

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE DE
RIBEIRÃO PRETO
DEPARTAMENTO DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA - ÁREA: ECONOMIA APLICADA

ANTONIO DANIEL RICARDO ENGRACIA CALUZ

O papel das habilidades socioemocionais no fluxo escolar: uma análise do Ensino Médio
brasileiro

Orientador: Prof. Dr. Daniel Domingues
dos Santos

RIBEIRÃO PRETO
2018

Prof. Dr. Vahan Agopyan
Reitor da Universidade de São Paulo

Prof. Dr. Dante Pinheiro Martinelli
Diretor da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto

Prof. Dr. Sérgio Kannebley Junior
Chefe do Departamento de Economia

Prof. Dr. Sergio Naruhiko Sakurai
Coordenador do Programa de Pós-Graduação em Economia - Área: Economia Aplicada

ANTONIO DANIEL RICARDO ENGRACIA CALUZ

O papel das habilidades socioemocionais no fluxo escolar: uma análise do Ensino Médio brasileiro

Dissertação de mestrado desenvolvida no Departamento de Economia da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, mestrado em Economia Aplicada do Programa de Pós-Graduação em Economia, para obtenção do título de Mestre em Ciências. Versão corrigida (a original encontra-se disponível no Serviço de pós-graduação da FEA-RP/USP).

Orientador: Prof. Dr. Daniel Domingues dos Santos

RIBEIRÃO PRETO

2018

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Caluz, Antonio Daniel Ricardo Engracia

O papel das habilidades socioemocionais no fluxo escolar: uma análise do Ensino Médio brasileiro / Antonio Daniel Ricardo Engracia Caluz – Ribeirão Preto, 2018.

50f.: il.; 30 cm

Dissertação de mestrado desenvolvida no Departamento de Economia da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, mestrado em Economia Aplicada do Programa de Pós-Graduação em Economia, para obtenção do título de Mestre em Ciências. Versão corrigida (a original encontra-se disponível no Serviço de pós-graduação da FEA-RP/USP). – Universidade de São Paulo

Orientador: Santos, Daniel Domingues dos

1. Escolhas Educacionais. 2. Evasão Escolar. 3. Capital Humano. 4. Habilidade Socioemocional.

ANTONIO DANIEL RICARDO ENGRACIA CALUZ

O papel das habilidades socioemocionais no fluxo escolar: uma análise do Ensino Médio brasileiro

Dissertação de mestrado desenvolvida no Departamento de Economia da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, mestrado em Economia Aplicada do Programa de Pós-Graduação em Economia, para obtenção do título de Mestre em Ciências. Versão corrigida (a original encontra-se disponível no Serviço de pós-graduação da FEA-RP/USP).

Área de Concentração: Economia Aplicada

Data de Aprovação:

____/____/____

Banca Examinadora:

Prof. Dr. Daniel Domingues dos Santos

Orientador e Presidente da Banca

Professor

Prof. Dr. Luiz Guilherme Dácar da Silva
Scorzafave

Professor

Prof. Dr. Naércio Aquino Menezes Filho

Professor

Prof. Dr. Ricardo Primi

Dedicado a Deus e a todos aqueles que fizeram este momento possível.

AGRADECIMENTOS

Primeiramente, é meu dever agradecer aqueles que participaram efetivamente da elaboração deste trabalho. Meu orientador, Daniel, grande mentor das ideias deste trabalho, a todos do Lepes, sem exceção, desde seus alunos membros, os professores envolvidos, os institutos que apoiam o laboratório e os funcionários e professores da FEA-RP que dão suporte ao funcionamento do grupo de pesquisa dentro da Universidade. O espaço é curto para agradecer a todos que participaram direta ou indiretamente do desenvolvimento deste trabalho, mas devo um agradecimento especial a Fernando, a Vitória e a Duda, pelas intensas ajudas na elaboração e no tratamento dos dados, na redação do texto e, acima de tudo, pelo apoio no dia-a-dia. Formamos uma grande equipe no Lepes e foi um período de muito aprendizagem da minha parte, devo bastante a vocês, mas agradeço principalmente pela amizade de vocês. Aos professores Luiz Guilherme, Elaine e Walter, pelas participações em minhas apresentações, coletando excelentes feedbacks para meu trabalho. Pela ajuda direta, agradeço Rafael, Bruna, Matheus e todas amizades e parcerias que fiz no Lepes.

Agradeço a todos do PPGE seus membros pelo esforço que dispenderam, em especial, ao Jefferson Bertolai, pelas provas extremamente desafiadoras, pelos cursos extremamente planejados e, principalmente, por ser um grande exemplo de um acadêmico de alto nível. Ao Alex e Luciano, foram os grandes responsáveis por eu estar na vida acadêmica, me ensinando muito sobre a ANPEC, sobre a carreira e sobre a curiosidade científica, me incentivando e me encaminhando para atravessar essa dura trajetória. A Elaine devo mais um agradecimento por toda a orientação dada, pelo imenso esforço que teve comigo, paciência e por todos os conselhos que me deu. Não poderia ter tido uma mentora melhor. Ao Daniel, também faço um novo agradecimento, por todos os debates de alto nível. Diversas vezes tive contato e aprendi coisas muito mais avançadas do que o que estava acostumado, sempre fui atrás dos assuntos e termos específicos que você me recomendava. Dentre os professores, finalmente agradeço ao Marcelo Ambrozini, por ser uma grande pessoa e um grande acadêmico que tive o prazer de ter como orientador. Entre outros, agradeço muito ao Vinicius, poder compartilhar dessa ambição de fazer o mestrado, da ambição de ter uma carreira acadêmica bem sucedida, me possibilitando ter um grande amigos. Lembro de diversos momentos que passamos nos esforçando, nunca chegamos despreparados em uma prova, e isso foi essencial para o meu (e sei que foi para o seu) desenvolvimento acadêmico. Agradeço meus amigos da vida, minha namorada e minha família, pelo sacrifício que fizeram por mim, reconheço cada um de vocês como parte deste título.

Agradeço aos amigos da ECEC VI, minha turma de graduação, a minha turma de mestrado do PPGE, todos alunos que tive, amigos que fiz na faculdade, entre eles, meu irmão da vida, que fiz no apto 101. Desenvolvi um amor incondicional pela toda FEA - Ribeirão Preto, espero que o indigerível fim deste vínculo seja breve e que eu possa contribuir muito ainda com o lugar que tanto me ensinou. Enfim, muito obrigado FEA-RP.

"No que diz respeito ao desempenho, ao compromisso, ao esforço, não existe meio termo. Ou você faz uma coisa bem feita ou não faz."
-Ayrton Senna,

RESUMO

CALUZ, A.D.R.E.. **O papel das habilidades socioemocionais no fluxo escolar: uma análise do Ensino Médio brasileiro**. Dissertação (Mestrado). Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2018.

O fluxo escolar brasileiro representa um problema crônico para o caso do Ensino Médio do país. A literatura mostra que o retorno do investimento escolar é atrativo no país, tanto por apresentar uma taxa média alta, como pelo fato de que o adicional de salário devido aos níveis educacionais mais altos são maiores do que nos estágios iniciais da educação, i.e., o retorno educacional brasileiro aparenta ser crescente e convexo, diferente do que se apresenta na literatura internacional. A explicação usual para a evasão se dá através das restrições orçamentárias e de crédito enfrentadas pelas famílias que, sendo restritas no acesso ao crédito, poderia fazer com que o jovem saísse da escola precocemente, mesmo que o aluno esperasse um salário futuro maior. Somado a este retorno atrativo da educação, o país expandiu abruptamente os gastos educacionais. Porém, apesar da expansão, a escolaridade e as medidas de fluxo no país não reagiram proporcionalmente, despertando, assim, a atenção da literatura para explicação desse *puzzle*. Apesar destes fatos, a literatura internacional avançou no sentido de mostrar dois fatos que auxiliam na investigação dessa questão: em primeiro lugar, o retorno da educação pode variar entre indivíduos, ainda que a média seja alta. Por exemplo, indivíduos com maior aptidão podem ser os que se beneficiam mais de uma escolaridade maior, explicando o motivo de alguns abandonarem a escola. Em segundo lugar, a literatura avançou em mostrar que um fator importante na previsão de resultados escolares são habilidades não-cognitivas, como as habilidades socioemocionais. Portanto, este presente trabalho buscou explorar uma coleta de dados realizada em Sertãozinho - SP, em 2008, 2012 e 2017, em que estão disponíveis dados socioemocionais dos estudantes, além de dados demográficos e cognitivos, de estudantes que estavam no segundo ano do Ensino Fundamental em 2008, e em 2017 idealmente estariam no Ensino Médio, possibilitando investigar se existe uma associação entre características socioemocionais e o fluxo escolar. Os resultados indicam que tais fatores têm poder preditivo relevante na explicação do fluxo escolar brasileiro, medidos pela probabilidade de os indivíduos permanecerem estudando e pela probabilidade de se atingir o Ensino Médio em 2017, sendo que a Conscienciosidade e a Amabilidade do estudante aumentam a chance do aluno persistir estudando, enquanto que a Extroversão reduz essa probabilidade, em linha com algumas evidências da literatura. Os resultados trazem como contribuição uma evidência empírica inicial acerca da associação entre habilidades não-cognitivas e o fluxo escolar brasileiro.

Palavras-chave: Escolhas Educacionais, Evasão Escolar, Capital Humano, Habilidade Socioemocional

JEL: C35, I20, J24.

ABSTRACT

CALUZ, A.D.R.E.. **The role of socio emotional skills on the school flow: an analysis of Brazilian high-school**. Dissertation (Master). Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2018.

The Brazilian school flow represents a chronic problem for the country's high school case. The literature shows that the return on school investment is attractive in the country, both for having a high average rate, and for the fact that the additional salary due to higher education levels is higher than in the early stages of education, ie, The Brazilian educational return appears to be growing and convex, different from what is presented in the international literature. The usual explanation for avoidance is through budget and credit constraints faced by families who, being restricted in access to credit, could cause the young person to leave school early, even if the student expects a larger future salary. Added to this attractive return of education, the country abruptly expanded educational spending. However, despite the expansion, the schooling and flow measures in the country did not react proportionally, thus awakening the attention of the literature to explain this textit puzzle. In spite of these facts, the international literature has advanced to show two facts that help in the investigation of this question: firstly, the return of education can vary among individuals, even if the average is high. For example, individuals with higher aptitude may be those who benefit most from higher schooling, explaining why some drop out of school. Second, the literature has advanced in showing that an important factor in predicting school outcomes are non-cognitive skills, such as social-emotional skills. Therefore, this study aimed to explore a data collection held in Sertãozinho - SP, in 2008, 2012 and 2017, in which are available socioemotional data of the students, as well as demographic and cognitive data, of students who were in the second year of Elementary Education in 2008, and by 2017 would ideally be in High School, making it possible to investigate if there is an association between socioemotional characteristics and the school flow. The results indicate that such factors have a relevant predictive power in the explanation of the Brazilian school flow, measured by the probability of individuals remaining studying and by the probability of reaching high school in 2017, and the student's Conscientiousness and Kindness increase the chance of while the Extroversion reduces this probability, in line with some evidence in the literature. The results contribute as an initial empirical evidence about the association between non-cognitive abilities and the Brazilian school flow.

Keywords: Educational Choices, School Dropout, Human Capital, Socioemotional Skill

JEL: C35, I26, J24.

