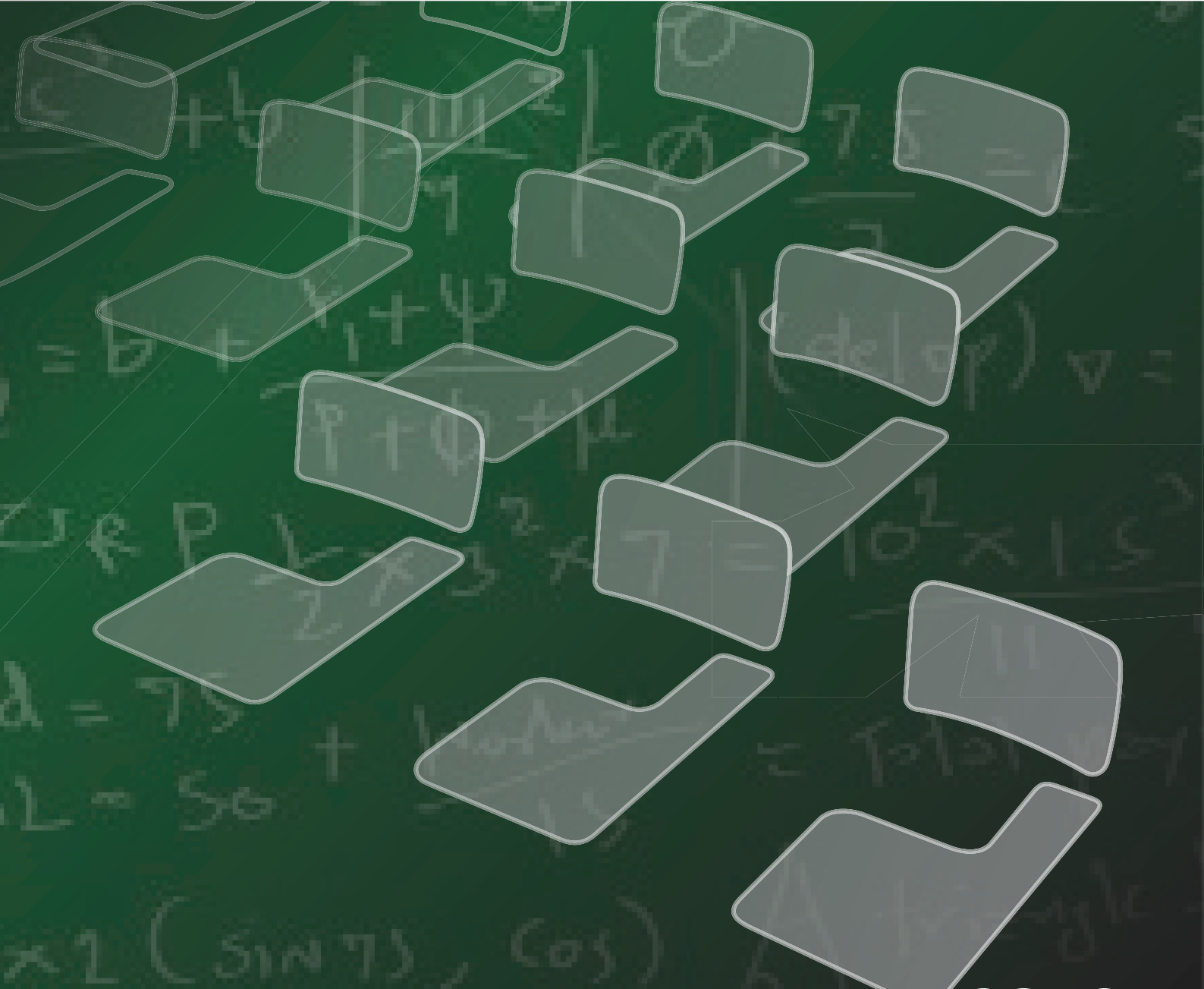




IJSN - Instituto Jones dos Santos Neves **2018**

58

TEXTO PARA DISCUSSÃO

**Modelo de predição do
Abandono Escolar**

Pesquisa Aplicada à Gestão de Educação

MODELO DE PREDIÇÃO DO ABANDONO ESCOLAR

Pesquisa Aplicada à Gestão de Educação

Instituto Jones dos Santos Neves

Modelo de predição do Abandono Escolar: Pesquisa Aplicada à Gestão da Educação.

Vitória, ES, 2018. 56 p. il. tab. (TD | 58)

1. Educação. 2. Importância da Educação para o Desenvolvimento Social e Econômico. 3. Plano Estadual de Educação. 4. Modelo Preditor de Abandono. 5. Resultados. 6. Espírito Santo (Estado)

I. Giuberti, Ana Carolina. II. Rangel, Lara Nassar. III. Castro, Magnus Willian de. *et al.* IV. Título.

As opiniões emitidas são exclusivas e de inteira responsabilidade do (os) autor (es), não exprimindo necessariamente, o ponto de vista do Instituto Jones dos Santos Neves ou da Secretária de Estado de Economia e Planejamento do governo do Estado do Espírito Santo.

GOVERNO DO ESTADO DO ESPÍRITO SANTO

Paulo César Hartung Gomes

VICE-GOVERNADORIA

César Colnago

SECRETARIA DE ESTADO DE ECONOMIA E PLANEJAMENTO - SEP

Regis Mattos Teixeira

SECRETARIA DE ESTADO DA EDUCAÇÃO - SEDU

Haroldo Corrêa Rocha

FUNDAÇÃO DE AMPARO À PESQUISA E INOVAÇÃO DO ESPÍRITO SANTO – FAPES

José Antônio Bof Buffon

SECRETARIA DE ESTADO DE ECONOMIA E PLANEJAMENTO - SEP

Regis Mattos Teixeira

INSTITUTO UNIBANCO

Marcelo Luis Orticelli

Instituto Jones dos Santos Neves

TD – 58

Diretora Presidente

Gabriela Gomes de Macêdo Lacerda

Diretora de Estudos e Pesquisas

Ana Carolina Giuberti

Coordenador da Pesquisa Aplicada à Gestão Escolar

Magnus William de Castro

Elaboração

Ana Carolina Giuberti

Lara Nassar Rangel

Magnus William de Castro

Márcio Rodrigues dos Santos

Marlen Vazzoler

Samuel Franco (Instituto Unibanco)

Taís Regina da Silva Ferreira

Thais Oliveira de Oliveira

Vânia Gomes

Revisão

Ana Carolina Giuberti

Editoração

Eugênio Herkenhoff

Assessoria de Relacionamento Institucional

Bibliotecário

Jair Rosário Filho

Assessoria de Relacionamento Institucional

Sumário

APRESENTAÇÃO.....	05
1. INTRODUÇÃO.....	07
1.1 Importância da educação para o desenvolvimento social e econômico.....	07
1.2 Abandono escolar: causas e consequências.....	09
2. CONTEXTUALIZAÇÃO.....	14
2.1 Direito à educação no Brasil e Plano Estadual de Educação.....	14
2.2 Panorama do rendimento escolar no Espírito Santo.....	17
3. MODELO PREDITOR DE ABANDONO.....	22
3.1 Justificativa para utilização da predição de abandono e aplicações na literatura internacional.....	22
3.2 Descrição do modelo de predição.....	27
3.3 Metodologia adotada na construção do modelo empírico.....	29
4. RESULTADOS.....	48
4.1 Previsão de abandono para 2015 a partir dos dados de 2014.....	48
4.2 Previsão de abandono para 2016 a partir dos dados de 2015.....	51
4.3 Previsão de abandono para 2017 a partir dos dados de 2016.....	51
5. CONSIDERAÇÕES FINAIS.....	55
REFERÊNCIAS BIBLIOGRÁFICAS.....	57

Apresentação

Compreendendo as iniciativas do poder público como um conjunto de ações que repercutem na economia e na sociedade, impactando significativamente o cotidiano do cidadão, este texto para discussão apresenta o desenho estratégico desenvolvido para auxiliar a Secretaria de Estado da Educação (SEDU) na formulação e condução de política para a redução do abandono escolar.

Este estudo faz parte do projeto “Pesquisa aplicada à Gestão da Educação”, desenvolvido pelo Instituto Jones dos Santos Neves (IJSN) e a Secretaria de Estado da Educação (SEDU), em parceria com a Fundação de Amparo à Pesquisa e Inovação do Espírito Santo (FAPES). A metodologia adotada para a construção do “Modelo de Predição do Abandono Escolar”, aqui descrita, contou com o apoio do Instituto Unibanco.

Em julho de 2016 teve início o estudo e desenvolvimento do modelo preditor do abandono escolar com o propósito de criar evidências para subsidiar a elaboração de Políticas Públicas para Educação no Estado do Espírito Santo, com vistas à reversão do alto índice de abandono escolar no ensino médio, terceira e última etapa da Educação Básica. A metodologia criada a partir de métodos estatísticos e à luz de questões sobre a Rede Pública Estadual permitirá às escolas identificar, ao final do primeiro trimestre letivo, os alunos do primeiro ano do ensino médio mais propensos ao abandono, e atuar antes que eles efetivamente abandonem a escola.

Esta identificação será o ponto de partida de um conjunto sistêmico de ações, que está sendo estruturado, para nortear as escolas no processo de detecção desses alunos com maior probabilidade de abandono. Para instrumentalizar e respaldar a equipe pedagógica para que esta possa monitorar e atuar de forma específica junto a esse aluno prevê-se ainda, a criação de uma matriz detalhada e descritiva para tomada de decisão e condução adequada a fim de atuar diante do problema identificado.

Além desta apresentação, esta nota técnica apresenta outras quatro seções. A seguir são discutidas a importância da educação para o desenvolvimento social e econômico e as causas e consequências do abandono. A seção 2 aborda as metas do Plano Estadual de Educação para o ensino médio e apresenta um breve panorama do rendimento escolar no Estado. A terceira seção apresenta a discussão internacional a respeito da predição de abandono e a construção do modelo de predição de abandono escolar para a rede estadual do Estado. A seguir são apresentados os resultados obtidos. O texto encerra-se com as considerações finais.

1. INTRODUÇÃO

1.1 Importância da educação para o desenvolvimento social e econômico

A educação é um dos determinantes fundamentais para o desenvolvimento social e econômico das nações. Devido sua importância, tem sido um tema amplamente pesquisado e discutido, bem como alvo de políticas públicas. A comunidade internacional, em assembleia da UNESCO, ratificou a importância da alfabetização para erradicar a pobreza, reduzir a mortalidade infantil, alcançar igualdade de gênero, assegurar o desenvolvimento sustentável, a paz e a democracia. Estudos demonstram que o ensino secundário traz diversos benefícios sociais, como o fortalecimento da democracia, aumento da coesão social, maior propensão para alcançar os Objetivos de Desenvolvimento do Milênio – principalmente aqueles relacionados à saúde, e contribui para a competitividade dos países em um cenário econômico cada vez mais globalizado (UNESCO, 2007).

Em média, uma maior escolaridade está relacionada a maiores salários. Analisando dados de 98 países, Psacharopoulos e Patrinos (2004) encontraram uma taxa média de retorno da educação de 10% para cada ano adicional de estudo. Os países de baixa e média renda são os que apresentam os maiores retornos à escolaridade. Países da América Latina, Caribe e África Subsaariana apresentam maiores retornos comparados aos países de alta renda pertencentes à OCDE, por exemplo.

Para além do retorno positivo em relação à renda, a educação também contribui para ganhos não pecuniários. Estudos apontam evidência de ligação causal entre educação e melhora da saúde, aumento da longevidade, maior satisfação pessoal, menor taxa de gravidez na adolescência, menor número de divórcios, aumento da paciência e confiança (BURGESS, 2016). Há também importantes retornos sociais de um maior nível de escolaridade da população, como menores taxas de criminalidade, maior participação cívica, e fortalecimento da democracia em relação à ditadura.

Em relação à política macroeconômica, Burgess (2016) argumenta que a educação é ponto chave para três questões cruciais de política, a saber, para o crescimento econômico – o estoque de habilidades de um país é imprescindível para o crescimento econômico em um ambiente de alta competitividade internacional; para a redução da desigualdade de renda – a distribuição de capital humano é determinante para a igualdade de renda; e para a mobilidade social – a relação entre capital humano e experiência pessoal é fundamental para a mobilidade social.

Considerando a importância da educação para aspectos relevantes da vida privada e social, esforços têm sido implementados em diversos países para ampliar o acesso à educação e melhorar sua qualidade. Em 2000, 164 países comprometeram-se com as metas¹ para educação estabelecidas no Fórum Mundial de Educação em Dakar, sendo o Brasil um dos signatários. Segundo dados da UNESCO para estes países, a taxa de escolarização na escola primária era 84%, em 1999, com estimativa de que atingisse 93%, em 2015. A taxa de escolarização do ensino médio aumentou de 71%, em 1999, para 85%, em 2012 (UNESCO, 2015).

Embora avanços tenham sido feitos em relação ao acesso à escola, os esforços para abordar as questões de qualidade ainda não correspondem às necessidades. Permanecem níveis significativos de evasão, distorção idade-série, problemas de progressão e conclusão do ensino formal em diversos países (UNESCO, 2015).

O abandono e a evasão escolar se destacam como problemas que necessitam de políticas específicas dadas as consequências negativas que geram do ponto de vista social e econômico. Segundo dados da Unesco (2017), a taxa bruta de admissão à última série sugere que 90% dos alunos chegaram ao final do ensino primário e 77% chegaram aos anos iniciais do ensino secundário em 2015. Considerando a taxa média de conclusão entre 2010-2015, 83% dos alunos concluíram o ensino primário, 69% concluíram os anos iniciais do ensino secundário e

¹ As metas estabelecem acesso gratuito e compulsório à educação primária de qualidade, eliminação da disparidade de gênero no ensino primário e secundário, avanço na alfabetização de adultos, entre outros (UNESCO, 2015).

45% dos alunos concluíram todo o ensino secundário. Esses dados sinalizam o problema que o abandono e evasão representam para diversos países, e o esforço que necessita ser feito por parte das sociedades e dos governos para reduzir essas taxas e evitar as consequências negativas que geram do ponto de vista social e econômico.

