	0,136	7,84	7,86	1,26%
DF	229,203	229,163	-0,040	5,02	5,02	0,98%

Fonte: Elaborado pela autora

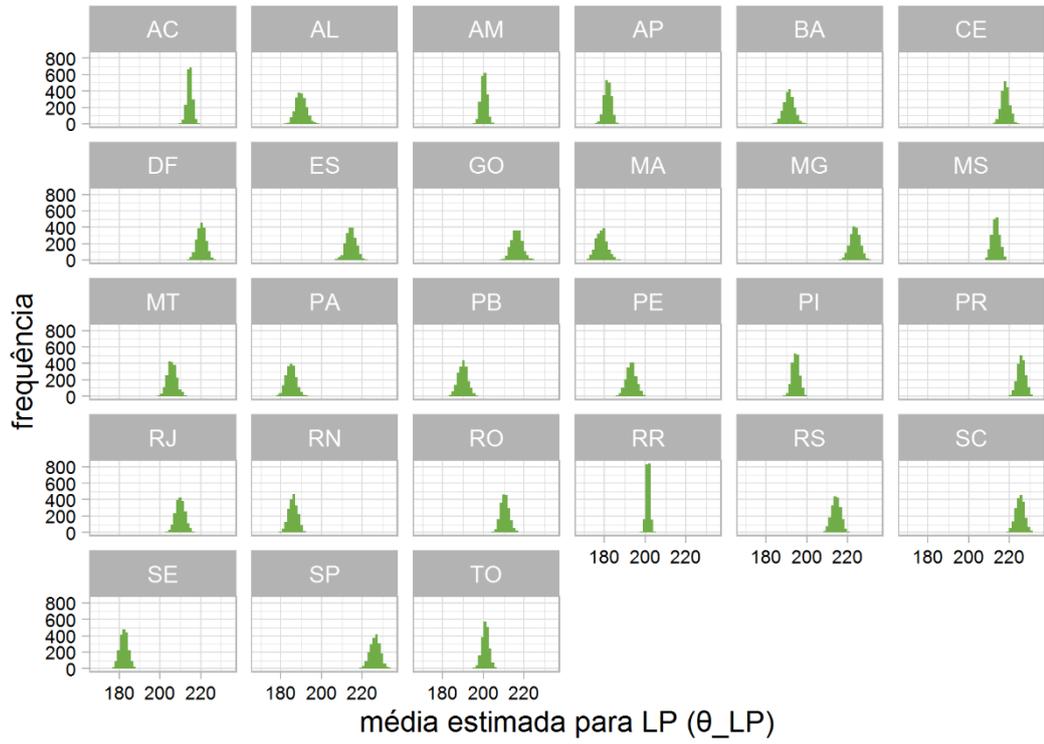
Gráfico 8 - Dispersão dos valores de $\hat{\theta}_{iLP}$ para o Brasil (Simulação $\hat{\rho}_c$)

Fonte: Elaborado pela autora

Gráfico 9 - Distribuição dos valores de $\hat{\theta}_{iLP}$ para o Brasil (Simulação $\hat{\rho}_c$)

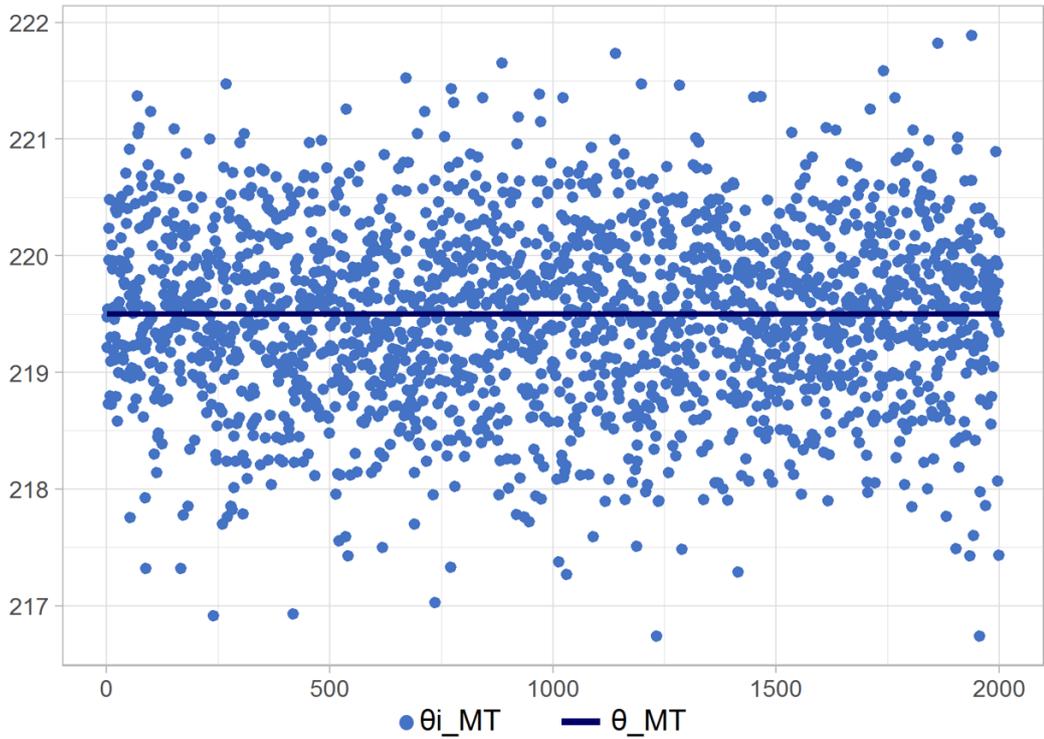
Fonte: Elaborado pela autora

Gráfico 10 - Distribuição dos valores de $\hat{\theta}_{iLP}$ por UF (Simulação $\hat{\rho}_c$)



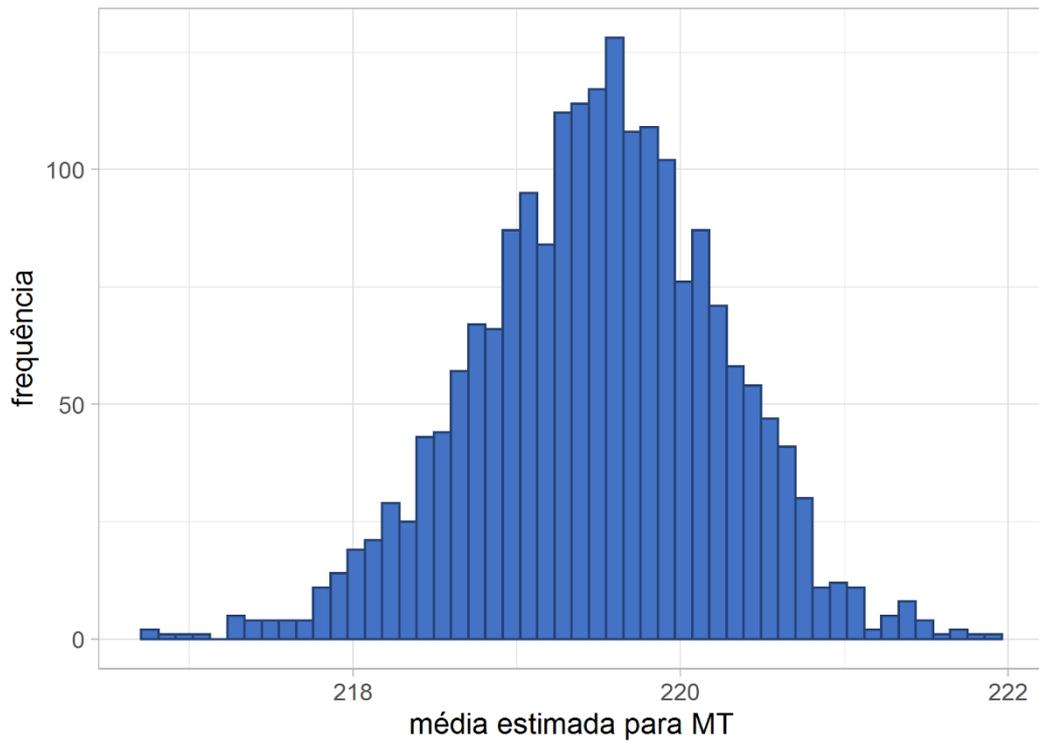
Fonte: Elaborado pela autora

Gráfico 11 - Dispersão dos valores de $\hat{\theta}_{iMT}$ para o Brasil (Simulação $\hat{\rho}_c$)



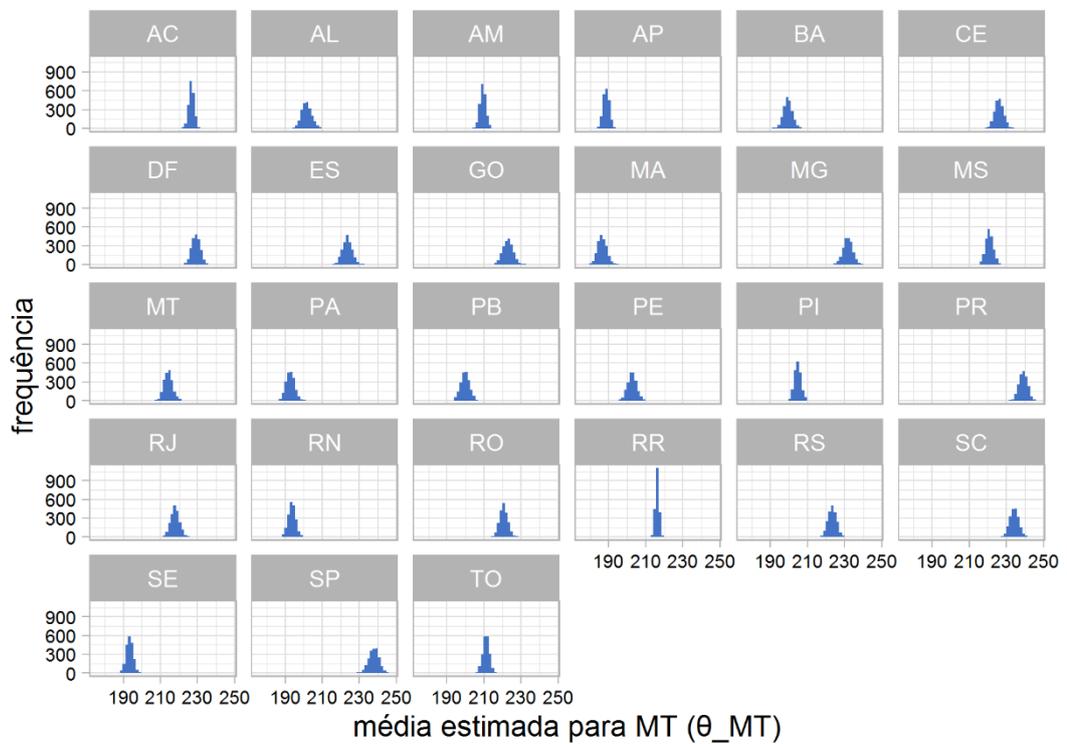
Fonte: Elaborado pela autora

Gráfico 12 - Distribuição dos valores de $\hat{\theta}_{iMT}$ para o Brasil (Simulação $\hat{\rho}_c$)



Fonte: Elaborado pela autora

Gráfico 13 - Distribuição dos valores de $\hat{\theta}_{iMT}$ por UF (Simulação $\hat{\rho}_c$)



Fonte: Elaborado pela autora

Tabela 19 - Cobertura dos intervalos de confiança e Margem de Erro (Simulação $\hat{\rho}_c$)

UF	Cobertura LP	Cobertura MT	$ME(\hat{\theta}_{LP})$	$ME(\hat{\theta}_{MT})$
BR	96,2%	95,8%	1,50	1,58
RO	97,3%	96,8%	4,73	4,98
AC	99,6%	99,3%	4,11	4,39
AM	97,8%	97,3%	3,48	3,59
RR	99,8%	99,9%	3,39	3,42
PA	95,3%	95,0%	5,18	4,80
AP	97,0%	97,6%	4,12	3,83
TO	98,3%	98,3%	4,45	4,43
MA	95,0%	94,9%	5,54	5,00
PI	97,5%	97,6%	4,13	4,18
CE	95,8%	96,0%	4,18	5,05
RN	96,5%	95,5%	4,71	4,22
PB	95,6%	95,1%	5,05	4,91
PE	95,4%	95,7%	4,98	5,04
AL	95,4%	94,8%	5,46	5,85
SE	95,5%	95,4%	4,21	3,88
BA	94,5%	95,0%	4,97	4,66
MG	95,5%	95,0%	5,19	5,30
ES	94,6%	94,8%	5,10	5,12
RJ	95,7%	95,2%	4,75	4,87
SP	94,6%	94,6%	4,97	5,43
PR	96,1%	96,6%	4,20	4,85
SC	96,3%	95,9%	4,66	5,09
RS	95,2%	95,3%	4,39	4,54
MS	96,9%	96,0%	4,05	4,35
MT	96,0%	95,8%	4,75	4,81
GO	94,8%	95,1%	5,40	5,81
DF	97,3%	97,9%	5,04	5,13

Fonte: Elaborado pela autora

APÊNDICE B – Resultados da simulação realizada utilizando ρ_{int_2}

Para verificar se o tamanho da amostra gerado com a correlação intraclasse calculada a partir das variâncias estimadas pelo modelo hierárquico de dois níveis promoveria melhores resultados que o modelo atual adotado pelo Saeb, foi realizada simulação considerando a fórmula de ρ_{int_2} para o dimensionamento amostral.