LISTA DE ILUSTRAÇÕES

Figura 1 – Investimento Público real por estudante e a taxa de matrícula na Escola e no Ensino Médio dos jovens 15 a 17 anos de idade	14
Figura 2 – Estrutura dos dados	26

LISTA DE TABELAS

Tabela 1 – Distribuição da série da amostra em 2012	33
Tabela 2 – Estatísticas descritivas do perfil socioeconômico dos estudantes de Sertãozinho em 2012	34
Tabela 3 – Estimação do Probit para probabilidade de se atingir o Ensino Médio com variáveis Socioemocionais	35
Tabela 4 – Estimação do Probit para a permanência escolar com variáveis Socioemocionais	36
Tabela 5 – Estimação do Probit para probabilidade de permanência na escola utilizando o Locus de Controle interno	38
Tabela 6 – Estimação do Probit para a Probabilidade de se atingir o Ensino Médio utilizando o Locus de Controle interno	39
Tabela 7 – Estimação do Probit para probabilidade de se atingir o Ensino Médio utilizando as Habilidades Sociais	39
Tabela 8 – Estimação do Probit para a permanência escolar utilizando as Habilidades Sociais	40
Tabela 9 – Regressão da permanência escolar pelo método Heckprobit com variáveis socioemocionais e Locus de Controle	41
Tabela 10 – Regressão da permanência escolar pelo método Heckprobit com Habilidades Sociais	42
Tabela 11 – Estimação do Probit para probabilidade de se atingir o Ensino Médio na sub-amostra do quinto ano do Ensino Fundamental de 2012	49
Tabela 12 – Estimação do Probit para a probabilidade de atingir o Ensino Médio na sub-amostra do sexto ano do Ensino Fundamental de 2012	49
Tabela 13 – Estimação do Probit para a permanência escolar na sub-amostra do quinto ano do Ensino Fundamental de 2012	50
Tabela 14 – Estimação do Probit para a permanência escolar na sub-amostra do sexto ano do Ensino Fundamental de 2012	50

SUMÁRIO

	INTRODUÇÃO	13
1	REVISÃO DA LITERATURA	17
1.1	O Retorno da Educação	17
1.2	Evidências empíricas sobre o retorno da educação	19
1.3	A evasão escolar no Brasil	21
1.4	Habilidades não cognitivas	23
2	METODOLOGIA	25
2.1	Dados	25
2.2	Estratégia Econométrica	27
2.3	Controlando para o viés de Seleção	28
3	RESULTADOS	32
3.1	Descritivas	32
3.2	Resultados do modelo Probit	35
3.3	Contornando o viés de seleção amostral utilizando dados de 2008	41
4	CONCLUSÃO	44
	REFERÊNCIAS	46
	APÊNDICE A – ESTIMAÇÕES DO PROBIT EM SUB-AMOSTRAS	49

INTRODUÇÃO

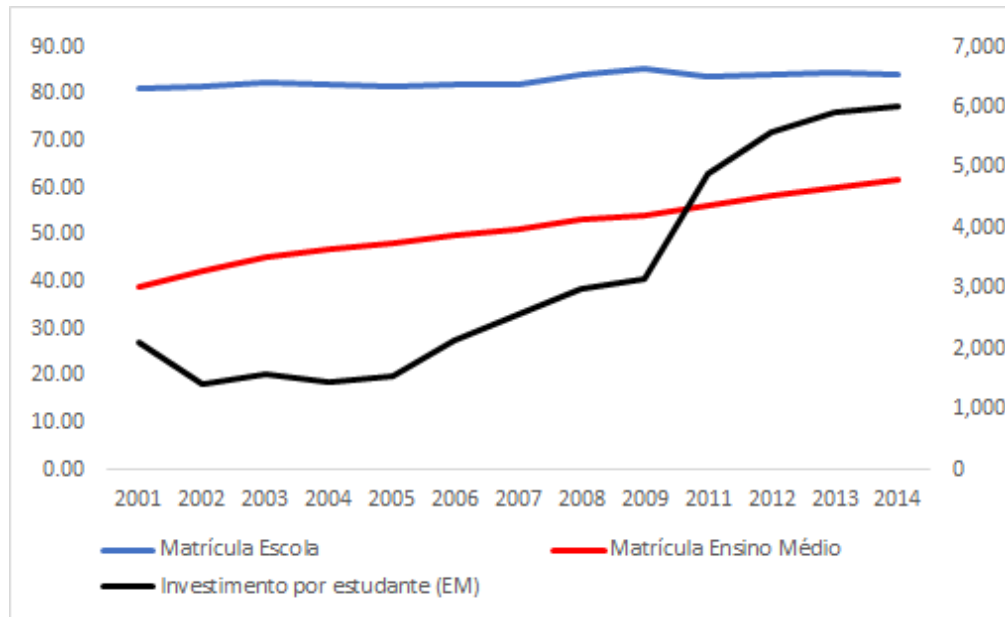
Durante a década de 2000 e o início da década de 2010 o Brasil atingiu um patamar elevado de frequência escolar no ensino fundamental, que passou de aproximadamente 84% de frequência para 98%, mostrando que o país evoluiu positivamente em um dos maiores problemas de sua educação, que é a inclusão educacional. Porém, essa universalização do ensino não foi acompanhada pelo Ensino Médio do país. De acordo com dados do IPEA (Instituto de Pesquisa Econômica Aplicada) e da PNAD, desde os anos 2000, a frequência escolar no ensino médio dos jovens de 15 a 17 anos – que são os que idealmente deveriam estar no ensino médio – é de aproximadamente 61%, próximo da porcentagem de jovens com 19 anos de idade que concluíram o ensino médio, pouco abaixo dos 60%.

O problema da evasão escolar em si é muito estudado na literatura em diversos países, mas no Brasil ainda existem questionamentos e atrativos científicos em torno dessa questão. Na literatura clássica dos anos 60, iniciada com Mincer (1958) e Becker (1964), entende-se que a evasão decorria principalmente das restrições de crédito enfrentadas pela família, indicando que, ainda que o retorno da educação fosse alto, não eram todos os indivíduos que possuíam condições de se aproveitar desse retorno. Assim, políticas públicas voltadas ao alívio financeiro das famílias economicamente vulneráveis pareciam ser formas eficazes para aumentar a escolaridade e, por conseguinte, a produtividade nacional.

A despeito do problema da restrição creditícia das famílias, a literatura internacional acumulou evidências nas últimas décadas da importância de que o capital humano era maleável, ou seja, apesar de existir um componente inato, o aprendizado era importante para o acúmulo de capital humano. Esse capital humano determina, em última instância, o retorno de um ano adicional no mercado de trabalho, que a priori a literatura fazia essa hipótese de que esse retorno era homogêneo entre os indivíduos. Mais recentemente, essa literatura chamou a atenção para a importância da existência de heterogeneidade dos retornos. Assim, mesmo que o retorno médio da educação seja alto é possível que a decisão de continuar ou não estudando pode não ser ótimo para alguns indivíduos, mesmo para indivíduos de famílias não vulneráveis (e que, a princípio, não sofreriam com a restrição de crédito para continuar estudando). Além disso, a partir dessa maleabilidade do capital humano, a literatura abriu espaço para entender como se dá a formação de capital humano nos indivíduos. É nesse ponto que surge a importância de fatores não cognitivos para o desenvolvimento (HECKMAN; RUBINSTEIN, 2001).

Adicionalmente, o caso brasileiro ainda apresentou um fato intrigante para a literatura tradicional do capital humano. Em particular, diversos estudos mostram que, no país, o retorno da educação é especialmente alto (BARROS et al., 2001; LEON; MENEZES-FILHO, 2002; NERI et al., 2015), sendo crescente e convexo na educação (FERREIRA; BARROS, 1999; FERREIRA; LITCHFIELD, 2001), i.e., níveis educacionais mais altos tendem a trazer um adicional de salário futuro maior do que os primeiros estágios da educação no Brasil. Assim, o

Figura 1 – Investimento Público real por estudante e a taxa de matrícula na Escola e no Ensino Médio dos jovens 15 a 17 anos de idade



Fonte – Elaboração própria. Dados disponibilizados pelo site do INEP. Investimento atualizado para valores de 2014.

cenário brasileiro apresenta, em conjunto com a alta taxa de evasão escolar, um retorno do investimento educacional atrativo. O que desperta ainda mais a atenção da literatura para este cenário é o abrupto aumento dos gastos governamentais com a educação. Ao aliviar a restrição orçamentária e creditícia das famílias, era de se esperar que mais indivíduos cursassem e completassem o Ensino Médio. Como resultado, deveria ocorrer uma resposta na escolaridade média e nas medidas de fluxo escolar do país que correspondesse a esse aumento dos gastos educacionais. Porém, em geral, isso não ocorreu, gerando um fato estilizado na literatura. A taxa de conclusão do Ensino Médio continuou baixa, assim como a taxa de matrícula, ainda mais quando comparada ao Ensino Fundamental. O gráfico 1 confirma tais informações, mostrando o aumento do investimento educacional real por estudante, mostrando que a taxa de matrícula no Ensino Médio apresentou uma melhora ligeira, reagindo pouco ao aumento dos gastos, sendo que a taxa de matrícula na escola, como um todo, ficou estagnada.

Com os desafios apresentados à literatura tradicional no caso brasileiro de evasão escolar, e com a literatura internacional cada vez mais apontando para a importância de fatores não cognitivos para a evasão escolar, se torna imprescindível produzir uma evidência mais completa do problema da evasão escolar incorporando a importância das habilidades não cognitivas. No país, da maneira pioneira na literatura nacional (e com dados nacionais) de Economia da Educação, pode-se estudar empiricamente como esses fatores socioemocionais afetam a evasão escolar, devido a duas coletas realizadas na cidade de Sertãozinho - SP, que possui um conjunto dessas características, que podem ser estudados em conjunto com dados socioeconômicos e de desempenho escolar presentes nas mesmas coletas. Esta oportunidade de estudar a evasão com

fatores cognitivos e não cognitivos em conjunto pode gerar resultados importantes para o direcionamento de políticas públicas voltadas ao Ensino Médio, visto que, como demonstrado por Heckman e Cameron (2001), em geral políticas voltadas a alívios de restrições de curto prazo são pouco eficazes em relação a políticas voltadas para a melhoria da renda de longo prazo dos indivíduos (que são políticas voltadas a alterar de fato os atributos dos estudantes).

Sendo assim, dado o potencial de contribuições de uma análise do papel das habilidades socioemocionais nas escolhas educacionais, o objetivo central da pesquisa proposta foi investigar os determinantes da permanência na escola para o Ensino Médio brasileiro a partir de características não-cognitivas, visando explorar a lacuna existente na atual literatura, que é insuficiente para explicar a estagnação da desigualdade de conclusão do ensino médio entre famílias mais ricas e mais pobres. Para isso, a pesquisa se valeu de elementos da teoria do Capital Humano apresentada Cunha, Heckman e Schennach (2010), verificando como habilidades cognitivas se combinam com as não cognitivas na explicação da permanência da escola. Ou seja, se a hipótese de que habilidades não cognitivas se complementam com as cognitivas para formar o capital humano de um indivíduo, deve-se esperar que o perfil socioemocional dos indivíduos e outras habilidades não cognitivas tenham poder preditivo sobre o retorno que um ano adicional de educação para seus resultados futuros e, portanto, sobre a evasão escolar.

Para tal investigação foi elaborado um modelo empírico de escolha discreta, avaliando como se dá a decisão educacional na passagem do Ensino Fundamental para o Ensino Médio para o segundo ano, seguindo a ideia de Heckman e Cameron (2001). O modelo segue uma escolha binária para a escolha de continuar ou não estudando, em que as variáveis explicativas utilizadas são fatores socioeconômicos, socioemocionais e habilidade cognitiva do aluno. Para aplicação da metodologia, foram unidas as três bases de dados já existentes provenientes de coleta de dados realizada na cidade de Sertãozinho – SP, em 2008 e 2012, sendo que em 2012 foram aplicados questionários psicológicos para mensurar habilidades sociais e características de traços de personalidade (Big Five Inventory), além de perguntas no questionário que permitem a mensuração de habilidades sociais (social skills) e de lócus de controle. Também se levantou, junto às secretarias educacionais de Ribeirão Preto¹ e Sertãozinho, as informações de quais alunos ainda faziam parte do sistema educacional no ano de 2017, ano no qual os alunos idealmente deveriam estar no Ensino Médio. Com isso, é possível inferir quais alunos evadiram o sistema educacional e quais permaneceram.²

O estudo partiu de duas hipóteses das quais os canais das habilidades socioemocionais poderiam influenciar a permanência na escola. Em primeiro lugar, a perseverança e responsabilidade, medidas pela Conscienciosidade, disponível nos dados deste trabalho, poderia predizer positivamente a permanência na escola. Em segundo lugar, em uma perspectiva distinta e me-

¹ Essa cidade foi consultada por ser a maior cidade em volta de Sertãozinho e, de fato, diversos alunos estudavam na cidade.

