

METAS SOCIAIS

Simulador do Indicador de Inclusão Escolar
Natura, Construído a Partir de um Modelo de
Fluxo e Desempenho Escolar

Relatório Técnico

Metas Sociais para Instituto Natura

04 de maio de 2023

Sumário

Apresentação.....	3
O Índice de Inclusão Educacional - IIE	4
O uso das coortes de nascimento vizinhas.....	7
O alinhamento idade-série no Censo Escolar.....	8
Metodologia para Previsão de Impacto de Mudanças no Fluxo e na Aprendizagem sobre o IIE	9
Modelo de Fluxo e Proficiência Escolar	11
Ingresso	12
Ensino Fundamental I	14
Ensino Fundamental II.....	19
Ensino Médio	22
Estrutura Lógica do Simulador.....	24
Etapa 1 – Ingresso.....	24
Etapa 2 – Ensino Fundamental 1	25
Etapa 3 – Ensino Fundamental 2	25
Etapa 4 – Ensino Médio	26
Cálculo das Probabilidades de transição	27
As Funções de Ligação	34
Simulações de variação nas taxas de aprovação.....	35
Simulações de variação na aprendizagem.....	35
Simulações de variação no IDEB	37
Simulações de variação na alfabetização (proficiência do 2º ano).....	39
a Implementação do simulador do IIE - Natura.....	40
Apêndice	42

Modelo de Fluxo e Desempenho Escolar e Simulador do Indicador de Inclusão Educacional Natura

Reynaldo Fernandes

Fabiana de Felicio

Apresentação

O Instituto Natura idealizou um indicador para monitorar a inclusão no nível básico da educação no Brasil, e em cada um de seus estados e Distrito Federal, que considerasse a idade e o nível de aprendizagem dos jovens ao concluir o ensino médio. A partir dessa definição, foi concebido o Índice de Inclusão Educacional – IIE, que traz uma proposta inovadora de indicador educacional definido como o percentual dos jovens nascidos em um determinado período de 12 meses, que conclui o ensino médio no ano em que completa 18 anos ou antes, tendo alcançado um nível de aprendizagem considerado adequado.

Feitas algumas escolhas e assumidas algumas hipóteses, foi possível calcular o IIE a partir dos dados nacionais, conforme detalhado ao longo deste relatório técnico. Mas, se este for o indicador escolhido para representar a qualidade das oportunidades da educação básica, surgem questões relevantes tais como: Qual a relação entre indicadores educacionais tradicionais, medidos ao longo do ensino fundamental e médio, e o IIE? Qual será o IIE de uma coorte de nascimento que está em idade escolar e que pode frequentar a escola em diferentes cenários, a depender das políticas públicas implementadas? Quais as políticas educacionais com maior impacto sobre o IIE?

Para viabilizar essas análises, foi desenvolvido um novo Modelo de Fluxo Escolar, que incorpora a aprendizagem e condiciona a probabilidade de aprovação em uma etapa de ensino ao desempenho escolar da etapa anterior, medidos pelo atraso escolar e nível

de proficiência. A partir desta modelagem, calcula-se a distribuição de uma geração entre os estados possíveis e calculam-se as probabilidades de transição ao longo da vida escolar. Essas informações alimentam um simulador, construído para emular inúmeros cenários possíveis, que podem ser experimentados por uma coorte de nascimento ao longo de sua vida escolar, e então calcular o IIE esperado para essa geração aos 18 anos.

A metodologia e os procedimentos adotados para calcular o IIE e alcançar o objetivo de viabilizar simulações de efeitos de diferentes cenários sobre o IIE estão apresentados nas próximas seções.

O Índice de Inclusão Educacional - IIE

O IIE tem por objetivo sintetizar em um só índice medidas dos principais fatores que caracterizam a inclusão educacional no nível básico: o ingresso das crianças no ensino fundamental na idade adequada, a evolução contínua entre as etapas escolares, considerando a evasão e a reprovação, e o nível de aprendizagem ao concluir o ensino médio. Por exemplo, o IIE da geração de 2000 representa a proporção dos nascidos no ano 2000, que concluíram o ensino médio com até 18 anos (até 2018), com o nível de proficiência básico ou mais (conforme o Saeb).

Essa definição é clara, completa e de fácil compreensão, porém seu cálculo não é simples. O IIE é um índice rigoroso e, por contemplar tanto o nível de aprendizagem quanto a idade adequada para conclusão do ensino médio de toda uma coorte de nascimento refere-se não a uma série escolar em um ano-calendário (como a 3ª série do ensino médio em 2017), mas sim a todos os nascidos em um mesmo ano calendário que, para simplificar, chamaremos de *geração*. Essa geração é composta por pessoas que seguem trajetórias muitas vezes diferentes ao longo da vida escolar. Em especial no Brasil, em que a taxa média de reprovação ainda é alta, bem como a evasão antes da

conclusão da educação básica, uma única geração pode ter pessoas percorrendo inúmeros caminhos diferentes e aos 18 anos podem estar fora da escola, tendo ou não concluído o ensino médio, ou na escola, em diferentes séries. O princípio do IIE de referir-se a uma geração, pressupõe conhecer os principais caminhos possíveis e localizar todos os componentes de uma geração de interesse.

Nenhuma base de dados oficial nacional dispõe de todos esses dados necessários. Foi desenvolvida, então, uma metodologia a fim de obter uma estimativa para o IIE a partir dos dados disponíveis: dados do Saeb e Censo Escolar, ambas bases coletadas pelo Inep/ MEC, e da PNADc, pesquisa conduzida pelo IBGE.

A construção se baseia na classificação dos componentes da geração de interesse conforme os dois componentes determinantes do IIE: (i) alinhamento idade-série e; (ii) nível de proficiência. A composição de ambas as classificações adotadas está apresentada no quadro abaixo. Os próximos quadros ilustram os dados necessários para a construção do IEE com os dados observados.

Como indicador de inclusão na educação básica, o IIE foi pensado para a idade adequada de conclusão do ensino médio. Entretanto a mesma metodologia foi estendida para as duas etapas do ensino fundamental, permitindo, assim, o acompanhamento ao longo da vida escolar e contribuindo para o melhor diagnóstico da questão educacional como um todo.

Quadro 1. Classificação de alinhamento idade-série

Adiantado	Componentes da geração de interesse que concluem uma cada etapa de ensino com um ano de idade a menos que o recomendado: concluem os anos iniciais no ano em que completam 9 anos; os anos finais no ano em que completam 13 anos e; o ensino médio aos 16 anos.
Em linha	Componentes da geração de interesse que concluem uma cada etapa de ensino com a idade recomendada: os anos iniciais, no ano em que completam 10 anos; os anos finais no ano em que completam 14 anos e; o ensino médio aos 17 anos.
Atrasado até 1 ano	Componentes da geração de interesse que concluem uma cada etapa de ensino com um ano de idade a mais que o recomendado: os anos iniciais, no ano em que completam 11 anos; os anos finais no ano em que completam 15 anos e; o ensino médio aos 18 anos.
Atrasado 2 anos ou mais	Componentes da geração de interesse que concluem uma cada etapa de ensino com dois anos de idade a mais que o recomendado, ou mais

velhos: os anos iniciais, no ano em que completam 12 anos ou mais; os anos finais no ano em que completam 16 anos ou mais e; o ensino médio aos 19 anos ou mais. Também são incluídos aqui aqueles que não completam o ensino médio.

Classificação de nível de proficiência (em pontos do Saeb)								
Anos Iniciais			Anos Finais			Ensino Médio		
Abaixo do básico	Básico	Adequado	Abaixo do básico	Básico	Adequado	Abaixo do básico	Básico	Adequado
<200	200 a 275	>275	<275	275 a 350	>350	<300	300 a 375	>375

Quadro 2. Ilustração dos dados necessários para construção do IIE, segundo a classificação de alinhamento idade série e níveis de proficiência e geração observada em uma edição do Saeb

	Abaixo do Básico	Básico	Adequado	Geração observada em uma edição do Saeb
Adiantados	X1	X2	X3	G1
Em Linha	Z1	Z2	Z3	G2
Atrasados 1	V1	V2	V3	G3
Atrasados 2	W1	W2	W3	G4
Total da Coorte de nascimento de interesse	Ab	B	A	Universo do Saeb

Em uma única edição do Saeb, apenas uma das linhas do Quadro 2 de uma dada geração (G) é observada. No caso do interesse ser na geração que deveria concluir o ensino médio no mesmo ano do Saeb, trata-se da geração G2, e apenas a 2ª linha, dos jovens *Em Linha*, refere-se à essa geração. Ainda que consideradas as ‘coortes vizinhas’, como detalhado a seguir, como uma aproximação dos quantitativos para os demais estados (linhas referentes à G1, G3 e G4), apenas uma parte da geração estaria representada no Universo do Saeb: a parte que está na escola.

As crianças e jovens que estão fora da escola e que não cursaram a última série do ensino médio (ou da etapa de ensino em questão) não participam do Saeb, porém compõem a geração de interesse. Não considerar tal parcela superestimaria o IIE, pois eles deveriam compor o denominador do indicador, contabilizados na parcela da geração que teria 2 anos ou mais de atraso escolar.

A solução adotada foi utilizar os dados da PNADc/ IBGE, em que se observa uma amostra de toda população do estado. Para isso, selecionam-se os dados da coorte de interesse – no exemplo acima, seriam todos os nascidos no ano 2000 - e verifica-se a proporção dessa coorte que está fora da escola sem nunca ter cursado o último ano da etapa de interesse (F).

Considera-se que aqueles que cursaram a última série estão representados no Saeb, ainda que por meio de ‘coortes vizinhas’. Já F não é observado no Saeb e, portanto, precisa ser contabilizado a partir da PNADc, sendo adicionado à W1, para compor o grupo de Atrasados 2-Abaixo do Básico, apenas por questão de organização, já que o nível de proficiência neste caso não afeta o índice.

Considerando, então, os dados de uma edição do Saeb, complementados com os dados da PNADc para os que estão fora da escola sem concluir o ensino médio, tem-se a geração toda representada, conforme o quadro a seguir.