Tabela 20 - Tamanho da amostra em número de alunos e turmas (Simulação ρ_{int_2})

UF	n_{LP}	n_{MT}	n	\bar{M}	n_t
BR	66.728	68.707	72.757	22	3.316
RO	2.218	2.341	2.383	23	105
AC	2.957	3.333	3.334	24	138
AM	5.043	4.878	5.098	25	207
RR	2.621	3.155	3.155	22	145
PA	2.442	1.869	2.442	22	112
AP	1.917	1.353	1.917	21	90
TO	3.432	3.563	3.587	22	166
MA	2.101	1.558	2.101	21	100
PI	3.773	3.980	3.999	21	192
CE	3.557	5.217	5.217	23	231
RN	2.349	1.823	2.349	20	116
PB	2.184	1.872	2.186	19	113
PE	3.257	3.573	3.582	23	157
AL	2.985	2.973	3.095	24	128
SE	2.025	1.625	2.025	21	98
BA	2.648	2.368	2.655	20	132
MG	2.580	2.607	2.673	22	125
ES	1.618	1.679	1.729	21	83
RJ	2.034	2.156	2.197	24	91
SP	1.774	2.004	2.020	24	85
PR	2.167	2.786	2.786	22	127
SC	1.778	2.045	2.049	21	98
RS	1.812	1.865	1.915	18	107
MS	2.183	2.367	2.386	24	102
MT	1.802	1.814	1.886	22	88
GO	2.126	2.590	2.591	23	112
DF	1.344	1.311	1.400	21	68

Fonte: Elaborado pela autora

Tabela 21 - Comparativo de quantidades entre censo e amostra (Simulação ρ_{int_2})

Quantidade	Saeb	Amostra Esperada	Média realizada
Alunos	2.594.335	73.551	73.632
Turmas	98.688	3.326	3.326
Escolas	48.165	2.543	2.543

Fonte: Elaborado pela autora.

Tabela 22 - Medidas de desempenho para Língua Portuguesa (Simulação ρ_{int_2})

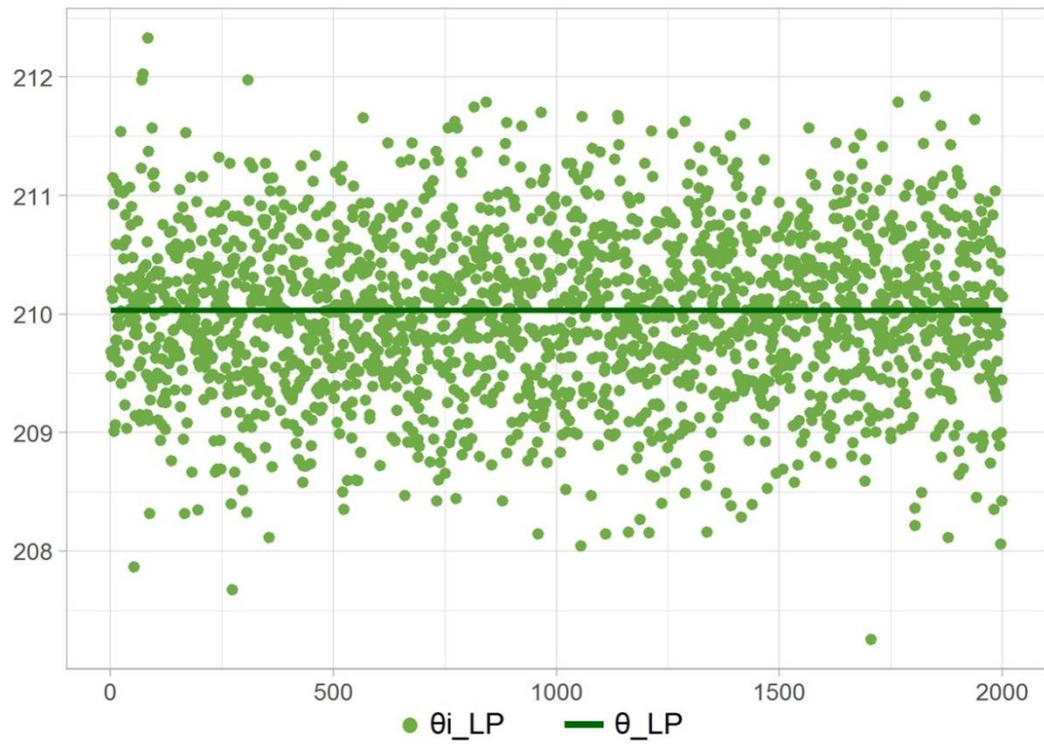
UF	θ_{LP}	$\bar{\theta}_{LP}$	Viés ($\hat{\theta}_{LP}$)	Var ($\hat{\theta}_{LP}$)	EQM ($\hat{\theta}_{LP}$)	RREQM ($\hat{\theta}_{LP}$)
BR	210,03	210,01	-0,03	0,47	0,47	0,33%
RO	210,47	210,57	0,10	4,38	4,39	1,00%
AC	214,62	214,58	-0,04	1,91	1,91	0,64%
AM	200,34	200,32	-0,03	2,16	2,16	0,73%
RR	200,92	201,41	0,50	0,84	1,09	0,52%
PA	185,04	184,97	-0,07	6,66	6,66	1,39%
AP	181,92	181,67	-0,25	4,06	4,12	1,12%
TO	200,97	200,92	-0,05	2,78	2,78	0,83%
MA	178,59	178,54	-0,05	7,64	7,65	1,55%
PI	194,56	194,55	-0,01	2,73	2,73	0,85%
CE	218,04	218,01	-0,02	3,19	3,20	0,82%
RN	185,93	185,88	-0,05	4,64	4,64	1,16%
PB	190,04	189,99	-0,05	5,67	5,67	1,25%
PE	193,17	193,17	0,01	3,69	3,69	0,99%
AL	189,82	189,82	0,01	8,61	8,61	1,55%
SE	182,48	182,39	-0,10	4,28	4,29	1,13%
BA	191,24	191,23	-0,02	4,83	4,83	1,15%
MG	223,61	223,58	-0,03	5,43	5,43	1,04%
ES	214,52	214,40	-0,12	5,97	5,98	1,14%
RJ	209,93	209,96	0,02	5,24	5,25	1,09%
SP	226,19	226,20	0,00	6,36	6,36	1,11%
PR	225,86	225,83	-0,04	3,69	3,69	0,85%
SC	225,32	225,28	-0,04	4,90	4,90	0,98%
RS	214,50	214,33	-0,17	4,89	4,92	1,03%
MS	213,26	213,21	-0,05	4,05	4,05	0,94%
MT	205,99	205,92	-0,07	5,36	5,36	1,12%
GO	216,33	216,33	0,00	6,13	6,13	1,14%
DF	220,39	220,43	0,03	5,08	5,08	1,02%

Fonte: Elaborado pela autora

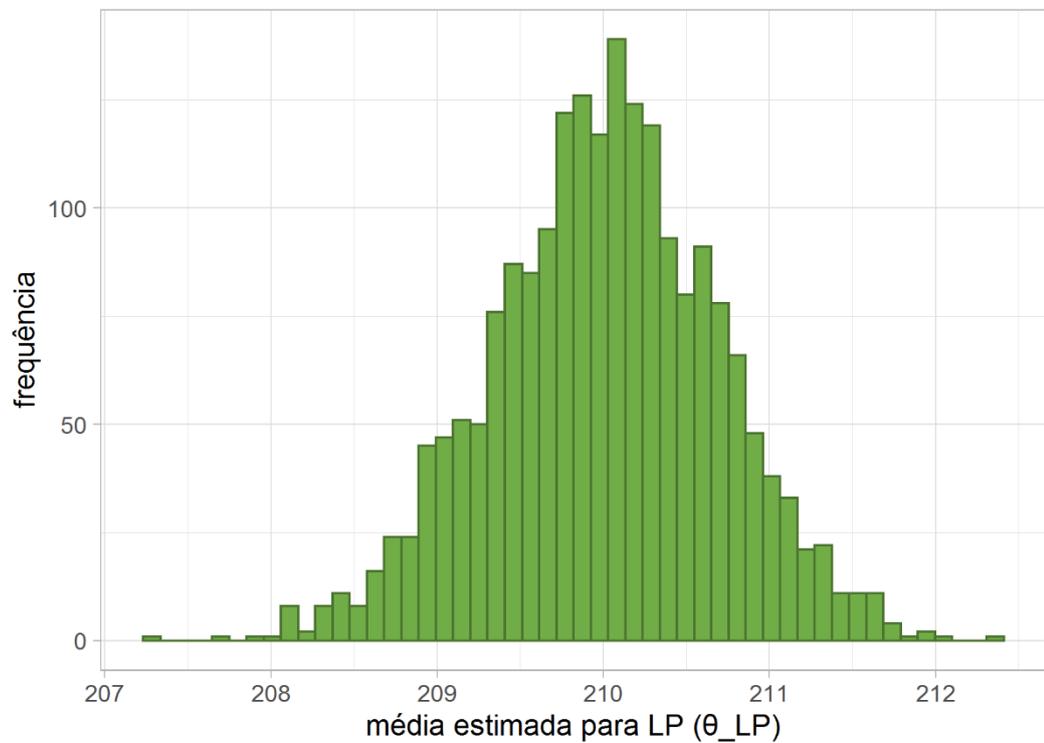
Tabela 23 - Medidas de desempenho para Matemática (Simulação ρ_{int_2})

UF	θ_{MT}	$\bar{\theta}_{MT}$	Viés ($\hat{\theta}_{MT}$)	Var ($\hat{\theta}_{MT}$)	EQM ($\hat{\theta}_{MT}$)	RREQM ($\hat{\theta}_{MT}$)
BR	219,50	219,47	-0,03	0,52	0,52	0,33%
RO	220,57	220,65	0,08	4,94	4,95	1,01%
AC	226,70	226,67	-0,02	2,27	2,28	0,67%
AM	209,52	209,45	-0,06	2,44	2,44	0,75%
RR	215,73	216,23	0,50	0,89	1,14	0,49%
PA	192,82	192,78	-0,04	5,65	5,65	1,23%
AP	188,88	188,65	-0,23	3,56	3,61	1,01%
TO	211,16	211,11	-0,05	2,63	2,64	0,77%
MA	186,58	186,53	-0,05	6,27	6,27	1,34%
PI	204,76	204,77	0,01	2,64	2,64	0,79%
CE	226,11	226,05	-0,05	4,49	4,49	0,94%
RN	193,59	193,56	-0,04	3,73	3,73	1,00%
PB	199,97	199,89	-0,08	5,26	5,27	1,15%
PE	202,69	202,66	-0,03	3,80	3,80	0,96%
AL	201,69	201,67	-0,02	10,28	10,28	1,59%
SE	193,42	193,33	-0,09	3,49	3,50	0,97%
BA	199,52	199,51	-0,01	4,31	4,31	1,04%
MG	231,76	231,75	-0,01	5,76	5,76	1,04%
ES	223,56	223,49	-0,07	6,00	6,01	1,10%
RJ	217,86	217,83	-0,02	5,43	5,43	1,07%
SP	237,83	237,80	-0,03	7,47	7,47	1,15%
PR	239,14	239,08	-0,06	4,80	4,81	0,92%
SC	234,34	234,30	-0,04	5,87	5,88	1,03%
RS	223,48	223,35	-0,13	5,12	5,13	1,01%
MS	220,88	220,82	-0,06	4,82	4,82	0,99%
MT	214,34	214,27	-0,07	5,36	5,36	1,085
GO	222,79	222,83	0,04	6,84	6,84	1,17%
DF	229,20	229,27	0,07	5,04	5,04	0,98%

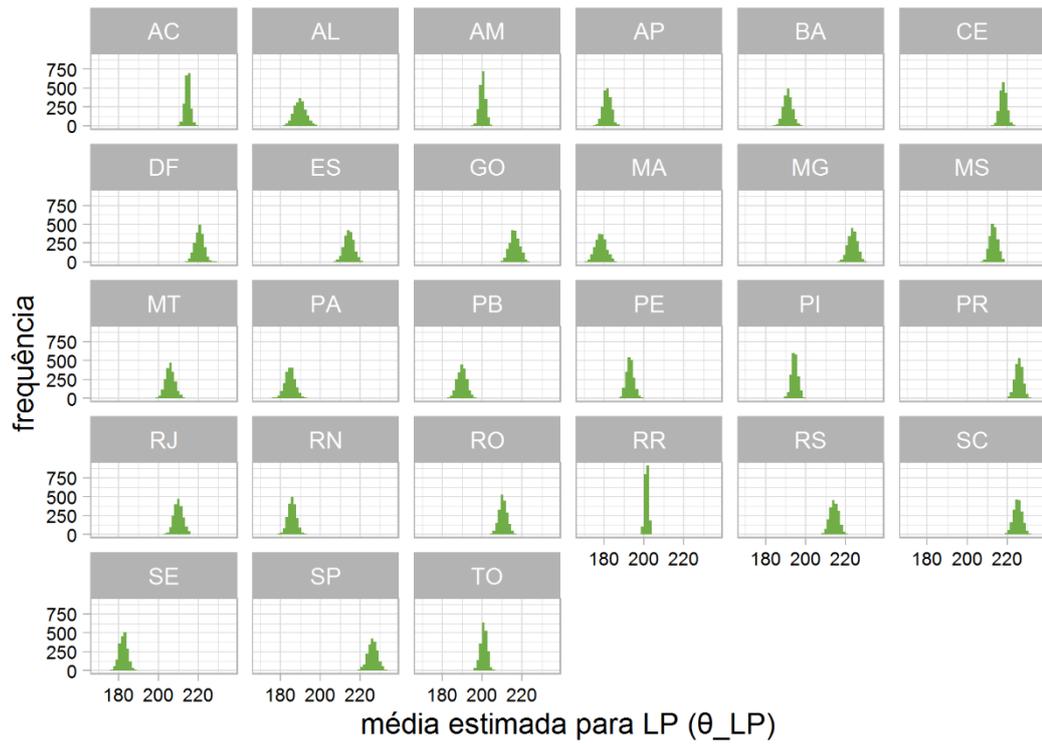
Fonte: Elaborado pela autora

Gráfico 14 - Dispersão dos valores de $\hat{\theta}_{iLP}$ para o Brasil (Simulação ρ_{int_2})