² De fato, mesmo que o aluno tenha reprovado, conseguimos obter a informação se ele constava no sistema educacional (e em qual série estava), por isso conseguimos identificar a evasão, não apenas progressão.

nos relacionada ao desempenho escolar, os jovens com maior sentimento de pertencimento ao grupo de colegas ou ao ambiente escolar em geral, poderiam fazer com que indivíduos mais extrovertidos e amáveis tivessem maior probabilidade de permanecer na escola. Em ambos os casos, mesmo com retorno alto da educação para suas características cognitivas, o retorno de um ano adicional de educação pode ser substancialmente menor em caso de habilidades socioemocionais piores.³ Se a hipótese for verdadeira, mesmo com restrição de crédito não representando um grande problema na família e mesmo com uma alta habilidade cognitiva, pode ser ótimo para o indivíduo evadir a escola, dado que pode existir uma complementariedade entre características socioemocionais e cognitivas.

A principal contribuição do trabalho foi mostrar que, entre os traços de personalidade, a Conscienciosidade e a Amabilidade se mostram habilidades cognitivas associadas a medidas de fluxo escolar, juntamente com a Extroversão que, neste caso, se mostrou um determinante negativamente associado ao fluxo do estudante na escola. Os resultados são mais condizentes com uma explicação que aponte para a importância do sentimento de pertencimento ao grupo para a permanência na escola. Quanto ao Locus de Controle, observou-se que quanto mais interno é essa habilidade no indivíduo, maior é a probabilidade de permanência na escola. Não foram encontrados resultados para as habilidades sociais.

Este trabalho está organizado da seguinte forma: na seção seguinte será apresentada a revisão literária, na terceira seção será apresentada a Metodologia, detalhando as coletas de dados que foram utilizadas e a escolha das variáveis socioemocionais, caracterizando-as; por fim, será realizada a Análise dos dados, para chegar a conclusão do trabalho.

³ Note que não há um consenso na literatura sobre a possibilidade de mais habilidades socioemocionais serem sempre melhores para resultados escolares.

1 REVISÃO DA LITERATURA

1.1 O Retorno da Educação

A literatura internacional que trata da decisão da escolaridade ótima ganhou força com os trabalhos de Mincer (1958), Mincer (1974), Becker (1964) e Becker e Chiswick (1966). Mincer com a formulação da equação que leva seu nome tinha um grande interesse em entender o retorno da escolaridade no mercado de trabalho, ou seja, considerando o ano adicional de estudo como um investimento em educação **qual é o retorno desse investimento em educação no mercado de trabalho?** Embora os dois trabalhos de Mincer impliquem na mesma formulação como resultado final - $\ln(w(e, x)) = \beta_0 + \beta_1 \cdot e + \beta_2 \cdot x + \beta_3 \cdot x^2 + \varepsilon$ ¹ - a motivação teórica de ambos os trabalhos era distinta.

No seu primeiro trabalho, Mincer (1958) parte da ideia de diferenciais compensatórios para entender o motivo pelo qual pessoas com diferentes escolaridades ganham salários diferentes. Em um contexto de mercado de crédito perfeito e sem incerteza, a ideia de diferenciais compensatórios diz basicamente que, em equilíbrio, o valor presente dos ganhos nos fluxos salariais futuros deve igualar o valor que o indivíduo abre mão de ganhar para estudar um ano a mais. Ou seja, o *trade off* que o indivíduo enfrenta está em parar de estudar para a iniciar seus ganhos laborais hoje ou estudar um ano a mais para obter rendimentos laborais maiores. Heckman, Lochner e Todd (2003) discutem três das principais implicações do trabalho de Mincer: o coeficiente do salário associado a escolaridade converge para a taxa de juros quando o tempo de vida é longo²; Pessoas com mais educação ganham um salário maior; a taxa de retorno por pessoa é crescente tanto na taxa de juros quando na idade de aposentadoria.

Em seu segundo trabalho, Mincer (1974) motiva a modelagem da escolaridade ótima por premissas distintas do estudo citado acima. A ideia importada dos estudos de Becker (1964) e Becker e Chiswick (1966) consolida um modelo de identidade contábil. Neste trabalho o modelo deixa claro o porquê do coeficiente da educação em uma regressão de salário ser interpretado como uma taxa de retorno de investimento. Partindo da identidade de que para cada instante do tempo o investimento educacional é igual ao produto entre o potencial de ganhos investidos e a fração efetivamente investida no instante, e utilizando a experiência de modo similar a Ben-Porath (1967), o autor encontra a mesma equação estimável, havendo duas possibilidades: a primeira é que o intercepto e a inclinação referente a educação são idênticas entre os indivíduos; a segunda flexibiliza isso, permitindo alguma heterogeneidade entre os efeitos para os indivíduos, o que com algumas simplificações acaba por desaparecer do modelo. Ou seja, já se vislumbrava a ideia de retornos heterogêneos para a educação, embora ela não tenha

¹ w é o salário do indivíduo, que depende da escolaridade e e da experiência x .

² O tempo T de vida pode ser interpretado como a idade de aposentadoria. Adicionalmente, no modelo, uma parte dos ganhos salariais que dependem da escolaridade dependem também de T , e essa parcela converge para zero conforme T converge para infinito.

sido efetivamente tratada.

Essa literatura é a base para entendimento do valor da educação e, portanto, da escolha da escolaridade por parte dos indivíduos. Assim, diversos ramos da literatura procuraram estender as ideias geradas por Becker (1964) e Mincer (1974). Os estudos nessa linha foram, então, buscando entender outras fontes pelas quais a educação pode trazer valor aos indivíduos, ligada às incertezas relacionadas ao futuro e a assimetria de informação presente no mercado de trabalho. Um marco teórico nesse sentido está em Spence (1973), que mostra que a educação pode ser utilizada como sinal para o mercado a respeito da produtividade do indivíduo, o que, em última instância, determina seu salário. Em um modelo simples, supõe-se que o custo da educação é maior para o indivíduo menos produtivo, o que pode gerar um equilíbrio separador, com o mercado identificando os indivíduos com maior produtividade, com um salário correspondente a sua produtividade marginal.

Uma parte do valor da educação pode ser explicada pela existência de uma “incerteza sequencial” nos momentos em que são tomadas as decisões educacionais. Ou seja, ainda que os indivíduos e famílias formem expectativas acerca do retorno de estudar um ano a mais, existe uma incerteza decorrida da temporalidade em que os anos de educação estão sendo tomados e o momento de ingresso no mercado de trabalho. Devido a este fato, a literatura evoluiu no sentido de incorporar os efeitos da incerteza, indicando que além do efeito sinalização da educação, ele pode prover um efeito credenciamento, que é entendida como o valor de opção da educação, análogo a uma opção no mercado de capitais. Ao continuar estudando, o indivíduo mantém a possibilidade futura de competir no mercado de trabalho por salários correspondente aos indivíduos mais escolarizados. Ao não ter a certeza de qual salário vigorará para indivíduos mais escolarizados no futuro, parte do valor atribuído ao ano adicional de educação se dá pelo credenciamento a este salário. Com isso, um componente do valor educacional segue uma fórmula análoga ao do preço de uma opção (WEISBROD, 1962; COMAY; MELNIK; POLLATSCHEK, 1973)³.

³ Heckman, Lochner e Tood (2003) promovem uma discussão profunda da fundamentação teórica para o valor da opção da educação. Porém, empiricamente, este componente não se mostra tão relevante.

1.2 Evidências empíricas sobre o retorno da educação

A estimação da equação de salário, utilizando a escolaridade, foi elaborada por diversos estudos a partir dos artigos pioneiros utilizando a equação *minceriana*, trazendo consigo diversas evidências empíricas e críticas acerca dessas estimativas. A literatura se concentrou em duas questões: como eliminar o viés de variável omitida na estimação do retorno da escolaridade; e como os retornos são heterogêneos nas características dos indivíduos. Apesar dos problemas econométricos, evidências internacionais sugerem que o retorno do investimento educacional é alto relativamente a outros investimentos (HECKMAN; LOCHNER; TODD, 2003; CARD, 1994; CARNEIRO, 2003).

Assim, apesar dos retornos educacionais se mostrarem altos na literatura, os números devem ser vistos com cuidado. Ao se estimar uma equação de salário, como apresentada no tópico anterior, pelo tradicional Método de Mínimos Quadrados Ordinários, percebe-se que um componente importante da determinação do salário está omitido da equação: a aptidão. Esse componente não observável pode ser correlacionado com a educação - como sugere, por exemplo, o modelo teórico de (SPENCE, 1973) - nesse caso, as estimativas seriam enviesadas e a correção desse viés de aptidão, por si só, deu origem a vários estudos sumarizados por Card (1994) e Heckman, Lochner e Todd (2003).

Em geral, nos dados, é possível observar o *background* familiar, como a educação da mãe, e utilizar essa informação para corrigir o viés de aptidão. Card (1995) discute que essa variável já foi utilizada na literatura tanto como uma *proxy* da aptidão do indivíduo, como variável instrumental para a educação, método que visa corrigir alguns tipos de endogeneidade em econometria. Outros tipos de variáveis instrumentais também já foram testadas na literatura e, em geral, são relacionados a fatores institucionais que promovem um efeito exógeno sobre a decisão de escolaridade. Ou seja, a literatura empírica se esforçou em entender o quanto os indivíduos esperam ter de retorno, marginalmente, ao estudar um ano a mais, pois, em última instância, é este retorno que determina a decisão de seguir ou não estudando.

As estimações realizadas a partir do método de variáveis instrumentais (VI) funcionam sob condições que nem sempre são válidas, como a exogeneidade da variável instrumental escolhida. Somando a este fato, Carneiro (2003) mostra que o retorno da escolaridade pode ser heterogêneo entre os diferentes tipos de indivíduos, sendo assim, uma variável instrumental não estimaria a média dos retornos marginais dessa escolaridade, ela estimaria um retorno marginal apenas para o grupo movido pelo instrumento, e isso poderia amplificar o viés das estimativas, em relação as estimações advindas do Método de Mínimos Quadrados. Ou seja, o estimador de VI para esse retorno pode ser ainda mais enviesado do que o estimador de MQO, visto que sofrerá do viés de variável omitida caso a variável instrumental escolhida não satisfaça as condições de um bom instrumento e a ainda estará estimando o efeito sobre um grupo específico. Apesar de todos os problemas na estimação da Equação de Mincer, os possíveis vieses não alteram drasticamente a estimativa do retorno da escolaridade (CARD, 1994). Adicionalmente,

outras metodologias, como o cálculo da Taxa Interna de Retorno do investimento escolar, podem ser utilizadas para o cálculo deste retorno. Essa diversidade de tipos de estudo mostra a quão amplo e desafiador pode ser o estudo da escolaridade dos indivíduos, gerando diversos questionamentos teóricos e metodológicos.