Quadro 3. Ilustração dos dados necessários para construção do IIE, segundo a classificação de alinhamento idade série e níveis de proficiência e geração observada em uma edição do Saeb, complementados pelos jovens fora da escola, relativamente ao total da geração de interesse (G)

	Proporção Abaixo do Básico	Proporção Básico	Proporção Adequado	Proporção Total na linha
Adiantados	X1/ G	X2/ G	X3/ G	G1/ G
Em Linha	Z1/ G	Z2/ G	Z3/ G	G2/ G
Atrasados 1	V1/ G	V2/ G	V3/ G	G3/ G
Atrasados 2	(W1+F4) / G	W2/ G	W3/ G	(G4+F4) / G
Total	(Ab+F4) / G	B/ G	A/ G	1 (ou G/G)

O uso das coortes de nascimento vizinhas

Dispondo de dados para compor esse quadro para a geração G, é possível calcular o IIE para os níveis de proficiência básico ou adequado, a partir das seguintes equações:

$$IEE_Adequado = (X3 + Z3 + V3) / G$$

$$\text{IEE_Básico} = (X2 + Z2 + V2 + X3 + Z3 + V3) / G$$

Porém, não se trata de um cálculo simples. Alguns desafios se apresentam, sendo o principal o fato de o Saeb ser uma avaliação bienal, de forma que não é possível verificar no Saeb a proficiência de todos os nascidos em um mesmo ano calendário, tampouco em diferentes edições da avaliação. No exemplo acima, seria possível observar a proficiência da coorte de 2000 que concluiria o ensino médio, se em linha, no ano de 2017, porém não se observariam os que concluem em 2018 (com um ano de atraso), tampouco os que concluíram adiantados, em 2016 – todos necessários para compor o numerador do índice.

É preciso utilizar, então, uma aproximação a partir da observação do que denominamos ‘coortes vizinhas’ representadas por G1, G3 e G4 no quadro acima e que, nesse exemplo, seriam: (i) nascidos em 2001 ou anos seguintes que fizeram Saeb em 2017 (16 anos ou menos), como aproximação da parcela da coorte 2000 que concluiu o ensino médio adiantado; (ii) nascidos em 1999 que fizeram Saeb em 2017 (18 anos), representando a parcela da coorte 2000 que concluiu o ensino médio com um ano de atraso e; (iii) nascidos em 1998 ou anos anteriores que fizeram o Saeb em 2017 (19 anos ou mais), representando a parcela da coorte 2000 que concluiu o ensino médio com dois anos de atraso ou mais.

O alinhamento idade-série no Censo Escolar

O Censo Escolar é a base considerada mais completa. Todas as crianças e jovens matriculados na escola devem ser reportadas ao Inep por todas as escolas brasileiras em dois momentos por ano. Na primeira declaração, muitas informações são preenchidas no nível individual de todos os alunos matriculados e professores, além de outras questões sobre a própria escola, referentes ao início daquele mesmo ano letivo. Na segunda coleta, as escolas reportam informações sobre o final do ano letivo

anterior, denominado Situação do Aluno, e as informações dizem respeito, na maior parte dos casos, ao rendimento escolar de cada aluno – se terminou o ano com aprovação, reprovação ou abandono¹.

Usando 3 edições subsequentes dos dados coletados de Situação do Aluno, é possível identificar todos os jovens de uma geração que concluem a etapa de ensino (o que não é observado no Saeb ou outra base de dados), qualquer que seja série que frequenta, contanto que esteja matriculado na escola. Assim, é possível identificar a distribuição da geração de interesse entre as classificações de alinhamento idade série, o que faz desta, a melhor fonte para essa informação. Por outro lado, não é possível identificar aqueles que evadiram, tampouco classificá-los quanto ao nível de aprendizagem, e continua sendo necessário utilizar os dados do Saeb, complementados pela PNADc.

Essa combinação de informações se dá, então, redistribuindo cada grupo de nível de aprendizagem observado no Saeb (representados pelas colunas 2, 3 e 4 do Quadro 3), conforme a distribuição de alinhamento idade série observada no Censo Escolar. Em ambos os casos, os jovens da geração de interesse observados como fora da escola na PNADc, são adicionados ao grupo com 2 ou mais anos de atraso.²

Metodologia para Previsão de Impacto de Mudanças no Fluxo e na Aprendizagem sobre o IIE

O IIE, por si só, tem diversas vantagens relativamente a outros indicadores de qualidade da educação, como ser um índice único para toda a educação básica (ainda

¹ Também constam como situações possíveis, os casos de alunos que tiverem sido transferidos de escola ou se houver caso de falecimento.

² Mais detalhes sobre essa abordagem e quanto ao caso especial do Saeb 2019 e a omissão da variável de idade (ou data de nascimento) estão no Apêndice deste relatório.

que ele possa ser calculado por etapa de ensino), utilizar poucas informações (o que o torna de fácil compreensão), e contemplar todos os jovens nascidos em um mesmo ano, já que considera os que frequentam a escola, os que evadiram e os que concluíram o ensino médio. Assim como outros índices, tais como IDEB e IOEB, também se presta a diversas aplicações, como acompanhamento da qualidade da educação em um território, identificação de casos de preocupação e casos de sistema educacional bem sucedido, entre outros. Entre os desafios, estão a dificuldade de calcular o índice para territórios menores que estados e Distrito Federal e a relação desconhecida com os indicadores educacionais tradicionalmente acompanhados.³

O objetivo aqui é apresentar a metodologia desenvolvida para avaliar o impacto indireto de programas e políticas públicas sobre o IIE. Ou seja, dado o impacto estimado das ações sobre a idade de ingresso no 1º ano do fundamental, sobre o nível de alfabetização ao final do 2º ano e, principalmente, sobre as taxas de rendimento escolar, a proficiência ou o IDEB, gerar uma estimativa de qual seria o efeito no curto ou médio prazo sobre o IIE.

Desenvolveu-se, então, uma estratégia que visa simular possíveis cenários em que alguma(s) medida(s) altera(m) as probabilidades de ingresso no ensino fundamental por idade, probabilidade de aprovação ou probabilidade de conclusão das etapas de ensino por nível de proficiência e, a partir dessas mudanças, calcular o IIE futuro das gerações que frequentarem a escola nessas novas condições.

Para a construção do que chamaremos de Simulador do IIE - SIIE, inicialmente é necessário definir o método de cálculo dessas probabilidades e então calcular quais seria o ponto de partida do simulador, ou seja, as probabilidades mais recentes

³ Para saber mais sobre os outros índices educacionais, ver o Texto para Discussão do Inep/MEC, Índice de Desenvolvimento da Educação Básica – Ideb, de autoria de Reynaldo Fernandes (2007) - disponível em: <http://td.inep.gov.br/ojs3/index.php/td/article/view/3850> e o site do IOEB, www.ioeb.org.br e Nota Técnica do IOEB 2021, de autoria de Reynaldo Fernandes e Fabiana de Felício <https://ioeb.org.br/nota-tecnica-ioeb-2021/>

observadas. Adicionalmente, algumas simulações de cenários de interesse partem de mudanças nas proporções de jovens em cada estado, ou seja, em cada categoria de alinhamento idade série e nível de proficiência. Assim, também é necessário dispor das proporções de crianças e jovens em cada estado, por etapa de ensino.

Para chegar nas estimativas desejadas para o IIE, foram desenvolvidas as etapas que seguem:

- a) O **Modelo de Fluxo & Aprendizagem** define o método de cálculo das probabilidades e proporções necessárias;
- b) A **Estrutura Lógica do Simulador**, para as adequações das equações do modelo a fim de implementar o simulador
- c) O **Cálculo das Probabilidades de Transição**, a fim de implementar os cálculos necessários com base no modelo;
- d) Análise e definição das **Funções de Ligação**, que estimam as relações entre as variações nas taxas de aprovação e nas proficiências, em mudanças das probabilidades de transição do modelo de fluxo, e entre as variações de aprovação e proficiência com as variações do IDEB.
- e) Construção do Simulador do IEE – **SIIE-Natura**, para implementar a automatização do cálculo do IIE a partir das simulações de mudanças das probabilidades ou das proporções de jovens em cada estado.

MODELO DE FLUXO & APRENDIZAGEM

Modelos de fluxo são modelos que ligam estados da natureza (estados educacionais, no nosso caso) por meio de probabilidades de transição. O modelo desenvolvido aqui considera uma geração hipotética, composta por todas as crianças nascidas em um intervalo de 12 meses, e os estados educacionais possíveis em que as crianças dessa geração poderão se encontrar estão divididos em 4 etapas: ingresso, término dos anos iniciais, término dos anos finais e término do ensino médio.

O objetivo do modelo é representar o fluxo escolar dos estudantes dessa geração entre as etapas de ensino, agrupadas em três níveis de proficiência e 4 grupos de alinhamento idade série, de forma a obter a proporção dos que concluem cada etapa de ensino com até um ano de atraso e sejam consideradas proficientes.

Os estados possíveis para cada etapa são:

- a) Quanto à etapa de ingresso no ensino fundamental, a geração está distribuída de acordo com o alinhamento idade-série. Tem-se, então, 4 estados: *adiantado, em linha, 1 ano de atraso e 2 anos ou mais de atraso*.
- b) Nas demais etapas, a geração está distribuída considerando a combinação entre o alinhamento idade-série e os 3 níveis de proficiência definidos para o final de cada uma das 3 etapas de ensino (anos iniciais e anos finais do ensino fundamental e ensino médio): *abaixo do básico, básico e adequado*. O que resulta em 12 estados para cada etapa de ensino.

Desse modo, o modelo considera 40 estados educacionais. As probabilidades são unidirecionais e assume-se que não há transição intragrupos: as transições vão do ingresso para os anos iniciais; dos anos iniciais para os anos finais; e dos anos finais para o EM. Há três conjuntos de probabilidades: probabilidades de ingresso, probabilidades de aprovação e probabilidades de conclusão das etapas em um dos três níveis de proficiência considerados (*abaixo do básico, básico e adequado*).

Ingresso

Seja $G(0)$ o número de crianças da geração que deveria ingressar na escola no tempo 0. Os alunos dessa geração podem ingressar na escola no período -1 (um ano adiantado), no período 0 (no período correto), no período 1 (com um ano de atraso) ou em outros períodos (com 2 ou mais anos de atraso). Definindo p_i^I como a probabilidade de

ingresso com i anos de atraso (o sinal negativo indica anos adiantados) e E_i como o número de alunos que ingressam com i anos de atraso, temos (1).

$$E_{-1} = G(0)p_{-1}^I \quad (1)$$

$$E_0 = G(0)p_0^I$$

$$E_1 = G(0)p_1^I$$

$$E_{2+} = G(0)p_{2+}^I = G(0) - E_{-1} - E_0 - E_1$$

Note que $G(0)$, o tamanho da coorte, pode ser estimado com base no Censo Escolar (que dá o número de estudantes na escola) e na PNADc (que dá a proporção de estudantes da geração fora da escola). O Censo Escolar também nos fornece E_i ($i = -1, 0, 1$ e 2) e, assim, as probabilidades de ingresso (p_i^I) são identificáveis.