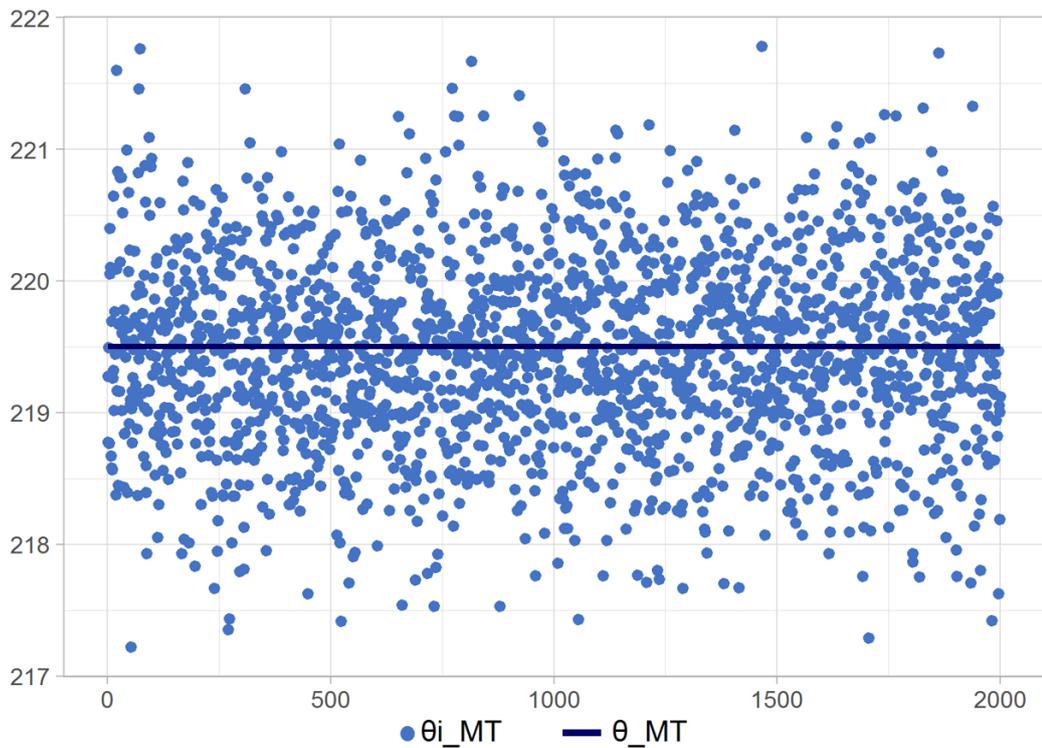
Fonte: Elaborado pela autora

Gráfico 15 - Distribuição dos valores de $\hat{\theta}_{iLP}$ para o Brasil (Simulação ρ_{int_2})

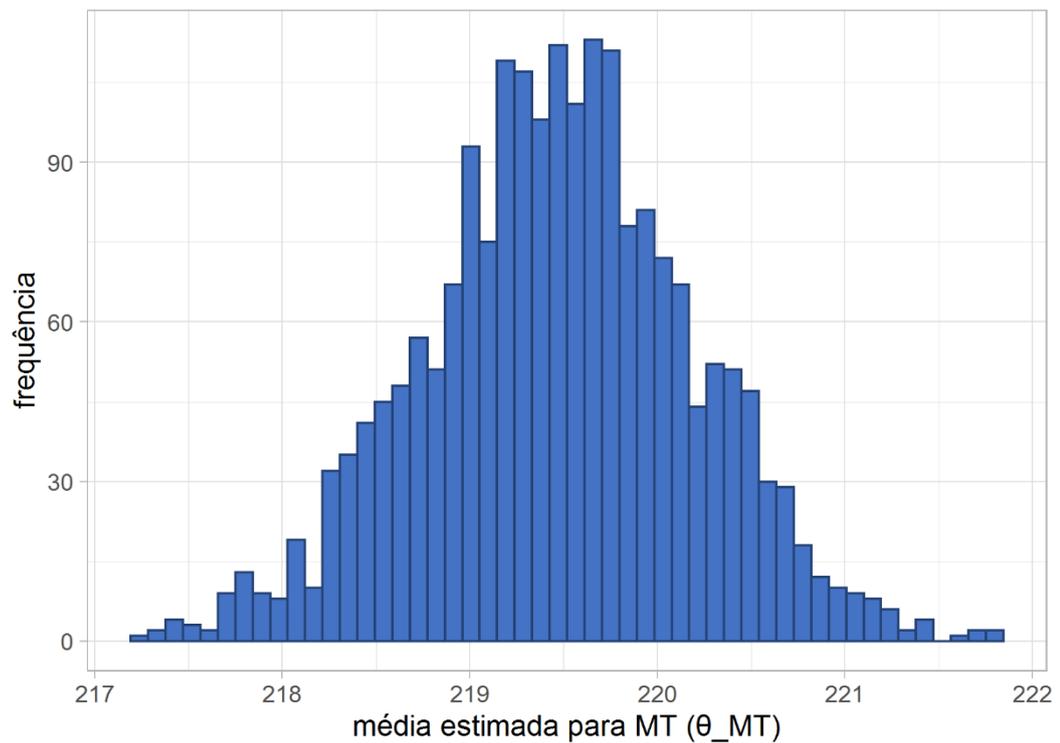
Fonte: Elaborado pela autora

Gráfico 16 - Distribuição dos valores de $\hat{\theta}_{iLP}$ por UF (Simulação ρ_{int_2})

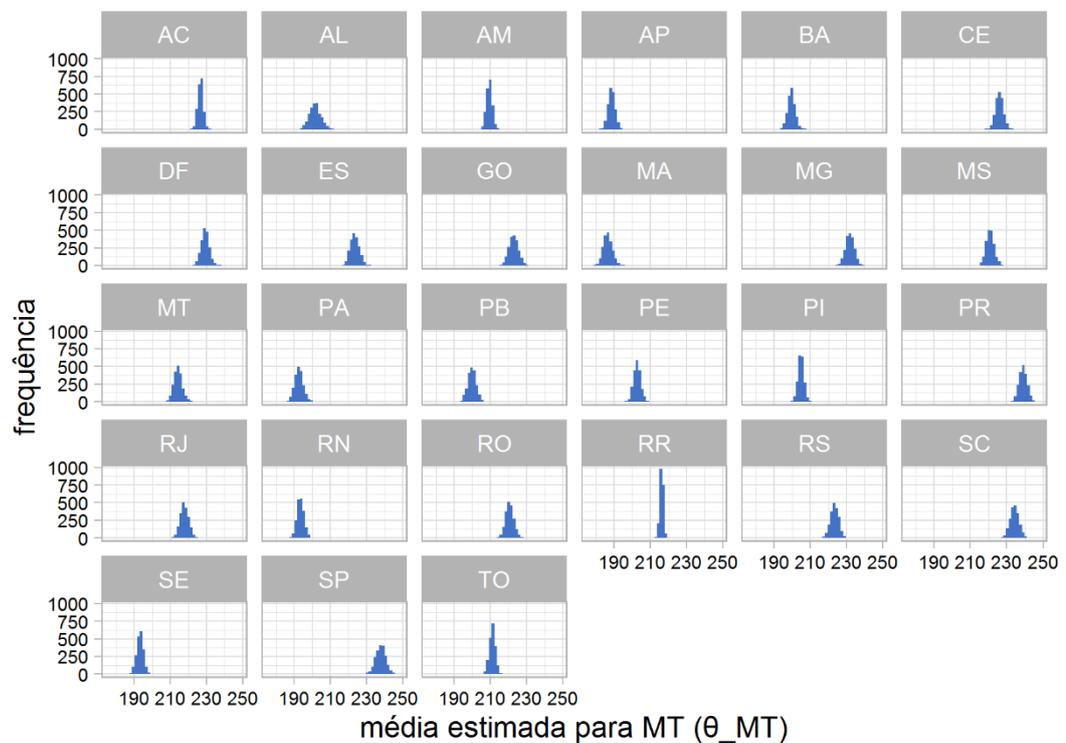
Fonte: Elaborado pela autora

Gráfico 17 - Dispersão dos valores de $\hat{\theta}_{iMT}$ para o Brasil (Simulação ρ_{int_2})

Fonte: Elaborado pela autora

Gráfico 18 - Distribuição dos valores de $\hat{\theta}_{iMT}$ para o Brasil (Simulação ρ_{int_2})

Fonte: Elaborado pela autora

Gráfico 19 - Distribuição dos valores de $\hat{\theta}_{iMT}$ por UF (Simulação ρ_{int_2})

Fonte: Elaborado pela autora

Tabela 24 - Cobertura dos intervalos de confiança e Margem de Erro (Simulação ρ_{int_2})

UF	Cobertura LP	Cobertura MT	$ME(\hat{\theta}_{LP})$	$ME(\hat{\theta}_{MT})$
BR	96,5%	96,2%	1,47	1,55
RO	97,4%	97,2%	4,73	4,98
AC	99,5%	99,3%	4,14	4,42
AM	98,0%	97,9%	3,60	3,72
RR	99,8%	99,8%	3,38	3,41
PA	94,9%	94,8%	5,24	4,85
AP	96,4%	96,3%	4,45	4,15
TO	98,7%	98,8%	4,27	4,25
MA	94,5%	94,6%	5,53	5,00
PI	97,4%	97,1%	3,74	3,74
CE	97,0%	96,8%	3,89	4,70
RN	96,2%	95,9%	4,62	4,13
PB	95,0%	95,1%	4,85	4,71
PE	95,8%	95,4%	3,94	3,98
AL	94,8%	93,4%	6,01	6,41
SE	95,0%	95,8%	4,23	3,89
BA	95,5%	94,7%	4,49	4,21
MG	95,9%	95,6%	4,82	4,93
ES	95,9%	95,9%	4,97	5,00
RJ	95,6%	95,9%	4,70	4,82
SP	93,9%	94,6%	5,00	5,45
PR	95,8%	96,4%	3,99	4,62
SC	96,0%	96,0%	4,68	5,12
RS	94,9%	95,2%	4,39	4,54
MS	96,7%	95,4%	4,26	4,59
MT	95,6%	95,7%	4,79	4,86
GO	95,6%	95,8%	5,10	5,47
DF	96,3%	97,1%	4,90	4,99

Fonte: Elaborado pela autora

APÊNDICE C – Scripts utilizados para processamento das simulações

PASSO 1_Universo_simulacao (R)

Limpeza da base dos microdados do Saeb 2017 para identificação do Universo da simulação

Habilitar pacotes

library (data.table)

library (dplyr)

configurar área de trabalho

setwd ('D:/MESTRADO/CALCULOS/DADOS')

Ler arquivo original dos microdados em .csv

```
ts_aluno_5ef <- fread ('TS_ALUNO_5EF.csv', header = TRUE, dec = ".", sep = ',', na.strings = "NA")
```

Confirmar variáveis existentes no arquivo

```
names (ts_aluno_5ef)
```

Criar tabela de universo para simulação

- excluir questionário para o arquivo ficar mais leve

- Filtrar alunos válidos

- Filtrar alunos de escolas estaduais e municipais

```
universo_simulacao <- ts_aluno_5ef %>%
```

```
  select(1:35) %>%
```

```
  filter(IN_PROFICIENCIA == 1, IN_SITUACAO_CENSO == 1,
```

```
  ID_DEPENDENCIA_ADM %in% c(2,3))
```

```
names (universo_simulacao)
```

Ordenar base por ID_UF

```
universo_simulacao <- arrange(universo_simulacao, ID_UF)
```

Exportar tabela como .csv

```
write.csv(universo_simulacao, "universo_simulacao.csv", row.names = FALSE)
```

PASSO 2_tamanho_amostra (Simulação 1 - SAS)

```
/******
```

```
/*          CÁLCULO DO TAMANHO DA AMOSTRA - SAEB          */
```

```
/******
```

```
/* Modelo hierárquico de 3 níveis, nulo, com estimação pelo método de máxima */
```

```
/* verossimilhança restrita. Variância "comum" considerando pesos.          */
```

```
/******
```

```
/******
```

```
libname SAEB "/home/laene/SAEB";
```

```
/*
```

Para rodar o programa foi utilizado o universo de referência da simulação gerado a partir dos alunos válidos dos microdados Saeb 2017 (IN_PROFICIENCIA = 1) das escolas públicas

```
*/
```

```

/* Calcula o número médio de alunos por turma */
PROC SQL;
  CREATE TABLE WORK.MEDIA_B AS
  SELECT DISTINCT
    t1.ID_PROVA_BRASIL,
    t1.ID_REGIAO,
    t1.ID_UF,
    COUNT (DISTINCT t1.ID_TURMA) AS TURMAS,
    COUNT (DISTINCT t1.ID_ALUNO)/CALCULATED TURMAS AS MEDIA_B
  FROM SAEB.UNIVERSO_SIMULACAO AS t1
  GROUP BY t1.ID_UF;
QUIT;

/** PORTUGUÊS ***/

/* Calcula as Variâncias considerando apenas os pesos dos alunos */

proc means data=SAEB.UNIVERSO_SIMULACAO var NOPRINT;
var proficiencia_lp_saeb;
output out=work.var_lp
      var(=);
class id_uf;
run;

/* Calcula as Variâncias considerando o modelo hierárquico */

proc mixed data=SAEB.UNIVERSO_SIMULACAO noclprint; /* default method=reml
(residual maximum likelihood) que é o que a Denise usou no R */
class id_turma id_escola;
model proficiencia_lp_saeb= /ddfm=res solution;
  random int /type=vc subject=id_turma*id_escola;
  random int /type=vc subject=id_escola;
ods output covparms=work.aluno;
by id_uf;
run;

/* organiza base de dados para facilitar a análise */

PROC SQL;
  CREATE TABLE WORK.CovParm AS
  SELECT t1.ID_UF,
    t1.CovParm,
    CASE WHEN t1.Subject IS MISSING THEN 'ALUNO' ELSE t1.Subject END AS
Subject,
    t1.Estimate
  FROM WORK.aluno AS t1;
QUIT;

PROC SORT