Apesar de todos os desafios metodológicos, as evidências internacionais sugerem que o retorno da educação é alto e ainda mostram que a taxa de retorno da educação pode depender de quão cedo é realizado o investimento em educação. Quanto mais jovem é o indivíduo maior tende a ser o benefício futuro do investimento presente na escolaridade. Porém, no Brasil, esse retorno é especialmente alto e atrativo frente aos outros países. Ueda e Hoffman estimam pelo método de VI o retorno da educação em torno de 10%, enquanto Sachsida, Loureiro e Mendonça (2004) fazem uma diversificação de métodos com dados da PNAD, usando MQO, VI e método de seleção de Heckman, obtém estimativas entre 10% e 20% para o retorno da educação. Também utilizando o método de VI, com dados da PNAD de 1997 e 2007, Teixeira e Menezes-Filho (2012) obtém coeficientes entre 5,5% e 12% para o retorno da educação. Barbosa e Pessoa utilizaram a Taxa Interna de Retorno do investimento educacional⁴, pois, como argumenta Moura (2008), o coeficiente da equação minceriana só é equivalente a essa taxa de retorno sob hipóteses nem sempre válidas, encontrando uma taxa superior a 15% a.a.⁵.

Ou seja, o retorno da educação no Brasil tem um valor especialmente alto, contando ainda com uma atratividade que a educação oferece nacionalmente e em outros países em desenvolvimento: o retorno aparenta ser crescente e convexo na educação (FERREIRA; BARROS, 1999; FERREIRA; LITCHFIELD, 2001). Ou seja, enquanto as evidências internacionais sugerem que os anos iniciais de educação são os que trazem o maior ganho futuro para o beneficiado, no Brasil, os anos finais de escolaridade trazem um acréscimo nos rendimentos futuros maior do que o acréscimo devido a educação inicial. E é essa evidência que torna o problema de fluxo no Brasil um *puzzle*. Somado a expansão da gratuidade da educação e do alívio das restrições orçamentárias das famílias mais pobres, essa atratividade do retorno educacionais poderia ser um fator que estimulasse os indivíduos mais necessitados a permanecer na escola. Ainda assim, essa o problema de fluxo no país é grave e, como será mostrado no próximo tópico, uma parcela desse problema ainda pode ser atribuída a fatores socioeconômicos.

⁴ Taxa que iguala o custo de um ano adicional de educação ao valor presente dos ganhos salariais futuros.

⁵ Moura (2008) testa a linearidade da equação nos anos de estudo e paralelismo, ou seja, separabilidade entre os anos de estudo e experiência, o que implica que o ganho marginal advindo da experiência não depende da escolaridade. O autor rejeita as hipóteses.

1.3 A evasão escolar no Brasil

O tema do fluxo escolar tem sido explorado e discutido no Brasil por se tratar de um fenômeno importante e relevante (NERI et al., 2009), sendo um dos principais problemas educacionais atuais no país. Nesta subseção pretende-se explorar os trabalhos que exploraram nacionalmente este fenômeno, elencando os principais fatores que corroboram para tal.

Um dos primeiros estudos empíricos que buscam entender quais são os fatores determinantes da evasão escolar no país advém da pesquisa de Neri et al. (2000). O trabalho teve como objetivo principal verificar se a mudança no status laboral do pai leva a criança a evasão escolar com o intuito de começar a trabalhar. A partir da base de dados de Pesquisa Mensal de Emprego (PME) do IBGE, os autores utilizaram uma metodologia não linear baseada em um modelo *logit* binomial, obtendo como resultados que não há uma relação significativa entre o pai ficar desempregado e a criança entrar para o mercado de trabalho, porém há uma relação positiva entre o pai ficar desempregado e a criança começar a trabalhar dentro de casa.

Além disso, o trabalho encontrou que características da criança como gênero e idade apresentaram características altamente significantes para garotos, sendo mais prováveis de começar a trabalhar, repetir de ano, e evadir a escola. Ainda sobre a evasão escolar, foram obtidos que pais que apresentam maior nível educacional, há uma menor probabilidade de seus filhos evadirem a escola; Regiões como Porto Alegre e Rio de Janeiro são áreas metropolitanas com maior probabilidade de crianças abandonarem as escolas; A renda do pai tem efeito negativo sobre a evasão escola mesmo quando controlado para a educação do pai; E, por fim, foi encontrado evidências fortes de que pais que passam a receber 0 de renda, as chances dos filhos evadirem a escola é de 1,23.

Leon e Menezes-Filho (2002) analisaram descritivamente os indicadores e determinantes da reprovação, avanço e evasão escolar condicional à reprovação no Brasil para a 4ª e 8ª séries do ensino fundamental e 3º ano do ensino médio, no período de 1984-1997. O estudo mostra que a melhora educacional verificada no Brasil nos últimos 20 anos, foi mais lenta do que nos demais países da América Latina, sendo que estudantes que concluíram a 4ª série, mais de 95% continuam estudando, e estudantes que concluíram a 8ª série, mais de 90% continuam estudando o ensino médio. Porém, ao observar a continuação do estudo para o ensino superior, apenas 60% dos estudantes que concluem o ensino médio, continuam estudando. Dessa forma, o trabalho sugere que uma possível medida de combate à evasão escolar, seria o crédito estudantil ou aumento de oferta de vagas para o ensino superior ou profissionalizante.

Posteriormente, Neri et al. (2015) busca entender as motivações que estão relacionadas às pessoas não estarem estudando no contexto educacional brasileiro. Para eles há um paradoxo, visto que há um alto retorno privado para investimentos na educação, cerca de 15% por ano a mais de estudo, e mesmo assim os brasileiros investem tão pouco na educação. Com dados da PNAD e da PME de 2003-2004 e 2006-2007, focando em faixas de idade entre 15 e 17 anos, o trabalho elenca três principais respostas para este paradoxo, dentre elas: a falta de interesse

dos pais e dos alunos sobre a oferta educacional, ou desconhecimento dos potenciais benefícios educacionais sobre a renda do cidadão; segundo, a miopia dos gestores de políticas públicas na restrição de oferta de serviços educacionais; e, por fim, operações de restrição da renda ou do mercado de crédito que impedem as pessoas de explorar os altos retornos oferecidos pela educação no longo prazo.

Além do mais, os autores ressaltam fatores relacionados à atratividade escolar, como por exemplo, regiões com maiores oportunidades de trabalho são as que mais atraem os jovens fora da escola, principalmente em grandes metrópoles. Outra motivação seria o auxílio à renda familiar, sendo mais afetados jovens dos quais os pais perderam o emprego e os quais as mães não possuem instrução. Adicionalmente, se combinados esses dois fatores - oportunidade de trabalho e carência da renda familiar - a evasão escolar é ainda mais afetada.

1.4 Habilidades não cognitivas

O foco dessa pesquisa reside nas habilidades socioemocionais. Os dados coletados pela aplicação do *Big Five Inventory* (JOHN; SRIVASTAVA, 1999) em Sertãozinho permitem o cálculo de escalas socioemocionais que podem ser importantes no processo educacional. A partir dele, pode-se extrair medidas de Abertura a Novas Experiências, Conscienciosidade, Extroversão, Amabilidade e Neuroticismo. Diversos estudos caracterizam como esses traços pessoais são relacionados a resultados educacionais.

Primi e Santos (2014) definem Abertura a novas experiências como um domínio relacionado à curiosidade, imaginação e tolerância. Segundo os autores, tal característica se revela como uma das habilidades socioemocionais mais correlacionadas com resultados escolares, porém, chamam a atenção para o fato de que essa correlação demanda cautela para interpretá-la como causal, já que esta característica é correlacionada significativamente à inteligência. O presente estudo trata este problema ao inserir medidas cognitivas na explicação da permanência e do progresso escolar, assim, o resultado dos socioemocionais são condicionais as habilidades cognitivas.

A Conscienciosidade está ligada à responsabilidade, disciplina, foco e racionalidade. De acordo com Primi e Santos (2014) este é o construto mais correlacionado com resultados educacionais dentre os socioemocionais, *Big Five*. Lleras (2008) mostra que o efeito de um aumento em um desvio-padrão de tal característica sobre os anos de estudo é bem próximo ao da inteligência, sendo então um importante preditor da evasão escolar. Tal resultado segue a mesma direção do estudo citado como “experiência do *Marshmallow*” (MISCHEL; SHODA; RODRIGUEZ, 1989), indicando que a capacidade de postergar recompensa estava ligada a uma melhora no sucesso educacional futuro. Tal fato está mais ligado à Conscienciosidade do que aos outros fatores socioemocionais.

Sobre a Extroversão, esse construto indica quanto o indivíduo é sociável, animado e quanto canaliza suas energias às outras pessoas e ao mundo externo. Essa característica é indicada na literatura por apresentar correlações ambíguas para os resultados educacionais, que por vezes indicam que a extroversão pode mudar seu efeito ao longo da idade, apresentando correlação negativa com resultados de estudantes quanto maior a idade destes (POROPAT, 2009). Além disso, resultados que indicam que, embora a Extroversão possa ser um fator menos relevante, em relação aos demais, para a nota dos alunos em Português e Matemática na idade adulta, ela pode afetar a decisão de permanência dos alunos na escola por mais tempo.

A Amabilidade é a maneira de agir cooperativamente e não egoísta. Duncan, Magnuson et al. (2011) indicam que tal variável é um preditor negativo para verificar se o aluno concluiu ou não o ensino médio.

Outro fator encontrado por Duncan, Magnuson et al. (2011) como preditor negativo da chance de se completar o ensino médio é o Neuroticismo, que é definido por Primi e Santos

(2014) como instabilidade emocional.

Uma das principais medidas não cognitivas disponíveis nos dados é o Locus de Controle de Rotter (1966). O Locus de controle pode ser visto como uma medida que diz quanto o indivíduo se acha protagonista de sua própria vida, ou seja, quanto tal indivíduo se julga responsável pelos eventos que acontecem com ele. O questionário utilizado é o Tel Aviv (MILGRAM; MILGRAM, 1975). O questionário é desenvolvido para crianças que têm entre 10 e 14 anos de idade.

De acordo com Findley e Cooper (1983) níveis maiores de Locus de Controle estão associados a maiores níveis de desempenho escolar⁶. Apesar dessa importância para o aprendizado, a medida citada tem importância per si, sendo que trabalhos como os de Prearo (2012) e Chikitani (2015) buscavam fatores que influenciassem o Locus de Controle⁷.

Por fim, outro conjunto importante de características psicológicas para resultados educacionais são as variáveis de Habilidades Sociais, que podem ser mensuradas pelo instrumento *Social Skills Rating System* (GRESHAM; ELLIOTT, 1990). As habilidades citadas formam fatores que foram moldados com o objetivo de prever resultados (educacionais, laborais etc.)⁸. As habilidades são: Assetividade, Cooperatividade, Autocontrole e Empatia.

A natureza dessa literatura mais conectada com as áreas de estudo da psicologia, para entender o processo em que ocorre evasão escolar, foi inicialmente estudado por Tinto (1975), que adaptou um modelo de suicídio para a evasão educacional. A ideia do autor é a de que a decisão de sair da escola pode ocorrer se o aluno não for suficientemente integrado à escola, academicamente e socialmente no ambiente educacional. Ou seja, a falta de interação social e a baixa integração acadêmica podem levar a um menor comprometimento com os objetivos que o aluno tem, elevando a probabilidade de decidir evadir a escola.

Portanto, essa natureza multidisciplinar que envolve a investigação acerca do fluxo escolar justifica um estudo que entenda como fatores socioeconômicos, fatores cognitivos e fatores não cognitivos se combinam na determinação do fluxo escolar no Brasil. A seção seguinte apresenta a metodologia para investigação deste problema, em que as dimensões utilizadas no estudo são embasadas na literatura aqui apresentada.

⁶ Em outra nomenclatura, diz-se também que mais interno é o Locus de Controle quanto maior o *score* de Locus de Controle.