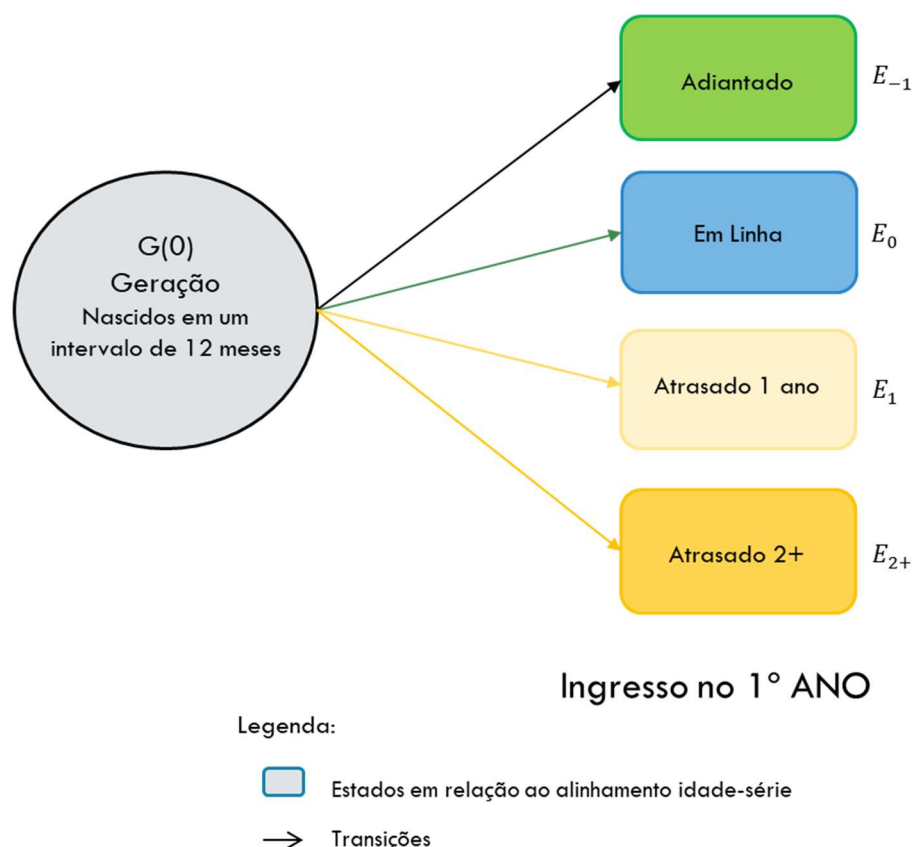


Figura 1. Ilustração da distribuição de uma geração entre os estados da etapa de Ingresso no ensino fundamental

Ensino Fundamental I

Algumas hipóteses simplificadoras são necessárias para viabilizar a abordagem das questões reais em um modelo. Neste caso, admitiu-se que nenhum estudante evade da escola sem atingir 2 ou mais anos de atraso escolar e que não existem programas de aceleração para reduzir o atraso escolar.

Definiu-se também FI_i como o número de alunos de $G(0)$ que completam a primeira etapa do ensino fundamental (EFI) com i anos de atraso e p_i^{AFI} a probabilidade de aprovação de uma série do EFI (que é invariante em relação à série) para os alunos que ingressaram na escola com i anos de atraso. Assim, pode-se obter (2).

$$FI_{-1} = E_{-1}(p_{-1}^{AFI})^5 \quad (2)$$

$$FI_0 = E_{-1}5(p_{-1}^{AFI})^5(1 - p_{-1}^{AFI}) + E_0(p_0^{AFI})^5$$

$$FI_1 = E_{-1}15(p_{-1}^{AFI})^5(1 - p_{-1}^{AFI})^2 + 5E_0(p_0^{AFI})^5(1 - p_0^{AFI}) + E_1(p_1^{AFI})^5$$

$$FI_{2+} = G(0) - FI_{-1} - FI_0 - FI_1$$

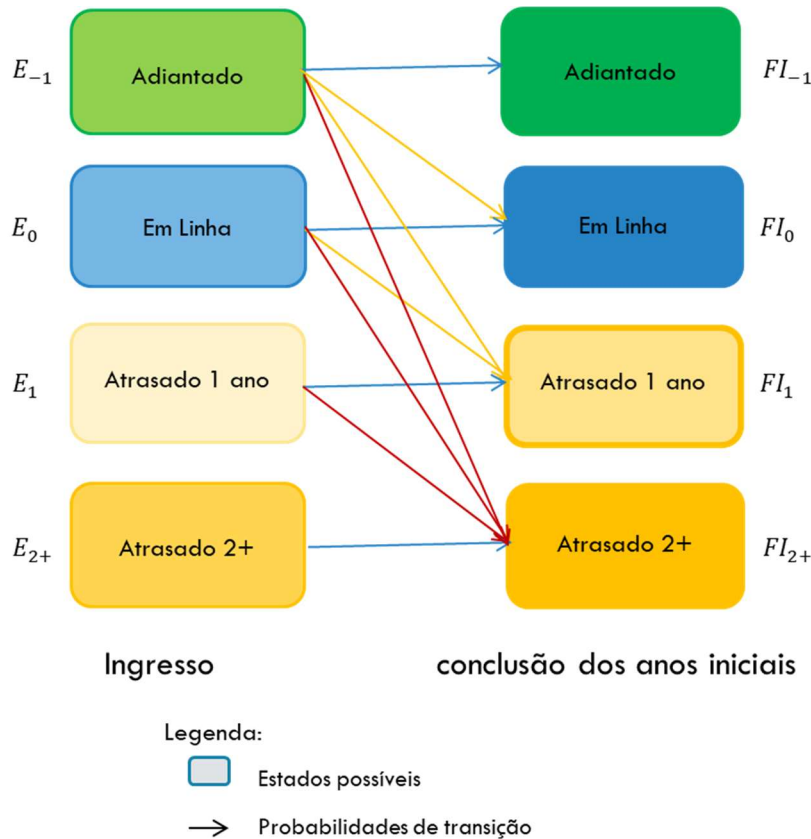


Figura 2. Ilustração das transições possíveis de para uma geração entre a etapa de ingresso e o final dos anos iniciais do ensino fundamental, considerando apenas as probabilidades de aprovação por alinhamento idade série

Em (2), FI_{2+} é o número de alunos de $G(0)$ que se encontra na escola com dois ou mais anos de atraso ou que já evadiu. Esses alunos não têm mais condições de terminar o ensino médio com, no máximo, um ano de atraso e, assim, não terão mais o fluxo escolar acompanhado.

O Censo Escolar fornece uma medida para FI_i , de modo que a probabilidade de transição p_{-1}^{AFI} [a primeira equação de (2)] pode ser diretamente estimada: $p_{-1}^{AFI} = \left(\frac{FI_{-1}}{E_{-1}}\right)^{\frac{1}{5}}$. Dado esse cálculo, obtém-se uma estimativa para de p_0^{AFI} , e assim por diante.

De acordo com os pressupostos do modelo, os estudantes que concluem o EFI podem ser considerados classificados em três níveis de proficiência: abaixo do básico, básico e proficiente. Para cada disciplina avaliada, a escala de proficiência tem duas notas de cortes (alta e baixa). Os estudantes proficientes são aqueles com pontuação igual ou superior à *nota de corte alta* nas duas disciplinas consideradas (matemática e leitura), enquanto os alunos com nível básico são aqueles que não atingem a *nota de corte alta* em pelo menos uma disciplina e que atingem a *nota de corte baixa* nas duas disciplinas.

Os alunos considerados abaixo do básico não atingem a *nota de corte baixa* em pelo menos uma das disciplinas. Considera-se FI_i^P , FI_i^B e FI_i^{AB} como o número de alunos do grupo i que concluem a etapa nos níveis *proficiente*, *básico* e *abaixo do básico*, respectivamente. Por fim, define-se como p_i^{PFI} a probabilidade de finalizar a etapa no nível proficiente, e p_i^{BFI} a probabilidade de finalizar a etapa no nível básico, para os alunos do grupo i , e obtém-se (3).

$$FI_i^P = FI_i p_i^{PFI} \quad (3)$$

$$FI_i^B = FI_i p_i^{BFI}$$

$$FI_i^{AB} = FI_i (1 - p_i^{PFI} - p_i^{BFI})$$

Para estimar p_i^{PFI} e p_i^{BFI} , faz-se uso do Saeb, no qual é possível identificar os alunos da geração de interesse em cada nível de proficiência - FI_i^P – nível adequado, p_i^{BFI} – nível

básico e FI_i^{AB} – abaixo do básico. Com base nesses quantitativos é possível estimar, então, p_i^{PFI} e p_i^{BFI} .⁴

A principal contribuição deste modelo de fluxo escolar é a incorporação dos níveis de aprendizagem afetando as probabilidades de transição. A Figura 3 ilustra essas transições possíveis e sua comparação com a Figura 2 reflete a principal diferença para o modelo de fluxo tradicional.

⁴ Aqui será preciso enfrentar duas questões. A primeira é que a prova é feita ao final do ano e alguns estudantes ainda serão reprovados. A segunda é que os atrasados e adiantados não serão observados fazendo a prova do SAEB.