```

```

DATA=WORK.CovParm(KEEP=ID_UF CovParm Subject Estimate)
OUT=WORK.SORTTempTableSorted;
BY ID_UF;
RUN;

PROC TRANSPOSE DATA=WORK.SORTTempTableSorted
OUT=WORK.VARIANCIAS_LP
PREFIX=VAR_
NAME=Source
LABEL=Label;
BY ID_UF;
ID SUBJECT;
VAR ESTIMATE;
RUN;

/**** MATEMÁTICA ****/

/* Calcula as Variâncias considerando apenas os pesos dos alunos */

proc means data=SAEB.UNIVERSO_SIMULACAO var NOPRINT;
var proficiencia_mt_saeb;
/* weight peso_aluno_mt; */
output out=work.var_mt
var(=);
class id_uf;
run;

/* Calcula as Variâncias considerando o modelo hierárquico */

proc mixed data=SAEB.UNIVERSO_SIMULACAO noclprint; /* default method=reml
(residual maximum likelihood) que é o que a Denise usou no R */
class id_turma id_escola;
model proficiencia_mt_saeb= /ddfm=res solution;
random int /type=vc subject=id_turma*id_escola;
random int /type=vc subject=id_escola;
ods output covparms=work.aluno;
by id_uf;
run;

/* organiza base de dados para facilitar a análise */

PROC SQL;
CREATE TABLE WORK.CovParm AS
SELECT t1.ID_UF,
t1.CovParm,
CASE WHEN t1.Subject IS MISSING THEN 'ALUNO' ELSE t1.Subject END AS
Subject,
t1.Estimate
FROM WORK.aluno AS t1;
QUIT;

```

```
PROC SORT
  DATA=WORK.CovParm(KEEP=ID_UF CovParm Subject Estimate)
  OUT=WORK.SORTTempTableSorted;
  BY ID_UF;
RUN;
```

```
PROC TRANSPOSE DATA=WORK.SORTTempTableSorted
  OUT=WORK.VARIANCIAS_MT
  PREFIX=VAR_
  NAME=Source
  LABEL=Label;
  BY ID_UF;
  ID SUBJECT;
  VAR ESTIMATE;
RUN;
```

/**/ Cálculo do tamanho da amostra em número de alunos e de turmas /**/

```
PROC SQL;
  CREATE TABLE WORK.LP_5EF AS
  SELECT DISTINCT t1.ID_UF,
                 t3.TURMAS,
                 t3.MEDIA_B,
                 /* tamanho da amostra em número de alunos para portugues */
                 t1.VAR_ALUNO AS VAR_ALUNO_LP,
                 t1.'VAR_ID_TURMA*ID_ESCOLA'n AS VAR_TURMA_LP,
                 t1.VAR_ID_ESCOLA AS VAR_ESCOLA_LP,
                 t2.PROFICIENCIA_LP_SAEB AS VAR_LP,

  (t1.'VAR_ID_TURMA*ID_ESCOLA'n+t1.VAR_ID_ESCOLA)/(t1.VAR_ALUNO+t1.'VAR
_ID_TURMA*ID_ESCOLA'n+t1.VAR_ID_ESCOLA) AS Cor_Intraclassa_LP,
  CALCULATED Cor_Intraclassa_LP*(t3.MEDIA_B-1)+1 AS EPA_LP,
  ((1.96**2)*t2.PROFICIENCIA_LP_SAEB)/(4**2) AS N_AAS_LP, /* 1,96
= 95% NA TABELA NORMAL (0,1); 4 = ERRO ESPERADO */
  CALCULATED EPA_LP * CALCULATED N_AAS_LP AS N_AC_LP

  FROM WORK.VARIANCIAS_LP AS t1, WORK.VAR_LP AS t2, WORK.MEDIA_B
AS t3
  WHERE (t1.ID_UF = t2.ID_UF AND t1.ID_UF = t3.ID_UF);
QUIT;
```

```
PROC SQL;
  CREATE TABLE WORK.MT_5EF AS
  SELECT DISTINCT t1.ID_UF,
                 t3.TURMAS,
                 t3.MEDIA_B,
                 /* tamanho da amostra em número de alunos para matemática */
                 t1.VAR_ALUNO AS VAR_ALUNO_MT,
```

```

t1.'VAR_ID_TURMA*ID_ESCOLA'n AS VAR_TURMA_MT,
t1.VAR_ID_ESCOLA AS VAR_ESCOLA_MT,
t2.PROFICIENCIA_MT_SAEB AS VAR_MT,

(t1.'VAR_ID_TURMA*ID_ESCOLA'n+t1.VAR_ID_ESCOLA)/(t1.VAR_ALUNO+t1.'VAR
_ID_TURMA*ID_ESCOLA'n+t1.VAR_ID_ESCOLA) AS Cor_Intraclassa_MT,
CALCULATED Cor_Intraclassa_MT*(t3.MEDIA_B-1)+1 AS EPA_MT,
((1.96**2)*t2.PROFICIENCIA_MT_SAEB)/(4**2) AS N_AAS_MT, /* 1,96
= 95% NA TABELA NORMAL (0,1); 4 = ERRO ESPERADO */
CALCULATED EPA_MT * CALCULATED N_AAS_MT AS N_AC_MT

FROM WORK.VARIANCIAS_MT AS t1, WORK.VAR_MT AS t2, WORK.MEDIA_B
AS t3
WHERE (t1.ID_UF = t2.ID_UF AND t1.ID_UF = t3.ID_UF);
QUIT;

```

```

PROC SQL;
CREATE TABLE WORK.N_5EF AS
SELECT DISTINCT *,

/* tamanho da amostra em número de turmas */
MAX(t1.N_AC_LP, t2.N_AC_MT) AS N_FINAL,
CALCULATED N_FINAL/t1.MEDIA_B AS N_TURMAS,
/*
CALCULATED N_TURMAS*1.25 AS N_TURMAS_FOLGA, /* INCLUI
FOLGA DE 25% COBRIR PARA EVENTUAIS PERDAS */
/*
CALCULATED N_TURMAS_FOLGA*0.4 AS TURMAS_PRIVADAS, /*
CEIL (CASE WHEN CALCULATED N_TURMAS >= t1.TURMAS THEN
t1.TURMAS ELSE CALCULATED N_TURMAS END) AS N_TURMAS_FINAL

FROM WORK.LP_5EF AS t1, WORK.MT_5EF AS t2
WHERE (t1.ID_UF = t2.ID_UF);
QUIT;

```

```

proc export data= WORK.N_5EF
OUTFILE="/home/laene/SAEB/N_5EF.xlsx" /* endereço e nome do arquivo */
DBMS=xlsx /* formato - deve coincidir com o do nome do arquivo */
REPLACE; /* substituir arquivo já existente*/
/*DELIMITER=';';*/

run;

```

PASSO 2_tamanho_amostra (Simulação 2 - R)

Calcular tamanho da amostra por UF - Variâncias estimadas

```

# Habilitar pacotes
library (data.table)
library (dplyr)

```

```

# configurar área de trabalho
setwd ('D:/MESTRADO/CALCULOS/DADOS')

```

```

# Ler arquivo universo_simulacao.csv
universo_simulacao <- fread ('universo_simulacao.csv', header = TRUE, dec = ".", sep = ',',
na.strings = "NA")

# Calcular variâncias dentro de cada escola

var_lp <- universo_simulacao %>%
  group_by(ID_UF, ID_ESCOLA) %>%
  summarise(var_i_lp = var(PROFICIENCIA_LP_SAEB, na.rm = TRUE))

var_mt <- universo_simulacao %>%
  group_by(ID_UF, ID_ESCOLA) %>%
  summarise(var_i_mt = var(PROFICIENCIA_MT_SAEB, na.rm = TRUE))

var_intra=inner_join(var_lp, var_mt)

qt_aluno_esc <- universo_simulacao %>%
  group_by(ID_UF, ID_ESCOLA) %>%
  count(ID_UF) %>%
  rename(qt_aluno_esc = "n")

var_intra=inner_join(var_intra, qt_aluno_esc)

# Calcular variâncias entre as escolas da UF

media_lp <- universo_simulacao %>%
  group_by(ID_UF, ID_ESCOLA) %>%
  summarise(media_lp = weighted.mean(PROFICIENCIA_LP_SAEB, PESO_ALUNO_LP,
na.rm = TRUE))

media_mt <- universo_simulacao %>%
  group_by(ID_UF, ID_ESCOLA) %>%
  summarise(media_mt = weighted.mean(PROFICIENCIA_MT_SAEB, PESO_ALUNO_MT,
na.rm = TRUE))

media_esc <- inner_join(media_lp, media_mt)

var_e_lp <- media_esc %>%
  group_by(ID_UF) %>%
  summarise(var_e_lp = var(media_lp, na.rm = TRUE))

var_e_mt <- media_esc %>%
  group_by(ID_UF) %>%
  summarise(var_e_mt = var(media_mt, na.rm = TRUE))

var_entre <- inner_join(var_e_lp, var_e_mt)

base_uf <- universo_simulacao %>%
  group_by(ID_UF) %>%

```

```

count(ID_UF)%>%
rename (qt_aluno = "n")

base_uf <- inner_join(base_uf, var_entre)

# Estimar a Correlação intraclasse - LP

temp <- var_intra %>%
  group_by (ID_UF) %>%
  summarise(t1_lp = sum(var_i_lp, na.rm = TRUE))
base_uf <- inner_join(base_uf, temp)

temp <- var_intra %>%
  group_by (ID_UF) %>%
  summarise(t2_lp = sum((qt_aluno_esc-1)*var_i_lp, na.rm = TRUE))
base_uf <- inner_join(base_uf, temp)

base_uf$num_lp <- base_uf$var_e_lp-(1/base_uf$qt_aluno)*base_uf$t1_lp
base_uf$den_lp <- base_uf$var_e_lp+(1/base_uf$qt_aluno)*base_uf$t2_lp
base_uf$corr_int_lp <- base_uf$num_lp/base_uf$den_lp

# Estimar a Correlação intraclasse - MT

temp <- var_intra %>%
  group_by (ID_UF) %>%
  summarise(t1_mt = sum(var_i_mt, na.rm = TRUE))
base_uf <- inner_join(base_uf, temp)

temp <- var_intra %>%
  group_by (ID_UF) %>%
  summarise(t2_mt = sum((qt_aluno_esc-1)*var_i_mt, na.rm = TRUE))
base_uf <- inner_join(base_uf, temp)

base_uf$num_mt <- base_uf$var_e_mt-(1/base_uf$qt_aluno)*base_uf$t1_mt
base_uf$den_mt <- base_uf$var_e_mt+(1/base_uf$qt_aluno)*base_uf$t2_mt
base_uf$corr_int_mt <- base_uf$num_mt/base_uf$den_mt

# Calcular número médio de alunos por turma (media_b) e número de turmas por UF

aluno_turma <- universo_simulacao %>%
  group_by(ID_UF, ID_TURMA) %>%
  count(ID_UF)%>%
  rename (qt_aluno = "n")

media_b <- aggregate(qt_aluno ~ ID_UF, aluno_turma, FUN = mean, na.rm = TRUE) %>%
  rename (media_b = "qt_aluno")

base_uf <- inner_join(base_uf, media_b)

turmas <- universo_simulacao %>%

```

```

group_by(ID_UF) %>%
  summarise(turmas = n_distinct(ID_TURMA))
base_uf <- inner_join(base_uf, turmas)

# Calcular variância dos alunos por UF

var_lp <- universo_simulacao %>%
  group_by(ID_UF) %>%
  summarise(var_lp = var(PROFICIENCIA_LP_SAEB, na.rm = TRUE))
base_uf <- inner_join(base_uf, var_lp)

var_mt <- universo_simulacao %>%
  group_by(ID_UF) %>%
  summarise(var_mt = var(PROFICIENCIA_MT_SAEB, na.rm = TRUE))
base_uf <- inner_join(base_uf, var_mt)

# Calcular EPA e tamanho da amostra
base_uf$sepa_lp <- base_uf$corr_int_lp*(base_uf$media_b-1)+1
base_uf$n_aas_lp <- ((1.96^2)*base_uf$var_lp)/(4^2) # 1,96 = 95% NA TABELA
NORMAL (0,1); 4 = ERRO ESPERADO
base_uf$n_ac_lp <- base_uf$sepa_lp*base_uf$n_aas_lp

base_uf$sepa_mt <- base_uf$corr_int_mt*(base_uf$media_b-1)+1
base_uf$n_aas_mt <- ((1.96^2)*base_uf$var_mt)/(4^2) # 1,96 = 95% NA TABELA
NORMAL (0,1); 4 = ERRO ESPERADO
base_uf$n_ac_mt <- base_uf$sepa_mt*base_uf$n_aas_mt

base_uf$n_final <- ifelse(base_uf$n_ac_lp>base_uf$n_ac_mt, base_uf$n_ac_lp,
base_uf$n_ac_mt)
base_uf$n_turmas <- base_uf$n_final/base_uf$media_b
base_uf$n_turmas_final <- ceiling (ifelse(base_uf$n_turmas>=
base_uf$turmas,base_uf$turmas,base_uf$n_turmas))

# Salvar arquivo final de resumo das variâncias
write.csv2(base_uf,'n_5ef_novo.csv',row.names = FALSE)

```

PASSO 2_tamanho_amostra (Simulação 3 - R)

```

# Calcular tamanho da amostra por UF

# Habilitar pacotes
library (data.table)
library (dplyr)
library(foreign)
library(lme4)
library("Hmisc")

# configurar área de trabalho
setwd ('D:/MESTRADO/CALCULOS/DADOS')