1.2 Abandono escolar: causas e consequências

Dada a importância da educação tanto em termos privados quanto sociais, o engajamento do indivíduo até completar o ciclo formal de ensino é um desafio para os sistemas educacionais. Os altos índices de abandono e evasão têm motivado o desenvolvimento de investigações sobre suas causas e consequências em diversos países.

Designado como condição do aluno que deixa de frequentar a escola durante o andamento do ano letivo, o abandono escolar difere-se da evasão escolar². Embora comumente utilizados como sinônimos, os dois conceitos possuem significados distintos. Entretanto, as causas que incidem sobre eles são muito similares.

Estudos relacionam as causas do abandono a diversos fatores, como características pessoais do indivíduo, características familiares e socioeconômicas, desempenho escolar, e as características da escola e professores. Algumas conclusões desses estudos são apresentadas a seguir. A literatura sobre educação compreende o abandono como resultado final de um processo de desligamento da escola que vai se desenvolvendo ao longo do tempo e têm múltiplos fatores causais. Dentre esses fatores podem-se elencar os aspectos socioeconômicos dos jovens e de suas famílias, a relação com os docentes, dificuldades de aprendizado, percepção equivocada sobre a importância da educação, dentre outros.

Analisando o abandono escolar para o ensino secundário na América Central, Adelman e Székely (2016) salientam que o benefício percebido de permanecer na escola reduz-se à medida que o avanço nas séries se torna mais difícil quando o aluno tem uma base

² Taxa de evasão: Na série k no ano t é a proporção de alunos matriculados na série k no ano t que não se matricularam no ano (t+1) (GLOSSÁRIO/IBGE, 2017).

relativamente fraca de habilidades adquiridas. Os custos de permanecer na escola podem levar em consideração a distância em relação à escola, a falta de vagas, ou mensalidade escolar. Por outro lado, a necessidade familiar de trabalho não remunerado (como cuidar de irmãos mais novos ou auxiliar a família em trabalho agrícola) pode representar um alto custo de oportunidade para a manutenção do jovem na escola. As oportunidades no mercado de trabalho também são relevantes para essa escolha: o crescimento da demanda por mão de obra não qualificada pode incentivar os jovens a sair da escola antes de completar o ciclo de ensino.

Leon e Menezes-Filho (2002) destacam que fatores relevantes para o desempenho do indivíduo na escola e sua escolha em relação a seu nível educacional estão relacionados à condição socioeconômica do estudante, a compatibilidade do estudo com o mercado de trabalho, as condições econômicas e sociais da região onde vive, e as características observáveis (idade, sexo) e não observáveis (talento, determinação e vontade de continuar estudando, por exemplo). Indivíduos em situação econômica precária tendem a substituir o estudo por trabalho por uma questão de sobrevivência. Por outro lado, se o mercado local demanda mão de obra pouco qualificada, há um incentivo para que os indivíduos deixem a escola e comecem a trabalhar mais cedo.

Segundo Duryea et al. (2007 *apud* UNESCO, 2015), os meninos mais pobres têm maior risco ao abandono. Os autores apontam que no Brasil os meninos de famílias pobres têm 46% a mais de chance de abandonarem a escola após uma súbita queda na renda familiar, comparados aos meninos de famílias não pobres.

Fortin et al (2006) apontam que o risco de abandono é influenciado por fatores pessoais, familiares e relativos à escola. Entre as características pessoais, são elencados fatores como dificuldades de aprendizado, baixo desempenho escolar, perda de motivação, níveis altos de ansiedade, depressão, e problemas de comportamento como agressividade, criminalidade, abuso de drogas e álcool.

Em relação aos fatores familiares, os referidos autores apresentam estudos que associam um maior risco de abandono para estudantes cujas famílias têm estilos de vida não convencional (como por exemplo, consumo de drogas em casa), pais separados ou famílias monoparentais, falta de apoio emocional por parte dos pais, falta de envolvimento dos pais nas atividades escolares, conflitos entre adolescentes e os pais, dentre outros. Estudantes cujas famílias dependem de assistência social ou cujos pais enfrentam frequentes mudanças de emprego apresentam maior risco de abandonar a escola do que estudantes cujos pais têm emprego estável (FORTIN *ET AL*, 2006).

O contexto escolar tem importante papel no desenvolvimento pessoal e social das crianças e adolescentes. A qualidade da interação entre professor e aluno é um aspecto fundamental para a permanência e desempenho do aluno na escola – uma postura de cooperação e interesse por parte do professor é muito importante para atingir tais objetivos. Por outro lado, um ambiente de sala de aula cujas regras não são claras ou são inconsistentes, assim como professores focados em intervenções punitivas e suspensões frequentes contribuem para aumentar o risco de abandono (FORTIN *ET AL*, 2006).

As características socioeconômicas das famílias são comumente analisadas e relacionadas ao nível de escolaridade dos indivíduos. A literatura sobre educação aponta que a renda familiar per capita e a escolaridade dos pais têm influência significativa na quantidade de anos de estudo dos filhos.

Em estudo com dados da PNAD para indivíduos entre 11 e 25 anos moradores de áreas urbanas no Nordeste e Sudeste do Brasil, Barros *et al* (2001) encontraram que a escolaridade dos pais, em particular a da mãe, foi o fator mais importante para o desempenho escolar dos filhos, dentre quatro conjuntos de variáveis analisadas: a qualidade e disponibilidade dos serviços educacionais; o custo de oportunidade do tempo; a disponibilidade de recursos familiares (onde se encaixa a escolaridade dos pais); e a disponibilidade de recursos da comunidade. Um ano adicional na escolaridade dos pais tem mais impacto do que a escolaridade dos professores e o nível de renda familiar per capita, por exemplo.

Estudos realizados na Argentina, Brasil, Colômbia, Índia, Marrocos, Peru e Turquia constataram que intervenções ainda no ensino primário tendem a ter efeito positivo sobre a probabilidade de matrícula, a redução da repetência e abandono, e sobre o desempenho dos alunos no ensino fundamental. Tais efeitos tendem a ser maiores no caso de crianças pobres e marginalizadas (MYERS, 1992; MYERS, 2004 *apud* UNESCO, 2015). Adelman *et al* (2017) afirmam que para reduzir as taxas de abandono e evasão, os formuladores de política devem primeiro abordar a questão de como direcionar recursos limitados de forma eficaz para tais intervenções.

Em relação às consequências do abandono, a literatura sobre educação aponta tanto aspectos econômicos quanto sociais. A evidência global sugere que, em média, os jovens que abandonam a escola prematuramente terão salários menores e enfrentarão mais desafios sociais e econômicos em comparação àqueles que terminam os estudos (PATRINOS and PSACHAROPOULOS, 2004; OREOPOULOS and SALVANES, 2011; BENTAOUET-KATTAN and SZÉKELY, 2014; *apud* ADELAMM, 2017).

Bowers, Sprott and Taff (2013) destacam que, conforme uma série de pesquisas sobre o tema, alunos que não concluem o ensino médio estão propensos a taxas mais altas de desemprego e encarceramento, menores ganhos monetários e menor expectativa de vida.

A partir de estudo realizado na América Latina, Cardenas, De Hoyos, and Székely (2015) apontam o abandono escolar como fator que concorre para o aumento da população de jovens não qualificados e não engajados na região, constituindo um desafio socioeconômico para seus governos.

Embora as meninas permaneçam menos propensas a ingressar na escola, em muitos países os meninos apresentam maior risco de não completar os estudos. Dentre as causas para o abandono deles estão a pobreza, a necessidade ou desejo de trabalhar, entrada tardia na escola, desempenho ruim, desinteresse pela escola (BARKER *et al.*, 2012; HUNT, 2008; JHA *et al.*, 2012 *apud* UNESCO, 2015).

Altas taxas de abandono entre os homens têm importantes repercussões em questões de gênero. Pesquisa conduzida entre 2009 e 2010 no Brasil, Chile, Croácia, Índia, México e Ruanda apontou que homens com menor nível educacional são mais propensos a expressar visões de gênero discriminatórias, a serem violentos em casa e menos propensos ao envolvimento nos cuidados com filhos (BARKER *et al*, 2011 *apud* UNESCO 2015). Em contrapartida, homens que completaram o ensino secundário demonstraram atitudes e práticas mais equitativas do ponto de vista de gênero.

Os indivíduos que por um motivo ou outro abandonam ou evadem farão parte de um grande contingente de cidadãos com baixa formação educacional e dificuldades de assumir questões fundamentais da vida em sociedade, tanto na esfera pessoal, profissional ou vinculada à cidadania. Sendo assim, o abandono e a evasão apresentam-se como um desafio social a ser transposto a fim de melhorar tanto a qualidade de vida do cidadão quanto à qualidade de vida da sociedade em geral.

2. CONTEXTUALIZAÇÃO

2.1 Direito à educação no Brasil e Plano Estadual de Educação

O Brasil apresenta elevadas taxas de retorno à educação (BARROS *et al*, 2001; PSACHAROPOULOS e PATRINOS, 2004), fato que deveria constituir um forte incentivo para investimentos em educação no país. Mesmo com os avanços nos últimos anos – como o aumento do número de alunos na escola e aumento da escolaridade média da população – muitos desafios persistem para o sistema educacional brasileiro.

Segundo Barros *et al* (2001), o fraco desempenho educacional no Brasil tem como determinantes fatores diversos como a disponibilidade e qualidade dos serviços educacionais, que podem desestimular o empenho dos alunos; a atratividade do mercado de trabalho local, que, caso demande mão de obra pouco qualificada, pode fazer com que o aluno abandone a escola a fim de trabalhar; a disponibilidade de recursos financeiros e não financeiros das famílias, que pode constituir incentivo ou entrave ao desempenho escolar; e o volume de recursos da comunidade em que o indivíduo vive. Sendo assim, não só aspectos estruturais e pedagógicos da educação importam para o bom desempenho escolar, como também aspectos socioeconômicos dos indivíduos.

Considerado um dos maiores desafios da sociedade brasileira para os próximos anos, a garantia do direito à educação abrange dois pontos principais: a oportunidade de acesso à escola e a garantia da permanência do aluno na instituição escolar, mediante a universalização do ensino com qualidade.

A busca por garantir esse direito social está associada ao modo de monitoramento, acompanhamento, avaliação e intervenção em relação ao ensino nas unidades escolares, de forma a assegurar melhorias no que tange ao acesso e à permanência do indivíduo na instituição de ensino. Indicadores que permitam avaliar os resultados e desempenho de programas numa perspectiva sistêmica auxiliam o alcance de tais melhorias.

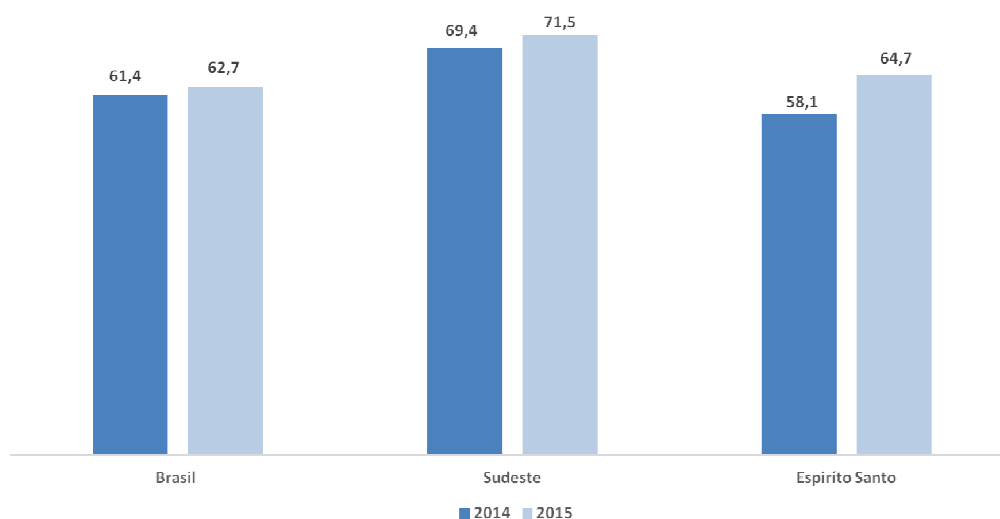
O Plano Estadual de Educação do Espírito Santo (PEE), aprovado pela Lei nº 10.382, de 24 de junho de 2015, estabelece diretrizes e metas para o ensino básico e para o ensino superior no Estado, construídas em consonância com o Plano Nacional de Educação de 2014. No rol das diretrizes constam, por exemplo, a universalização do atendimento escolar, melhoria da qualidade da educação e estabelecimento de meta de aplicação de recursos públicos em educação como proporção do Produto Interno Bruto estadual.

Dentre as metas do PEE, destacam-se duas relacionadas ao ensino médio: universalizar o atendimento escolar para a população na faixa etária entre 15 e 17 anos até 2016, e elevar a Taxa Líquida de Matrículas³ no ensino médio para 85% até o ano de 2025. Garantir o cumprimento dessas metas requer um avanço expressivo nas políticas de acesso e permanência na escola conjuntamente à melhoria da qualidade do ensino.

Em 2015, a taxa líquida de matrícula no Estado foi de 64,7% ante 58,1% em 2014 (gráfico 1). O desempenho do ES nesse indicador ficou acima da média nacional (62,7%), mas abaixo do registrado para o Sudeste (71,5%). Mesmo com a ampliação do atendimento escolar registrado em 2015, o Espírito Santo precisa continuar avançando para alcançar a meta proposta pelo PEE até 2025.

³ A Taxa Líquida de Matrícula - representa a razão entre o número de matrículas de alunos com idade prevista (15 a 17 anos) para estar cursando determinada etapa de ensino e a população total na mesma faixa etária (OBSERVATÓRIO, 2017).