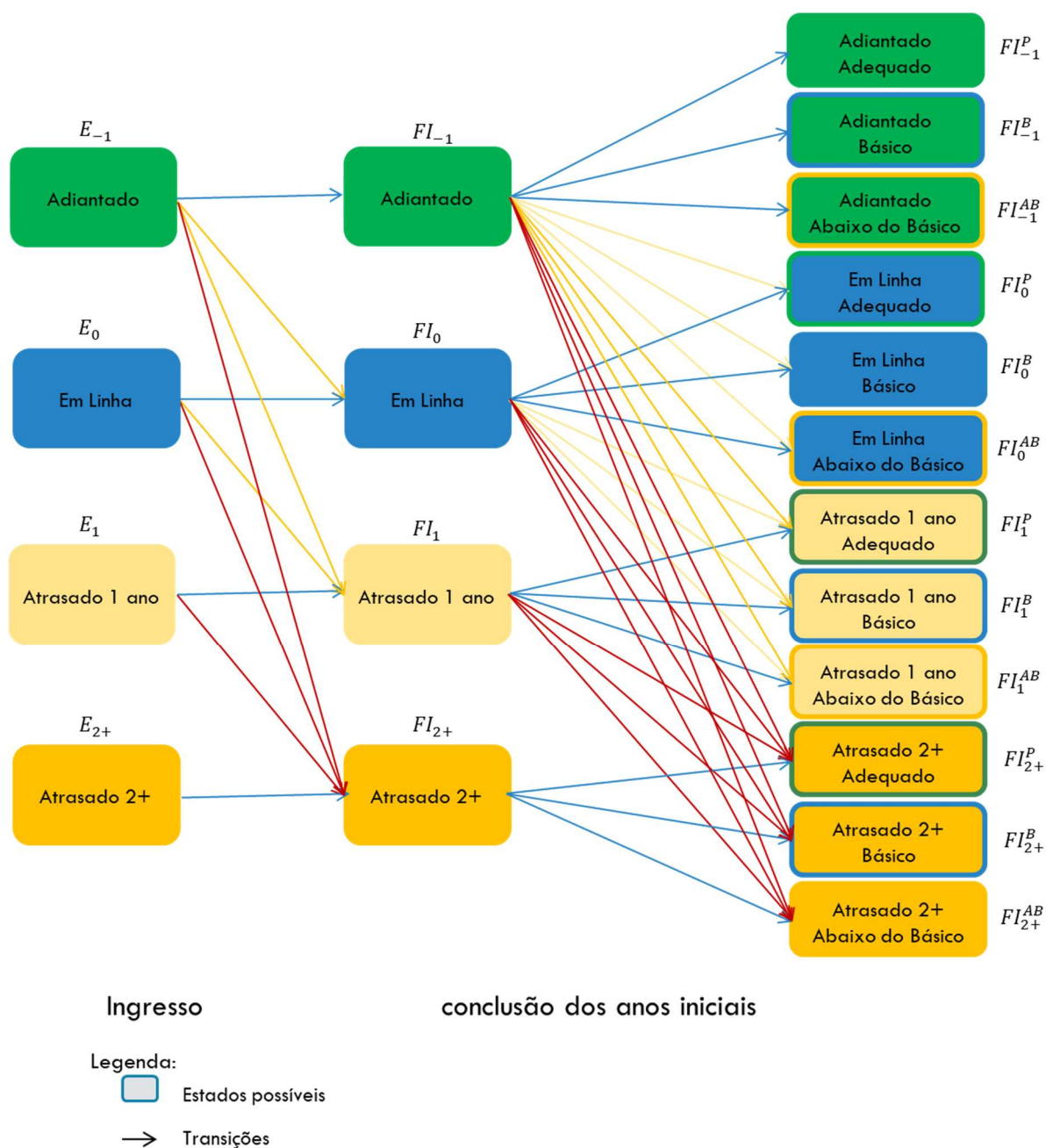


Figura 3. Ilustração das transições possíveis de para uma geração entre a etapa de ingresso e o final dos anos iniciais do ensino fundamental, considerando as probabilidades de aprovação por alinhamento idade série e de conclusão por nível de aprendizagem

Ensino Fundamental II

De forma análoga ao realizado para a primeira etapa de ensino fundamental, define-se FII_i como o número de alunos de $G(0)$ que completam a segunda etapa do ensino fundamental (EFII) com i anos de atraso. Neste caso, admite-se que os estudantes que terminam o EFI no nível adequado ou no nível básico de proficiência (FI_i^P e FI_i^B) não sofrem repetência no EFII. Define-se também p_i^{AFII} como a probabilidade de aprovação de uma série do EFII (que é invariante em relação à série) para os alunos que terminaram o EFI com i anos de atraso e estavam no nível de proficiência abaixo do básico. Então, obtém-se (4):

$$FII_{-1} = FI_{-1}^P + FI_{-1}^B + FI_{-1}^{AB}(p_{-1}^{AFII})^4 \quad (4)$$

$$FII_0 = FI_0^P + FI_0^B + FI_{-1}^{AB}4(p_{-1}^{AFII})^4(1 - p_{-1}^{AFII}) + FI_0^{AB}(p_0^{AFII})^4$$

$$FII_1 = FI_1^P + FI_1^B + FI_{-1}^{AB}10(p_{-1}^{AFII})^4(1 - p_{-1}^{AFII})^2 + FI_0^{AB}4(p_0^{AFII})^4(1 - p_0^{AFII}) + FI_1^{AB}(p_1^{AFII})^4$$

$$FII_{2+} = G(0) - FII_{-1} - FII_0 - FII_1$$

De modo similar ao EFI, considera-se FII_i^P , FII_i^B e FII_i^{AB} como o número de alunos do grupo i que concluem o EFII com nível adequado, básico e abaixo do básico, respectivamente. Define-se também que $p_i^{FII^P}$ é a probabilidade de terminar o EFII proficiente, dado que o estudante terminou o EFI no nível básico; e $p_i^{FII^{B+}}$ é a probabilidade de terminar o EFII no nível básico, dado que o estudante terminou o EFI no nível adequado.

Para simplificar o modelo e permitir a identificação, assumem-se as seguintes relações entre essas probabilidades: (i) se $p_i^{FII^P} > 0$, então $p_i^{FII^{B+}} = 0$; e (ii) se $p_i^{FII^P} = 0$, então $p_i^{FII^{B+}} > 0$. Deste modo, FII_i^P é dado por (5).

Similarmente, fica definido que p_i^{FIIAB} é a probabilidade de terminar o EFII no nível abaixo do básico, dado que o estudante terminou o EFI no nível básico; e $p_i^{FII B^-}$ é a probabilidade de terminar o EFII no nível básico, dado que o estudante terminou o EFI abaixo do básico. Assume-se também as seguintes relações entre essas probabilidades: (i) se $p_i^{FII} > 0$, então $p_i^{FIIAB} = 0$; e (ii) se $p_i^{FII} = 0$, então $p_i^{FIIAB} > 0$. Assim, FII_i^{AB} é dado por (6). Evidentemente, temos $FII_i^B = FII_i - FII_i^P - FII_i^{AB}$.

$$FII_i^P = \begin{cases} FII_i^P (1 - p_i^{FII B^-}) \\ \text{ou} \\ FII_i^P + FII_i^B p_i^{FII P} \end{cases} \quad (5)$$

$$FII_i^{AB} = \begin{cases} (FII_i - FII_i^P - FII_i^B)(1 - p_i^{FII B^-}) \\ \text{ou} \\ (FII_i - FII_i^P - FII_i^B) + FII_i^B p_i^{FIIAB} \end{cases} \quad (6)$$

Da mesma forma que no EFI, as probabilidades de aprovação (p_i^{AFII}) são obtidas com base nos dados do Censo Escolar, diretamente de (4) (e.g. $p_{-1}^{AFII} = \left(\frac{FII_{-1} - FII_{-1}^P}{FII_{-1}^N} \right)^{\frac{1}{4}}$). Os dados do SAEB nos permitem identificar FII_i^P , FII_i^B e FII_i^{AB} .

Em (5) consideram-se dois casos: (i) se $FII_i^P > FII_1^P$, então $p_i^{FII P} > 0$; e (ii) se $FII_i^P < FII_1^P$, então $p_i^{FII B^+} > 0$. Em (6) também são considerados dois casos: (i) se $FII_i^{AB} > FII_1^{AB}$, então $p_i^{FIIAB} > 0$; e (ii) se $FII_i^{AB} < FII_1^{AB}$, então $p_i^{FII B^-} > 0$. Assim, todas as probabilidades de (4), (5) e (6) são perfeitamente identificáveis.

A Figura 4 ilustra as transições possível entre a conclusão nos anos iniciais e a conclusão dos anos finais do ensino fundamental. A ilustração é similar para as transições possíveis entre o final do ensino fundamental e a conclusão do ensino médio.

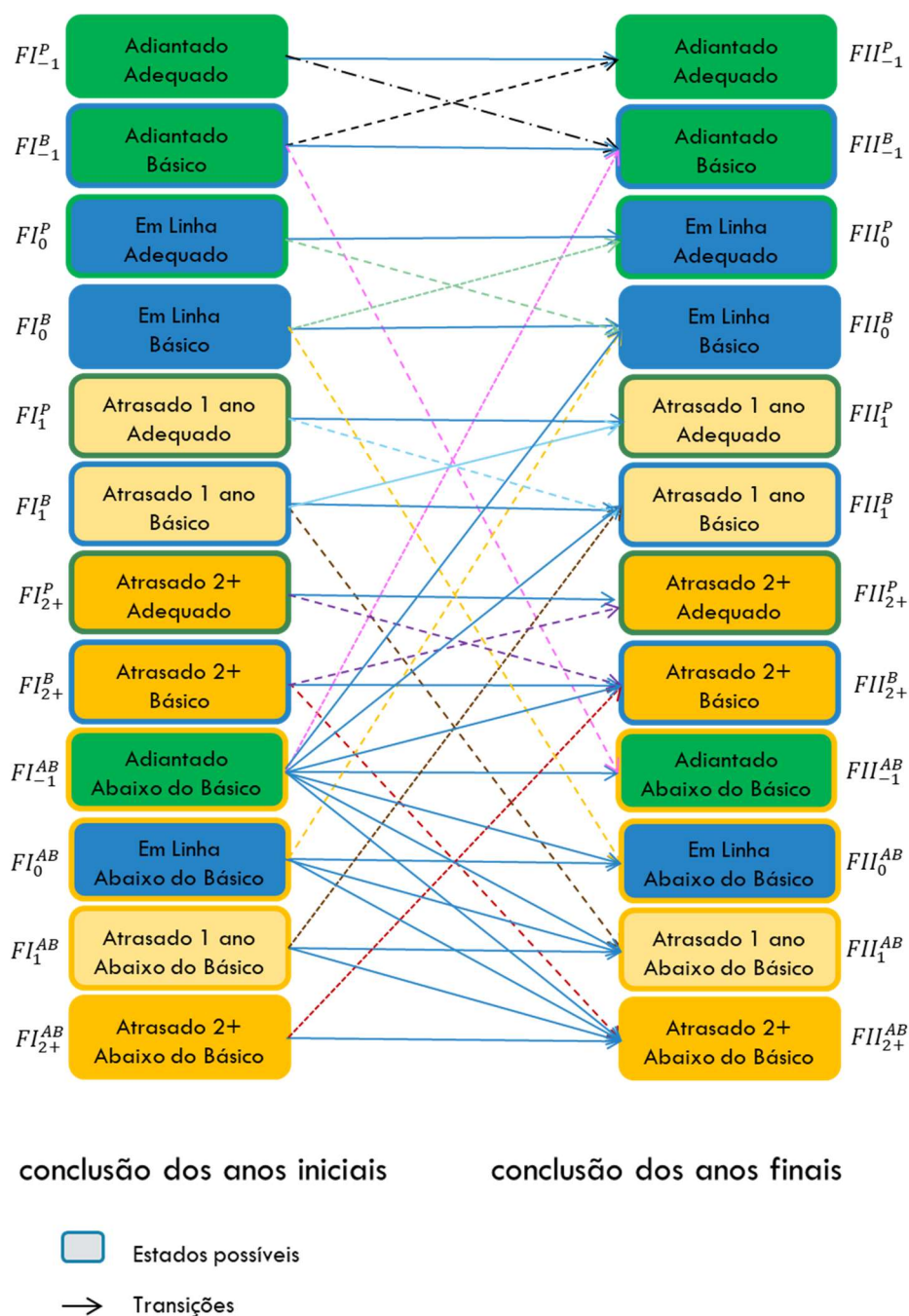


Figura 4. Ilustração das transições possíveis de para uma geração entre a etapa do final dos anos iniciais e a conclusão dos anos finais do ensino fundamental, considerando as probabilidades de aprovação por alinhamento idade série e de conclusão por nível de aprendizagem

Destaca-se na ilustração que os estados em que o nível de aprendizagem é abaixo do básico foram agrupadas ao final, para ressaltar a diferença dos demais. Entre estes estados, o número de transições possíveis é maior que para os demais, pois para eles está prevista a possibilidade de reprovação, além da mudança de um nível de proficiência para cima ou para baixo, que é a única possível para os demais estados. As cores das setas pontilhadas indicam os pares de transições em que apenas uma dela é possível de ser observada, conforme assumido pelo modelo. Por exemplo, entre a transição de alunos do Adiantado-Adequado nos anos finais para Adiantado-Básico no ensino médio e a de Adiantado-Básico nos anos finais para Adiantado-Adequado no ensino médio, apenas uma probabilidade será possível. Na aplicação aos dados, a probabilidade observada será aquela que gerar uma variação positiva entre o estado no final do fundamental e o correspondente do final do ensino médio.