```

```

# Ler arquivo universo_simulacao.csv
universo_simulacao <- fread ('universo_simulacao.csv', header = TRUE, dec = ".", sep = ',',
na.strings = "NA")

# Calcular variancias por UF

# Preparar bases alimentadas no lopp
variancias_lp <- data.frame(i=NA,ID_UF=NA,Var_escola_lp=NA,Var_aluno_lp=NA,
var_lp=NA,turmas=NA,media_b=NA)

variancias_mt <- data.frame(i=NA,ID_UF=NA,Var_escola_mt=NA,Var_aluno_mt=NA,
var_mt=NA,turmas=NA,media_b=NA)

##### iniciar loop principal (amostras) #####
n <- 2000 #número de repetições da simulação

for (j in 1:n){

# Preparar base dos alunos da amostra com peso e proficiencia
setwd ('D:/MESTRADO/CALCULOS/PESOS')
peso <- fread (paste0('peso',j,'.csv'), dec=',', header = TRUE, sep = ';', na.strings = "")

# Ler arquivo da amostra .csv e preparar base de proficiências dos alunos
setwd ('D:/MESTRADO/CALCULOS/AMOSTRAS')
amostra <- fread (paste0('sample',j,'.csv'), dec=',', header = TRUE, sep = ';', na.strings = "")
profic <- inner_join(amostra, universo_simulacao, by = c("ID_ESCOLA", "ID_TURMA"))
%>%
  select (ID_ALUNO,
ID_TURMA,ID_ESCOLA,IN_PROFICIENCIA,IN_SITUACAO_CENSO,PROFICIENCIA
_LP_SAEB,PROFICIENCIA_MT_SAEB)

notas <- left_join(peso,profic, by = c("ID_ESCOLA")) %>%
  select(ID_ALUNO, ID_TURMA, ID_ESCOLA, ID_UF,
IN_PROFICIENCIA,IN_SITUACAO_CENSO,PROFICIENCIA_LP_SAEB,PROFICIENCIA
A_MT_SAEB,peso_ac) %>%
  arrange(ID_UF, ID_ESCOLA, ID_TURMA, ID_ALUNO)

##### iniciar segundo loop (UFs - LP) #####

# Para Língua Portuguesa (LP)

iduf<-names(table(notas$ID_UF))
variancias<-matrix(0,nrow=27,ncol=7)

for(i in 1:27){
  print(paste(i,"- LP - ID_UF =",iduf[i]))
  base_uf<-notas[notas$ID_UF==as.numeric(iduf[i]),]
  m4<-lmer(PROFICIENCIA_LP_SAEB ~ (1|ID_ESCOLA), base_uf)
  vc2<-print(VarCorr(m4),comp=c("Variance"))
  variancias[i,1]<-j
}

```

```

variancias[i,2]<-as.numeric(iduf[i])
variancias[i,3:4]<-as.data.frame(vc2)[,4]
variancias[i,5]<-wtd.var(x=base_uf$PROFICIENCIA_LP_SAEB)
variancias[i,6]<-length(table(base_uf$ID_TURMA))
variancias[i,7]<-mean(table(base_uf$ID_TURMA))
}
##### encerrar segundo loop #####

variancias_i<-as.data.frame(variancias)
names(variancias_i)<-
c("i","ID_UF","Var_escola_lp","Var_aluno_lp","var_lp","turmas","media_b")

variancias_lp <- rbind(variancias_lp, variancias_i)

# salvar base variancias_lp
setwd ('D:/MESTRADO/CALCULOS/RESULTADOS/simulacao_3')
variancias_lp <- filter(variancias_lp, !is.na(i))
write.csv2(variancias_lp,'variancias_lp.csv',row.names = FALSE)

##### iniciar terceiro loop (UFs - MT) #####
# Para Matemática (MT)
iduf<-names(table(notas$ID_UF))
variancias<-matrix(0,nrow=27,ncol=7)

for(i in 1:27){
  print(paste(i,"-MT-ID_UF =",iduf[i]))
  base_uf<-notas[notas$ID_UF==as.numeric(iduf[i]),]
  m4<-lmer(PROFICIENCIA_MT_SAEB ~ (1|ID_ESCOLA), base_uf)
  vc2<-print(VarCorr(m4),comp=c("Variance"))
  variancias[i,1]<-j
  variancias[i,2]<-as.numeric(iduf[i])
  variancias[i,3:4]<-as.data.frame(vc2)[,4]
  variancias[i,5]<-wtd.var(x=base_uf$PROFICIENCIA_MT_SAEB)
  variancias[i,6]<-length(table(base_uf$ID_TURMA))
  variancias[i,7]<-mean(table(base_uf$ID_TURMA))
}
##### encerrar terceiro loop #####

variancias_i<-as.data.frame(variancias)
names(variancias_i)<-
c("i","ID_UF","Var_escola_mt","Var_aluno_mt","var_mt","turmas","media_b")

variancias_mt <- rbind(variancias_mt, variancias_i)

# salvar base variancias_mt
variancias_mt <- filter(variancias_mt, !is.na(i))
write.csv2(variancias_mt,'variancias_mt.csv',row.names = FALSE)

}

```

```

# Calcular número médio de alunos por turma (media_b) e número de turmas por UF

aluno_turma <- universo_simulacao %>%
  group_by(ID_UF, ID_TURMA) %>%
  count(ID_UF) %>%
  rename (qt_aluno = "n")

media_b <- aggregate(qt_aluno ~ ID_UF, aluno_turma, FUN = mean, na.rm = TRUE) %>%
  rename (media_b = "qt_aluno")

turmas <- universo_simulacao %>%
  group_by(ID_UF) %>%
  summarise(turmas = n_distinct(ID_TURMA))

# usar as variâncias calculadas para calcular o tamanho da amostra

n <- inner_join (variâncias_lp, variâncias_mt, by= c("i", "ID_UF")) %>%
  select(i, ID_UF, Var_escola_lp, Var_aluno_lp, var_lp, Var_escola_mt, Var_aluno_mt, var_mt)
n <- inner_join(n, media_b)
n <- inner_join(n, turmas)

n$scor_int_lp <- (n$Var_escola_lp)/(n$Var_aluno_lp + n$Var_escola_lp)
n$sepa_lp <- n$scor_int_lp*(n$media_b-1)+1
n$n_aas_lp <- ((1.96^2)*n$var_lp)/(4^2) # 1,96 = 95% NA TABELA NORMAL (0,1); 4 =
ERRO ESPERADO
n$n_ac_lp <- n$sepa_lp*n$n_aas_lp

n$scor_int_mt <- (n$Var_escola_mt)/(n$Var_aluno_mt + n$Var_escola_mt)
n$sepa_mt <- n$scor_int_mt*(n$media_b-1)+1
n$n_aas_mt <- ((1.96^2)*n$var_mt)/(4^2) # 1,96 = 95% NA TABELA NORMAL (0,1); 4 =
ERRO ESPERADO
n$n_ac_mt <- n$sepa_mt*n$n_aas_mt

n$n_final <- ifelse(n$n_ac_lp > n$n_ac_mt, n$n_ac_lp, n$n_ac_mt)
n$n_turmas <- n$n_final/n$media_b
n$n_turmas_final <- ceiling (ifelse(n$n_turmas >= n$Turmas, n$Turmas, n$n_turmas))

# Salvar arquivo final de tamanhos da amostra
write.csv2(n, 'n_simulacao.csv', row.names = FALSE)

# Resumo das medidas por UF

# Variâncias LP
var_lp <- n %>%
  group_by (ID_UF) %>%
  summarise (var_lp = mean(var_lp, na.rm = TRUE))

var_aluno_lp <- n %>%
  group_by (ID_UF) %>%

```

```

  summarise (var_aluno_lp = mean(Var_aluno_lp, na.rm = TRUE))
resumo_var <- inner_join(var_lp, var_aluno_lp)

var_escola_lp <- n %>%
  group_by (ID_UF) %>%
  summarise (var_escola_lp = mean(Var_escola_lp, na.rm = TRUE))
resumo_var <- inner_join(resumo_var, var_escola_lp)

# Variâncias MT
var_mt <- n %>%
  group_by (ID_UF) %>%
  summarise (var_mt = mean(var_mt, na.rm = TRUE))
resumo_var <- inner_join(resumo_var, var_mt)

var_aluno_mt <- n %>%
  group_by (ID_UF) %>%
  summarise (var_aluno_mt = mean(Var_aluno_mt, na.rm = TRUE))
resumo_var <- inner_join(resumo_var, var_aluno_mt)

var_escola_mt <- n %>%
  group_by (ID_UF) %>%
  summarise (var_escola_mt = mean(Var_escola_mt, na.rm = TRUE))
resumo_var <- inner_join(resumo_var, var_escola_mt)

# Correlações intraclass e EPA
cor_int_lp <- n %>%
  group_by (ID_UF) %>%
  summarise (cor_int_lp = mean(cor_int_lp, na.rm = TRUE))
resumo_var <- inner_join(resumo_var, cor_int_lp)

cor_int_mt <- n %>%
  group_by (ID_UF) %>%
  summarise (cor_int_mt = mean(cor_int_mt, na.rm = TRUE))
resumo_var <- inner_join(resumo_var, cor_int_mt)

epa_lp <- n %>%
  group_by (ID_UF) %>%
  summarise (epa_lp = mean(epa_lp, na.rm = TRUE))
resumo_var <- inner_join(resumo_var, epa_lp)

epa_mt <- n %>%
  group_by (ID_UF) %>%
  summarise (epa_mt = mean(epa_mt, na.rm = TRUE))
resumo_var <- inner_join(resumo_var, epa_mt)

# Tamanho da amostra
n_lp <- n %>%
  group_by (ID_UF) %>%
  summarise (n_lp = mean(n_ac_lp, na.rm = TRUE))
resumo_var <- inner_join(resumo_var, n_lp)

```

```
n_mt <- n %>%
  group_by (ID_UF) %>%
  summarise (n_mt = mean(n_ac_mt, na.rm = TRUE))
resumo_var <- inner_join(resumo_var, n_mt)
```

```
n_final <- n %>%
  group_by (ID_UF) %>%
  summarise (n_final = mean(n_final, na.rm = TRUE))
resumo_var <- inner_join(resumo_var, n_final)
```

```
n_turmas_final <- n %>%
  group_by (ID_UF) %>%
  summarise (n_turmas_final = mean(n_turmas_final, na.rm = TRUE))
resumo_var$n_turmas_final <- ceiling(n_turmas_final$n_turmas_final)
```

```
# Salvar arquivo final de resumo das variâncias
write.csv2(resumo_var,'resumo_var.csv',row.names = FALSE)
```

PASSO 3 INSUMO ESTRATIFICACAO (R)

```
# Gera contagens para referência do cálculo de estratificação da amostra
```

```
# Habilitar pacotes
library (data.table)
library (dplyr)
```

```
# configurar área de trabalho
setwd ('D:/MESTRADO/CALCULOS/DADOS')
```

```
# Ler arquivo do universo da simulacao e base com CEP das escolas em .csv
universo_simulacao <- fread ('universo_simulacao.csv', header = TRUE, dec = ".", sep = ',',
na.strings = "NA")
```

```
# Confirmar variáveis existentes nos arquivos
names (universo_simulacao)
```

```
# Unificar as bases importadas, selecionando as variáveis de interesse
universo <- universo_simulacao %>%
  select(ID_UF, ID_DEPENDENCIA_ADM, ID_LOCALIZACAO, ID_AREA,
ID_ESCOLA, ID_TURMA, ID_ALUNO)
```

```
# Criar identificador de porte
# - CONTAR TURMAR POR ESCOLA
# - CRIAR VARIÁVEL IDENTIFICADORA (IF)
agg <- aggregate(data=universo, ID_TURMA ~ ID_ESCOLA, function(x) length(unique(x)))
%>%
  rename(qt_turma_esc=ID_TURMA)
universo <- inner_join(universo, agg, by = "ID_ESCOLA")
universo$porte <- ifelse(universo$qt_turma_esc <3, 1, 2)
```

```

# Criar estratos
universo$estrato <- paste0 (universo$ID_UF, universo$ID_DEPENDENCIA_ADM,
universo$ID_LOCALIZACAO, universo$ID_AREA)

# Criar lista dos estratos
info_estrat <- universo %>%
  select(ID_UF, ID_DEPENDENCIA_ADM, ID_LOCALIZACAO, ID_AREA, estrato) %>%
  unique()

# Contar alunos, turmas e escolas por estrato
agg <- aggregate(data=universo, ID_ESCOLA ~ estrato, function(x) length(unique(x))) %>%
  rename(escolas=ID_ESCOLA)
info_estrat <- inner_join(info_estrat, agg, by = "estrato")

agg <- aggregate(data=universo, ID_TURMA ~ estrato, function(x) length(unique(x))) %>%
  rename(turmas=ID_TURMA)
info_estrat <- inner_join(info_estrat, agg, by = "estrato")

agg <- aggregate(data=universo, ID_ALUNO ~ estrato, function(x) length(unique(x))) %>%
  rename(alunos=ID_ALUNO)
info_estrat <- inner_join(info_estrat, agg, by = "estrato")

# Contar alunos, turmas e escolas por estrato em escolas porte = 1
universo_p <- universo %>% filter(porte == 1)

agg <- aggregate(data=universo_p, ID_ESCOLA ~ estrato, function(x) length(unique(x)))
%>%
  rename(escolas_p=ID_ESCOLA)
info_estrat <- left_join(info_estrat, agg, by = "estrato")

agg <- aggregate(data=universo_p, ID_TURMA ~ estrato, function(x) length(unique(x)))
%>%
  rename(turmas_p=ID_TURMA)
info_estrat <- left_join(info_estrat, agg, by = "estrato")

agg <- aggregate(data=universo_p, ID_ALUNO ~ estrato, function(x) length(unique(x)))
%>%
  rename(alunos_p=ID_ALUNO)
info_estrat <- left_join(info_estrat, agg, by = "estrato")

# Contar alunos, turmas e escolas por estrato em escolas porte = 2
universo_g <- universo %>% filter(porte == 2)

agg <- aggregate(data=universo_g, ID_ESCOLA ~ estrato, function(x) length(unique(x)))
%>%
  rename(escolas_g=ID_ESCOLA)
info_estrat <- left_join(info_estrat, agg, by = "estrato")

agg <- aggregate(data=universo_g, ID_TURMA ~ estrato, function(x) length(unique(x)))
%>%

```

```

  rename(turmas_g=ID_TURMA)
info_estrat <- left_join(info_estrat, agg, by = "estrato")

agg <- aggregate(data=universo_g, ID_ALUNO ~ estrato, function(x) length(unique(x)))
%>%
  rename(alunos_g=ID_ALUNO)
info_estrat <- left_join(info_estrat, agg, by = "estrato")

# Exportar tabela como .csv
setwd ('D:/MESTRADO/CALCULOS/RESULTADOS')
write.csv(info_estrat, "info_estrat.csv", row.names = FALSE)

```

Observação: Entre os passos 3 e 4 o tamanho da amostra é ajustado pelo processo de estratificação realizado por alocação proporcional utilizando planilha em Excel. O número de zonas em cada estrato foi calculado dividindo o número de escolas em cada estrato por dois, em função da seleção de duas escolas por zona.