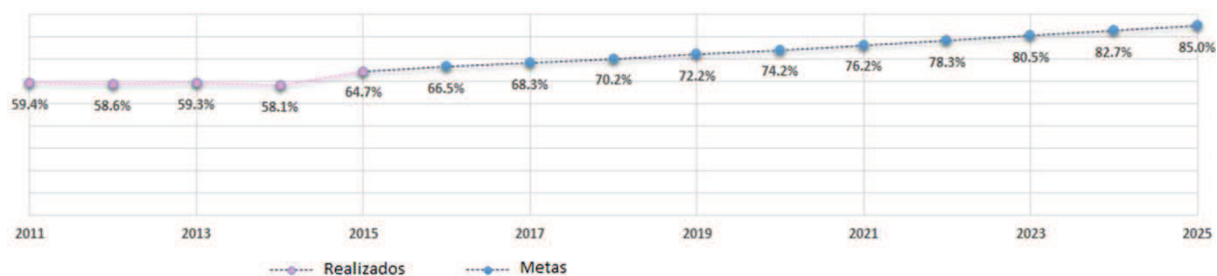
Gráfico 1 – Taxa Líquida de Matrícula no Ensino Médio em relação à população de 15 a 17 anos



Fonte: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP)

Assim sendo, constitui-se um desafio do Estado atingir, em média, uma taxa mínima de crescimento anual de 2,76%, dentro de dez anos, entre os anos de 2016 a 2025 (gráfico 2).

Gráfico 2– Projeção da Taxa Líquida de Matrícula no Ensino Médio - Espírito Santo



Fonte: elaboração própria, baseado em informações do Relatório de Acompanhamento do Plano Estadual de Educação/IJSN, 2017.

A taxa de abandono escolar tem se tornado uma das deficiências no ensino médio, prejudicando a permanência do aluno na escola, aumentando a evasão e dificultando o cumprimento das metas para a educação. Não só aumentar o acesso à escola é importante, como garantir a permanência do aluno na escola até completar o ciclo básico de ensino. Sendo assim, é de extrema importância acompanhar e desenvolver estratégias para reduzir o abandono e a evasão escolar.

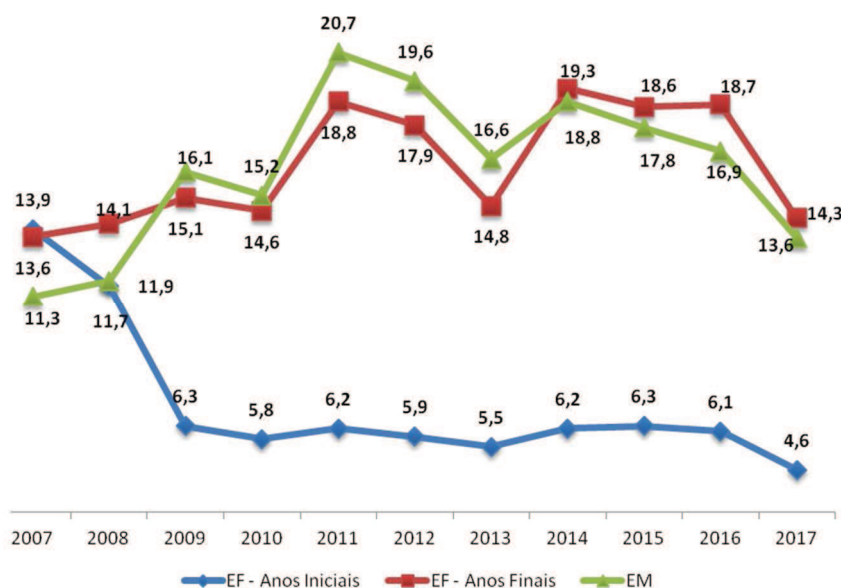
2.2 Panorama do rendimento escolar no Espírito Santo

Segundo dados da PNAD para 2017, dos indivíduos com 18 anos ou mais residentes no Espírito Santo, 35,7% não completaram o ensino fundamental e 13,5% tem o ensino fundamental completo ou ensino médio incompleto. Considerando que a idade ideal para concluir o ensino básico é 18 anos, constata-se que uma parcela importante da população tem baixo nível de escolaridade. Embora a escolaridade média da população acima de 25 anos aumentou entre 2012 e 2017, passando de 7,6 anos para 8,3 anos, ainda é menor do que o período do atual ensino fundamental (9 anos).

Como indicadores quantitativos do rendimento escolar, serão apresentados dados referentes à reprovação, abandono e distorção idade-série para escolas estaduais de ensino fundamental e médio do Estado.

As taxas de reprovação nos anos finais do ensino fundamental e para o ensino médio apresentam crescimento expressivo entre 2007 até 2011 alcançando 18,8% e 20,7%, respectivamente. A partir de 2012 apresentam queda aproximando-se do patamar inicial, mas ainda elevadas – 14,3% e 13,6%, respectivamente (gráfico 3).

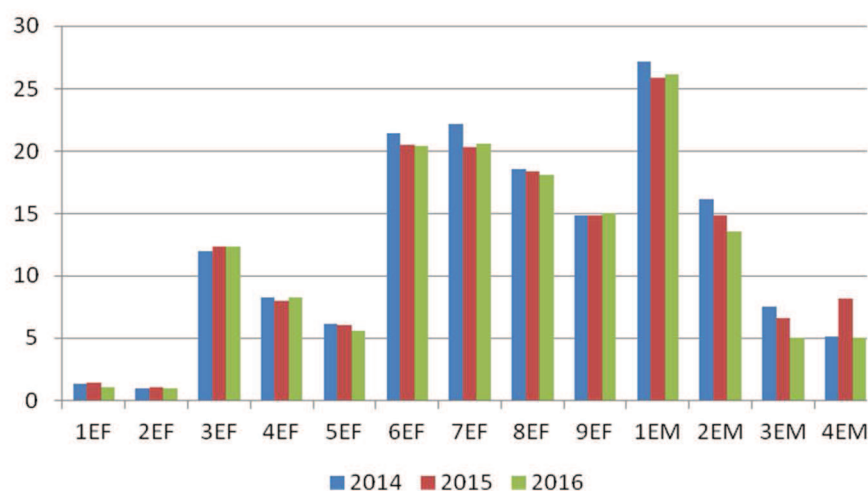
Gráfico 3 - Taxa de Reprovação – Escolas Estaduais - ES



Fonte: Elaboração própria a partir dos dados do INEP.

Ao analisar a reprovação ao longo do ensino básico para as escolas estaduais do ES entre os anos 2014 a 2016, percebe-se um padrão de aumento na taxa de reprovação em três períodos específicos (gráfico 4): do 2º para o 3º ano e do 5º para o 6º ano do ensino fundamental, e do 9º ano do ensino fundamental para o 1º ano do Ensino Médio.

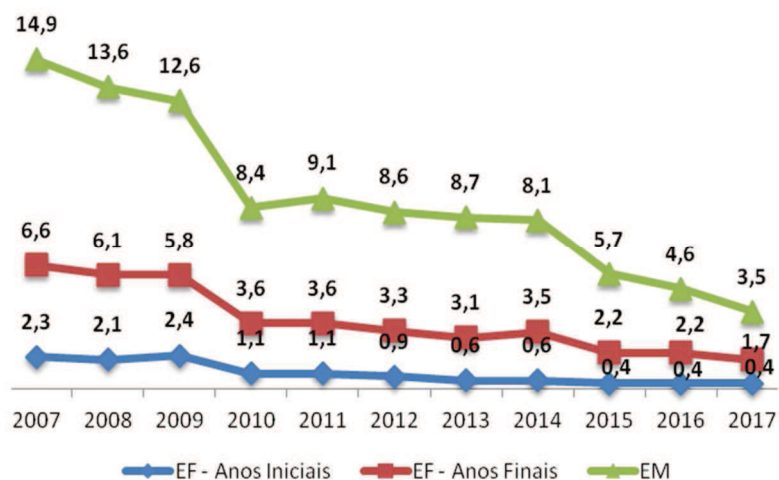
Gráfico 4 – Taxa de Reprovação - Ensino Básico - Escolas Estaduais - ES



Fonte: Elaboração própria a partir dos dados do INEP.

O abandono escolar vem sendo reduzido nas escolas estaduais do ES ao longo dos últimos anos. Embora haja um movimento de convergência entre as taxas, nota-se que o abandono é mais elevado para o ensino médio do que para os anos iniciais e finais do ensino fundamental (gráfico 5).

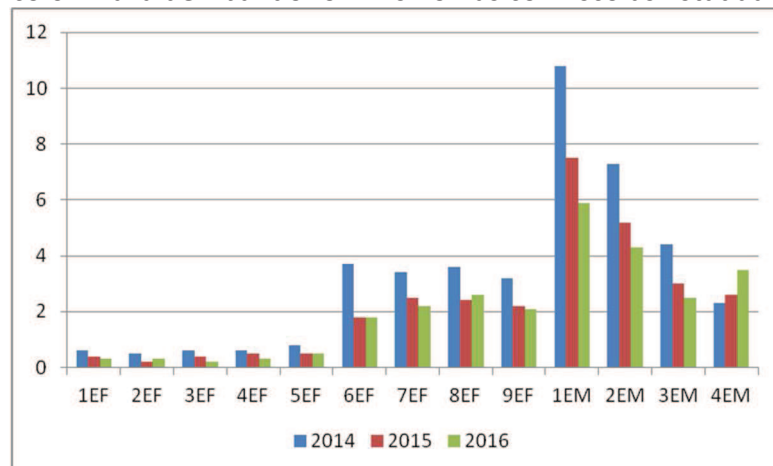
Gráfico 5 - Taxa de Abandono - Escolas Estaduais - ES



Fonte: Elaboração própria a partir dos dados do INEP.

Considerando a taxa de abandono especificamente para cada ano do ensino básico, constata-se que o abandono é mais elevado nos anos finais em relação aos anos iniciais do ensino fundamental, e cresce no ensino médio especialmente no 1º ano (gráfico 6). Em 2016, a taxa de abandono no quinto ano do ensino fundamental foi 0,5%, passando para 2,1% no nono ano e atingindo o pico de 5,9% no primeiro ano do ensino médio.

Gráfico 6 – Taxa de Abandono - Ensino Básico - Escolas Estaduais - ES



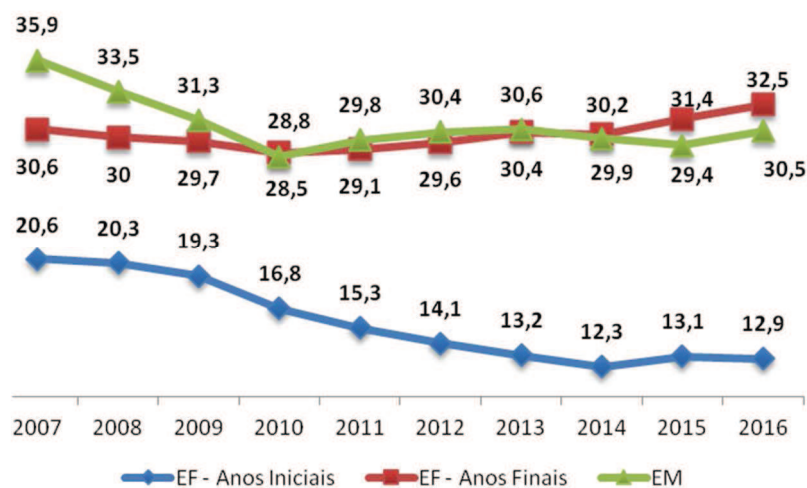
Fonte: Elaboração própria a partir dos dados do INEP.

A Taxa de Distorção Idade-Série⁴ (TDIS) permite avaliar o percentual de alunos com idade superior à idade recomendada para determinada série. São considerados alunos com

⁴ Para o indicador idade série, 2016 é o último ano com disponibilidade de dados.

distorção idade-série aqueles indivíduos que, num determinado ano, apresentam 2 anos ou mais de idade em relação à idade recomendada para a série.

Gráfico 7 – Taxa de Distorção Idade Série - Escolas Estaduais - ES

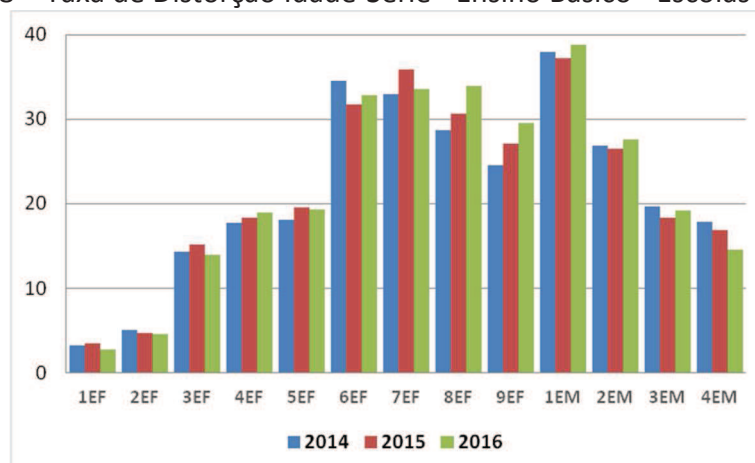


Fonte: Elaboração própria a partir dos dados do INEP.

Observando os dados do ensino fundamental, nota-se que a TDIS para os anos iniciais (1º ao 5º anos) reduziu-se de 20,6% para 12,9% no período. Para os anos finais, o indicador apresentou alta de 1,9 p.p., e para o ensino médio houve redução 5,4% (gráfico 7). Assim como ocorre nas taxas de reprovação e abandono, as TDIS para os anos finais do ensino fundamental são mais elevadas do que para os anos iniciais.

Considerando o ciclo de ensino básico em 2016, nota-se concentração de maiores TDIS entre o 6º ano do Ensino Fundamental e o 1º ano do Ensino Médio, sendo este o ano que apresenta o ápice das taxas – 38% em 2014, 37,2% em 2015 e 38,8% em 2016 (gráfico 8). Esse dado corresponde à expectativa, uma vez que as maiores taxas de reprovação também acontecem no mesmo intervalo de séries do ensino básico.

Gráfico 8 – Taxa de Distorção Idade-Série - Ensino Básico - Escolas Estaduais



Fonte: Elaboração própria a partir dos dados do INEP.

3. MODELO PREDITOR DE ABANDONO

3.1 Justificativa para utilização da predição de abandono e aplicações na literatura internacional sobre educação

Embora no dia-a-dia da escola os educadores tenham percepção dos alunos mais propensos ao abandono escolar, tal informação ainda não está sistematizada ou instrumentalizada na rede Estadual de ensino, a ponto de subsidiar e favorecer a tomada de ações conjuntas e permitir um acompanhamento mais próximo do abandono por outros setores e órgãos da educação. Além disso, não há respaldo para o empenho de recursos em ações de prevenção ao abandono escolar em ampla escala, medidas estas que demandam grande volume de recursos financeiros.