Ensino Médio

Definindo EM_i como o número de alunos de $G(0)$ que completam o ensino médio com i anos de atraso e p_i^{AM} como a probabilidade de aprovação de uma série do ensino médio (que é invariante em relação à série) para os alunos que terminaram o EFII com i anos de atraso e com nível de proficiência abaixo do básico (não há reprovação para alunos com nível proficiente e básico), a exemplo do EFII, obtém-se (7).

$$EM_{-1} = FII_{-1}^P + FII_{-1}^B + FII_{-1}^{AB}(p_{-1}^{AM})^3 \quad (7)$$

$$EM_0 = FII_0^P + FII_0^B + FII_{-1}^{AB}3(p_{-1}^{AM})^3(1 - p_{-1}^{AM}) + FII_0^{AB}(p_0^{AM})^3$$

$$EM_1 = FII_1^P + FII_1^B + FII_{-1}^{AB}6(p_{-1}^{AM})^3(1 - p_{-1}^{AM})^2 + FII_0^{AB}3(p_0^{AM})^3(1 - p_0^{AM}) + FII_1^{AB}(p_1^{AM})^3$$

$$EM_{2+} = G(0) - EM_{-1} - EM_0 - EM_1$$

Da mesma forma que no EFII, considera-se EM_i^P , EM_i^B e EM_i^{AB} como o número de alunos do grupo i que concluem o ensino médio no nível adequado, básico e abaixo do básico, respectivamente. Pelo mesmo raciocínio do caso anterior, obtêm-se (8) e (9), onde p_i^{MP} é a probabilidade de terminar o ensino fundamental proficiente, dado que o estudante terminou o EFII no nível básico; p_i^{MB} é a probabilidade de terminar o ensino médio no nível básico, dado que o estudante terminou o EFII proficiente; p_i^{MAB} é a probabilidade de terminar o ensino médio no nível abaixo do básico, dado que terminou o EFII no nível básico; e p_i^{MB-} é a probabilidade de terminar o ensino médio no nível básico, dado que terminou o EFII no nível abaixo do básico.

$$EM_i^P = \begin{cases} FII_1^P (1 - p_i^{MB+}) \\ ou \\ FII_1^P + FII_i^B p_i^{MP} \end{cases} \quad (8)$$

$$EM_i^{AB} = \begin{cases} (EM_i - FII_i^P - FII_i^B)(1 - p_i^{MB-}) \\ ou \\ (EM_i - FII_i^P - FII_i^B) + FII_i^B p_i^{MAB} \end{cases} \quad (9)$$

As estimações de p_i^{MP} , p_i^{MB+} , p_i^{MAB} e p_i^{MB-} são similares ao caso do EFII. Assim, todas as probabilidades de transição do modelo de fluxo são perfeitamente identificáveis. A proporção de estudantes que termina o ensino médio proficiente com, no máximo, um ano de atraso e dado por $P = \frac{EM_{-1}^P + EM_0^P + EM_1^P}{G(0)}$.

A ESTRUTURA LÓGICA DO SIMULADOR

O Simulador do IIE é um modelo numérico que operacionaliza o Modelo de Fluxo&Aprendizagem, tomando as probabilidades de transição como variáveis exógenas e os estados educacionais como variáveis endógenas. Assim, dadas as probabilidades de transição, o simulador produz todos os estados educacionais do modelo. As simulações consistem em mudar as probabilidades de transição (exógenas) e verificar o que ocorre com os estados educacionais (endógenos).

Para viabilizar as funções do simulador é preciso ter um ponto de partida de referência para as probabilidades e as proporções de jovens por estado educacional, como aquelas experimentadas pela geração que concluiu o ensino médio mais recentemente ou para a última geração que concluiu cada etapa de ensino mais recentemente. Abaixo seguem as equações que formam a estrutura lógica do simulador, implementadas no SIIE-Natura, a fim de obter esse ponto de partida. Essas equações são, na verdade, uma forma diferente de escrever o Modelo de Fluxo&Aprendizagem apresentado na seção anterior.

Etapa 1 – Ingresso

Nesta etapa são conhecidos o tamanho da geração ($G = 100$) e as probabilidades de ingresso (p_{-1}^I , p_0^I e p_1^I) e determina-se a proporção dos que entram adiantado, em linha e com 1 ano de atraso (E_{-1} , E_0 , E_1).

$$Gp_{-1}^I = E_{-1}$$

$$Gp_0^I = E_0$$

$$Gp_1^I = E_1$$

Etapa 2 – Ensino Fundamental 1

Nesta etapa são conhecidas a proporção dos que entraram adiantados, em linha e com 1 ano de atraso (E_{-1}, E_0 e E_1), as probabilidades de aprovação e as probabilidades de finalizar os anos iniciais no nível adequado, no nível básico e abaixo do básico e determina-se a proporção dos alunos da geração que finalizam os anos iniciais em um dos 9 estados de interesse: adiantado e nível adequado, em linha e nível adequado, com 1 ano de atraso e nível adequado, adiantado e nível básico, em linha e nível básico, com 1 ano de atraso e nível básico, adiantado e nível abaixo do básico, em linha e nível abaixo do básico, e com 1 ano de atraso e nível abaixo do básico.

$$E_{-1}(p_{-1}^{AFI})^5 p_{-1}^{PFI} = FI_{-1}^P$$

$$E_{-1}(p_{-1}^{AFI})^5 p_{-1}^{BFI} = FI_{-1}^B$$

$$E_{-1}(p_{-1}^{AFI})^5 (1 - p_{-1}^{PFI} - p_{-1}^{BFI}) = FI_{-1}^{AB}$$

$$[E_{-1}5(p_{-1}^{AFI})^5 (1 - p_{-1}^{AFI}) + E_0(p_0^{AFI})^5] p_0^{PFI} = FI_0^P$$

$$[E_{-1}5(p_{-1}^{AFI})^5 (1 - p_{-1}^{AFI}) + E_0(p_0^{AFI})^5] p_0^{BFI} = FI_0^B$$

$$[E_{-1}5(p_{-1}^{AFI})^5 (1 - p_{-1}^{AFI}) + E_0(p_0^{AFI})^5] (1 - p_0^{PFI} - p_0^{BFI}) = FI_0^{AB}$$

$$[E_{-1}15(p_{-1}^{AFI})^5 (1 - p_{-1}^{AFI})^2 + 5E_0(p_0^{AFI})^5 (1 - p_0^{AFI}) + E_1(p_1^{AFI})^5] p_1^{PFI} = FI_1^P$$

$$[E_{-1}15(p_{-1}^{AFI})^5 (1 - p_{-1}^{AFI})^2 + 5E_0(p_0^{AFI})^5 (1 - p_0^{AFI}) + E_1(p_1^{AFI})^5] p_1^{BFI} = FI_1^B$$

$$[E_{-1}15(p_{-1}^{AFI})^5 (1 - p_{-1}^{AFI})^2 + 5E_0(p_0^{AFI})^5 (1 - p_0^{AFI}) + E_1(p_1^{AFI})^5] (1 - p_1^{PFI} - p_1^{BFI}) = FI_1^{AB}$$

Etapa 3 – Ensino Fundamental 2

Nesta etapa são conhecidas a proporção dos alunos da geração que finalizam os anos iniciais em um dos 9 estados de interesse ($FI_{-1}^P, FI_{-1}^B, FI_{-1}^{AB}, FI_0^P, FI_0^B, FI_0^{AB}, FI_1^P, FI_1^B$ e FI_1^{AB}), as probabilidades de aprovação e as probabilidades de transição entre os níveis de proficiência (proficiente, básico e abaixo do básico) e determina-se a

proporção dos alunos da geração que finalizam os anos finais em um dos 9 estados de interesse.

$$FI_{-1}^P(1 - p_{-1}^{FII B+}) + FI_{-1}^B p_{-1}^{FII P} = FII_{-1}^P$$

$$FI_0^P(1 - p_0^{FII B}) + FI_0^B p_0^{FII P} = FII_0^P$$

$$FI_1^P(1 - p_1^{FII B+}) + FI_1^B p_1^{FII P} = FII_1^P$$

$$FI_{-1}^{AB}(p_{-1}^{AFII})^4(1 - p_{-1}^{FII B-}) + FI_{-1}^B p_{-1}^{FII AB} = FII_{-1}^{AB}$$

$$[FI_{-1}^{AB} 4(p_{-1}^{AFII})^4(1 - p_{-1}^{AFII}) + FI_0^{AB}(p_0^{AFII})^4](1 - p_0^{FII B-}) + FI_0^B p_0^{FII AB} = FII_0^{AB}$$

$$[FI_{-1}^{AB} 10(p_{-1}^{AFII})^4(1 - p_{-1}^{AFII})^2 + FI_0^{AB} 4(p_0^{AFII})^4(1 - p_0^{AFII}) + FI_1^{AB}(p_1^{AFII})^4](1 - p_1^{FII B-}) + FI_1^B p_1^{FII AB} = FII_1^{AB}$$

$$[FI_{-1}^P + FI_{-1}^B + FI_{-1}^{AB}(p_{-1}^{AFII})^4] - FII_{-1}^P - FII_{-1}^{AB} = FII_{-1}^B$$

$$[FI_0^P + FI_0^B + FI_{-1}^{AB} 4(p_{-1}^{AFII})^4(1 - p_{-1}^{AFII}) + FI_0^{AB}(p_0^{AFII})^4] - FII_0^P - FII_0^{AB} = FII_0^B$$

$$[FI_1^P + FI_1^B + FI_{-1}^{AB} 10(p_{-1}^{AFII})^4(1 - p_{-1}^{AFII})^2 + FI_0^{AB} 4(p_0^{AFII})^4(1 - p_0^{AFII}) + FI_1^{AB}(p_1^{AFII})^4] - FII_1^P - FII_1^{AB} = FII_1^B$$

Note que $p_i^{FII B}$ e $p_i^{FII P}$ entram na definição de FII_i^P , mas temos que (i) se $p_i^{FII P} > 0$, então $p_i^{FII B} = 0$; e (ii) se $p_i^{FII P} = 0$, então $p_i^{FII B+} > 0$. O mesmo vale para $p_i^{FII B-}$ e $p_i^{FII AB}$ na definição de FII_i^{AB} .