⁷ Ambos os trabalhos utilizaram a mesma fonte de dados deste presente projeto de pesquisa

⁸ Ou seja, ao contrário dos *Big Five* que emergiram empiricamente de análises estatísticas, o SSRS foi um instrumento construído de maneira inversa, pensando na predição *ex-ante*

2 METODOLOGIA

2.1 Dados

A base de dados utilizada no trabalho pertence ao Instituto Ayrton Senna e advém de uma coleta realizada na cidade de Sertãozinho - SP, contendo informações em três pontos no tempo: 2008, 2012 e 2017. Em 2008, a amostra representa os dados dos estudantes do Segundo Ano do Ensino Fundamental da cidade. Nos anos seguintes, se buscou seguir esses estudantes, agregando também, na amostra, os outros alunos que eram colegas de sala dos estudantes iniciais, expandindo, dessa forma, a quantidade de alunos na base de dados. O exercício principal deste trabalho se vale da amostra de 2012, ano em que a maior parte dos alunos estava na quinta ou na sexta série, analisando o que ocorreu com tais alunos em 2017, ano em que idealmente eles deveriam atingir o ensino médio.

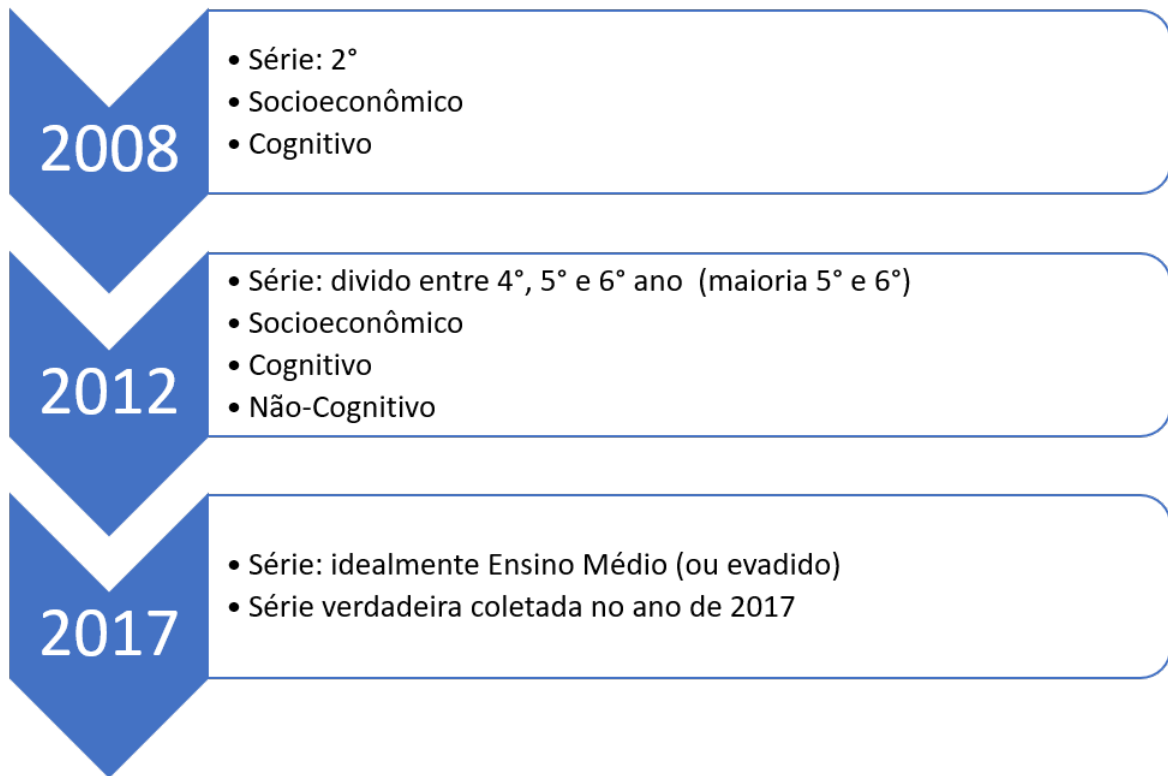
A coleta de 2012 é continuação da coleta de 2008, também na cidade de Sertãozinho - SP, com o intuito de obter dados educacionais dos estudantes do segundo ano do Ensino Fundamental. No caso de 2008, no conjunto de alunos do Segundo Ano do Ensino Fundamental, foram aplicados o questionário socioeconômicos, mas sem a aplicação de questionário socioemocional, de interesse deste trabalho. Para o ano de 2017, a coleta ainda estava em andamento durante a elaboração deste trabalho, porém, foi possível identificar a informação sobre a série que os estudantes estavam até então. Na coleta de 2012 estão presentes informações socioeconômicas, prova de Matemática e Língua Portuguesa, informações dos pais, professores e diretores, além do questionário de habilidades Socioemocionais, enquanto a coleta de 2008 não conta com tais dados. O questionário socioemocional permite realizar um estudo empírico que incorpore habilidades não-cognitivas para o entendimento do fluxo escolar no país, que é uma lacuna na literatura. A figura 2 apresenta resumidamente a estrutura dos dados

Assim, foi possível explorar essas habilidades não cognitivas presentes nos dados em 2012, associando ao fluxo escolar verificado a partir das informações de 2017. As principais habilidades não cognitivas que o questionário permite mensurar são os Traços de Personalidade¹. O questionário aplicado foi o *Big Five Inventory* (BFI) de John e Srivastava (1999). Outras habilidades não cognitivas que puderam ser mensuradas são as habilidades sociais (GRESHAM; ELLIOTT, 1990) e do Lócus de Controle (ROTTER, 1966).

As duas medidas de fluxo que serão utilizadas como variável dependente, construídas a partir desses dados, são: a permanência escolar, que é uma indicadora que diz quais alunos matriculados em 2012 que estavam matriculados em 2017; e a outra medida é uma indicadora se o estudante atingiu ou não o ensino médio. As principais variáveis dependentes medem fatores que a literatura apresentada julga importante na decisão de onde e quanto estudar, mensurando as condições socioeconômicas (para capturar restrição creditícia das famílias) e estrutura fami-

¹ conhecido como *Big Five*, de Goldberg (1993)

Figura 2 – Estrutura dos dados



Fonte – elaboração própria

liar, habilidade cognitiva e variáveis socioemocionais. Dentre as variáveis não cognitivas, testamos os três grupos separadamente, as Habilidades Sociais (Assertividade, Autocontrole, Cooperatividade e Empatia), Traços de personalidade (Amabilidade, Extroversão, Neuroticismo, Abertura a Novas Experiências e Conscienciosidade), além do Locus de Controle interno.

2.2 Estratégia Econométrica

A estratégia econométrica consiste na estimação de um Probit, em que as variáveis explicativas são as características coletadas em 2012 e a variável dependente é uma *dummy* que represente o fluxo escolar para cada indivíduo. A primeira medida de fluxo é uma variável *dummy* igual a 1 se o indivíduo atingiu o Ensino Médio. A outra medida é igual a 1 caso o indivíduo esteja estudando em 2017, como será caracterizado mais a frente.

Suponha que exista uma variável latente não observável y^* (por exemplo, o *payoff* da educação) que depende das características X do indivíduo e de características não observáveis, sumarizadas por um escalar ε , supondo separabilidade aditiva e linearidade em parâmetros, podemos escrever $y^* = x'\beta + \varepsilon$. Porém, nos dados, apenas observamos uma *dummy* que é a medida de fluxo escolar. Ou seja, para cada indivíduo i , define-se:

$$y_i = \begin{cases} 1 & \text{se } \varepsilon_i > -x'_i\beta \\ 0 & \text{caso contrário} \end{cases} \quad (2.1)$$

O objetivo foi estimar a média condicional da medida de fluxo escolar, condicionada as características dos estudantes, $E(y|X) = Pr(y = 1|X)$. Se ε tem distribuição normal padrão então a probabilidade a se estimar tem distribuição acumulada normal:

$$\begin{aligned} E(y|X) &= Pr(y_1 = 1|X) \\ &= \Phi_1(x'_1\beta_1) \end{aligned} \quad (2.2)$$

Os parâmetros do modelo serão estimados por máxima verossimilhança, cujo processo de estimação pode ser escrito pela seguinte forma (seguindo um Probit tradicional):

$$\hat{\beta} = \arg \max_{\beta} \prod_{i=1}^N \Phi(x'_i\beta)^{y_i} (1 - \Phi(x'_i\beta))^{1-y_i} \quad (2.3)$$

Com a estimação do modelo pode-se realizar alguns procedimentos para melhor entendimento do papel das habilidades não cognitivas. Primeiramente, verificar individualmente e conjuntamente a significância das variáveis. Em segundo lugar, verificar o quanto um aumento de desvio-padrão em uma habilidade socioemocional pode reduzir o *gap* entre a permanência escolar prevista pelo modelo e o 100%. Também será feita a comparação entre estes fatores não cognitivos com os cognitivos, comparando a dimensão desses distintos papéis exercidos sobre o fluxo escolar.

2.3 Controlando para o viés de Seleção

A metodologia elaborada acima, até então, se vale da coleta realizada em 2012 e da informação sobre a série dos estudantes em 2017 para verificar como as habilidades socioemocionais podem ser associadas as chances dos alunos permanecerem nas escolas. Porém, pode ser útil explorar os dados de 2008, que podem nos indicar a respeito de algum tipo de seleção amostral ocorrida entre 2008 e 2012, o que poderia prejudicar a realização de inferência.

Para isso, foi elaborado um modelo Probit, porém, controlando para o atrito ocorrido na primeira passagem dos alunos, entre 2008 e 2012. Essencialmente, o Probit com viés de seleção (que também indicaremos aqui como *Heckprobit*²) se torna um Probit sequencial, com duas escolhas.

Para poder incorporar o viés de seleção serão utilizadas características desses indivíduos no segundo ano do Ensino Fundamental, coletadas em 2008, como será apresentado nesta seção. Desenhando a escolha em dois períodos, podemos modelar tais decisões de uma forma estimável suponha que a função objetivo da escolha educacional tomada pela família venha de um vetor de características observáveis X e de características não observáveis $\tilde{\varepsilon}$. Assim, suponha que a utilidade latente do indivíduo i associada a continuar estudando em t seja

$$U_{i,t} = U(X_{i,t}, \tilde{\varepsilon}_{i,t}), \quad i = 1, \dots, N \text{ e } t = 1, 2. \quad (2.4)$$

Suponha que os efeitos das observáveis e não observáveis são aditivamente separáveis, e suponha linearidade em parâmetro. Seja ε um escalar que captura todo o efeito das não observáveis, temos então:

$$U_{i,t} = X'_{i,t}\beta_t + \varepsilon_{i,t}, \quad i = 1, \dots, N \text{ e } t = 1, 2. \quad (2.5)$$

Nos dados, temos características observáveis em 2008 (que definimos aqui como período 1), no qual os alunos estavam na segunda série do ensino fundamental, e 2012 (no qual variavam entre quarta, quinta e sexta série). O interesse reside em verificar quais desses alunos de 2012 que estavam estudando ainda em 2017, ano que idealmente os estudantes dessa base de dados deveriam estar no ensino médio. A decisão de interesse só é observada se o indivíduo seguir estudando de 2008 para 2012. É nesse ponto que surge a seleção amostral, que pode ser corrigida modelando corretamente a probabilidade dos indivíduos, em 2012, seguirem estudando em 2017, condicional a permanecerem na escola entre 2008 e 2012 .

Analogamente, a escolha de 2012 é seguir até 2017 estudando, ano que idealmente os alunos deveriam estar no ensino médio³. Também é necessário notar que só é possível observar

² O nome, utilizado em softwares estatísticos, se deve ao fato de que a modelagem do controle por viés de seleção se torna análoga ao proposto por (HECKMAN, 1977), porém, adequado a modelos de escolha binária.

³ Note que sabemos se em 2017 o aluno estava estudando mesmo se o mesmo reprovar, ou seja, só não temos

a segunda escolha caso. De fato, o interesse reside na passagem do ensino fundamental (em 2012) para o ensino médio (ou continuar estudando mesmo em caso de reprovação, até 2017). Assim, com um conjunto parcimonioso de hipóteses, pode-se corrigir o viés de seleção presente nessa escolha. Note ainda que se dividirmos ambos os lados das desigualdades (1) e (2) temos um modelo que é observacionalmente equivalente, para lidar com tal problema, vamos supor que ambos os fatores não observáveis têm variância unitária.