Bowers, Sprott and Taff (2013) argumentam que é imprescindível conhecer os fatores que influenciam o abandono tanto para identificar quem tem maior chance de abandonar a escola quanto para auxiliar a decisão da escola em fornecer intervenções aos alunos considerados em risco. A precisão do preditor de abandono é muito importante, pois alunos erroneamente identificados como propensos a abandonar podem receber intervenção de forma inócua, enquanto alunos realmente propensos e não identificados não irão receber a intervenção. Tal situação exemplifica um gerenciamento ineficiente dos limitados recursos públicos. Alguns estudos estimam que modelos de predição de abandono só identificam com precisão cerca de 50% a 60% dos alunos que realmente abandonam a escola, o que significa um grande percentual de erros de identificação e de não identificação (BALFANZ, HERZOG e MACIVER, 2007; GLEASON e DYNARSKI, 2002; JANOSZ, ARCHAMBAULT, MORIZOT, e PAGANI, 2008, *apud* BOWERS, SPOTT e TAFF, 2013).

Adelman *et al* (2017) ressaltam que identificar com precisão alunos em risco de abandono é importante para direcionar intervenções eficazes particularmente em contextos de recursos limitados e prioridades concorrentes. As pesquisas sobre previsão de abandono são amplamente disseminadas nos EUA e Canadá e têm sido cada vez mais implementadas em outros países. A maioria dos países europeus mantém sistemas de monitoramento de

potencial abandono através do gerenciamento de informações. Os modelos de previsão de abandono comumente baseiam-se em indicadores do engajamento e desempenho do aluno, como presença, problemas disciplinares, notas, e em aspectos socioeconômicos, como raça e qualificação para programas sociais.

Os preditores de abandono mais precisos utilizam a modelagem *Growth Mixture Modeling* (GMM), uma forma de análise longitudinal multivariada que considera a trajetória acadêmica dos alunos, modelando os resultados de aprendizagem ao longo do tempo. Esse método de previsão atinge cerca de 80% a 90% de identificação correta dos alunos que irão desistir e cerca de 10% a 20% de falso-positivos (alunos identificados como propensos ao abandono, mas que concluíram os estudos). O poder de previsão dessa metodologia corrobora a afirmação da literatura de que o abandono é a manifestação de um processo de desligamento e não uma decisão discreta.

Bowers, Sprott and Taff (2013) analisaram 110 indicadores contidos em 36 diferentes estudos sobre previsão de abandono. Os autores recalcularam, a partir dos dados amostrais de cada estudo, as medidas de precisão, especificidade e sensibilidade para cada indicador de forma a comparar sua acurácia.

A precisão de um indicador de abandono é o quantitativo de verdadeiro-positivos (estudantes previstos como abandono e que de fato abandonaram) dividido pelo número total de estudantes previstos como abandono. A proporção de verdadeiro-positivos, ou sensibilidade, é a divisão dos verdadeiro-positivos pelo número total dos abandonos ocorridos (previstos e não previstos). A proporção de verdadeiro-negativos, ou especificidade, é a divisão dos verdadeiro-negativos pelo total de alunos que não abandonaram a escola. Por fim, a proporção de falso-positivos ($1 - \text{especificidade}$) é o resultado da divisão entre os falso-positivos e o número total de graduados.

O quadro 2 descreve a tabela de contingência na qual o evento em questão é se o aluno abandonou a escola ou concluiu os estudos (colunas). O indicador de abandono prevê se o aluno vai abandonar ou não abandonar (linhas). A previsão pode resultar nos seguintes casos:

a) verdadeiro-positivos – previsão correta de abandonos; b) verdadeiro-negativos – previsão correta de não abandono; c) falso-positivos – alunos previstos como abandono, mas que não abandonaram (erro tipo I); d) falso-negativos – alunos previstos como não abandono, mas que abandonaram (erro tipo II).

Embora a precisão seja a métrica mais relatada nos estudos analisados, Bowers, Sprott and Taff (2013) argumentam que o foco exclusivo nesta medida é limitante uma vez que a precisão sozinha não fornece indicação da proporção de todos os alunos que foram identificados como propensos ao abandono. Um indicador acurado é aquele que identifica a maioria dos alunos que irão abandonar a escola, sem identificar erroneamente os alunos que concluirão os estudos como potenciais desistentes. Assim sendo, a medida de acurácia considera a proporção de verdadeiro-positivos (sensibilidade) gerados pelo indicador com sua proporção de falso-positivos (1 – especificidade) – quanto maior a proporção de verdadeiro-positivos e menor a proporção de falso-positivos, mais acurado o indicador.

Quadro 2 – Eventos para o cálculo das proporções de contingência de abandono

		Evento	
		Abandono	Conclusão
Previsão	Abandono	a Verdadeiro-Positivo Correto	b Falso-Positivo Erro Tipo I
	Conclusão	c Falso-Negativo Erro Tipo II	d Verdadeiro-Negativo Correto
Precisão = $a/(a+b)$			
Sensibilidade = $a/(a+c)$			
Especificidade = $d/(b+d)$			
1-Especificidade = $b/(b+d)$			

Fonte: Bowers, Sprott e Taff (2013).

Dentre os indicadores analisados pelos referidos autores, os três mais acurados foram obtidos a partir de análise longitudinal multivariada através do método Growth Mixture Modeling (tabela 1). A constatação de que estes indicadores incorporam uma forma de modelagem

longitudinal apoia a literatura que sugere que o melhor modo de descrever o processo de abandono não é com dados em corte transversal, mas com dados longitudinais que captam a trajetória do aluno na escola ao longo do tempo.

Tabela 1 – Indicadores de Abandono mais Acurados

Estudo	Indicador de Abandono	Tamanho da Amostra	Proporção de Verdadeiro-Positivo (Sensibilidade)	Proporção de Falso-Negativo (1 - Especificidade)
Muthén (2004)	GMM utilizando a evolução da nota em matemática	2.757	89,6%	6,8%
Bowers e Sprott (2012)	GMM utilizando a nota do teste GPA	5.400	91,8%	18,0%
Janosz et al. (2008)	GMM utilizando medidas de engajamento	13.300	78,7%	6,8%

Fonte: Bowers, Sprott e Taff (2013).

Embora os modelos que utilizam GMM se mostrem os mais acurados, os altos requisitos de dados e a complexidade relativa dessa abordagem limita a possibilidade de utilização pela maior parte dos sistemas escolares (ADELMAN *ET AL*, 2017). Sendo assim, outros métodos menos acurados, mas que demonstram bons resultados são utilizados em diversos estudos.

Como alternativa ao método GMM, alguns métodos mais simples e fáceis de serem aplicados apresentam bom desempenho. As escolas públicas de Chicago utilizam um método chamado “*On Track Indicator*”, que consiste em um indicador binário de duas medidas de desempenho dos alunos no primeiro ano do ensino médio – se o aluno obteve os créditos mínimos para a promoção para a próxima série e se não teve mais do que um “F” (nota mínima) em alguma disciplina por semestre. Este indicador atinge 75% de sensibilidade e 16% de falso-positivo (ALLENSWORTH e EASTON 2005, 2007, *apud* ADELMAN *ET AL*, 2017).

Regressões logísticas incluindo indicadores de repetência e a nota média anual para algumas cidades dos EUA foram capazes de obter 81% de especificidade e 25% de falso-positivos (BOWERS, 2010 *apud* ADELMAN *ET AL*, 2017). Utilizando dados administrativos sobre frequência, violações de comportamento, notas e características demográficas do estado de

Wisconsin, Knowles (2015 *apud* ADELMAN *ET AL*, 2017), estimou uma gama de modelos e descobriu que os modelos de regressão logística apresentam desempenho muito bom quando comparados a modelos com algoritmos mais complexos.

Na Guatemala e em Honduras, o substancial avanço no escopo e confiabilidade dos dados administrativos de educação permitiu a aplicação do modelo de predição de abandono com ótimo desempenho. Uma importante inovação no sistema de informação da Guatemala foi a implementação de um identificador único por aluno que permite acompanhar a trajetória dos alunos ao longo do tempo – cada escola envia ao Ministério da Educação dados relativos à situação do aluno no início do ano letivo, gênero do aluno, série cursada, dentre outros. Estes microdados administrativos básicos são combinados com dados sobre as características dos alunos e das escolas provenientes da Avaliação Nacional de Alunos (ADELMAN *ET AL*, 2017).

Foram estimadas, para os referidos países, três especificações de modelos de probabilidade linear para os três períodos onde a taxa de abandono se apresenta mais elevada: 1) a transição do ensino primário para o secundário, 2) o secundário inferior, e 3) a transição do secundário inferior para o secundário superior. Para cada especificação foi utilizada uma combinação diferente de dados administrativos e características dos alunos a partir de pesquisas nacionais de educação.

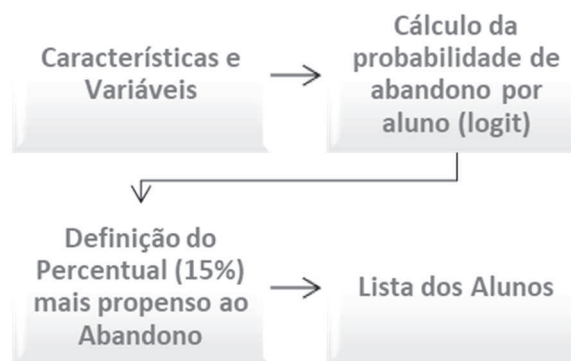
Como exemplo, os modelos mais acurados para a transição entre ensino primário e secundário apresentam os seguintes resultados: para Guatemala, o modelo mais acurado alcançou taxa de sensibilidade de 80% e proporção de falso-positivo de 21,2%, enquanto para Honduras o método mais acurado alcançou taxa de sensibilidade de 78,2% e proporção de falso-positivo de 19,8% (ADELMAN *et al*, 2017). Os autores ressaltam que dados provenientes de um sistema completo e consistente de informações administrativas, assim como a combinação desses dados com outras pesquisas, melhora a qualidade da previsão dos modelos, sendo essenciais para enfrentar o problema do abandono e para a utilização eficiente e eficaz dos recursos públicos direcionados às ações preventivas.

3.2 Descrição do modelo de predição

Buscando uma maior abrangência quanto à identificação dos alunos com maior propensão ao abandono e a inclusão dos mesmos nas ações preventivas a serem adotadas pelas escolas, foi desenvolvido o Modelo Preditor de Abandono Escolar para as escolas da rede pública do ES.

O “Modelo de Predição do Abandono Escolar” apresenta-se como uma proposta inovadora e desafiadora (figura 1), que deverá contar com a participação das instituições de ensino, parceiros e a própria Secretaria de Estado da Educação para a identificação dos alunos, assim como para a execução das ações preventivas a serem adotadas, a partir de um protocolo de ações específicas. Tal iniciativa apresenta-se como fundamental para a consolidação da política de redução do abandono. Se bem-sucedido, o modelo será um importante instrumento para a política de educação no Estado, auxiliando o cumprimento das metas do Plano Estadual de Educação.

Figura 1 – Síntese do “Modelo de Predição do Abandono Escolar”

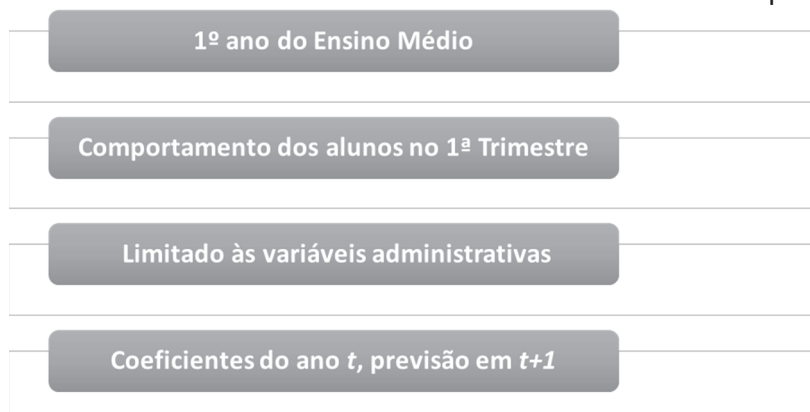


Fonte: elaboração própria.

Para estimar a probabilidade dos alunos abandonarem a escola, adotou-se a regressão logística múltipla (logit). O modelo foi estimado utilizando variáveis referentes às características pessoais e de desempenho dos alunos do primeiro ano do Ensino Médio, no primeiro trimestre do ano letivo, adotando para cálculo da previsão no período futuro ($t+1$) o valor estimado dos coeficientes da cross-section no tempo t . A escolha do modelo *logit* também ocorreu considerando os dados disponíveis para o estudo (ver seção 3.3.3).

A decisão de escolher o 1º ano deve-se ao fato de ser a série com maior taxa de abandono escolar no Ensino Médio, fato que ocorre não só no sistema escolar do Estado, mas que também é referenciado na literatura internacional sobre abandono escolar conforme discutido anteriormente. Como o objetivo é impedir o abandono destes alunos, optou-se por utilizar as informações mais próximas ao início do ano letivo, ou seja, o comportamento dos alunos no 1º trimestre, tendo assim a possibilidade de aplicar o protocolo de ações para reduzir o abandono nos dois trimestres seguintes do mesmo ano (figura 2). Foram testados dois modelos – um utilizando somente variáveis administrativas e outro com a inclusão de variáveis relacionadas à infraestrutura da escola e características socioeconômicas por escola. Conforme discutido na subseção 3.3.3, o modelo que utiliza somente as variáveis administrativas mostrou-se mais adequado.

Figura 2 - Características das variáveis utilizadas no modelo de previsão



Fonte: Elaboração própria.