Etapla 4 – Ensino Médio

Nesta etapa são conhecidas a proporção dos alunos da geração que finalizam os anos iniciais em um dos 9 estados de interesse ($FII_{-1}^P, FII_{-1}^B, FII_{-1}^{AB}, FII_0^P, FII_0^B, FII_0^{AB}, FII_1^P, FII_1^B$ e FII_1^{AB}), as probabilidades de aprovação e as probabilidades de transição entre os níveis de proficiência (proficiente, básico e abaixo do básico) e determina-se a

proporção dos alunos da geração que finalizam os anos finais em um dos 9 estados de interesse.

$$FII_{-1}^P(1 - p_{-1}^{MB+}) + FII_{-1}^B p_{-1}^{MP} = EM_{-1}^P$$

$$FII_0^P(1 - p_0^{MB}) + FII_0^B p_0^{MP} = EM_0^P$$

$$FII_1^P(1 - p_1^{MB+}) + FII_1^B p_1^{MP} = EM_1^P$$

$$FII_{-1}^{AB}(p_{-1}^{AM})^3(1 - p_{-1}^{MB}) + FII_{-1}^B p_{-1}^{MAB} = EM_{-1}^{AB}$$

$$[FII_{-1}^{AB} 3(p_{-1}^{AM})^3(1 - p_{-1}^{AM}) + FII_0^{AB}(p_0^{AM})^3](1 - p_0^{MB}) + FII_0^B p_0^{MAB} = EM_0^{AB}$$

$$[FII_{-1}^{AB} 6(p_{-1}^{AM})^3(1 - p_{-1}^{AM})^2 + FII_0^{AB} 3(p_0^{AM})^3(1 - p_0^{AM}) + FII_1^{AB}(p_1^{AM})^3](1 - p_1^{MB}) + FII_1^B p_1^{MAB} = EM_1^{AB}$$

$$[FII_{-1}^P + FII_{-1}^B + FII_{-1}^{AB}(p_{-1}^{AM})^3] - EM_{-1}^P - EM_{-1}^{AB} = EM_{-1}^B$$

$$[FII_0^P + FII_0^B + FII_{-1}^{AB} 3(p_{-1}^{AM})^3(1 - p_{-1}^{AM}) + FII_0^{AB}(p_0^{AM})^3] - EM_0^P - EM_0^{AB} = EM_0^B$$

$$[FII_1^P + FII_1^B + FII_{-1}^{AB} 6(p_{-1}^{AM})^3(1 - p_{-1}^{AM})^2 + FII_0^{AB} 3(p_0^{AM})^3(1 - p_0^{AM}) + FII_1^{AB}(p_1^{AM})^3] - EM_1^P - EM_1^{AB} = EM_1^B$$

CÁLCULO DAS PROBABILIDADES DE TRANSIÇÃO

Como visto, as probabilidades de transição são consideradas exógenas no simulador. Investiu-se, portanto, em obter probabilidades de transições que sejam as bases para as simulações, com aderência com a realidade. O procedimento adotado foi calcular as probabilidades de transição de uma determinada etapa tendo como base o Modelo de Fluxo&Aprendizagem, considerando os estados educacionais observados para a última coorte de nascimento, para a qual todos os dados necessários estavam disponíveis, em cada etapa de ensino.

Por exemplo, para calcular as probabilidades de transição do ensino médio, considerou-se a geração que deveria terminar o ensino médio no ano de 2019 se todos os alunos estivessem ‘em linha’, ou seja, nem adiantados e nem atrasados, e por esse motivo esse é o ano referência para a geração nascida no ano de 2002.⁵ Para essa geração, é preciso obter dados da distribuição nos estados educacionais (conforme Quadro 4) em duas etapas: na conclusão do ensino médio e na conclusão dos anos finais. Conforme o exemplo ilustrativo do Quadro 4, 5% da geração terminou a etapa (ensino médio ou anos finais) adiantada e no nível básico de proficiência; 4% terminaram com um ano de atraso e no nível adequado de proficiência.

Quadro 4. Distribuição da Geração nos Estados Educacionais (ensino médio ou anos finais)

	Abaixo do Básico	Básico	Adequado	Total
Adiantados	0	0,05	0,02	0,07
Em Linha	0,1	0,2	0,04	0,34
Atrasados 1	0,2	0,15	0,04	0,39
Atrasados 2	0,2	0	0	0,2
Total	0,5	0,4	0,1	1

Nota: os números no Quadro são hipotéticos, não refletem os dados observados para qualquer geração.

Assim, de posse dos dados do Quadro 4 para a geração de 2002, e com base nas equações definidas no modelo, é possível obter todas as probabilidades de transição referentes à etapa do ensino médio (o modelo é perfeitamente identificável). Os dados referentes ao ensino médio fornecem os estados educacionais finais da etapa, enquanto os dados referentes aos anos finais fornecem os estados educacionais de linha de base. Note que a geração de 2002, que terminaria o ensino médio em linha em 2019, terminaria os anos finais em linha no ano de 2016.

⁵ A geração de 2002, que concluiria o ensino médio em 2019 se estivesse em linha, é a última para qual todos os dados estavam disponíveis no nível de aluno, durante o desenvolvimento deste projeto.

Procedimento semelhante é feito para se obter as probabilidades de transição das demais etapas. Para as probabilidades de transição dos anos finais são necessários os dados referentes aos anos finais e anos iniciais para a geração que concluiria os anos finais em 2019 se estivessem em linha. Trata-se, então, da geração nascida ao longo do ano de 2005 e, portanto, concluiria os anos iniciais em linha em 2015.

Seguindo raciocínio semelhante, para as probabilidades de transição dos anos iniciais são necessários os dados para a geração nascida em 2009, referentes aos anos iniciais do ano de 2019, além das probabilidades de ingresso no ano de 2015. Por fim, as probabilidades de ingresso para o cenário de base, utilizou-se os dados de da geração que deveria ingressar no sistema em 2019, correspondendo à geração de 2013.

Como tratado na seção sobre o cálculo do IIE, os dados do Quadro 4 não são diretamente observáveis: não há fonte de dados que permita extrair diretamente todos os dados desse quadro. Assim, é necessário adotar alguns procedimentos que permitem obtê-los, ainda que de forma aproximada e utilizando mais de uma fonte. Os dados do Quadro 4 podem ser obtidos pela combinação dos dados dos Quadros 5 e 6, abaixo.

Quadro 5. Distribuição da Geração por Nível de Proficiência ao Final de Determinada Etapa (anos iniciais, anos finais ou ensino médio)

	Abaixo do Básico	Básico	Adequado	Total
Adiantados	X1	X2	X3	1
Em Linha	Z1	Z2	Z3	1
Atrasados 1	V1	V2	V3	1

Quadro 6. Distribuição da Geração Segundo a Adequação do Período do Término da Etapa (anos iniciais, anos finais ou ensino médio)

Adiantados	G1
Em Linha	G2
Atrasados 1	G3
Atrasados 2	G4
TOTAL	1

O Quadro 5 é construído para somar um em cada linha (uma geração completa em cada linha, distribuída entre os estados de nível de aprendizagem), enquanto o Quadro 6 é feito para somar 1 na coluna (uma geração completa em cada coluna, distribuída entre os estados de alinhamento idade-série). Como o IIE só considera alunos com até um ano de atraso, não precisamos obter dados para aqueles com 2 ou mais anos de atraso. A proporção de uma geração que termina uma determinada etapa adiantada e abaixo do básico é obtida pela multiplicação de G1 (Quadro 6) por X1 (Quadro 5), conforme o Quadro 7. O Quadro 7 é exatamente o Quadro 4, só que ao invés de números contém o cálculo necessário para chegar nos dados necessários, a partir dos Quadros 5 e 6.

Quadro 7. Distribuição da Geração nos Estados Educacionais de Cada Etapa

	Abaixo do Básico	Básico	Adequado	Total
Adiantados	G1X1	G1X2	G1X3	G1
Em Linha	G2Z1	G2Z2	G2Z3	G2
Atrasados 1	G3V1	G3V2	G3V3	G3
Atrasados 2	-	-	-	G4
Total	-	-	-	1

Note que é possível obter diretamente o IIE do Quadro 7 (tal como apresentado no Quadro 3 - seção do IIE). Considerando o nível adequado, o IIE é dado por $G1X3 + G2Z3 + G3V3$, enquanto para o nível básico o IIE é dado por $G1X3 + G2Z3 + G3V3 + G1X2 + G2Z2 + G3V2$. Então, para a estimação das probabilidades de transição e para o cálculo do IIE é necessário obter os dados contidos nos Quadros 5 e 6.

Para preencher o Quadro 5 foi utilizado o SAEB. Para o cálculo das probabilidades do ensino médio da geração nascida em 2002, seria necessário dispor dos SAEBs de 2019 e 2016 para observar os resultados dos alunos em linha que,

respectivamente, seriam obtidos os dados para o final do ensino médio e para o final do ensino fundamental. Porém o SAEB só é realizado em anos ímpares. Então, a distribuição da geração por nível de proficiência ao final dos anos finais foi aproximada com base no SAEB 2015. Ou seja, utilizou-se a distribuição da geração um ano mais velha, que realizou o SAEB um ano antes, como aproximação da distribuição da geração de interesse. Esta aproximação não é necessária para as demais etapas. Para os anos finais os estados educacionais ao início da etapa requerem o SAEB de 4 anos antes, que está disponível. Para os anos iniciais, os estados iniciais são as condições de ingresso e, portanto, não requerem o SAEB para serem calculados.

Mesmo quando se dispõe do SAEB para os anos requeridos, os dados do Quadro 5 não são observados diretamente. Isto porque nem todos os alunos da geração cursam a última série da etapa no mesmo ano (em linha). Como descrito na seção sobre o cálculo do IIE, utiliza-se as *gerações vizinhas* para aproximar os dados das linhas 1 e 3.

Ainda que se admita que a utilização das gerações vizinhas produza uma boa estimativa da distribuição nos níveis de proficiência para a geração de referência, restam dois problemas a lidar. O primeiro é que a realização do SAEB se dá antes que a situação de aprovação e reprovação esteja definida. Assim, alguns daqueles representados no SAEB podem não concluir a etapa, e as probabilidade que se deseja estimar são para a etapa completa. Como a reprovação deve, provavelmente, ser mais elevada para os de baixa proficiência, alguém pode alegar que o procedimento adotado subestima a proporção de pessoas da geração nos níveis mais elevados de proficiência (adequado é básico).

O segundo é que cerca de 20% dos escalados para fazer o SAEB (25% no ensino médio) não realizam o exame. Como os que faltam à prova devem, provavelmente, estar super-representados por aqueles de baixa proficiência (muitos já deixaram de

frequentar a escola porque já não possuem chances de aprovação), isso geraria uma distorção que vai no sentido oposto àquele da questão anterior.