Hipótese 1 (H1):

$$\begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right) \quad (2.6)$$

H1 auxilia a modelagem da dependência entre as escolhas, assim, testar o coeficiente ρ significa mostrar se há ou não uma seleção amostral relevante nos exercícios empíricos e permite então aplicar uma correção de Heckman para um modelo de escolha binária.

Com a utilidade latente, observa-se apenas uma indicadora se o indivíduo prosseguiu estudando ou não. Assim definimos:

$$y_{it} = \mathbb{I}(U_{it} > 0), \quad i = 1, \dots, N \text{ e } t = 1, 2. \quad (2.7)$$

Para verificar como as habilidades socioemocionais se relacionam com a escolha de seguir ou não estudando até 2017, deseja-se estimar

$$\begin{aligned} \mathbb{E}(y_2|y_1 = 1, X) &= Pr(y_2 = 1|y_1, X) \\ &= \frac{1}{Pr(y_1 = 1|X)} Pr(y_1 = 1, y_2 = 1|X) \\ &= \frac{\Phi(x'_1\beta_1, x'_2\beta_2, \rho)}{\Phi_1(x'_1\beta_1)}. \end{aligned} \quad (2.8)$$

Para ver a última passagem, note que $y_{2i} = \mathbb{I}_2(U_2 > 0) = \mathbb{I}_2(\varepsilon_2 > x'_2\beta_2)$. Logo pela hipótese 1 é possível calcular os termos acima:

$$\begin{aligned} Pr(y_2 = 1, y_1 = 1|X) &= Pr(\varepsilon_1 > -x'_1\beta_1, \varepsilon_2 > -x'_2\beta_2|X) \\ &= Pr(\varepsilon_1 \leq x'_1\beta_1, \varepsilon_2 \leq x'_2\beta_2|X) \\ &= \Phi(x'_1\beta_1, x'_2\beta_2, \rho) \end{aligned} \quad (2.9)$$

e note ainda que

$$\begin{aligned}
Pr(y_1 = 1|X) &= Pr(\varepsilon_1 > -x'_1\beta_1|X) \\
&= Pr(\varepsilon_1 \leq x'_1\beta_1|X) \\
&= \Phi_1(x'_1\beta_1).
\end{aligned} \tag{2.10}$$

A probabilidade acima ainda é útil para calcular a contribuição para a função de verossimilhança do modelo, para aqueles que não se autoselecionaram no primeiro período. Por fim, é possível calcular a contribuição da verossimilhança daqueles que estavam em 2012 estudando e em 2017 - quando idealmente deveriam estar no ensino médio - evadiram.

$$\begin{aligned}
Pr(y_2 = 0|y_1 = 1, X) &= \frac{1}{Pr(y_1 = 1|X)} Pr(y_1 = 1, y_2 = 0|X) \\
&= \frac{1}{Pr(y_1 = 1|X)} Pr(\varepsilon_1 > -x'_1\beta_1, -\varepsilon_2 > x'_2\beta_2|X) \\
&= \frac{1}{Pr(y_1 = 1|X)} Pr(\varepsilon_1 \leq x'_1\beta_1, -\varepsilon_2 \leq -x'_2\beta_2|X) \\
&= \frac{\Phi(x'_1\beta_1, -x'_2\beta_2, -\rho)}{\Phi_1(x'_1\beta_1)}.
\end{aligned} \tag{2.11}$$

Assim, podemos modelar a função de verossimilhança para encontrar o vetor de parâmetros $\theta \equiv [\beta_1 \ \beta_2 \ \rho]'$.

$$\begin{aligned}
\theta^* &= \arg \max_{\beta, \rho} \prod_{i=1}^N (1 - \Phi_1(x'_{1i}\beta_1))^{1-y_{1i}} * \\
&\quad \Phi(x'_{1i}\beta_1, -x'_{2i}\beta_2, -\rho)^{y_{1i}(1-y_{2i})} * \\
&\quad \Phi(x'_{1i}\beta_1, x'_{2i}\beta_2, \rho)^{y_{1i}y_{2i}}
\end{aligned} \tag{2.12}$$

Como o modelo é não linear, os coeficientes β não são iguais aos efeitos marginais, que podem ser calculados pela equação

$$\begin{aligned}
\frac{\partial Pr(y_{i2} = 1|y_{i1} = 1, X)}{\partial x_{ji2}} &= \frac{\partial}{\partial x_{ji2}} \left(\frac{\Phi(x'_{1i}\beta_1, x'_{2i}\beta_2, \rho)}{\Phi_1(x'_{1i}\beta_1)} \right) \\
&= \frac{1}{\Phi_1(x'_{1i}\beta_1)} \frac{\partial \Phi(x'_{1i}\beta_1, x'_{2i}\beta_2, \rho)}{\partial x_{ji2}} \beta_{j2}
\end{aligned} \tag{2.13}$$

A razão para não se fazer um Probit condicional simples se deve ao fato de que as estimativas são enviesadas, enquanto a proposta acima é consistente e assintoticamente eficiente (GREENE, 2000). Para ver isso, basta notar que se $\rho = 0$ então podemos escrever a normal bivariada como o produto das marginais, que são normais, logo

$$\begin{aligned}
\frac{\partial Pr(y_{i2} = 1|y_{i1} = 1, X)}{\partial x_{ji2}} &= \frac{\partial}{\partial x_{ji2}} \left(\frac{\Phi(x'_{1i}\beta_1, x'_{2i}\beta_2, \rho)}{\Phi_1(x'_{1i}\beta_1)} \right) \\
&= \frac{\partial}{\partial x_{ji2}} \left(\frac{\Phi_1(x'_{1i}\beta_1)\Phi_2(x'_{2i}\beta_2)}{\Phi_1(x'_{1i}\beta_1)} \right) \\
&= \varphi_2(x_{2i}\beta_2)\beta_{j2}
\end{aligned} \tag{2.14}$$

que é o efeito marginal calculado para o Probit tradicional. Note ainda que esse efeito depende das características.

A partir da metodologia aplicada, pode-se responder a pergunta do trabalho a partir de alguns exercícios. O primeiro a ser realizado é considerar um modelo mais restrito, com coeficientes zero para as características socioemocionais. Nesse caso, o modelo de escolha fica restrito a um caso particular da versão mais antiga na literatura, considerando apenas um conjunto de cognitivos e socioeconômicos como fatores decisivos para a evasão escolar. Ao estimar o modelo mais restrito, é possível comparar o poder explicativo de ambos os modelos. A ideia é testar se o grau de ajuste do modelo aos dados melhora consideravelmente ao introduzir as características não cognitivas.

Um exercício essencial para apresentar como este modelo representa uma mudança na explicação da evasão escolar brasileira é verificar a heterogeneidade de efeitos para Socioemocionais e socioeconômicos. Para isso, deseja-se verificar se mesmo quando restrito a grupos de melhores condições de renda, ainda existe uma diferença de evasão explicada pelas características socioemocionais. O mesmo seria realizado para os mais pobres. Se observada a heterogeneidade, isso explicaria o porquê de, no Brasil, mesmo aliviando a restrição orçamentária dos indivíduos mais vulneráveis, o país não consegue verificar uma resposta correspondente na escolaridade média dos indivíduos. Nesse caso, o uma explicação que não levasse em conta habilidades não cognitivas seriam um caso particular quando este tipo de habilidade fosse constante entre os indivíduos, o que se pretende mostrar que não é o caso, na maioria das vezes.

Foi realizado um teste de razão de verossimilhança para testar se as equações são independentes. Existe uma motivação para corrigir a seleção existente em um modelo de escolhas dadas sequencialmente, porém, a baixa taxa de evasão antes do ensino médio não ocasiona seleção amostral relevante, como será mostrado nos resultados, reforçando a robustez dos exercícios econométricos anteriores. Ou seja, o teste sobre o parâmetro ρ basicamente nos diz se o coeficiente ρ da equação é zero, podendo decidir entre uma metodologia mais simples, que não necessite de uma análise sequencial da coorte de 2008, ou uma metodologia que leve em conta a seleção.

A seção seguinte apresenta os resultados econométricos combinada com uma análise qualitativa dos mesmos. Os resultados seguem a mesma ordem lógica da seção metodológica, visando apresentar as principais implicações das estimações do trabalho, bem como apresentar quão robusto são esses exercícios empíricos.

3 RESULTADOS

3.1 Descritivas

Os exercícios empíricos foram aplicados sobre a base de dados de Sertãozinho - SP, que contém informações em três pontos do tempo: 2008, 2012 e 2017. Em 2008, o estudo abrangia uma aplicação de questionários socioeconômicos e cognitivos sobre o universo de alunos do segundo ano do Ensino Fundamental da cidade, sendo que efetivamente foram coletados 2076 observações. Em 2012, foram realizados dois campos: um no primeiro semestre e outro no segundo semestre. Os alunos que pertenceram a primeira coleta se encontravam divididos entre Quarto, Quinto e Sexto ano do Ensino Fundamental, tais estudantes foram seguidos no primeiro semestre para uma nova coleta, agora com informações não-cognitivas dos estudantes. Além disso, a amostra aumentou, pois, em vários casos, não foram coletadas apenas as informações dos estudantes que participaram em 2008, como também dos colegas de sala.

A prova cognitiva (Matemática e Língua Portuguesa), em 2012, tinha uma versão para o Quinto Ano e outra para o Sexto Ano do ensino fundamental, assim, os estudantes do Quarto Ano, parcela pequena da amostra, foram convidados a fazer a prova do Quinto Ano. Já no segundo semestre, ocorreu uma coleta sobre todo o Quarto Ano do Ensino Fundamental da cidade, porém, não foram aplicadas provas de Matemática e Língua Portuguesa. Portanto, este trabalho se valeu apenas da coleta do primeiro semestre de 2012, não levando em conta os alunos do Quarto Ano de 2012 que não estavam previamente na coleta de 2008. Em 2012 estão disponíveis as informações socioemocionais dos estudantes, foco da presente pesquisa.

O último ponto de observação da coleta corresponde a 2017, sendo que, neste ano, idealmente, os estudantes do estudo estariam no Ensino Médio. Neste ponto do tempo estão disponíveis a informação que diz se o estudante estava matriculado em alguma rede de ensino e, caso afirmativo, em qual série tais estudantes se encontravam¹.

Com a união de todas as informações dos questionários de 2012 foram obtidas 3349 observações. Se forem analisados as medidas de fluxo presentes na base de dados, pode-se notar indicadores elevados, como é o caso brasileiro. Desses 3349 estudantes, 2918 foram encontrados no sistema escolar em 2017, representando uma taxa de permanência desses estudantes de 87,2%. Para efeito de comparação, 84,27% é a taxa de atendimento escolar entre jovens de 15 a 17 anos no Brasil, de acordo com dados da PNAD/IBGe, em 2014. Ainda nos dados, desses 3349 estudantes, 2419 estavam no Ensino Médio, representando uma taxa de 72,3%, sendo que, no Brasil, a taxa de jovens de 15 a 17 anos no Ensino Médio é de 61,51%.

A composição da amostra, em termos da série em que cada estudante se encontrava em

¹ Foram buscadas as informações sobre os estudantes em todas as escolas de Sertãozinho e de Ribeirão Preto, que é uma cidade próxima para onde muitos estudantes migram, porém, se o estudante permanece estudando em outra cidade, supomos, por simplicidade, que este evadiu.

2012, pode ser vista na tabela 1, ressaltando que nos exercícios econométricos, inicialmente, não se utiliza os indivíduos do Quarto Ano do Ensino Fundamental, pois, a princípio, estes não podem ter alcançado o Ensino Médio em 2017, objeto de estudo deste trabalho.

Tabela 1 – Distribuição da série da amostra em 2012

Série em 2012	Frequência	Proporção
4ºano	126	3.76%
5ºano	1680	50.16%
6ºano	1543	46.07%
Total	3349	100.00%

Fonte: elaboração própria. A estatística se baseia apenas nos alunos que participaram da coleta do primeiro semestre de 2012.