Na construção do modelo *logit*, o resultado abandono é uma variável dependente que assume apenas dois valores:

$$y = \begin{cases} 1 & \text{com probabilidade } p, \\ 0 & \text{com probabilidade } 1 - p. \end{cases} \quad (1)$$

Seguindo os autores Cameron e Trivedi (2005, p. 466-469), o modelo *logit* ou modelo de regressão logística especifica p da seguinte forma:

$$p = \Lambda(x'\beta) = \frac{e^{x'\beta}}{1 + e^{x'\beta}}, \text{ onde} \quad (2)$$

- x é a matriz de regressores de dimensão $n \times K$, sendo n a quantidade de alunos e $K - 1$ a quantidade de variáveis utilizadas;
- β é o vetor de parâmetros $K \times 1$ e
- $\Lambda(\cdot)$ é a função distribuição acumulada da logística.

Assim, é garantido que p sempre assumirá valores entre 0 e 1, como desejado.

A partir do cálculo dessas probabilidades, é identificado o grupo de alunos com propensão ao abandono, e é gerada uma lista com os 15% mais propensos dentro desse grupo. O ponto de corte de 15% será discutido mais adiante, entretanto cabe ressaltar que um dos critérios para decisão foi minimizar o erro tipo I e erro tipo II, que a literatura internacional referencia como falso-positivos e falso-negativos. Neste trabalho os termos são referenciados, respectivamente, como *erro de inclusão* – alunos previstos ao abandono, mas que não abandonam de fato; e *erro de exclusão* – alunos previstos como não abandono, mas que abandonaram.

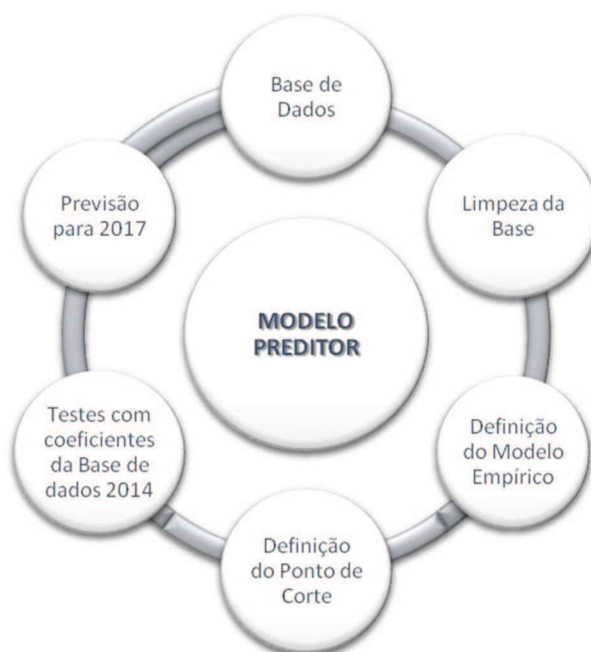
A lista gerada a partir do modelo preditor será entregue para as instituições de ensino para que esses alunos recebam assistência individual e participem de ações e intervenções que cooperem para sua permanência na escola.

3.3 Metodologia adotada na construção do modelo empírico

A metodologia adotada na construção do “**Modelo de Predição do Abandono Escolar**”, contemplou cinco ações primordiais: coleta de dados, limpeza da base de dados para aplicação do modelo, definição do modelo empírico, definição do ponto de corte para a lista

dos alunos mais propensos ao abandono, teste dos coeficientes e previsão para 2017 (figura 3).

Figura 3 – Metodologia utilizada na Construção do Modelo Preditor



Fonte: elaboração própria.

3.3.1 Base de Dados

Para desenvolvimento e construção do modelo de predição foram realizados testes iniciais com informações obtidas a partir de bases de dados disponibilizadas pelos seguintes órgãos: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), Instituto Brasileiro de Geografia e Estatística (IBGE), Instituto de Tecnologia da Informação e Comunicação do Estado do Espírito Santo (PRODEST), Secretaria de Estado da Segurança Pública e Defesa Social (SESP) e da própria Secretaria de Estado da Educação (SEDU).

3.3.2 Limpeza da Base

A partir do recebimento dos dados foi necessário um tratamento cuidadoso e exaustivo para a eliminação das informações duplicadas. O tratamento descrito nessa etapa refere-se à utilização de dados dos anos de 2014, 2015, 2016 e 2017⁵.

Inicialmente tentou-se utilizar como código identificador do aluno, o código do INEP, entretanto, verificou-se que tal informação não existia para alguns alunos, o que impediria seu uso como código identificador. Outra tentativa foi a identificação do aluno por meio do código fornecido pela SEDU, o qual denominamos *idaluno*. Entretanto, notou-se que em determinados casos havia mais de um *idaluno* associado a um mesmo código INEP, o que inviabilizou o seu uso.

Sendo assim, verificou-se a necessidade de realizar diversas fases de limpeza visando eliminar as duplicações tanto com relação ao *idaluno*, quanto ao código INEP, utilizando para tratamento dos dados o *software Stata*. O processo de busca por maior fidedignidade da base de dados demandou tempo e dedicação consideráveis diante de alguns desafios encontrados e dos ajustes necessários. Para isso, foram realizadas quatro etapas de limpeza da base, que podem ser adotadas como padrão nos casos em que há duplicações e impurezas na base de dados inicial (figura 4).

Figura 4 – Etapas da Limpeza da Base de Dados

Etapa 1	• <i>Eliminação de linhas</i>
Etapa 2	• <i>Eliminação das duplicações por idaluno e disciplina</i>
Etapa 3	• <i>Eliminação das duplicações por nome do aluno e data de nascimento</i>
Etapa 4	• <i>Eliminação das duplicações por código inep</i>

Fonte: Elaboração própria.

⁵ A previsão para o ano X considera os dados do ano X-1 e os dados do primeiro trimestre do ano X. Sendo assim, a seção 4 apresentará a previsão de abandono para os anos 2015, 2016 e 2017.

A) Etapa 1- Eliminação de linhas

A base de dados apresenta-se em um formato em que cada linha corresponde a um aluno e uma disciplina respectivamente. Sendo assim, um aluno que cursa 10 disciplinas aparece em 10 linhas.

São consideradas no modelo apenas as disciplinas de Matemática, Língua Portuguesa, Física, Química, Biologia, História e Geografia. Portanto, as linhas correspondentes às demais disciplinas foram eliminadas.

O modelo somente inclui os alunos dos turnos matutino, vespertino e noturno. Os alunos do turno integral foram excluídos da base, por não fazerem parte da população foco desse estudo. A tabela 2 apresenta o quantitativo de *idaluno* inicial e os remanescentes nas duas fases de eliminação de linhas: eliminando disciplinas não incluídas no modelo e eliminando alunos do turno integral.

Tabela 2 – Quantidade de *idaluno*

	2016			2017		
	Sem código INEP	Com código INEP	Total	Sem código INEP	Com código INEP	Total
Inicial	2661	46108	48769	6389	36414	42803
Eliminando disciplinas	2658	46080	48738	6389	36385	42774
Eliminando integral	2642	45540	48182	6097	34432	40529

Fonte: elaboração própria.

B) Etapa 2 - Eliminação das duplicações por *idaluno* e disciplina

A segunda etapa iniciou o processo de eliminação das duplicações de fato. Foram consideradas duplicadas as linhas que possuem o mesmo *idaluno* e a mesma disciplina. As eliminações foram realizadas em quatro fases, na seguinte ordem: **Nota final; Data de Matrícula; Nota do Primeiro Trimestre e Sorteio.**

1ª Fase - Nota Final

No caso de duplicação existem três cenários possíveis quanto à ausência ou não de resposta para a variável “Nota Final”:

1. Todas as duplicações têm informação.	2. Nenhuma das duplicações possui informação.	3. Algumas duplicações têm informação e outras não.
---	---	---

Nos dois primeiros cenários, as linhas duplicadas são mantidas. No terceiro devem ser eliminadas as linhas sem informação (tabela 3).

Tabela 3 - Exemplo Prático Nota Final

Idaluno	Disciplina	Nota Final	Opção Adotada
1	Matemática	-	EXCLUIR
1	Matemática	10	CONTINUAR
2	Matemática	-	CONTINUAR
2	Matemática	-	CONTINUAR
3	Matemática	10	CONTINUAR
3	Matemática	08	CONTINUAR

Fonte: elaboração própria.

2ª Fase - Data de Matrícula

Verificar se as linhas duplicadas possuem a variável referente à “Data de Transferência”.

Cenários possíveis:

1. Alguma das repetições possui linha sem data.	2. Linhas com mesma Data.	3. Linhas com Datas diferentes.
---	---------------------------	---------------------------------

Nos dois primeiros cenários, as linhas duplicadas são mantidas. No terceiro, as linhas referentes às datas mais recentes devem ser mantidas e as demais eliminadas.

A tabela 4 apresenta exemplo da eliminação de linhas duplicadas por “Data de Matrícula” do *idaluno*.

Tabela 4 - Exemplo Prático - Data de Matrícula

Idaluno	Disciplina	Nota Final	Data de Matrícula	Opção Adotada
1	Matemática	10	01.04	CONTINUAR
2	Matemática	-	28.03	EXCLUIR
2	Matemática	-	02.04	CONTINUAR
3	Matemática	10	02.04	CONTINUAR
3	Matemática	08	02.04	CONTINUAR
4	Matemática	08	03.05	CONTINUAR
4	Matemática	09	-	CONTINUAR

Fonte: elaboração própria.

3ª Fase – Primeiro Trimestre

No caso de duplicação existem três cenários possíveis quanto à ausência ou não de resposta para a variável “Nota do Primeiro Trimestre”:

1. Todas as duplicações têm informação.	2. Nenhuma das duplicações possui informação.	3. Algumas duplicações possuem informação e outras não.
---	---	---

Nos dois primeiros cenários, as linhas duplicadas são mantidas. No terceiro, devem ser eliminadas as linhas sem informação.

Segue exemplo, na tabela 5, da eliminação de linhas duplicadas por “Nota do Primeiro Trimestre” do *idaluno*.

Tabela 5 - Exemplo Prático – Nota do Primeiro Trimestre

Idaluno	Disciplina	Nota do Primeiro Trimestre	Nota Final	Opção Adotada
1	Matemática	-	-	EXCLUIR
1	Matemática	10	-	CONTINUAR
2	Matemática	-	-	CONTINUAR
2	Matemática	-	-	CONTINUAR
3	Matemática	10	10	CONTINUAR
3	Matemática	08	08	CONTINUAR

Fonte: elaboração própria.

4ª Fase – Sorteio

Realizar sorteio das linhas duplicadas, para manter apenas uma linha com informação do aluno, conforme passos a seguir:

- Ordenar a base de dados por idaluno e disciplina.
- Escolher uma semente para gerar variável aleatória N, com distribuição uniforme, definindo um valor para cada linha.

Assim, a base será reordenada em ordem crescente, considerando a chave formada por Idaluno, disciplina e a variável N. Cenários possíveis:

1. Quando não houver mais duplicação.

2. Quando houver duplicação.

No primeiro cenário não há necessidade de intervenção. Já no segundo, deverão ser mantidas na base apenas a repetição que apresenta o maior valor para a Variável N (número de linha). A tabela 6 apresenta a eliminação de linhas duplicadas por “Variável N” do *idaluno*.

Tabela 6 - Exemplo Prático - Sorteio

Idaluno	Disciplina	Nota do Primeiro Trimestre	Variável Aleatória	Variável N	Opção Adotada
11	Matemática	10	0.0915535	30506	EXCLUIR
11	Matemática	10	0.3047738	102448	EXCLUIR
11	Matemática	10	0.5675986	190734	EXCLUIR
11	Matemática	10	0.8338892	280577	CONTINUAR
12	Física	9	0.1772932	59421	EXCLUIR
12	Física	9	0.4360625	146420	CONTINUAR
13	Química	8	0.4650271	156182	EXCLUIR
13	Química	8	0.7551267	253924	EXCLUIR
13	Química	8	0.9296933	312335	CONTINUAR

Fonte: elaboração própria.

Após a realização do sorteio das linhas duplicadas, não restará duplicação de linhas relacionadas às variáveis de *idaluno* e disciplina.

A tabela 7 apresenta a quantidade de eliminações em cada uma das fases descritas. Nota-se que o total de *idaluno* se mantém o mesmo, uma vez que foram eliminadas apenas as duplicações.

Tabela 7 - Duplicação por idaluno e disciplina

Duplicações Eliminadas						
	2016			2017		
	Sem código INEP	Com código INEP	Total	Sem código INEP	Com código INEP	Total
Nota final	61	1289	1350	Não se aplica		
Data de matrícula	10	150	160	0	0	0
Nota do 1º trimestre	10	150	160	0	0	0
Sorteio	0	1	1	0	0	0
Quantidade de idaluno remanescentes						
Nota final	2642	45540	48182	Não se aplica		
Data de matrícula	2642	45540	48182	6097	34432	40529
Nota do 1º trimestre	2642	45540	48182	6097	34432	40529
Sorteio	2642	45540	48182	6097	34432	40529

Fonte: elaboração própria.

Devido ao fato de que cada aluno aparecia na base várias vezes conforme o número de disciplinas, foi necessário reorganizar a base de forma que as informações referentes a cada aluno aparecessem em uma única linha. Assim, a base passou a assumir a seguinte forma:

Tabela 8 – Modelo de Formatação da base – Alunos por linha

idaluno	codinep	nota_mt	nota_lp	nota_bio	nota_qui	nota_fis	nota_hist	nota_geo
1	120							
2	125							
3	130							

Fonte: elaboração própria.

Em alguns casos foram observados termos distintos para o mesmo conceito de interesse em relação à situação do aluno. Para o objetivo do projeto em questão, os termos foram agregados em quatro categorias relacionadas à situação do aluno: matriculado, matrícula suplementar, transferido ou desistente, abandono. Tal medida evita a ocorrência de equívocos principalmente durante a realização das duas etapas (3 e 4) de limpeza da base. O quadro abaixo apresenta a categorização utilizada para alinhar os termos existentes na base de dados.

Quadro 2 – Redefinição da situação do aluno para utilização no modelo

IJSN	SEDU	DESTINO:
Matriculado	"Aprovado" "Reprovado " "Matriculado"	Limpeza das Bases
Matrícula Suplementar	"Matr. Suplementar"	
Transferido ou Desistente	"Abandonou o curso" "Deixou de frequentar" "Desistente de matrícula" "Transferido" "Transferido de modalidade" "Abandono" "Desistente Matrícula "	
Abandono	1 "Abandonou o curso" "Deixou de frequentar" "Desistente de matrícula" "Abandono"	Modelo
	0 "Aprovado" "Reclassificado" "Reclassificado por avaliação" "Reprovado" "Matr. Suplementar" "Matriculado"	

Fonte: elaboração própria.