É difícil dizer qual o efeito líquido dessas duas distorções e os dados disponíveis dificultam a realização de estudos que permitam um esclarecimento da questão. Pouco pode ser feito para lidar com esses potenciais problemas no Simulador do IIE, exceto pelo fato de que os jovens da geração de interesse, que não estavam frequentando a escola e foram identificados na PNADc, foram considerados entre os que tinham 2 anos ou mais de atraso, no nível de proficiência abaixo do básico, o que ameniza em parte o problema.

Com relação ao SAEB-2019, ainda foi necessário lidar com um outro problema. Nesta edição, o INEP não divulgou a informação de idade dos estudantes, impossibilitando a classificação dos jovens nos estados de alinhamento idade-série - adiantados, em linha e atrasados (com 1 ou mais de 1 ano de atraso). Assim, o SAEB-2019 só nos permite obter uma estimativa para a última linha do Quadro 4. Para lidar com isso, o procedimento adotado foi, em primeiro lugar, distribuir o total de cada coluna do Quadro 4, em 2019, de acordo com a distribuição obtida para 2017 e, em segundo lugar, fazer com que cada linha do Quadro some 1, mas mantendo as proporções obtidas pela primeira transformação. O procedimento e os exemplos estão reunidos no Apêndice deste relatório técnico.

Se todos os indivíduos da geração frequentassem continuamente a escola até a conclusão do ensino médio, o Censo Escolar bastaria para preenchemos o Quadro 6. No entanto, uma parcela dos membros da geração de interesse nunca ingressa na escola ou evade antes de concluir o ensino médio. Para corrigir o tamanho da geração, incluindo aqueles fora da escola sem ter concluído cada etapa, a PNAD é utilizada.

Para preencher o Quadro 6 foram utilizados o Censo Escolar e a PNAD. O Censo Escolar permite observar os alunos no 1º ano do ensino fundamental e na última série de cada etapa e, no acompanhamento de rendimento escolar, verifica-se o desfecho em relação as situações de aprovação, reprovação ou abandono.

Com base em dois censos consecutivos - do ano em que a parte em linha da geração conclui a última série das duas etapas do ensino fundamental, e do ano imediatamente posterior, em que é possível localizar os que concluem cada etapa com 1 atraso, bem como os que concluirão com 2 anos ou mais, incluindo os que não concluirão. Os que concluem os anos iniciais adiantados são observados no mesmo censo em que se observam os alunos em linha, neste caso eles são observados nos anos finais. De forma similar os adiantados dos anos finais, serão observados matriculados no ensino médio. O caso do ensino médio é similar para quantificar os jovens em linha e os atrasados 1 ano e atrasados 2 ou mais anos. Já os adiantados do ensino médio precisam ser observados como concluintes da etapa no Censo Escolar imediatamente anterior àquele em que se observa os concluintes em linha.

Para finalizar, vale destacar que os estados educacionais resultantes das probabilidades obtidas por esse procedimento não representam os estados observados para nenhuma geração específica: simula uma geração que seria observada caso frequentasse a educação básica com as mesmas probabilidades de ingresso da geração que deveria ingressar em 2019, as probabilidades de transição nos anos iniciais da geração que deveria terminar os anos iniciais em 2019, assim como para os anos finais e ensino médio. Essa geração “híbrida” forma a linha de base para se realizar as simulações.

AS FUNÇÕES DE LIGAÇÃO

Note que, de posse das probabilidades de transição da linha de base, o SIIE estaria concluído. As simulações consistiriam em mudar as probabilidades de transição da linha de base e verificar os resultados nos estados educacionais e, consequentemente, no IIE.

A questão é que as probabilidades de transição do modelo não são usualmente empregadas nas análises educacionais. Por exemplo, nas avaliações de impacto de políticas educacionais, os indicadores de impacto usuais são as medidas de rendimento escolar (aprovação, reprovação e abandono), notas em exames padronizados (e.g. SAEB, ENEM e avaliações estaduais) e indicadores que combinam tais dimensões, a exemplo do IDEB. Então, para que o SIIE seja operacional, é necessário encontrar uma forma de transformar os impactos sobre tais indicadores, em variações nas probabilidades de transição empregadas pelo Modelo de Fluxo & Aprendizagem e pelo SIIE.

As estratégias adotadas para implementar tal conversão foram denominadas “funções de ligação”. Mais especificamente, elaborou-se funções de ligação para: (i) as probabilidades de aprovação, (ii) o desempenho no SAEB (matemática e língua portuguesa); (iii) o IDEB e; (iv) adicionalmente, foi estimada uma relação entre os impactos sobre a alfabetização dos alunos do 2º ano do ensino fundamental (mais especificamente sobre a distribuição dos estudantes em 3 níveis de alfabetização (Básico, Intermediário e Adequado).

Simulações de variação nas taxas de aprovação

No caso das taxas de aprovação a transformação é direta. Para um dado aumento na probabilidade de aprovação (dP) na etapa (anos iniciais, anos finais ou ensino médio), vamos admitir que as probabilidades de aprovação de todos os grupos da geração (adiantados, em linha e atrasados) tiveram um aumento similar em suas probabilidades. Por exemplo, a probabilidade de aprovação nos anos iniciais para os adiantados é dada por p_{-1}^{AFI} . Então, para um aumento na probabilidade de aprovação de dP na etapa, a nova probabilidade de aprovação dos adiantados da geração seria $p_{-1}^{AFI*} = p_{-1}^{AFI} + dP$. Ilustrativamente, se a taxa média de aprovação nos anos iniciais sobe de 90 para 95%, então adicionou-se 5 pontos percentuais às probabilidades de aprovação nos anos iniciais (adiantados, em linha e com 1 ano de atraso) da linha de base do simulador.

No caso dos anos finais e ensino médio é preciso fazer uma transformação, dado que as probabilidades de aprovação daqueles com nível básico e adequado é considerada 1 (100% aprovam). É preciso calcular o aumento de probabilidade dos abaixo do básico de modo que a probabilidade média de aprovação da geração seja igual ao aumento de probabilidade de aprovação da etapa.

Simulações de variação na aprendizagem

O recomendável para simulações de mudança de aprendizagem em uma etapa de ensino, de acordo com o modelo desenvolvido, seria que os inputs fossem dados como novas probabilidades de conclusão por nível de proficiência ($p_i^{PFI}, p_i^{BFI}, p_i^{ABFI}$, no caso dos anos iniciais, e $p_i^{PE}, p_i^{B+E}, p_i^{B-E}, p_i^{ABE}$, nos anos finais e ensino médio, em que E indica a etapa de ensino) da etapa em questão. Entretanto, esses não são indicadores familiares aos projetos e estudos, e o impacto das políticas sobre essas probabilidades não é conhecido.

O que se conhece, relativamente à aprendizagem, é o impacto sobre as proficiências médias do Saeb (ou de avaliação comparável). Assim, neste caso o input não tem uma transformação direta nas mudanças das probabilidades para mensuração do efeito sobre o IIE.

Foram necessárias, então, duas etapas adicionais para esta transformação: (i) transformar a variação média de proficiência em variação média da proporção de estudantes em cada nível de proficiência e; (ii) cálculo das novas probabilidades de conclusão por nível de proficiência, com base na nova distribuição.

As relações entre a variação na média das proficiências e a variação na distribuição dos estudantes entre os três níveis adotados foram estimadas com base nos dados observados do Saeb 2015, 2017 e 2019. Estimou-se, para cada etapa a seguinte equação:

$$E_i^{AB} = \beta_1 \Delta N_i^t$$

$$E_i^B = \beta_2 \Delta N_i^t$$

$$E_i^P = \beta_3 \Delta N_i^t$$

Em que ΔN_i^t é a variação da média das proficiências em língua portuguesa e matemática, entre os anos t e $t-1$, para a unidade da federação i , E é a proporção de estudantes em cada nível de aprendizagem (AB- abaixo do básico, B- básico e P- adequado) e os β são os coeficientes a serem estimados a fim de determinar as relações procuradas.

Estimando, por MQO, o coeficiente β que define a relação entre ΔFI_i^P e ΔN e entre ΔFI_i^B e ΔN , aplicado aos dados das variações bienais do Saeb entre 2013 e 2017, para o todo o Brasil, chega-se os resultados reportados no Quadro 4, abaixo.

Quadro 8. Coeficientes estimados para identificar a relação entre as variações da média de proficiência (N), e da proporção de alunos no nível de proficiência de interesse (abaixo do básico, básico e adequado) [†]

	Anos Iniciais	Anos Finais	Ensino Médio
β_{abaixo}	-0.0970***	-0.0880***	-0.1037***
$\beta_{básico}$	0.0820***	0.0815***	0.1002***
$\beta_{adequado}$	0.0150***	0.0065***	0.0035***

[†] Relações estimadas por MQO, aplicados a dados de variações bienais de N e das proporções no nível de interesse (abaixo do básico, básico ou adequado) entre 2013 e 2015 e entre 2015 e 2017.

*** - todos os coeficientes são estatisticamente significativos ao nível de 1%.

A mesma relação é aplicada a todos os grupos de alinhamento idade série, e obtêm-se assim os novos estados da etapa de interesse. De posse desses novos estados, faz-se o que poderia ser chamado de processo inverso, e calculam-se as novas probabilidades de conclusão por nível de proficiência, que serão simuladas para obter os potenciais novos IIEs.

Simulações de variação no IDEB

Assim como no caso das proficiências, o IDEB é um índice que também é comumente utilizado para acompanhar efeito de políticas e outras mudanças. Foi preciso adotar uma estratégia que permitisse que inputs feitos no SIIE via mudanças no IDEB fossem transformados em mudanças nas probabilidades descritas pelo modelo para que o novo IIE pudesse ser simulado.

Como o IDEB é composto pelos dois fatores – aprovação média por etapa de ensino e proficiências médias, além das etapas já descritas, nas respectivas seções acima, é necessária uma etapa que as precede, a fim de estabelecer a relação entre variações no

IDEB e variações nos fatores que o compõe. Essa relação não é direta pois diversas combinações dos dois componentes levam ao mesmo IDEB.

A estratégia adotada aqui foi utilizar a decomposição do IDEB, conforme apresentado a seguir:

$$I = NP$$

$$dI/I = dN/N + dP/P, \text{ ou}$$

$$\Delta\% I = \Delta\% P + \Delta\% N$$

$$(\Delta\% P)/(\Delta\% I) + (\Delta\% N)/(\Delta\% I) = 1$$

Com base na evolução prévia do IDEB podemos obter o comportamento médio de $(\Delta\% P)/(\Delta\% I)$ e $(\Delta\% N)/(\Delta\% I)$ e usá-los para simular dN e dP , para uma dada variação no IDEB (dI). Substituindo por dados das variações bienais de N , P e I deb entre 2013 e 2019, para o total do Brasil, obtêm-se os resultados reportados no quadro, abaixo.