PASSO 4_divisao_zonas (R)

Distribuição das escolas em zonas de amostragem

```

# Habilitar pacotes
library (data.table)
library (dplyr)

# configurar área de trabalho
setwd ('D:/MESTRADO/CALCULOS/DADOS')

# Ler arquivo do universo da simulação e base com CEP das escolas em .csv
universo_simulacao <- fread ('universo_simulacao.csv', header = TRUE, dec = ".", sep = ',',
na.strings = "NA")
cadastro_escola <- fread('lista_escolas_CEP.csv', header = TRUE, sep = ';', na.strings = "")

setwd ('D:/MESTRADO/CALCULOS/RESULTADOS/simulacao')
n_zonas <- fread('n_zonas.csv', header = TRUE, sep = ';', na.strings = "NA")

# Confirmar variáveis existentes nos arquivos
names (universo_simulacao)
names (cadastro_escola)
names (n_zonas)

# Unificar as bases importadas, selecionando as variáveis de interesse
universo <- left_join(universo_simulacao, cadastro_escola, by = c("ID_ESCOLA" =
"CÃ³digo INEP")) %>%
  select(ID_UF, ID_DEPENDENCIA_ADM, ID_LOCALIZACAO, ID_AREA,
ID_ESCOLA, CEP, ID_TURMA)

# Criar identificador de porte
# - CONTAR TURMAR POR ESCOLA
# - CRIAR VARIÁVEL IDENTIFICADORA (IF)

```

```

agg <- aggregate(data=universo, ID_TURMA ~ ID_ESCOLA, function(x) length(unique(x)))
%>%
  rename(qt_turma_esc=ID_TURMA)
universo <- left_join(universo, agg, by = "ID_ESCOLA")
universo$porte <- ifelse(universo$qt_turma_esc < 3,1,2)

# Criar estratos
universo$estrato <- as.integer (paste0 (universo$ID_UF,
universo$ID_DEPENDENCIA_ADM, universo$ID_LOCALIZACAO,
universo$ID_AREA))
universo$estrato_tamanho <- as.integer (paste0(universo$estrato,universo$porte))

# Juntar informações do n_zonas
universo <- left_join(universo, n_zonas, by = "estrato_tamanho")

# Criar base por escola
zonas <- universo %>%
  select(CEP, ID_ESCOLA, estrato.x, estrato_tamanho, qt_zona,) %>%
  rename(estrato = estrato.x) %>%
  unique()

# Contar Escolas por estrato e turmas por escola
agg <- aggregate(data=universo, ID_ESCOLA ~ estrato_tamanho, function(x)
length(unique(x))) %>%
  rename(qt_escolas=ID_ESCOLA)
zonas <- left_join(zonas, agg, by = "estrato_tamanho")

agg <- aggregate(data=universo, ID_TURMA ~ ID_ESCOLA, function(x) length(unique(x)))
%>%
  rename(nu_turmas_esc=ID_TURMA)
zonas <- left_join(zonas, agg, by = "ID_ESCOLA")

# Excluir escolas de estratos de tamanho que não tem nenhuma zona
zonas <- zonas %>%
  filter(qt_zona != 0)

# Ordenar base por estrato de tamanho e CEP
zonas <- arrange(zonas,estrato_tamanho, CEP, ID_ESCOLA)

# Criar contador de ordenação por estrato de tamanho
zonas <- zonas %>%
  group_by(estrato_tamanho) %>%
  mutate(seq_escolas = 1:n())

# Distribuir as escolas em zonas
zonas$aux_1 <- (zonas$qt_escolas/zonas$qt_zona)+0.01
zonas$aux_2 <- (zonas$seq_escolas/zonas$aux_1)
zonas$zona <- trunc(zonas$aux_2)+1

# Exportar tabela como .csv

```

```
write.csv(zonas, "zonas.csv", row.names = FALSE)
```

PASSO 5_selecao_amostras (R)

```
# Seleção das amostras para simulação
```

```
# Habilitar pacotes
```

```
library (data.table)
```

```
library (dplyr)
```

```
library(stringr)
```

```
# configurar área de trabalho
```

```
setwd ('D:/MESTRADO/CALCULOS/DADOS')
```

```
# Ler arquivo da distribuição por zonas em .csv
```

```
universo_simulacao <- fread ('universo_simulacao.csv', header = TRUE, dec = ".", sep = ',',  
na.strings = "NA")
```

```
setwd ('D:/MESTRADO/CALCULOS/RESULTADOS/simulacao')
```

```
zonas <- fread ('zonas.csv', header = TRUE, sep = ',', na.strings = "")
```

```
# Organizar base de amostragem das escolas
```

```
zonas$estrato_amostragem <- paste0(zonas$estrato_tamanho, zonas$zona)
```

```
escolas <- zonas %>%
```

```
  select(ID_ESCOLA,estrato,estrato_tamanho,zona,estrato_amostragem,nu_turmas_esc)
```

```
nu_turmas_estrato_amostragem <- data.frame(nu_turmas_estrato_amostragem =
```

```
rowsum(zonas$nu_turmas_esc, zonas$estrato_amostragem)) %>%
```

```
  mutate(estrato_amostragem=row.names(.))
```

```
escolas <- merge(escolas, nu_turmas_estrato_amostragem, by = "estrato_amostragem")
```

```
escolas$prob <- (escolas$nu_turmas_esc/escolas$nu_turmas_estrato_amostragem)
```

```
# organizar base para sorteio das turmas
```

```
turmas <- universo_simulacao %>% distinct(ID_ESCOLA, ID_TURMA)
```

```
# Contar alunos válidos por turma
```

```
matriculas_realizadas <- universo_simulacao %>%
```

```
  group_by(ID_ESCOLA, ID_TURMA) %>%
```

```
  count(ID_TURMA) %>%
```

```
  rename (matriculas_realizadas = "n")
```

```
turmas <- left_join(turmas, matriculas_realizadas, by = c("ID_ESCOLA", "ID_TURMA"))
```

```
turmas$matriculas_realizadas[is.na(turmas$matriculas_realizadas)] <- 0
```

```
# Criar lista de sementes para permitir replicação
```

```
set.seed(2019);
```

```
n <- 2000 #número de repetições da simulação
```

```
m <- 1 #número mínimo
```

```
h <- 10000 #número máximo
```

```
a <- (runif(n,m,h))
```

```
# Criar tabela para armazenar números aleatórios e contagens de perdas
```

```
aleatorios <- zonas %>% select(ID_ESCOLA)
```

```

perdas <- data.frame(i=NA,n=NA, qt_turmas=NA)

# Loop de seleção das amostras pelo método de Amostragem Sequencial de Poisson
setwd ('D:/MESTRADO/CALCULOS/RESULTADOS/simulacao/AMOSTRAS')
for (i in 1:n){

  # gerar e armazenar números aleatórios
  set.seed(a[i]);
  numero_aleatorio <- runif(nrow(escolas))
  aleatorios <- cbind.data.frame(aleatorios,numero_aleatorio)
  names(aleatorios)[i+1] <- paste0('x',i)

  # calcular quociente de seleção
  escolas$numero_aleatorio <- numero_aleatorio
  escolas$q_i <- escolas$numero_aleatorio/escolas$prob

  # Ordenar base por estrato de amostragem e selecionar amostra (2 por estrato de
  amostragem)
  ordenado <- escolas[order(escolas$estrato_amostragem,escolas$q_i),]
  r <- lapply(split(ordenado, ordenado$estrato_amostragem), function(y) head(y,2))
  escolas_sorteadas <- do.call(rbind, r)
  rownames(escolas_sorteadas) <- c()

  # Sortear turmas, conforme porte da escola
  escolas_turmas <- left_join(escolas_sorteadas, turmas, by=
c("ID_ESCOLA"="ID_ESCOLA"))
  ordenado2 <- escolas_turmas[order(escolas_turmas$ID_ESCOLA),]%>%
filter(!is.na(ID_TURMA))
  t <- lapply (split(ordenado2, ordenado2$ID_ESCOLA), function(y) y[sample(nrow(y),ifelse
(unique(y$nu_turmas_esc) < 3, 1, 2)),])
  turmas_sorteadas <- do.call(rbind, t)
  rownames(turmas_sorteadas) <- c()
  lista_perdas <- escolas_turmas[order(escolas_turmas$ID_ESCOLA),] %>%
filter(is.na(ID_TURMA))
  turmas_sorteadas <- rbind(turmas_sorteadas, lista_perdas)
  turmas_sorteadas$turmas_realizadas <- ifelse
(is.na(turmas_sorteadas$ID_TURMA),0,(ifelse(turmas_sorteadas$nu_turmas_esc<3, 1, 2)))
  turmas_sorteadas$matriculas_realizadas[is.na(turmas_sorteadas$matriculas_realizadas)] <- 0

  # salvar arquivo da amostra
  write.csv2(turmas_sorteadas,paste0('sample',i,'.csv'), row.names = FALSE)

  # contar perdas da amostra
  perdas_i <- data.frame(i,count(lista_perdas), qt_turmas= sum(lista_perdas$nu_turmas_esc))
  perdas <- rbind(perdas,perdas_i)

}

# salvar base de números aleatórios de cada repetição
write.csv2(aleatorios,'numeros_aleatorios.csv', row.names = FALSE)

```

```
# Salvar base com perdas da amostra
perdas <- rename(perdas, qt_escolas = "n") %>% filter(!is.na(i))
write.csv2(perdas,'perdas.csv', row.names = FALSE)
```

PASSO 6_insumos_pesos (R)

```
# Prepara insumos para cálculo dos pesos
```

```
# Habilitar pacotes
library (data.table)
library (dplyr)
```

```
# configurar área de trabalho
setwd ('D:/MESTRADO/CALCULOS/DADOS')
```

```
# Ler arquivo univeso_simulacao.csv
universo_simulacao <- fread ('universo_simulacao.csv', header = TRUE, dec = ".", sep = ',',
na.strings = "NA")
```

```
names (universo_simulacao)
```

```
# Criar estratos
universo_simulacao$estrato <- as.integer (paste0 (universo_simulacao$ID_UF,
universo_simulacao$ID_DEPENDENCIA_ADM, universo_simulacao$ID_LOCALIZACAO,
universo_simulacao$ID_AREA))
```

```
# Preparar base por escola
insumos_pesos <- universo_simulacao %>%
  select(ID_ESCOLA, ID_UF, ID_DEPENDENCIA_ADM, ID_LOCALIZACAO,
ID_AREA, estrato) %>%
  unique()
```

```
# Contar turmas por escola e criar identificador de porte
agg <- aggregate(data=universo_simulacao, ID_TURMA ~ ID_ESCOLA, function(x)
length(unique(x))) %>%
  rename(turmas_saeb_2017=ID_TURMA)
insumos_pesos <- left_join(insumos_pesos, agg, by = "ID_ESCOLA")
insumos_pesos$porte_saeb_2017 <- ifelse(insumos_pesos$turmas_saeb_2017 < 3,1,2)
```

```
# Preparar base por estrato
auxiliar_calibracao <- universo_simulacao %>%
  select(estrato) %>% unique()
```

```
# Contagens do Universo de referência por estrato
agg <- aggregate(data=universo_simulacao, ID_ALUNO ~ estrato, function(x)
length(unique(x))) %>%
  rename(matriculas_estrato=ID_ALUNO)
auxiliar_calibracao <- left_join(auxiliar_calibracao, agg, by = "estrato")
```

```
agg <- aggregate(data=universo_simulacao, ID_TURMA ~ estrato, function(x)
length(unique(x))) %>%
  rename(turmas_estrato=ID_TURMA)
auxiliar_calibracao <- left_join(auxiliar_calibracao, agg, by = "estrato")
```

```
agg <- aggregate(data=universo_simulacao, ID_ESCOLA ~ estrato, function(x)
length(unique(x))) %>%
  rename(escolas_estrato=ID_ESCOLA)
auxiliar_calibracao <- left_join(auxiliar_calibracao, agg, by = "estrato")
```

```
# salvar base de números aleatórios de cada repetição
setwd ('D:/MESTRADO/CALCULOS/PESOS')
write.csv2(insumos_pesos,'insumos_pesos.csv', row.names = FALSE)
write.csv2(auxiliar_calibracao,'auxiliar_calibracao.csv', row.names = FALSE)
```