C) Etapa 3 - Eliminação das duplicações por nome do aluno e data de nascimento

Na base de dados existem alunos que possuem mais de um *idaluno* e não possuem código INEP, de forma que a duplicação não seria detectada nem quando considerado o *idaluno* e nem quando considerado o código do INEP (tabela 9). Sendo assim, foi necessário eliminar as duplicações por nome do aluno e data de nascimento.

Tabela 9 - Duplicações por nome do aluno e data de nascimento

idaluno	codinep	Nome	data-nascimento
1	125544	João da Silva	19/05/2000
25		João da Silva	19/05/2000

Nota: os nomes utilizados são fictícios.

Fonte: elaboração própria.

Nessa etapa, a sequência de limpeza da base é feita conforme especificado abaixo:

1) Matriculado

Se alguma das repetições possuírem status “Matriculado”, ela irá prevalecer sobre as demais, que deverão ser eliminadas (tabela 10).

Tabela 10 - Exemplo Prático - Matriculado

Nome	Nascimento	Status	Opção adotada
João Sá	25/01/2000	Matriculado	PERMANECER
João Sá	25/01/2000	Transferido	EXCLUIR
João Sá	25/01/2000		EXCLUIR
Maria Santana	01/03/2001	Abandonou o curso	EXCLUIR
Maria Santana	01/03/2001	Matriculado	PERMANECER
Julia Sousa	28/09/1998	Matriculado	PERMANECER
Julia Sousa	28/09/1998	Matriculado	PERMANECER

Nota: os nomes utilizados no exemplo prático são fictícios.

Fonte: elaboração própria.

2) Matrícula Suplementar

As repetições que possuem status de “Matrícula Suplementar” prevalecem sobre as que possuem status de “Transferido”, “Desistente” e “Abandono”, que deverão ser eliminadas.

Tabela 11 - Exemplo Prático – Matriculado suplementar

Nome	Nascimento	Status	Opção adotada
João Sá	25/01/2000	Matriculado	PERMANECER
Maria Santana	01/03/2001	Matriculado	PERMANECER
Antônio Silva	08/08/2000	Matr. Suplementar	PERMANECER
Antônio Silva	08/08/2000	Transferido	EXCLUIR
Antônio Silva	08/08/2000	Matr. Suplementar	PERMANECER
Carlos Santos	13/05/1999	Matr. Suplementar	PERMANECER
Carlos Santos	13/05/1999	Desistente	EXCLUIR

Nota: os nomes utilizados no exemplo prático são fictícios.

Fonte: elaboração própria.

3) Transferido, Desistente de Matrícula e Abandono

Se as repetições possuem status variando entre “Transferido”, “Desistente” ou “Abandono” todas as repetições devem ser eliminadas da base (tabela 12). Como as datas não permitem identificar o que ocorreu primeiro, optou-se por deixar de considerar algum abandono real em detrimento de classificar como abandono alguém que não abandonou. Nessa fase da limpeza perdem-se observações. Até o momento eram eliminadas apenas duplicações, escolhendo uma das opções entre elas.

Tabela 12 - Exemplo Prático – Transferido, desistente de Matrícula e Abandono

Nome	Nascimento	Status	Opção adotada
Paulo Santos	27/09/1996	Transferido	EXCLUIR
Paulo Santos	27/09/1996	Abandonou o curso	EXCLUIR
Carla Costa	12/10/1998	Desistente	EXCLUIR
Carla Costa	12/10/1998	Abandonou o curso	EXCLUIR
Fernando Gil	06/07/1999	Desistente	EXCLUIR
Fernando Gil	06/07/1999	Transferido	EXCLUIR
Juliana Silveira	02/02/2000	Abandonou	PERMANECER
Juliana Silveira	02/02/2000	-	PERMANECER

Nota: os nomes utilizados no exemplo prático são fictícios.

Fonte: elaboração própria.

4) Data de Matrícula

Se as duplicações possuírem datas de matrícula diferentes, deve-se manter aquela com a data mais recente e eliminar as demais (tabela 13).

Tabela 13 - Exemplo Prático – Data de Matrícula

Nome	Nascimento	Status	Data de matrícula	Opção adotada
Julia Sousa	28/09/1998	Matriculado	01/02/2016	EXCLUIR
Julia Sousa	28/09/1998	Matriculado	06/02/2016	PERMANECER
Antônio Silva	08/08/2000	Matr. Suplementar	01/02/2016	PERMANECER
Antônio Silva	08/08/2000	Matr. Suplementar	01/02/2016	PERMANECER
Antônio Pedro	04/05/2000	Abandono	01/02/2016	EXCLUIR
Antônio Pedro	04/05/2000	-	01/05/2016	PERMANECER

Nota: os nomes utilizados no exemplo prático são fictícios.

Fonte: elaboração própria.

5) Sorteio

Sorteiam-se as linhas duplicadas, para manter apenas uma linha com informação do aluno. Ordena-se a base de dados por Nome do aluno e Data de nascimento. Escolhe-se uma semente para gerar variável aleatória, com distribuição uniforme, definindo um valor para cada linha. A base reordenada por essa variável uniforme, e com essa ordenação, cria-se uma nova variável com o número da linha, possibilitando os seguintes cenários:

- Não há duplicação (nada a ser feito);
- Identifica-se duplicação (são mantidas na base apenas aquelas duplicações com maior número de linha).

A seguir, a tabela 14 aponta o resultado da limpeza das duplicações até o momento. Cada linha corresponde à quantidade de *idaluno* remanescentes em cada etapa de limpeza, separando-os em três subgrupos: alunos que não abandonaram a escola (“Não”), alunos que abandonaram a escola (“Sim”) e os alunos para os quais não se tem informação a esse respeito (“Missing”).

Tabela 14 – Duplicação por nome do aluno e data de nascimento

Fonte: elaboração própria.

Quantidade de idaluno remanescentes – Todos					
	2016				2017
Abandono	Não	Sim	Missing	Total	Total
Matriculado	40508	2942	4555	48005	Não se aplica
Matrícula suplementar	40508	2942	4555	48005	Não se aplica
Transferido, Desistente de Matrícula e Abandono	40508	2927	4540	47975	Não se aplica
Data de matrícula	40507	2926	4539	47972	40489
Sorteio	40506	2926	4539	47971	40487
Quantidade de idaluno remanescentes - Com código INEP					
	2016				2017
Abandono	Não	Sim	Missing	Total	Total
Matriculado	38375	2756	4254	45385	Não se aplica
Matrícula suplementar	38375	2756	4254	45385	Não se aplica
Transferido, Desistente de Matrícula e Abandono	38375	2747	4242	45364	Não se aplica
Data de matrícula	38374	2746	4241	45361	34407
Sorteio	38374	2746	4241	45361	34407

Ainda a respeito da tabela 14, o ano de 2017 não apresenta as subcategorias citadas, pois estas só podem ser determinadas ao fim do aluno letivo, de forma que na época da construção do preditor tais informações não estavam disponíveis.

D) Parte 4 – Eliminação das duplicações por código Inep

Nessa última etapa da limpeza, são realizados exatamente os mesmos procedimentos da etapa anterior, sendo a única diferença que as duplicações são definidas com base no código do INEP. Os resultados são mostrados na tabela 15.

Tabela 15 – Duplicação por código INEP

Quantidade de idaluno remanescentes - Todos					
Abandono	2016				2017
	Não	Sim	Missing	Total	Total
Matriculado	40508	2942	4555	48005	Não se aplica
Matrícula suplementar	40508	2942	4555	48005	Não se aplica
Transferido, Desistente de Matrícula e Abandono	40508	2927	4540	47975	Não se aplica
Data de matrícula	40507	2926	4539	47972	40489
Sorteio	40506	2926	4539	47971	40487
Quantidade de idaluno remanescentes - Com código INEP					
Abandono	2016				2017
	Não	Sim	Missing	Total	Total
Matriculado	38374	2745	4209	45328	Não se aplica
Matrícula suplementar	38374	2745	4209	45328	Não se aplica
Transferido, Desistente de Matrícula e Abandono	38374	2742	4208	45324	Não se aplica
Data de matrícula	38374	2742	4208	45324	34402
Sorteio	38373	2742	4208	45323	34402

Fonte: elaboração própria.

3.3.3 Definição do Modelo Empírico

Foram testados dois modelos: um modelo ampliado, com uma gama maior de variáveis, e um modelo parcimonioso, utilizando somente os dados administrativos fornecidos pela Secretaria de Estado da Educação. A tabela 2 apresenta as variáveis utilizadas em cada modelo. No modelo parcimonioso foram incluídas as variáveis idade, sexo, notas médias da escola, da turma e nota individual nas disciplinas de português e matemática, além da proporção de faltas. O modelo ampliado buscou estimar a propensão de abandono incluindo outras variáveis, apresentadas na tabela 16. Conforme discutido anteriormente, as características socioeconômicas dos alunos influenciam em sua decisão de permanecer ou não na escola. As bases de dados disponíveis não possibilitavam acesso a informações socioeconômicas ao nível do aluno. Sendo assim, optou-se por testar a variável ‘Nível Socioeconômico da Escola’, considerando que ela reflete, em média, as condições socioeconômicas dos alunos. Foram incluídas também variáveis relacionadas à infraestrutura e localização da escola (se área rural ou urbana), variáveis relacionadas ao corpo docente e à administração da escola.

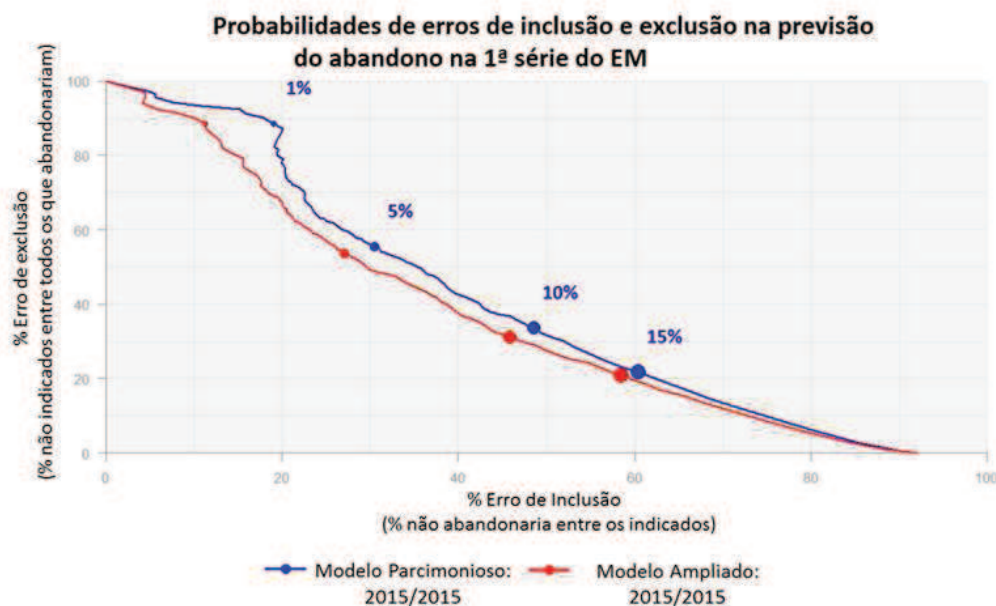
Tabela 16 – Variáveis Utilizadas nos Modelos Parcimonioso e Ampliado

VARIÁVEIS UTILIZADAS	Modelo Parcimonioso	Modelo Ampliado
Idade	x	x
Sexo	x	x
Nível Socioeconômico da escola (INSE)		x
Regional (SRE)		x
Área urbana ou rural		x
Infraestrutura		x
Índice de Complexidade da Gestão (ICG)		x
Indicador de Esforço Docente (IED)		x
Indicador de Adequação da Formação do Docente (AFD)		x
Nota Média da Escola - Português e Matemática	x	x
Turno (Matutino, Vespertino e Noturno)	x	x
Nota Média da Turma - Português e Matemática	x	x
Nota nas disciplinas de Língua Portuguesa, Matemática, História, Geografia, Biologia, Química, Física	x	x
Proporção de Faltas nas disciplinas de Língua Portuguesa, Matemática, História, Geografia, Biologia, Química, Física	x	x
Dummy Mudança de Falta	x	x
Dummy de Mudança de Nota	x	x

Fonte: Elaboração própria.

Os modelos foram estimados com a especificação *logit* discutida acima. Para compará-los, ponderou-se a probabilidade de dois erros: *Erro de Inclusão* (erro tipo I) ocorre quando o aluno foi previsto como potencial ao abandono, porém não abandona; e *Erro de Exclusão* (erro tipo II) acontece quando o aluno não foi previsto como potencial ao abandono, mas abandona a instituição de ensino.

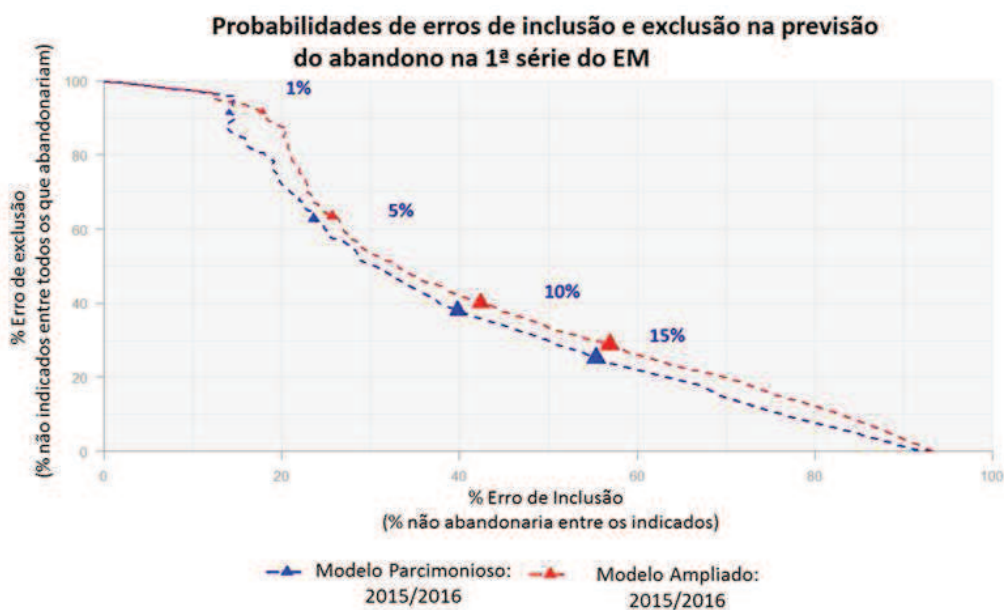
Gráfico 9 – Comparação Modelos com Previsão 2015/2015



Fonte: Elaboração própria.