Quadro 9. Quadro 1. Média das relações calculadas entre as variações bienais dos indicadores de proficiência e de aprovação, e do Ideb, segundo etapa de ensino

	Anos Iniciais	Anos Finais	Ensino Médio
$(\Delta\% P)/(\Delta\% I)$	0.19	0.40	0.426
$(\Delta\% N)/(\Delta\% I)$	0.81	0.60	0.583

(*) Relações calculadas a partir da média das variações bienais entre 2013 e 2019, para o ensino fundamental, e 2015 a 2019 para o ensino médio - não foram utilizados os dados de 2013 para o ensino médio devido à variação zero no Ideb.

Simulações de variação na alfabetização (proficiência do 2º ano)

Todo o Modelo de Fluxo&Aprendizagem foi desenvolvido por etapas de ensino, utilizando a divisão tradicional do ensino fundamental e médio. Entretanto, medidas que afetam a qualidade da educação e a inclusão educacional podem ter seus impactos medidos em indicadores que não estão diretamente relacionados a essas etapas, como indicadores da educação infantil ou da alfabetização, ou mesmo outros indicadores relativos ao ensino fundamental e médio, que não estão expressos como componentes do modelo, nem mesmo indiretamente, como qualificação docente e jornada escolar. Isso não impede que seja estimado o impacto de tais medidas sobre o IIE. Mas é preciso adicionar mais algumas etapas à estratégia metodológica.

A aprendizagem observada no 2º ano, ou o nível de alfabetização é um tema que vem sendo alvo de muitas políticas e investimento no país, sendo assim, foram estudadas formas de incorporar esse tema ao simulador. A abordagem escolhida foi a de dividir as unidades da federação em 3 grupos de desempenho escolar no 5º ano do ensino fundamental e calcular, para cada um desses grupos uma taxa de transição entre os estados de alfabetização (básico, intermediário e adequado) e os estados de proficiência do final dos anos iniciais (abaixo do básico, básico e adequado).

Utilizando dados do SPAECE, em que é possível localizar os mesmos alunos quando avaliados no 2º e no 5º ano do fundamental, um estudo realizado pela Universidade Federal do Ceará - UFC e pelo Laboratório de Estudos e Pesquisas em Economia Social - LEPES/USP, estimou essas taxas, a pedido do Instituto Natura. A partir dessas taxas de conversão, foi construído um método para possibilitar transformar os inputs de

variação da distribuição dos alunos do 2º ano por nível de alfabetização em variações na distribuição dos concluintes dos anos iniciais, segundo nível de proficiência.⁶

Para isso, assumiu-se que as taxas de conversão seriam as mesmas em qualquer alinhamento idade-série. E, a fim de minimizar o erro, apenas a variação de cada nível de proficiência na alfabetização (a diferença entre o indicador observado para 2019 e o input da simulação) segue a regra das taxas de conversão para o 5º ano.

É sempre bom ressaltar que quanto mais hipóteses, aproximações, generalizações e estimações são necessárias para se chegar nos indicadores de interesse, menor tende a ser a precisão em sua determinação. Por ordem de confiabilidade da simulação, pode-se listar as simulações possíveis conforme segue: probabilidade de ingresso, probabilidade de aprovação, nível de proficiência, IDEB e distribuição nos níveis de alfabetização.

A IMPLEMENTAÇÃO DO SIMULADOR DO IIE: SIIE -NATURA

Em linhas gerais, o Simulador do IEE-Natura consiste nos seguintes passos, realizados para cada etapa de ensino (exceto as mudanças no ingresso que afeta apenas a

⁶ A metodologia e os dados utilizados pela equipe estão descritos na Nota Técnica referente ao estudo sobre o tema, coordenado por Guilherme Irffi, Luiz Guilherme Scorzafave e Roberta Loboda Biondi Nastari, **Efeito da Alfabetização sobre os resultados nos Anos Iniciais do Ensino Fundamental: Análises dos dados longitudinais do Ceará (SPAEC-Alpha e do 5º ano EF) entre 2016 e 2019**, elaborada para o Instituto Natura.

probabilidade de ingresso, e a simulação de mudança na alfabetização que só afetam os anos iniciais do fundamental), para o Brasil, todos os estados e o Distrito Federal:

- (i) inputs de variações nas probabilidades de ingresso por alinhamento idade série, nas taxas de aprovação, nas proficiências médias (português e matemática), e no Ideb, para cada etapa de ensino, além da mudança na distribuição dos alunos do 2º ano por nível de alfabetização;
- (ii) transformação dos inputs de variação do Ideb em variação de taxa de aprovação e das proficiências médias, a partir da decomposição do índice em seus dois fatores;
- (iii) cálculo das novas taxas de aprovação – a partir da linha de base e das possíveis simulações de variação da aprovação ou no IDEB;
- (iv) cálculo dos novos quantitativos por estado de alinhamento idade série e nível de proficiência – a partir da linha de base e das simulações de variação nas proficiências médias, no IDEB e na distribuição dos estudantes do 2º ano por nível de alfabetização;
- (v) cálculo das novas probabilidades de conclusão por nível de proficiência e;
- (vi) cálculo das novas proporções de jovens por alinhamento idade série e nível, decorrentes das mudanças nas probabilidades.
- (vii) cálculo do no IIE.

O Simulador do IIE conta ainda com uma estimativa de qual seria o IDEB de cada etapa de ensino que corresponderia ao IIE estimado.⁷

⁷ Essa estimativa foi obtida a partir de dados observados para 2019 e de dados simulados pelo próprio SIIE, para o Brasil, estados e DF, simulando variações no IDEB de -2 a 2, de meio em meio ponto. A partir desses IDEBs e dos IIEs (observados e estimados), estimou-se por MQO os coeficientes da seguinte equação: $IIE_i = c + \beta_1 IIE_i + \beta_2 IIE_i^2 + \beta_3 IIE_i^3 + d_{UF}\beta_4 + \varepsilon$, em que c é a constante do modelo, d_{UF} é o vetor de variáveis dicotômicas de unidades da federação e ε é o termo aleatório.

Apêndice

O caso do Saeb 2019 e o IIE

No caso de o IIE para o ano de referência ser o de 2019, ou seja, a coorte de nascimento em questão ser a nascida em 2002, há um desafio maior, pois o Saeb 2019 não coletou, em seu questionário de contexto, a informação de ano e mês de nascimento (ou idade). Assim, é possível encontrar a distribuição dos alunos entre as 3 faixas de proficiência, mas não é possível cruzá-las com o alinhamento idade série a fim de possibilitar a construção dos componentes do índice em Adiantados, Em linha, Atrasados 1 ano e Atrasados 2 ou mais.

Para obter a proporção de alunos que atingiram o nível básico e nível adequado com até 1 ano de atraso, foi aplicada, aos dados do Saeb 2019, a mesma distribuição de alunos em cada faixa de alinhamento idade série observada em 2017, por nível de proficiência.⁸

Como exemplo do procedimento, admita que os dados do Quadro 4 são as estimativas dos estados educacionais ao final do ensino médio obtidos para aqueles que deveriam terminar o ensino médio em 2017. Para 2019 só é possível observar a última linha do Quadro. O primeiro procedimento adotado foi distribuir a parcela do total de estudantes do SAEB-2019 em cada nível de proficiência (em preto) de acordo com a distribuição das colunas obtidas pelo SAEB-2017 (em vermelho), conforme o Quadro 10. A distribuição da coluna total do Quadro 10 produz a distribuição de 2017, não a de 2019. Então, o segundo procedimento consiste em, mantendo as proporções das linhas do Quadro 10, fazer cada linha somar 1, conforme mostrado no Quadro 11. Os dados dessa último Quadro são combinados com os dados do Quadro 5 para obter

⁸ Não foi utilizada nenhuma tendência devido à mudança de metodologia do Saeb - 3a série do médio, a partir do ano 2017, primeiro ano de universalização da Saeb no ensino médio da rede pública.

a distribuição nos estados educacionais de 2019. Ou seja, o Quadro 11 dispõe dos dados referentes à Quadro 5.

Quadro 10. Distribuição do SAEB-2019 Mantendo as Proporções das Colunas de 2017

	Abaixo do Básico	Básico	Adequado	Total
Adiantados	0	0,0625	0,04	0,1025
Em Linha	0,06	0,25	0,08	0,39
Atrasados 1	0,12	0,1875	0,08	0,3875
Atrasados 2	0,12	0	0	0,12
Total	0,3	0,5	0,2	1

Quadro 11. Distribuição do SAEB-2019 Ajustada com Base no SAEB-2017

	Abaixo do Básico	Básico	Adequado	Total
Adiantados	0	0,6098	0,3902	1
Em Linha	0,15384	0,64103	0,20513	1
Atrasados 1	0,30968	0,48387	0,20645	1
Atrasados 2	1	0	0	1

A distribuição dos jovens por distorção idade-série, conforme o Censo Escolar

A estratégia adotada para suprir lacunas de dados no Saeb incorporam às análises algumas distorções relativas às informações para as quais se dispõe de dados mais fiéis à realidade no Censo Escolar. Suponha que quatro coortes de nascimento vizinhas em uma unidade da federação tenham, por motivo que não são relevantes aqui, tamanho significativamente diferentes. Ainda que a distribuição dos jovens ao final do ensino médio, observada no Censo Escolar, fosse igual para ambas, as proporções seriam observadas no Saeb como apresentado no quadro abaixo.

Quadro 12. Proporção dos jovens por alinhamento idade séria, segundo coorte de nascimento no Censo Escolar, Quantidade de jovens por alinhamento idade séria no Saeb, segundo coorte de nascimento e Proporção de jovens por alinhamento idade séria no Saeb

Ilustração da observação de um ano de Saeb com jovens de apenas 2 coortes	Distribuição dos jovens nas faixas de alinhamento idade-série na 3ª série do ensino médio				Total de jovens
	adiantados	em linha	com até 1 ano de atraso	com 2 anos ou mais de atraso	
Coorte nascida em t, em % da coorte, no CE	1	50	30	19	80
Coorte nascida em t+1, em % da coorte, no CE	1	50	30	19	90
Coorte nascida em t+2, em % da coorte, no CE	1	50	30	19	100
Coorte nascida em t+3, em % da coorte, no CE	1	50	30	19	110
Coorte nascida em t, em quantidade de jovens no Saeb do ano t+2				15,2	15,2
Coorte nascida em t+1, em quantidade de jovens no Saeb do ano t+2			27		27
Coorte nascida em t+2 em quantidade de jovens no Saeb do ano t+2		50			50
Coorte nascida em t+3 em quantidade de jovens no Saeb do ano t+2	1,1				1,1
Distribuição percentual observada no Saeb 3ª série do EM no ano t+18	1,2%	53,6%	29%	16,3%	100%