PASSO 7_pesos (R)

```
# Cálculo dos pesos para cada uma das amostras sorteadas
```

```
# Habilitar pacotes
library (data.table)
library (dplyr)
```

```
# configurar área de trabalho
setwd ('D:/MESTRADO/CALCULOS/RESULTADOS/simulacao/PESOS')
```

```
# Ler arquivo da base de alunos participantes no Saeb 2017 em .csv
aux_c <- fread ('auxiliar_calibracao.csv', header = TRUE, sep = ';', na.strings = "")
insumo <- fread ('insumos_pesos.csv', header = TRUE, sep = ';', na.strings = "")
```

```
# criar arquivo de conferência dos cálculos
confere <- data.frame(i=NA, total_estimado=NA,total_censo=NA,diferenca=NA,
valida_estrato = NA,total_fora=NA,valida=NA)
```

```
##### começar loop #####
```

```
n <- 2000 #número de repetições da simulação
```

```
for (i in 1:n){
```

```
  # Ler arquivo da amostra em .csv
  setwd ('D:/MESTRADO/CALCULOS/RESULTADOS/simulacao/AMOSTRAS')
  amostra_i <- fread (paste0('sample',i,'.csv'), dec=',', header = TRUE, sep = ';', na.strings =
  "")
```

```
  #limpar duplicidades de escola nas bases da amostra (escolas com duas turmas sorteadas)
  amostra <- amostra_i %>%
    group_by(ID_ESCOLA) %>%
    summarise(matriculas_realizadas = sum(matriculas_realizadas)) %>%
    filter(matriculas_realizadas!=0) %>%
```

```

data.frame()
temp <- amostra_i %>%
  distinct(estrato_amostragem,ID_ESCOLA,estrato,estrato_tamanho,nu_turmas_esc,
    nu_turmas_estrato_amostragem,prob,qi,turmas_realizadas) %>%
  filter(turmas_realizadas!=0)
amostra <- left_join(amostra, temp, "ID_ESCOLA")

# Buscar insumos das escolas selecionadas na amostra i
peso <- inner_join(amostra, insumo, by = c ("ID_ESCOLA"= "ID_ESCOLA", "estrato" =
"estrato")) %>%
  select(ID_ESCOLA,ID_UF,estrato, porte_saeb_2017,turmas_saeb_2017,
    prob,matriculas_realizadas,turmas_realizadas)

# incluir quantidades do saeb por estrato
peso <- left_join(peso, aux_c, by = "estrato") %>%
  select (ID_ESCOLA,ID_UF,estrato, porte_saeb_2017,turmas_saeb_2017,
    prob,matriculas_realizadas,turmas_realizadas, estrato,
    matriculas_estrato, turmas_estrato, escolas_estrato)

# Calcular quantas escolas participaram da aplicação por estrato (qt_esc_p)
qt_esc_p <- peso %>%
  group_by(estrato) %>%
  count(estrato) %>%
  rename (qt_esc_p = "n")
peso <- left_join(peso, qt_esc_p, "estrato")

# Calcular número de escolas selecionadas na amostra por estrato (qt_esc_s)
qt_esc_s <- amostra %>%
  group_by(estrato) %>%
  count(estrato) %>%
  rename (qt_esc_s = "n")
peso <- left_join(peso, qt_esc_s, "estrato")

# Calcular peso_escola
peso$peso_sel_esc = ifelse (peso$prob < 1, 1/(2*peso$prob), 1/peso$prob) # cada escola
tem duas oportunidades de ser sorteada porque seleciona-se duas escolas por zona, a menos
que a escola seja a única da zona
peso$taxa_part <- peso$qt_esc_p/peso$qt_esc_s
peso$peso_nr1 <- 1/peso$taxa_part
peso$peso1 <- peso$peso_sel_esc*peso$peso_nr1

# calcular calibração por escola (compensação das escolas ausentes)
total_pesol <- group_by(peso, estrato) %>%
  summarise(total_pesol = sum(peso1)) %>%
  data.frame()
peso <- left_join(peso, total_pesol, by = "estrato")
peso$calibacao_esc <- peso$escolas_estrato/peso$total_pesol
peso$peso_ec <- peso$peso1*peso$calibacao_esc

# Calcular peso_turma

```

```

peso$prob_sel_turma <- peso$porte_saeb_2017/peso$turmas_saeb_2017 # quantidades do
universo de referencia para seleção da amostra
peso$peso_sel_turma <- 1/peso$prob_sel_turma
peso$peso2 <- peso$peso_ec*peso$peso_sel_turma
peso$aux_peso_tc <- peso$peso2*peso$turmas_realizadas

# calcular calibração por turma (compensação das turmas ausentes)
total_aux_peso_tc <- group_by(peso, estrato) %>%
  summarise(total_aux_peso_tc = sum(aux_peso_tc)) %>%
  data.frame()
peso <- left_join(peso, total_aux_peso_tc, by = "estrato")
peso$calibricao_turmas <- peso$turmas_estrato/peso$total_aux_peso_tc
peso$peso_tc <- peso$peso2*peso$calibricao_turmas

# calcular peso_aluno
peso$prob_sel_aluno <- 1
peso$peso_sel_aluno <- 1 # 1/PROB_SEL_ALUNO = PESO_SEL_ALUNO
peso$peso3 <- peso$peso_tc*peso$peso_sel_aluno # COMO PESO_SEL_ALUNO = 1
ENTÃO PESO_3 = PESO_TC
peso$aux_peso_ac <- peso$peso3*peso$matriculas_realizadas

# calcular calibração por aluno (compensação dos alunos ausentes)
total_aux_peso_ac <- group_by(peso, estrato) %>%
  summarise(total_aux_peso_ac = sum(aux_peso_ac)) %>%
  data.frame()
peso <- left_join(peso, total_aux_peso_ac, by = "estrato")
peso$calibricao_alunos <- peso$matriculas_estrato/peso$total_aux_peso_ac
peso$peso_ac <- peso$peso3*peso$calibricao_alunos

# salvar arquivo do peso
setwd ('D:/MESTRADO/CALCULOS/RESULTADOS/simulacao/PESOS')
write.csv2(peso,paste0('peso',i,'.csv'), row.names = FALSE)

# conferir peso
peso$total_escola <- peso$matriculas_realizadas*peso$peso_ac
total_exp_ac <- group_by(peso, estrato) %>%
  summarise(total_exp_ac = sum(total_escola)) %>%
  data.frame()
total_exp_ac <- left_join(total_exp_ac, aux_c, by = "estrato")
total_exp_ac$dif_estrato <- round(total_exp_ac$total_exp_ac -
total_exp_ac$matriculas_estrato,0)
valida_estrato <- ifelse(sum(total_exp_ac$dif_estrato)==0,"ok","problema")

total_estimado <- sum(peso$total_escola)
total_censo <- sum(aux_c$matriculas_estrato)
diferenca <- total_censo - total_estimado

estratos_fora <- left_join(aux_c, amostra, "estrato") %>%
  distinct(estrato, matriculas_estrato, matriculas_realizadas) %>%
  filter(is.na(matriculas_estrato)|is.na(matriculas_realizadas))

```

```

total_fora <- sum(estratos_fora$matriculas_estrato)
valida <- ifelse((round(total_fora - diferenca,0))==0,"ok","problema")

confere_i <- c(i, total_estimado,total_censo,diferenca,valida_estrato, total_fora,valida)
confere <- rbind(confere,confere_i)

}

##### encerrar loop #####

# salvar arquivo de conferência
setwd ('D:/MESTRADO/CALCULOS/RESULTADOS/simulacao/PESOS')
confere <- filter(confere, !is.na(i))
write.csv2(confere,'confere.csv', row.names = FALSE)

# checar conferências
table(confere$valida_estrato)
table(confere$valida)

PASSO 8_medias (R)
# Calcular resultados para cada amostra selecionada considerando seus respectivos pesos

# Habilitar pacotes
library (data.table)
library (dplyr)
library(survey)

# Ler arquivo univeso_simulacao.csv
setwd ('D:/MESTRADO/CALCULOS/DADOS')
universo_simulacao <- fread ('universo_simulacao.csv', header = TRUE, dec = ".", sep = ',',
na.strings = "NA")

names (universo_simulacao)

# criar tabela para armazenar resultados
resultados <- data.frame(i=NA,ID_UF=NA,media_lp=NA, se_lp=NA,dp_lp= NA,cv_lp=NA,
media_mt=NA, se_mt=NA, dp_mt= NA, cv_mt=NA)
n <- data.frame (i=NA, n_esc=NA, n_tur=NA, n_alu=NA)

##### iniciar loop #####
s <- 2000 #número de repetições da simulação

for (i in 1:s){

# Ler arquivo de pesos .csv
setwd ('D:/MESTRADO/CALCULOS/RESULTADOS/simulacao/PESOS')
peso <- fread (paste0('peso',i,'.csv'), dec=',', header = TRUE, sep = ';', na.strings = "")

# Ler arquivo da amostra .csv e preparar base de proficiências dos alunos

```

```

setwd ('D:/MESTRADO/CALCULOS/RESULTADOS/simulacao/AMOSTRAS')
amostra <- fread (paste0('sample',i,'.csv'), dec=',', header = TRUE, sep = ';', na.strings = "")
expansao <- inner_join(amostra, universo_simulacao, by = c("ID_ESCOLA",
"ID_TURMA")) %>%
  select (ID_ALUNO, ID_TURMA,
ID_ESCOLA,IN_PROFICIENCIA,IN_SITUACAO_CENSO,PROFICIENCIA_LP_SAEB,P
ROFICIENCIA_MT_SAEB) %>%
  arrange(ID_ESCOLA, ID_TURMA, ID_ALUNO)

fpc <- peso %>% group_by(estrato) %>% summarise(fpc = sum(matriculas_estrato))

# Preparar arquivo para cálculo dos resultados
notas <- left_join(peso,expansao, by= "ID_ESCOLA") %>%

select(estrato,ID_UF,ID_ESCOLA,ID_TURMA,ID_ALUNO,PROFICIENCIA_LP_SAEB,P
ROFICIENCIA_MT_SAEB,peso_ac)
notas <- left_join(notas, fpc, "estrato")

notas$ID_UF <- factor(notas$ID_UF,levels=c(11:17,21:29,31:33,35,41:43,50:53))

#Configurar desenho amostral
design <-
svydesign(ids=~ID_ESCOLA,strata=~estrato,nest=TRUE,weights=~peso_ac,fpc=~fpc,
data=notas)
# Configurar opções de tratamento de estratos com apenas 1 escola
options(survey.lonely.psu="adjust")

# Calcular médias e erros Brasil
medias_lp <- data.frame (svymean(~PROFICIENCIA_LP_SAEB,design))
names(medias_lp) <- c ("media_lp","se_lp")
medias_lp$dp_lp <- sqrt(sum(notas$peso_ac*((notas$PROFICIENCIA_LP_SAEB -
medias_lp$media_lp)^2))/(sum(notas$peso_ac)-1))
medias_lp$scv_lp <- (medias_lp$dp_lp/medias_lp$media_lp)*100

medias_mt <- data.frame (svymean(~PROFICIENCIA_MT_SAEB,design))
names(medias_mt) <- c("media_mt","se_mt")
medias_mt$dp_mt <- sqrt(sum(notas$peso_ac*((notas$PROFICIENCIA_MT_SAEB -
medias_mt$media_mt)^2))/(sum(notas$peso_ac)-1))
medias_mt$scv_mt <- (medias_mt$dp_mt/medias_mt$media_mt)*100

media_br <- data.frame (i,ID_UF = 0,medias_lp,medias_mt)

# Calcular médias e erros por UF
medias_lp_uf <-
svyby(~PROFICIENCIA_LP_SAEB,~ID_UF,design,svymean,drop.empty.groups=F)
names(medias_lp_uf) <- c("ID_UF","media_lp","se_lp")

medias_mt_uf <-
svyby(~PROFICIENCIA_MT_SAEB,~ID_UF,design,svymean,drop.empty.groups=F)
names(medias_mt_uf) <- c("ID_UF","media_mt","se_mt")

```

```

dp <- notas %>% group_by(ID_UF) %>%
  summarise(total = sum(peso_ac))
notas <- left_join(notas, medias_lp_uf, by = "ID_UF")
notas$dq_lp <- notas$peso_ac*((notas$PROFICIENCIA_LP_SAEB - notas$media_lp)^2)
dq_lp_t <- notas %>% group_by(ID_UF) %>%
  summarise(dq_lp_t = sum(dq_lp))
dp$dp_lp <- sqrt(dq_lp_t$dq_lp_t/dp$total)

notas <- left_join(notas, medias_mt_uf, by = "ID_UF")
notas$dq_mt <- notas$peso_ac*((notas$PROFICIENCIA_MT_SAEB - notas$media_mt)^2)
dq_mt_t <- notas %>% group_by(ID_UF) %>%
  summarise(dq_mt_t = sum(dq_mt))
dp$dp_mt <- sqrt(dq_mt_t$dq_mt_t/dp$total)

media_uf <- data.frame(i,inner_join(medias_lp_uf,medias_mt_uf, "ID_UF"))
media_uf <- inner_join(media_uf, dp, by="ID_UF") %>%
  select(1,2,3,4,8,5,6,9)

media_uf$cv_lp <- (media_uf$dp_lp/media_uf$media_lp)*100
media_uf$cv_mt <- (media_uf$dp_mt/media_uf$media_mt)*100

media_uf <- select(media_uf, 1,2,3,4,5,9,6,7,8,10)

# Consolidar arquivo de resultados
resultados_i <- rbind(media_br,media_uf)
rownames(resultados_i) <- c()
resultados <- rbind(resultados, resultados_i)

# contar tamanho da amostra
n_esc = nrow(peso)
n_tur = sum(peso$turmas_realizadas)
n_alu = sum(peso$matriculas_realizadas)

n_i <- c(i, n_esc, n_tur, n_alu)
n <- rbind(n, n_i)

}

##### encerrar loop #####

# salvar arquivo de resultados
setwd ('D:/MESTRADO/CALCULOS/RESULTADOS/simulacao')
resultados <- filter(resultados, !is.na(i))
write.csv2(resultados,'resultados.csv', row.names = FALSE)

# salvar arquivo de tamanho das amostras
n <- filter(n, !is.na(i))
write.csv2(n,'n.csv', row.names = FALSE)