Considerando os dados de 2015 para prever 2015, o modelo ampliado resulta em menores erros de inclusão e exclusão (gráfico 9). Entretanto, utilizando os dados de 2015 para prever 2016, o modelo parcimonioso demonstra melhor resultado (gráfico 10). Dado que a ideia do modelo preditor é gerar os coeficientes com dados do ano X para a previsão do abandono no ano X+1, o modelo parcimonioso se mostra mais adequado.

Gráfico 10 – Comparação Modelos com Previsão 2015/2016



Fonte: Elaboração própria.

O modelo parcimonioso acerta mais na previsão e se mostra mais fácil de ser implementado. Ele utiliza somente dados administrativos da SEDU e por não depender de dados de outros órgãos, torna o processo de previsão mais célere, o que é importante visto que pretende-se prever o abandono após o primeiro trimestre do ano letivo para implementar as ações ao longo do mesmo ano.

Os testes realizados para a escolha do modelo, a partir das bases de dados de 2014, 2015 e 2016 chamaram atenção para duas observações: a) quanto mais simples as variáveis utilizadas, maior a assertividade do modelo; b) necessidade de delimitar o estudo conforme o ano de referência das informações, pois, nem todas as bases possuem informações referentes ao mesmo período. Torna-se importante ressaltar, mais uma vez, que dentro do próprio ano os melhores resultados em termos de predição apareciam nos modelos mais completos, entretanto, para a previsão no ano seguinte, os resultados melhores se davam com os modelos mais parcimoniosos.

Sendo assim, foi escolhido o modelo parcimonioso para ser implementado nas previsões de abandono, pois este modelo fornecia bons resultados e maior aplicabilidade.

3.3.4 Definição do Ponto de Corte

Para a construção da lista de alunos com maior probabilidade de abandono, definiu-se como ponto de corte o percentual de 15%, ou seja, foram incluídos 15% dos alunos com maior probabilidade de abandonar no ano letivo em curso. Dois fatores influenciaram diretamente nessa escolha, levando em consideração um maior equilíbrio entre eles:

- *Trade-off* risco/custo – busca por equilíbrio entre incluir o maior número possível de alunos e os custos desta inclusão. Entende-se que optar por atuar com um número maior de alunos pode demandar um custo maior para o programa, associado à dificuldade de dar um tratamento específico para cada aluno para as diferentes possibilidades do protocolo de ações de acordo com os problemas encontrados. No caso do modelo preditor do abandono, buscou-se o equilíbrio entre “custo e benefício”;
- As taxas referentes aos erros de Inclusão e erros de Exclusão obtidos durante os testes realizados. Buscou-se errar menos com relação ao erro de exclusão, para que o aluno não deixe de receber o tratamento relativo ao protocolo de ações, o qual poderia impedir o abandono escolar deste aluno.

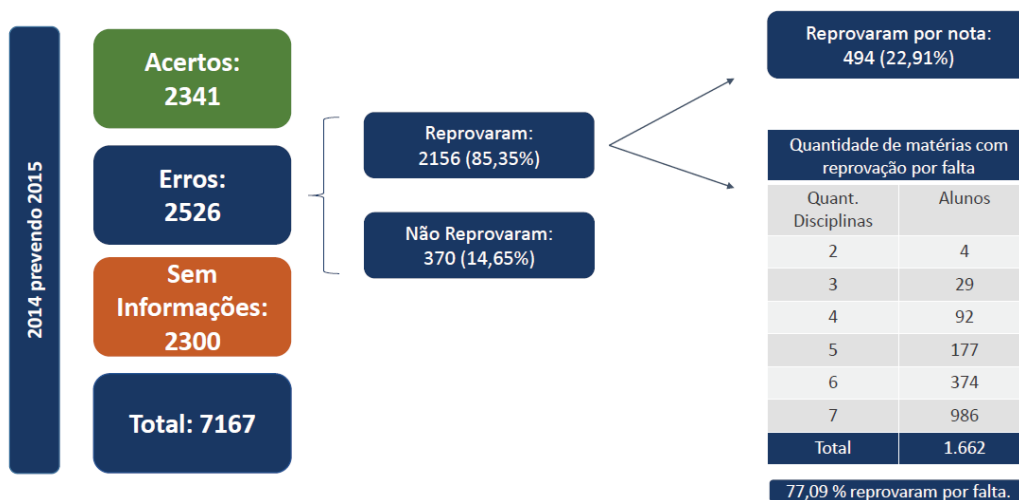
4. RESULTADOS

Nessa seção são apresentados os resultados da previsão para o ano de 2015, 2016 e 2017. Foram consideradas as variáveis e o ponto de corte definidos anteriormente. A partir da estimação dos coeficientes do modelo *logit* com os dados do ano *X*, identificou-se a probabilidade de abandono para cada aluno, ao utilizar os dados referentes ao primeiro trimestre do ano *X+1*. Após o ordenamento decrescente das probabilidades de abandono, foi selecionado o grupo de 15% dos alunos mais propensos ao abandono em cada ano.

4.1 Previsão de abandono para 2015 com coeficientes a partir dos dados de 2014

Para 2015, a lista dos alunos selecionados para receber o protocolo de ações, caso ele estivesse em vigor, totalizou 7.167 indivíduos. Deste total, 2.341 de fato abandonaram a escola, o que significa acerto de previsão. O erro de previsão totalizou o quantitativo de 2.526 alunos que foram previstos como abandono e não abandonaram (figura 5). Para 2.300 casos não há informações do status de matrícula, representando casos que não se enquadram em abandono (abandono, desistente de matrícula e deixou de frequentar) nem como aluno efetivamente matriculado (aprovado, reclassificado, reprovado, transferência do aluno para outro estado, transferência do aluno para outra rede, matrícula cancelada, morte, etc).

Figura 5 – Resultados de previsão para 2015 com os coeficientes de 2014

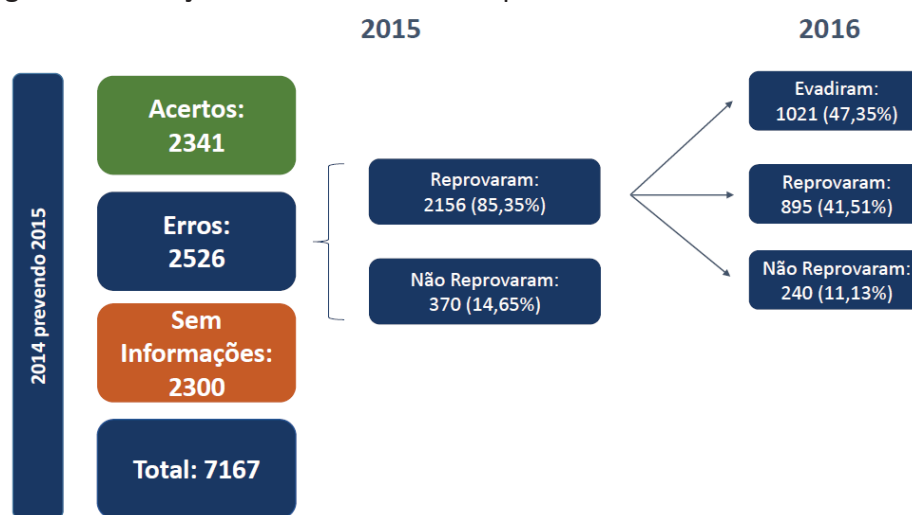


Fonte: Elaboração própria.

O modelo resultou em elevado número de erros de previsão, entretanto, ao acompanhar os alunos ao longo do tempo, quando analisada a situação ao final do período letivo de 2015 e todo o período letivo de 2016, a discussão torna-se mais aprofundada. Dentre os alunos que foram previstos como abandono (Figura 6), mas de fato não abandonaram em 2015, a maior parte reprovou - 85,3%. Destes, 77,1% reprovaram por falta em pelo menos duas disciplinas e 22,9% reprovaram por nota. Este fato demonstra que a inclusão dos alunos que reprovaram por falta no protocolo de ações não representa um desperdício de recursos.

Ao analisar a situação dos 2.156 alunos que reprovaram, mas não abandonaram em 2015, 1.021 (47,3%) evadiram e 895 (41,5%) reprovaram novamente em 2016 (figura 6). Esses dados mostram que apesar de não terem abandonado em 2015, a maioria evadiu no ano seguinte ou enfrentou nova reprovação. Uma vez que a reprovação é um forte indicativo de propensão ao abandono, mesmo que estes alunos não tenham abandonado é relevante que recebam as ações preventivas. As ocorrências verificadas confirmam que a magnitude do erro de inclusão, para a previsão em 2015, não representa uma questão preocupante para o modelo.

Figura 6 – Situação em 2016 dos alunos previstos como abandono em 2015



Fonte: Elaboração própria.

4.2 Previsão de abandono para 2016 com coeficientes a partir dos dados de 2015

A previsão para 2016 resultou em 7.389 alunos previstos para abandonar a escola (figura 7). Deste total, 2.181 abandonaram, 2.708 não abandonaram e para 2.500 não havia informações conclusivas. Do total de falso-positivos, 82,7% (2.241) reprovaram, o que confirma o ocorrido em 2015 mostrando que, embora com um quantitativo de erros considerável, o modelo está prevendo bem os alunos propensos ao abandono. Para o ano de 2016 não foi possível observar o tipo de reprovação, se por nota ou falta, dado a não existência desta informação.

Figura 7 – Resultados de previsão para 2016 com os coeficientes de 2015



Fonte: Elaboração própria.

4.3 Previsão de abandono para 2017 com coeficientes a partir dos dados de 2016

Utilizando a base de 2016 para prever o abandono em 2017, foi obtido um total de 6.540 alunos propensos ao abandono (considerando o ponto de corte de 15%). Deste total, 1.579 realmente abandonaram, 1.409 não abandonaram e 3.552 não apresentaram informações em relação ao status da matrícula (figura 8). Do total de falso-positivos (1.409), 79,1% (1.115) reprovaram, demonstrando a continuidade da tendência ocorrida nos anos anteriores (2015 e 2016), com a maior parcela dos erros ocorrendo para alunos com elevada propensão à reprovação.

Figura 8 – Resultados de previsão para 2017 com os coeficientes de 2016



Fonte: Elaboração própria.

Para avaliar o modelo, foram calculadas as medidas de acurácia discutidas anteriormente considerando como previsão o grupo de alunos que representa os 15% mais propensos ao abandono (tabela 17). O modelo estimado com os coeficientes gerados a partir dos dados de 2016 alcançou proporção de verdadeiro-positivos (sensibilidade) de 65,3% (do total de abandonos ocorridos, 65,3% foram previstos pelo modelo) e proporção de falso-positivos (1 – especificidade) de 4%.

Embora os modelos de predição analisados por Bowers, Sprott and Taff (2013) variem consideravelmente em relação ao nível de ensino analisado, ao tamanho da amostra, à forma como cada indicador é calculado e às taxas de abandono, apenas 27 dos 110 indicadores analisados alcançaram sensibilidade acima de 60%. Verifica-se no modelo apresentado uma sensibilidade consideravelmente alta com uma proporção de falso-negativos muito baixa, mostrando uma boa acurácia do mesmo.

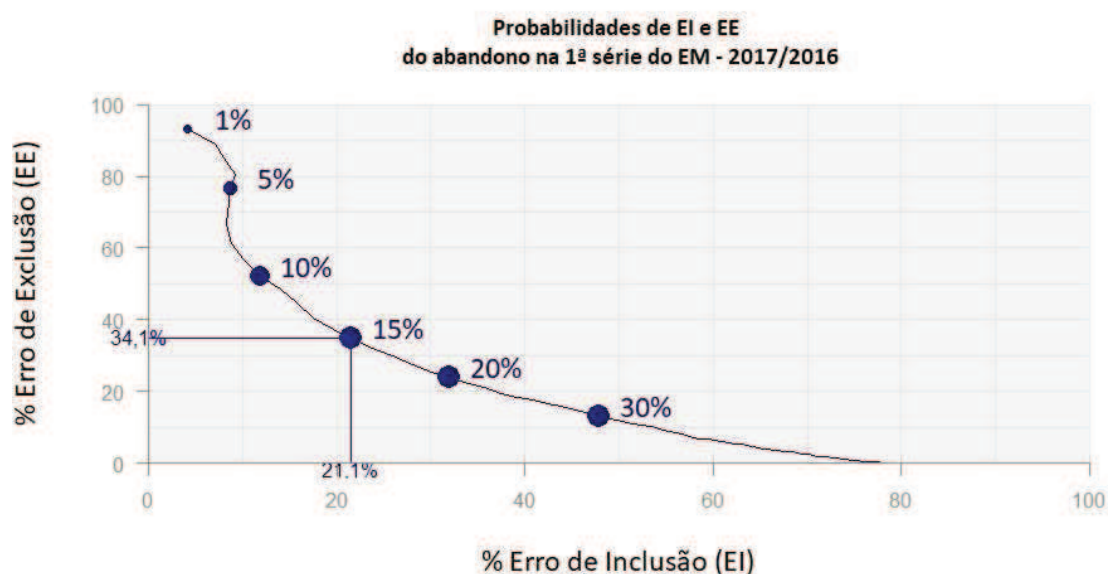
Tabela 17 – Métricas para Avaliação da Predição 2016/2017

Métrica	%
Precisão	24,1
Sensibilidade	65,3
Especificidade	96,0
1-Especificidade	4,0

Fonte: Elaboração própria.