Como citado, essa amostra é composta por estudantes que em 2008 estavam no segundo ano, com outros estudantes da cidade de Sertãozinho do Quinto e Sexto ano de 2017. Assim, os 126 estudantes que estavam no Quarto Ano do Ensino Fundamental são indivíduos que deixaram de progredir na escola por duas vezes entre as duas coletas de dados.

Um fator importante a se descrever nos dados é estrutura socioeconômica dos estudantes, que é tida pela literatura como um fator importante na determinação do fluxo e desempenho escolar, visto que esta característica, em última instância, representa as condições de restrição de crédito das famílias, que como mostrado na literatura, são tidas como as fontes primordiais na explicação do fluxo escolar. Assim, uma medida desta situação é dada pela educação da mãe, que é uma *proxy* da renda de longo prazo da família, sendo utilizada no modelo econométrico em conjunto com outros fatores socioeconômicos como cor, sexo e idade. Adicionalmente, como apresentado na literatura, a capacidade cognitiva do estudante determina também o retorno do estudo sobre o salário no mercado de trabalho. Assim, através da prova de Matemática e Língua Portuguesa, em conjunto com Teoria da Resposta ao Item para o cálculo dos *scores*, foram calculadas duas medidas cognitivas para utilização no modelo. Por fim, foco deste trabalho, estão as medidas socioemocionais, além de habilidades sociais e do *locus* de controle, como medidas não cognitivas. Como essas medidas são, a priori, difíceis de interpretar, tais variáveis foram padronizadas. A Tabela 2 descreve as variáveis socioeconômicas utilizadas nos exercícios econométricos.

Os números indicam que a maioria da amostra é composta por homens e por estudantes não brancos. Os números indicam que 37% dos alunos já reprovaram em alguma etapa de ensino e a mesma proporção de estudantes têm a mãe com Ensino Médio completo. A reprovação, por si, já é um indicador do problema de fluxo no Brasil, essa alta taxa de reprovação, que faz com que os indivíduos atrasem a chegada ao Ensino Médio, também pode induzir a uma decisão de evadir a escola, sendo portanto, importante controlar as estimações por este fator.

Tabela 2 – Estatísticas descritivas do perfil socioeconômico dos estudantes de Sertãozinho em 2012

Variável	N	Média	Desvio Padrão	Mínimo	Máximo
Permanência	3346	0.87	0.33	0	1
Ensino Médio	3346	0.72	0.45	0	1
Reprovação	3320	0.37	0.48	0	1
Abandono	3328	0.04	0.20	0	1
Pai_EM	2167	0.37	0.48	0	1
Mae_EM	2402	0.37	0.48	0	1
Homem	3332	0.52	0.50	0	1
Branco	3292	0.40	0.49	0	1

Fonte: elaboração própria. Permanência indica se o aluno estava estudando em 2017, Ensino Médio indica se o aluno atingiu o Ensino Médio em 2017, 'Pai_EM' indica se o pai tem ou não Ensino Médio Completo, 'Mãe_EM' indica se a mãe tem ou não Ensino Médio Completo.

As variáveis descritas acima serão utilizadas em conjunto com os fatores não cognitivos e com as habilidades em Matemática e Língua Portuguesa para descrição de duas medidas de fluxo: uma indicadora se o indivíduo permanece estudando em 2017; uma indicadora se o indivíduo atingiu o ensino médio. Como vimos nos dados, os indicadores de fluxo apresentam números desfavoráveis, como ocorre para o caso brasileiro como um todo. Assim, as estimações podem trazer sinais das causas do problema a nível nacional.

Portanto, com a descrição estatística das variáveis dependentes e as explicativas dos exercícios empíricos, pode-se realizar os exercícios econométricos, que são apresentadas nas seções seguintes.

3.2 Resultados do modelo Probit

Esta seção apresenta dos resultado econométricos obtidos a partir da estimação do modelo Probit, utilizando as informações coletadas no primeiro semestre de 2012 em Sertãozinho, além da alocação dos estudantes em 2017. Para tal estimação, os conjuntos de habilidades não cognitivas foram utilizadas em modelos separadamente (Socioemocionais, *Social Skills* e Locus de Controle) como explicativas para as medidas de fluxo, que são: indicadora se o estudante permaneceu na escola entre 2012 e 2017; indicadora se o aluno progrediu para o Ensino Médio até o ano de 2017. É importante ressaltar que foram utilizados nessa seção os estudantes que estavam no quinto e sexto ano do Ensino Fundamental em 2012, pois os estudantes que estavam no quarto ano em 2012 não atingiriam o Ensino Médio num progresso normal na escola.

Assim, as estimações apresentam os resultados para a probabilidade do indivíduo atingir o Ensino Médio e para a probabilidade de permanência escolar, sendo reportado o efeito marginal de cada variável explicativa do modelo, calculada sobre a média. São também reportadas estatísticas úteis para avaliação da adequação do modelo, como a probabilidade de se atingir o Ensino Médio, sobre o indivíduo médio, predita pelo modelo Probit, o teste LR (*Likelihood Ratio*) de existência do modelo, o p-valor associado, o Pseudo- R^2 e o número de observações utilizadas no modelo, verificando como o conjunto de habilidades socioemocionais prediz o fluxo escolar. Para mostrar a robustez dos resultados, diversas formulações foram estimadas.

A tabela 3 apresenta a estimação do efeito marginal das habilidades socioemocionais, que estão padronizadas na regressão, sobre a chance de se atingir o Ensino Médio, com controle para fatores socioeconômicos e cognitivos.

Tabela 3 – Estimação do Probit para probabilidade de se atingir o Ensino Médio com variáveis Socioemocionais

Pr(y_EM=1)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
matem	0.128***		0.097***	0.080***	0.121***		0.097***		0.099***	0.081***	0.083***
port		0.108***	0.044***	0.030*		0.100***	0.037**	0.076***		0.027*	0.026*
mae_EM	0.009	0.012	0.003	-0.015	0.008	0.012	0.004	-0.006	-0.009	-0.011	-0.013
Branco	0.010	0.016	0.008	0.005	0.003	0.012	0.004	0.008	0.001	0.002	0.001
Homem	-0.121***	-0.070***	-0.101***	-0.080***	-0.110***	-0.064***	-0.098***	-0.050**	-0.088***	-0.080***	-0.079***
Reprovada				-0.153***				-0.152***	-0.135***	-0.131***	-0.109***
Am					0.010	0.004	0.005	0.002	0.007	0.004	0.003
Cc					0.029**	0.032**	0.028**	0.030**	0.028**	0.027**	0.027*
Ex					-0.001	0.002	-0.004	0.003	0.000	-0.002	-0.001
Ne					-0.004	-0.001	-0.004	0.003	0.001	-0.000	0.001
Op					-0.008	-0.007	-0.008	-0.012	-0.013	-0.013	-0.013
Idade											-0.016
Pr(y=1 x)	0.758	0.760	0.759	0.760	0.781	0.782	0.782	0.783	0.782	0.783	0.783
LR χ^2	243.855	193.060	256.357	327.036	228.463	175.850	237.437	242.991	281.368	287.290	290.572
p-valor	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R^2	0.099	0.078	0.104	0.134	0.110	0.085	0.115	0.118	0.137	0.140	0.142
Observações	2235	2234	2231	2217	1968	1969	1966	1959	1958	1956	1956

Fonte: elaboração própria. ***significante a 1%. **significante a 5%. *significante a 10%. Valores indicam o efeito marginal da variável explicativa sobre a Pr(y=1). Am: Amabilidade. Cc: conscienciosidade. Ex: extroversão. Nr: Neuroticismo. Op: Abertura ao novo. Matem: habilidade de matemática. Port: habilidade de língua portuguesa. mae_EM: mãe com Ensino Médio Completo. Modelo aplicado apenas nos alunos que participaram da coleta do primeiro semestre de 2012, restrito ao quinto e sexto ano.

A tabela mostra que a conscienciosidade tem um poder preditivo robusto a formulação do modelo, positivo sobre a probabilidade de se atingir o Ensino Médio, com Efeito Marginal

em torno de 3%. A probabilidade prevista de se atingir o Ensino Médio a amostra é de aproximadamente 78%, ou seja, para conseguir que todos os estudantes que pudessem atingir o Ensino Médio conseguissem, seriam necessários um aumento de 22 pontos percentuais. Comparando com o efeito marginal da conscienciosidade, este, ainda que aumentasse em 1 desvio-padrão para todos os indivíduos, resolveria aproximadamente 1/7 deste problema de fluxo. Além disso, o efeito positivo da conscienciosidade é pequeno perto do efeito negativo de se ter uma reprovação anterior sobre a chance de atingir o Ensino Médio, sendo que, naturalmente, uma reprovação pode piorar as habilidades socioemocionais.

É necessário destacar o também importante papel que as habilidades cognitivas têm sobre o fluxo escolar. As habilidades de matemática e língua portuguesa, também padronizadas, são preditores fortes do fluxo. O efeito marginal da habilidade de matemática tem uma magnitude similar a da reprovação. Quando ao conjunto socioeconômico de variáveis, a variável indicadora do sexo masculino apresenta um efeito marginal negativo para a probabilidade de se atingir o Ensino Médio ². Nota-se ainda, que mesmo quando utilizada as duas medidas cognitivas, que a princípio são fortemente correlacionadas, ambas se mostram significantes na predição da probabilidade de se atingir o ensino médio.

O mesmo processo de estimação é utilizado para a prever a probabilidade do aluno de 2012 permanecer estudando em 2017, os resultados são sintetizados na tabela 4

Tabela 4 – Estimação do Probit para a permanência escolar com variáveis Socioemocionais

Pr(y ₂₀₁₇ =1)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
mae_EM	0.006	0.006	0.003	-0.008	0.006	0.008	0.005	-0.004	-0.004	-0.005	-0.007
Branco	0.018	0.021	0.018	0.014	0.018	0.022	0.019	0.020	0.017	0.018	0.017
Homem	-0.028*	-0.009	-0.019	-0.005	-0.020	-0.004	-0.016	0.003	-0.008	-0.006	-0.005
Matem	0.046***		0.031**	0.020*	0.043***		0.035***		0.030***	0.024*	0.027**
Port		0.041***	0.021*	0.012		0.035***	0.013	0.022**		0.008	0.004
Reprovada				-0.102***				-0.089***	-0.083***	-0.083***	-0.042*
Am					0.018*	0.015	0.016	0.013	0.015	0.014	0.013
Cc					-0.002	-0.001	-0.002	-0.001	-0.002	-0.002	-0.004
Ex					-0.007	-0.006	-0.008	-0.005	-0.005	-0.006	-0.004
Nr					0.001	0.002	0.001	0.003	0.002	0.002	0.003
Op					-0.003	-0.001	-0.002	-0.005	-0.006	-0.005	-0.006
Idade											-0.027***
Pr(y=1 x)	0.88	0.88	0.88	0.88	0.90	0.90	0.90	0.90	0.90	0.90	0.90
LR χ^2	52.31	46.29	57.27	107.12	53.22	41.43	54.82	79.99	85.76	87.09	106.49
p-valor	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
R ²	0.032	0.029	0.035	0.067	0.041	0.032	0.042	0.062	0.066	0.068	0.083
Observações	2235	2234	2231	2217	1968	1969	1966	1959	1958	1956	1956

Fonte: elaboração própria. ***significante a 1%. **significante a 5%. *significante a 10%. Valores indicam o efeito marginal da variável explicativa sobre a Pr(y=1|x). Am: Amabilidade. Cc: conscienciosidade. Ex: extroversão. Nr: Neuroticismo. Op: Abertura ao novo. Matem: habilidade de matemática. Port: habilidade de língua portuguesa. mae_EM: mãe com Ensino Médio Completo. Modelo aplicado apenas nos alunos que participaram da coleta do primeiro semestre de 2012, restrito ao quinto e sexto ano.