```

PASSO 9_analise_resultados (R)

```
# Análise dos resultados da simulação
```

```
# Habilitar pacotes
```

```
library (data.table)
```

```
library (dplyr)
```

```
# Ler arquivo de resultados da simulação em .csv
```

```
setwd ('D:/MESTRADO/CALCULOS/RESULTADOS/simulacao')
```

```
sim <- fread ('resultados.csv', header = TRUE, dec = ",", sep = ';', na.strings = "")
```

```
# Ler arquivo original de alunos participantes no Saeb 2017 e universo da simulação em .csv
setwd ('D:/MESTRADO/CALCULOS/DADOS')
```

```
ts_aluno_5ef <- fread ('TS_ALUNO_5EF.csv', header = TRUE, dec = ".", sep = ',', na.strings = "NA")
```

```
universo_simulacao <- fread ('universo_simulacao.csv', header = TRUE, dec = ".", sep = ',', na.strings = "NA")
```

```
# Informações básicas sobre o Universo da simulação
```

```
names (universo_simulacao)
```

```
class (universo_simulacao$PROFICIENCIA_LP_SAEB)
```

```
table (universo_simulacao$ID_DEPENDENCIA_ADM)
```

```
qt_tur = ts_aluno_5ef %>% summarise(qt_tur = n_distinct(ID_TURMA))
```

```
qt_alu = nrow(universo_simulacao)
```

```
m_tur = qt_alu/qt_tur
```

```
m_tur
```

```
# carregar arquivo de tamanho da amostra
```

```
setwd ('D:/MESTRADO/CALCULOS/RESULTADOS/simulacao')
```

```
n <- fread ('n.csv', header = TRUE, dec = ",", sep = ';', na.strings = "")
```

```
# Medidas descritivas do tamanho da amostra
```

```
media_n_esc <- mean(n$n_esc, na.rm = TRUE)
```

```
media_n_tur <- mean(n$n_tur, na.rm = TRUE)
```

```
media_n_alu <- mean(n$n_alu, na.rm = TRUE)
```

```
media_n <- c(media_n_esc, media_n_tur, media_n_alu)
```

```
media_n
```

```
# Verificar perdas em relação ao número de turmas esperado
```

```
# Tamanho da amostra escolas = 2.546 turmas = 3.335 alunos = 73.830
```

```
n$perda <- 3335 - n$n_tur
```

```
n$percent_perda <- (n$perda/3335)*100
```

```
media_perda <- mean(n$perda, na.rm = TRUE)
```

```
media_percent_perda <- mean(n$percent_perda, na.rm = TRUE)
```

```
# quantidades originais do Saeb 2017
```

```
ts_aluno_5ef <- ts_aluno_5ef %>% filter (ID_DEPENDENCIA_ADM %in% c(2,3))
```

```
table (ts_aluno_5ef$ID_DEPENDENCIA_ADM)
```

```
n_esc = ts_aluno_5ef %>% summarise(n_esc = n_distinct(ID_ESCOLA))
```

```
n_tur = ts_aluno_5ef %>% summarise(n_tur = n_distinct(ID_TURMA))
```

```

n_alu = nrow(ts_aluno_5ef)
n_saeb <- data.frame(i=0, n_esc, n_tur, n_alu)
n_saeb

# calcular médias aritméticas do Saeb
Saeb_lp <- universo_simulacao %>%
  group_by (ID_UF) %>%
  summarise (media_lp_saeb = mean(PROFICIENCIA_LP_SAEB, na.rm = TRUE))

Saeb_mt <- universo_simulacao %>%
  group_by (ID_UF) %>%
  summarise (media_mt_saeb = mean(PROFICIENCIA_MT_SAEB, na.rm = TRUE))
saeb_uf <- inner_join(Saeb_lp, Saeb_mt, "ID_UF")

MEDIA_LP <- mean(universo_simulacao$PROFICIENCIA_LP_SAEB, na.rm = TRUE)
MEDIA_MT <- mean(universo_simulacao$PROFICIENCIA_MT_SAEB, na.rm = TRUE)
saeb_br <- c(0, MEDIA_LP, MEDIA_MT)

saeb <- rbind(saeb_br, saeb_uf)

# Calcular estimativa de teta médio (média das médias)
medias_lp <- sim %>%
  group_by(ID_UF) %>%
  summarise(media_lp_media = mean(media_lp))

medias_mt <- sim %>%
  group_by(ID_UF) %>%
  summarise(media_mt_media = mean(media_mt))

# Criar tabela de indicadores por UF
indic <- inner_join(saeb, medias_lp, "ID_UF")
indic <- inner_join(indic, medias_mt, "ID_UF")

# Criar tabela completa de resultados
result <- inner_join(sim, indic, "ID_UF")

# Calcular viés
result$diferenca_lp <- (result$media_lp- result$media_lp_saeb)
result$diferenca_mt <- (result$media_mt- result$media_mt_saeb)

max(abs(result$diferenca_lp))
min(abs(result$diferenca_lp))

max(abs(result$diferenca_mt))
min(abs(result$diferenca_mt))

vies_lp <- result %>%
  group_by(ID_UF) %>%
  summarise(vies_lp = mean(diferenca_lp))

```

```

vies_mt <- result %>%
  group_by(ID_UF) %>%
  summarise(vies_mt = mean(diferenca_mt))

indic <- inner_join(indic, vies_lp, "ID_UF")
indic <- inner_join(indic, vies_mt, "ID_UF")

# Calcular estimativa do Erro Quadrático Médio (EQM)
result$dif_quad_lp <- (result$diferenca_lp)^2
erro_lp <- result %>%
  group_by(ID_UF) %>%
  summarise(eqm_lp = mean(dif_quad_lp))

result$dif_quad_mt <- (result$diferenca_mt)^2
erro_mt <- result %>%
  group_by(ID_UF) %>%
  summarise(eqm_mt = mean(dif_quad_mt))

indic <- inner_join(indic, erro_lp, "ID_UF")
indic <- inner_join(indic, erro_mt, "ID_UF")

# Calcular variâncias em função do EQM e do Viés
indic$var_lp <- indic$eqm_lp - (indic$vies_lp^2)
indic$var_mt <- indic$eqm_mt - (indic$vies_mt^2)

# Calcular variâncias pela fórmula original (Para conferência)
result$erro_quad_lp <- (result$media_lp-result$media_lp_media)^2
variancia_lp <- result %>%
  group_by(ID_UF) %>%
  summarise(var2_lp = mean (erro_quad_lp))

result$erro_quad_mt <- (result$media_mt-result$media_mt_media)^2
variancia_mt <- result %>%
  group_by(ID_UF) %>%
  summarise(var2_mt = mean (erro_quad_mt))

indic <- inner_join(indic, variancia_lp, "ID_UF")
indic <- inner_join(indic, variancia_mt, "ID_UF")

# Calcular Margem de erro
result$me_lp <- 1.96*(result$sse_lp)
result$me_mt <- 1.96*(result$sse_mt)

me_lp_media <- result %>%
  group_by(ID_UF) %>%
  summarise(me_lp_media = mean(me_lp))

me_mt_media <- result %>%
  group_by(ID_UF) %>%
  summarise(me_mt_media = mean(me_mt))

```

```

indic <- inner_join(indic, me_lp_media, "ID_UF")
indic <- inner_join(indic, me_mt_media, "ID_UF")

# intervalos de confiança
result$li_lp <- result$media_lp - result$me_lp
result$ls_lp <- result$media_lp + result$me_lp
result$c_lp <- ifelse((result$media_lp_saeb)>result$li_lp &
(result$media_lp_saeb)<result$ls_lp,1,0)
table(result$c_lp, result$ID_UF)

result$li_mt <- result$media_mt - result$me_mt
result$ls_mt <- result$media_mt + result$me_mt
result$c_mt <- ifelse((result$media_mt_saeb)>result$li_mt &
(result$media_mt_saeb)<result$ls_mt,1,0)
table(result$c_mt, result$ID_UF)

# Contar intervalos de confiança que contém a média real (c_lp e c_mt = 1)
cobertura_lp <- result %>%
  group_by(ID_UF) %>%
  summarise/qt_cobre_lp = sum(c_lp))
indic <- inner_join(indic, cobertura_lp, "ID_UF")

cobertura_mt <- result %>%
  group_by(ID_UF) %>%
  summarise/qt_cobre_mt = sum(c_mt))
indic <- inner_join(indic, cobertura_mt, "ID_UF")

# Organizar tabela indic e salvar em .csv
indic <- select(indic,"ID_UF",

"media_lp_saeb","media_lp_media","vies_lp","eqm_lp","var_lp","var2_lp","me_lp_media",
qt_cobre_lp",

"media_mt_saeb","media_mt_media","vies_mt","eqm_mt","var_mt","var2_mt","me_mt_med
ia","qt_cobre_mt")

setwd ('D:/MESTRADO/CALCULOS/RESULTADOS/simulacao')
write.csv2(indic,'indic.csv',row.names = FALSE)

PASSO 10_graficos (R)
# Gráficos dos resultados da simulação

# Habilitar pacotes
library (data.table)
library (dplyr)
library(ggplot2)

# Ler arquivo de resultados da simulação e siglas da UF's em .csv
setwd ('D:/MESTRADO/CALCULOS/RESULTADOS/simulacao')

```

```

resultados_orig <- fread('resultados.csv', header = TRUE, dec = ",", sep = ';', na.strings = "")
uf <- fread("UF.csv",header = TRUE, dec = ",", sep = ';', na.strings = "")
indic <- fread('indic.csv', header = TRUE, dec = ",", sep = ';', na.strings = "")

# Incluir siglas e nomes das UFs na base resultados
resultados <- inner_join(resultados_orig, uf, by = "ID_UF")
resultados <- inner_join(resultados, indic, by = "ID_UF") %>%
  select(i,ID_UF,SG_UF, media_lp, media_mt, media_lp_saeb, media_mt_saeb)

names(resultados)

# resumo dos dados de resultados
resultados %>% glimpse

# Histograma das médias_lp por UF
resultados %>%
  filter(ID_UF!= 0) %>%
  ggplot() +
  geom_histogram(aes(x = media_lp),bins=50, fill='#70AD47')+
  facet_wrap(~SG_UF)+
  theme_light()+
  labs(x = "média estimada para LP ( $\theta_{LP}$ ", y = "frequência",size=9.0))+
  theme(axis.text.x=element_text(colour = "black", size=7.0))+           #eixo x - TEXTO
  theme(axis.text.y=element_text(colour = "black", size=7.0))+           #eixo Y - TEXTO
  ggsave('hist_medias_uf_lp.tiff', plot = last_plot(), dpi = 320)

# Histograma das médias_mt por UF
resultados %>%
  filter(ID_UF!= 0) %>%
  ggplot() +
  geom_histogram(aes(x = media_mt),bins=50, fill='#4472C4')+
  facet_wrap(~SG_UF)+
  theme_light()+
  labs(x = "média estimada para MT ( $\theta_{MT}$ ", y = "frequência",size=9.0))+
  theme(axis.text.x=element_text(colour = "black", size=7.0))+           #eixo x - TEXTO
  theme(axis.text.y=element_text(colour = "black", size=7.0))+           #eixo y - TEXTO
  ggsave('hist_medias_uf_mt.tiff', plot = last_plot(), dpi = 320)

# Histograma das médias por regioao
resultados %>% filter(ID_UF %in% 11:17) %>%
  ggplot() +
  geom_histogram(aes(x = media_lp),bins=50, fill='#70AD47')+
  facet_wrap(~SG_UF)+
  theme_light()+
  labs(x = "média estimada para LP ( $\theta_{LP}$ ", y = "frequência")
  ggsave('medias_norte_lp.tiff', plot = last_plot(), dpi = 320)

resultados %>% filter(ID_UF %in% 11:17) %>%
  ggplot() +
  geom_histogram(aes(x = media_mt),bins=50, fill='#4472C4')+

```

```

facet_wrap(~SG_UF)+
theme_light()+
labs(x = "média estimada para MT ( $\theta_{MT}$ )", y = "frequência")
ggsave('medias_norte_mt.tiff', plot = last_plot(), dpi = 320)

# Histograma das médias_lp Brasil
resultados %>%
  filter(ID_UF == 0) %>%
  ggplot() +
  geom_histogram(aes(x = media_lp),bins=50, fill='#70AD47', color = '#4C7430')+
  theme_light()+
  labs(x = "média estimada para LP ( $\theta_{LP}$ )", y = "frequência")
ggsave('medias_brasil_LP.tiff', plot = last_plot(), dpi = 320)

# Histograma das médias_mt Brasil
resultados %>%
  filter(ID_UF == 0) %>%
  ggplot() +
  geom_histogram(aes(x = media_mt),bins=50, fill='#4472C4',color = '#254275')+
  theme_light()+
  labs(x = "média estimada para MT ( $\theta_{MT}$ )", y = "frequência")
ggsave('medias_brasil_MT.tiff', plot = last_plot(), dpi = 320)

# Grafico de dispersão - LP
resultados %>%
  filter(ID_UF == 0) %>%
  ggplot() +
  geom_point(aes(x = i, y = media_lp), color = '#70AD47')+
  geom_line(aes(x = i, y = media_lp_saeb), size = 1, color = '#006600')+
  theme_light()+
  theme(legend.position="bottom")+
  xlab(" $\theta_{LP}$        $\theta_{LP}$ ") +
  ylab(NULL)
ggsave('dispersao_brasil_lp.tiff', plot = last_plot(), dpi = 320)

# Grafico de dispersão - MT
resultados %>%
  filter(ID_UF == 0) %>%
  ggplot() +
  geom_point(aes(x = i, y = media_mt), color = '#4472C4')+
  geom_line(aes(x = i, y = media_mt_saeb), size = 1, color = '#000066')+
  theme_light()+
  xlab(" $\theta_{MT}$        $\theta_{MT}$ ") +
  ylab(NULL)
ggsave('dispersao_brasil_mt.tiff', plot = last_plot(), dpi = 320)

```