O gráfico 11 apresenta o *trade-off* entre o erro de inclusão (EI) e o erro de exclusão (EE) para o modelo estimado. Busca-se uma previsão que retorne menores EI e EE, considerando também que o número de alunos a ser atendidos seja factível. Para o ponto de corte de 15%, tem-se EI de 21% (do total de abandonos previstos, 21% não abandonaram) e EE 34% (do total de abandonos ocorridos, 34% não foram previstos). Este ponto de corte coloca 6.540 alunos como alvo das ações preventivas. Comparativamente, o ponto de corte de 10% resulta em menor erro de inclusão, mas maior erro de exclusão. Acima de 15%, o erro de inclusão cresce acentuadamente, assim como o número de alunos a serem incluídos nas ações preventivas. Por exemplo, para o ponto de corte de 20%, deveriam ser incluídos 2.024 alunos a mais nas ações. Uma vez que também é necessário ponderar a quantidade de alunos incluídos nas ações para que as escolas não sejam sobrecarregadas, dado que as ações serão planejadas especificamente para cada aluno, confirma-se a melhor adequação do ponto de corte de 15%.

Gráfico 11 – Probabilidades de EI e EE na previsão de abandono para 2017



Fonte: Elaboração própria.

Com os resultados do modelo será possível identificar quem são os alunos com maior propensão ao abandono escolar, selecionar os 15% com maior probabilidade, e esses receberão acompanhamento e tratamento específico, a partir de um protocolo de ações que será construído junto com as instituições de ensino e parceiros do programa.

Tabela 17 – Quantitativo de alunos a ser incluído no programa em 2017 para cada ponto de corte

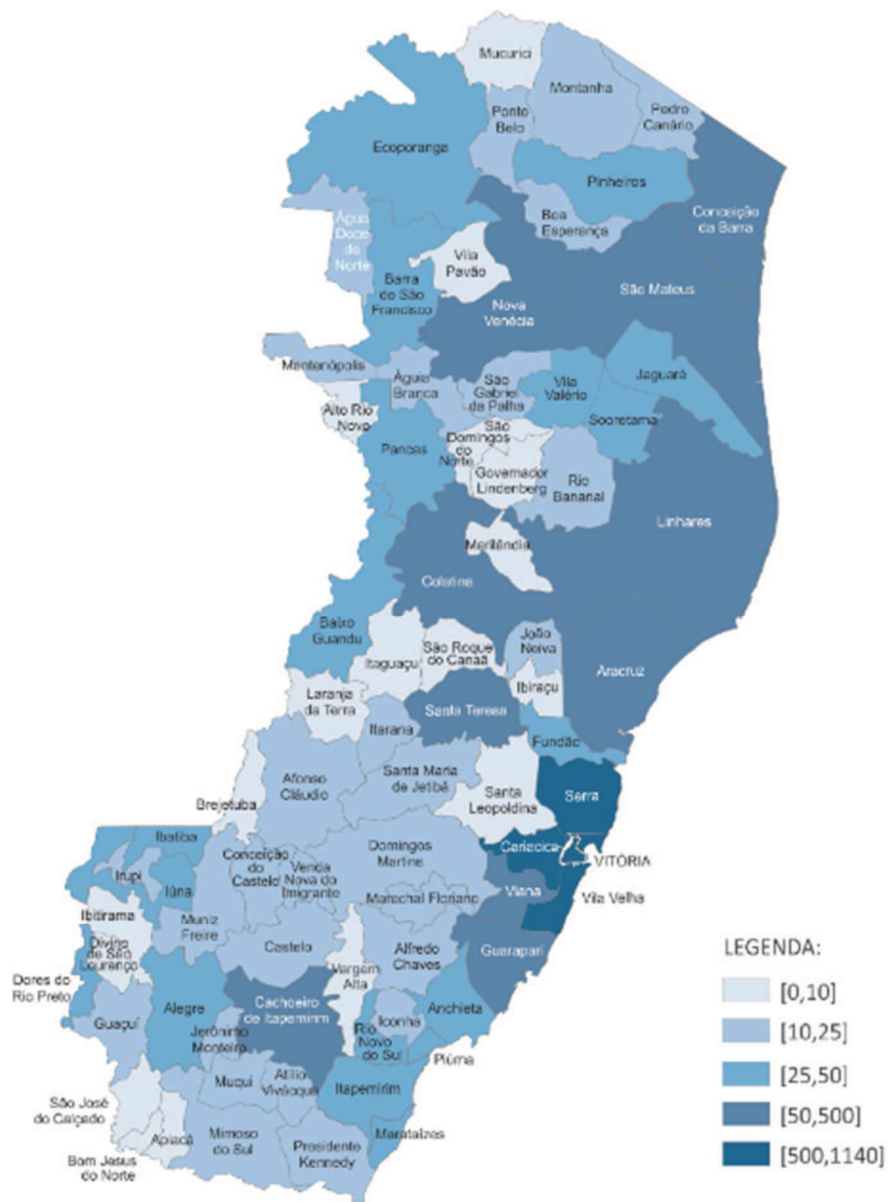
Ponto de Corte	Quantidade de alunos incluídos
1%	436
5%	2.180
10%	4.360
15%	6.540
20%	8.721
30%	13.081
40%	17.441
50%	21.802
60%	26.162
70%	30.522
80%	34.882
90%	39.243
99%	43.167

Fonte: elaboração própria.

Criado a partir dos dados obtidos com o Modelo Preditor, o mapa 1 apresenta os quantitativos de abandono previstos para 2017 para alunos do primeiro ano do ensino Médio da Rede Pública Estadual.

Os quatro municípios do estado do Espírito Santo em que se concentra o maior quantitativo de abandono previsto no ano de 2017 (entre 500 a 1.140 casos) estão localizados na Região Metropolitana, sendo eles, Cariacica, Serra, Vila Velha e Vitória.

Mapa 1 – Abandonos previstos por município do ES em 2017



Fonte: Núcleo de Estudos da Educação/IJSN

5. CONSIDERAÇÕES FINAIS

O propósito deste Texto para Discussão foi apresentar de forma objetiva o processo de construção do “Modelo de Predição do Abandono Escolar” a partir dos dados e informações disponíveis, de forma que possa ser replicado tanto para outras séries do ensino básico quanto para outros estados da Federação.

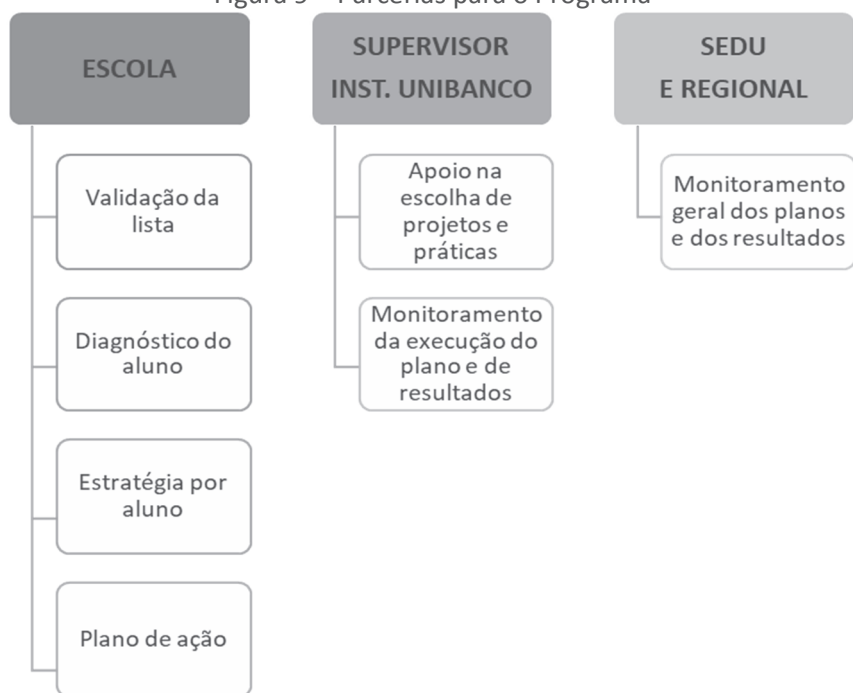
Com base na literatura discutida e no resultado das estimações, considera-se que a inclusão de dados socioeconômicos ao nível do aluno permitiria um melhor ajuste do modelo e melhor capacidade preditiva. Uma sugestão para aprimoramento do modelo preditor é a inclusão dessas variáveis. Ainda, seria possível acompanhar o aluno ao longo do tempo, com a disseminação em toda a base administrativa de dados do código INEP, o qual sempre é o mesmo para cada aluno em todos os anos. O histórico passado do aluno possibilitaria aumentar muito a acurácia em modelos futuros.

Considera-se que a utilização do modelo preditor pela SEDU e pelas escolas estaduais de ensino médio subsidiará a política de redução de abandono ao:

1. permitir a formalização e instrumentalização do processo de inserção dos alunos propensos ao abandono no programa de ações preventivas;
2. promover o acompanhamento da situação dos alunos propensos ao abandono de forma mais fácil e eficaz pela SEDU;
3. possibilitar a avaliação dos resultados de predição e aprimoramento do modelo;
4. possibilitar a avaliação das ações preventivas implementadas junto aos alunos.

Para isso, o projeto contará com o apoio de parceiros na realização das atividades requeridas (figura 9). Às escolas, caberá validar a lista de alunos mais propensos ao abandono e implementar as ações preventivas. O Instituto Unibanco prestará consultoria para escolha de projetos e práticas que visem prevenir o abandono, assim como monitorar a implementação das ações. A SEDU, enquanto gestora da educação no Estado, irá monitorar os planos implementados pelas escolas e os resultados obtidos com o intuito de avaliar e aprimorar a política de prevenção ao abandono escolar na rede pública do Estado.

Figura 9 – Parcerias para o Programa



Fonte: elaboração própria.

Como próximas etapas do projeto, está previsto o teste do Modelo Preditor do Abandono nas escolas participantes do Projeto Jovem de Futuro, do Instituto Unibanco, em que o Supervisor deste referido projeto realizará o papel de Monitoramento, ao acompanhar a aplicação do protocolo de ações e boas práticas nas escolas.

REFERÊNCIAS BIBLIOGRÁFICAS

ADELMAN, Melissa Ann *et al.* **Predicting school dropout with administrative data: new evidence from Guatemala and Honduras.** Policy Research Working Paper, The World Bank, 2017.

ADELMAN, Melissa Ann; SZÉKELY, Miguel. **School Dropout in Central America - An Overview of Trends, Causes, Consequences, and Promising Interventions.** Policy Research Working Paper, The World Bank, 2016.

BARROS, Ricardo Paes de *et al.* **Determinantes do desempenho educacional no Brasil.** Texto para Discussão, n. 834, IPEA, Rio de Janeiro, 2001.

BOWERS, A.J., SPROTT, R., TAFF, S.A. Do we Know Who Will Drop Out? A Review of the Predictors of Dropping out of High School: Precision, Sensitivity and Specificity. **The High School Journal.** 96(2), 2013.

BRAVA. **Políticas públicas para redução do abandono e evasão escolar de jovens.** 2017. Disponível em: <http://gesta.org.br/wp-content/uploads/2017/09/Brava_COMPLETA_V11.pdf>. Acesso em 17 out. 2017.

BURGESS, Simon. Human capital and education: The state of the art in the economics of education. **Institute for the Study of Labour (IZA)** Discussion Paper No. 9885, Bonn, 2016.

CAMERON, A. C.; TRIVEDI, P. K. **Microeconometrics: methods and applications.** Cambridge university press, 2005.

CÁRDENAS, Mauricio; HOYOS, Rafael; SZÉKELY, Miguel. Out-of-school and out-of-work youth in latin america: a persistent problem in a decade of prosperity . **Economía**, v. 16, n. 1, 2015.

ESPÍRITO SANTO. Lei n. 10.382, de 25 de junho de 2015. **Aprova o Plano Estadual de Educação do Espírito Santo – PEE/ES.** Disponível em: http://portal.sedu.es.gov.br:83/PDFs/Leis/LEI%2010382_2015-2025.pdf. Acesso em: 19/07/2018.

FORTIN, Laurier *et al.* Typology of students at risk of dropping out of school: Description by personal, family and school factors. **European Journal of Psychology of Education**, v. 21, n. 4, December, 2006.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Glossário - Conceitos e definições de pesquisas sociais.** 2017.

INSTITUTO JONES DOS SANTOS NEVES. **Relatório de Acompanhamento do Plano Estadual de Educação**. 2017. Disponível em: <http://www.ijsn.es.gov.br/artigos/4827-acompanhamento-do-plano-estadual-de-educacao-do-espirito-santo-pee-2017?highlight=WyJwZWUiXQ==>. Acesso em: 15 set. 2017

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP). **Censo Escolar**. 2014.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP). **Censo Escolar**. 2015.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP). **Censo Escolar**. 2016.

LEON, Fernanda Leite Lopez de; MENEZES-FILHO, Naércio Aquino. **Reprovação, avanço e evasão escolar no Brasil**. Pesquisa e Planejamento Econômico, v.32, n.3, dezembro, 2002.

SECRETÁRIA DA EDUCAÇÃO DO ESTADO DO ESPÍRITO SANTO. **Base de dados do registro administrativo**. 2014 - 2017.

OBSERVATÓRIO DA CRIANÇA E DO ADOLESCENTE. **Taxa de escolarização bruta e líquida no ensino médio**. Disponível em: <<https://observatoriocrianca.org.br/cenario-infancia/temas/ensino-medio/565-taxa-de-escolarizacao-bruta-e-liquida-no-ensino-medio?filters=1,134>>. Acesso em: 11 ago. 2017.

PSACHAROPOULOS, George; PATRINOS, Harry Anthony. **Returns to investment in education: a further update**, Education Economics, v. 12, n. 2, agosto 2004.

UNESCO. **Strong foundations**: Early childhood care and education. Education for All Global Monitoring Report, 2007.

UNESCO. **Education for all 2000-2015**: achievements and challenges. Education for All Global Monitoring Report, 2015.

UNESCO. **Accountability in education**: meeting our commitments. Education for All Global Monitoring Report, 2017.