No aspecto cognitivo, os resultados mostram que a habilidade de matemática é um preditor significativo em diversas formulações. A habilidade de Língua Portuguesa deixa de ser estatisticamente significativa quando controlado pela própria habilidade cognitiva de matemática e pela indicadora de reprovação anterior, embora também mostre ter poder preditivo significativo

² O efeito marginal para variáveis discretas, como uma *dummy* d , são calculados pela estimativa do objeto $\hat{P}r(y = 1|x, d = 1) - \hat{P}r(y = 1|x, d = 0)$.

em diversas formas funcionais estimadas. A habilidade não cognitiva Amabilidade é significativa a 10% apenas no modelo (5) da tabela 4, porém, apesar da falta de robustez, tal modelo apresenta controle por habilidade cognitiva e por características socioeconômicas, ainda assim traz mais poder preditivo para o modelo, tendo um efeito marginal positivo sobre a permanência escolar. No modelo (6), se controlar pela habilidade de língua portuguesa, o efeito da amabilidade deixa de ser significativa. Note que isso pode indicar que a correlação da habilidade de matemática não necessariamente é correlacionada com as mesmas habilidades não cognitivas que a habilidade de língua portuguesa.

O efeito de se acrescentar o conjunto de habilidades socioemocionais pode ser visto pelo pseudo R^2 . Se comparar o modelo (1) com o modelo (5), e o (2) com o (6), se observa um aumento do coeficiente de 3,2% para 4,1% no primeiro caso, e de 2,9% para 3,2%, mesmo com uma precisão menor das estimações devido a uma queda no número de observações do modelo. Em sua versão mais completa, com controle por todas as habilidades socioemocionais e cognitivas, além das variáveis socioeconômicas, além da idade e da indicadora de reprovação anteriores, o pseudo R^2 chega a 8,3%. Ainda se observa que a variável indicadora de reprovação anterior se mostrou aquela com maior efeito sobre as medidas de fluxo construídas, o que já é esperado, visto que a própria reprovação anterior é uma medida de fluxo.

O efeito das habilidades socioemocionais pode se alterar de acordo com a etapa da vida dos jovens, assim, pode ser interessante avaliar o resultado das estimações em subamostras de série e idade. Utilizando a indicadora se o aluno atingiu o Ensino Médio como variável dependente, a tabela 11 e 12, disponível no Apêndice, apresentam estimações que indicam que o efeito da Conscienciosidade é significativa e positivo para o sexto ano, sem efeito para o quinto ano, que ainda apresenta que a Amabilidade em uma das estimativas, é significativa a 10% e tem efeito marginal positivo sobre a probabilidade de se atingir o Ensino Médio. Em subamostras de idade não se encontrou efeito das habilidades socioemocionais. Isso pode ser um indício que a potencial heterogeneidade de efeito das habilidades socioemocionais estão mais relacionados com a etapa escolar do que com a idade em si, embora ambas as medidas são intimamente relacionadas.

Para verificar heterogeneidade das habilidades socioemocionais sobre a permanência escola utilizou-se o mesmo procedimento citado no acima, alterando a variável dependente da regressão. Nesse caso, quando observado os indivíduos que estavam no sexto ano do Ensino Fundamental em 2012, a Amabilidade se mostra significativa a 5% no no modelo (5) da tabela 14 e a 10% nas demais formulações, com sinal positivo para o efeito marginal desta variável, enquanto não foi observado algum efeito para o quinto ano do Ensino Fundamental³. Em dois modelos da tabela 14, modelos (5) e (7), a estimação do efeito marginal da Extroversão sobre a permanência escolar se tornou negativa. Primi e Santos (2014) reforçam que não é trivial o

³ Também foram realizados exercícios por subamostras de idade, não encontrando efeito das habilidades socioemocionais nessas subamostras.

sinal da relação da extroversão com os resultados educacionais não é monotônica (ou seja, nem sempre um *score* maior dessa característica é mais favorável ao resultado educacional).

Os resultados obtidos até aqui são importantes, pois mostram que, mesmo controlado pela habilidade cognitiva dos indivíduos e por indicadores socioeconômicos, as habilidades socioemocionais acrescentam um poder explicativo para o fluxo escolar brasileiro, em especial, no estágio crítico que é o da entrada para o Ensino Médio, até então, a maioria dos estudos estava concentrada no papel das habilidades socioemocionais para a proficiência e aprendizado. Adicionalmente, os resultados mostram depender da amostra, em particular, da etapa que os estudantes se encontram.

Os dados dos estudantes de Sertãozinho ainda contém outros *scores* não cognitivos, como o Locus de Controle e as habilidades sociais. Assim, os exercícios econométricos seguintes visaram verificar como os outros conjuntos de habilidades não cognitivas estão associadas ao fluxo escolar.

As tabelas 5 e 6 apresentam os exercícios econométricos para verificar o papel do Locus de Controle interno sobre o fluxo escolar, enquanto as tabelas 7 e 8 apresentam os resultados análogos utilizando como medida não cognitiva às Habilidades Sociais.

Tabela 5 – Estimação do Probit para probabilidade de permanência na escola utilizando o Locus de Controle interno

Pr(y ₂₀₁₇₌₁)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Mãe Ensino Médio	0.006	0.006	0.003	-0.008	0.005	0.007	0.004	-0.004	-0.004	-0.005	-0.007
Branco	0.018	0.021	0.018	0.014	0.017	0.022	0.018	0.019	0.016	0.017	0.016
Homem	-0.028*	-0.009	-0.019	-0.005	-0.021	-0.006	-0.017	0.001	-0.008	-0.007	-0.006
Matemática	0.046***		0.031**	0.020*	0.040***		0.032***		0.027***	0.022*	0.026**
Língua Portuguesa		0.041***	0.021*	0.012		0.031***	0.012	0.019*		0.006	0.003
Reprovada				-0.102***				-0.087***	-0.082***	-0.082***	-0.044**
Locus Interno					0.013*	0.016**	0.012	0.013**	0.011	0.010	0.008
Idade											-0.026***
Pr(y=1 x)	0.88	0.88	0.88	0.88	0.898	0.898	0.898	0.898	0.897	0.898	0.898
LR χ^2	52.31	46.29	57.27	107.12	50.57	40.86	52.20	79.34	83.87	85.26	102.34
p-valor	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
R ²	0.03	0.03	0.04	0.07	0.04	0.03	0.04	0.06	0.06	0.07	0.08
Observações	2235	2234	2231	2217	1980	1980	1977	1970	1970	1967	1967

Fonte: elaboração própria. ***significante a 1%. **significante a 5%. *significante a 10%. Valores indicam o efeito marginal da variável explicativa sobre a Pr(y=1). Mat-tem: habilidade de matemática. Port: habilidade de língua portuguesa. mae_EM: mãe com Ensino Médio Completo. Modelo aplicado apenas nos alunos que participaram da coleta do primeiro semestre de 2012, restrito ao quinto e sexto ano.

Nesta primeira estimação a habilidade em matemática se mostra um preditor significativo e positivamente relacionada a probabilidade de se atingir o Ensino Médio, enquanto que a habilidade de português deixa de ser significativa quando todas as variáveis de controle são utilizadas. Reprovações anteriores persistem em ser determinantes que afetam negativamente o fluxo escolar, sendo o Locus de Controle positivamente relacionado ao fluxo, porém, também perde significância estatística nas versões com mais covariadas.

Os resultado da tabela 6 sobre a probabilidade de permanência na escola são análogos aos da tabela anterior, com a diferença de que nessa estimação, a *dummy* que indica se o estudante é do sexo masculino mostrou-se significância estatística e ainda negativamente relacionada ao fluxo escolar.

Tabela 6 – Estimação do Probit para a Probabilidade de se atingir o Ensino Médio utilizando o Locus de Controle interno

Pr(y_EM=1)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Mae_Em	0.009	0.012	0.003	-0.015	0.004	0.008	0.000	-0.009	-0.012	-0.015	-0.017
Branco	0.010	0.016	0.008	0.005	0.001	0.010	0.001	0.005	-0.001	-0.001	-0.002
Homem	-0.121***	-0.070***	-0.101***	-0.080***	-0.114***	-0.068***	-0.100***	-0.055**	-0.094***	-0.083***	-0.083***
Matem	0.128***		0.097***	0.080***	0.119***		0.093***		0.097***	0.079***	0.081***
Port		0.108***	0.044***	0.030*		0.096***	0.038**	0.074***		0.028*	0.026*
Reprovada				-0.153***				-0.150***	-0.134***	-0.131***	-0.109***
Locus Interno					0.023*	0.029**	0.019*	0.023*	0.018	0.015	0.013
Idade											-0.015
Predict	0.758	0.760	0.759	0.760	0.780	0.781	0.781	0.782	0.781	0.781	0.781
LR χ^2	243.855	193.060	256.357	327.036	224.867	176.872	232.821	242.817	277.967	283.568	286.790
p-valor	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R ²	0.099	0.078	0.104	0.134	0.108	0.085	0.112	0.117	0.134	0.137	0.139
Observações	2235	2234	2231	2217	1980	1980	1977	1970	1970	1967	1967

Fonte: elaboração própria. ***significante a 1%. **significante a 5%. *significante a 10%. Valores indicam o efeito marginal da variável explicativa sobre a Pr(y=1). Matem: habilidade de matemática. Port: habilidade de língua portuguesa. mae_EM: mãe com Ensino Médio Completo. Modelo aplicado apenas nos alunos que participaram da coleta do primeiro semestre de 2012, restrito ao quinto e sexto ano.

No que diz respeito as habilidades sociais, a Cooperatividade dos estudantes possui um efeito marginal significativo e positivo sobre as duas medidas de fluxo utilizadas, tendo resultado análogo ao das demais estimações.

Tabela 7 – Estimação do Probit para probabilidade de se atingir o Ensino Médio utilizando as Habilidades Sociais

Pr(y_EM=1)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Mae_Em	0.009	0.012	0.003	-0.015	0.001	0.006	-0.003	-0.011	-0.014	-0.017	-0.019
Branco	0.010	0.016	0.008	0.005	-0.007	0.002	-0.007	-0.002	-0.008	-0.008	-0.009
Homem	-0.121***	-0.070***	-0.101***	-0.080***	-0.108***	-0.062***	-0.095***	-0.050**	-0.088***	-0.079***	-0.078***
Matem	0.128***		0.097***	0.080***	0.121***		0.094***		0.099***	0.079***	0.081***
Port		0.108***	0.044***	0.030*		0.101***	0.040**	0.078***		0.030*	0.028*
Reprovada				-0.153***				-0.148***	-0.132***	-0.128***	-0.106***
Cooperatividade					0.053***	0.055***	0.051***	0.050***	0.049***	0.047***	0.046***
Assertividade					-0.009	-0.009	-0.008	-0.010	-0.010	-0.009	-0.009
Empatia					-0.018	-0.022	-0.023	-0.019	-0.017	-0.021	-0.020
Auto Conceito					-0.003	-0.003	-0.001	-0.004	-0.004	-0.002	-0.003
Idade											-0.015
Pr(y=1 x)	0.736	0.737	0.736	0.738	0.757	0.758	0.757	0.759	0.758	0.759	0.759
LR χ^2	243.855	193.060	256.357	327.036	243.301	193.097	251.565	258.679	295.677	301.119	304.420
p-valor	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R ²	0.099	0.078	0.104	0.134	0.116	0.093	0.121	0.125	0.143	0.146	0.147
Observações	2235	2234	2231	2217	1979	1979	1976	1969	1969	1966	1966

Fonte: elaboração própria. ***significante a 1%. **significante a 5%. *significante a 10%. Valores indicam o efeito marginal da variável explicativa sobre a Pr(y=1). Matem: habilidade de matemática. Port: habilidade de língua portuguesa. mae_EM: mãe com Ensino Médio Completo. Modelo aplicado apenas nos alunos que participaram da coleta do primeiro semestre de 2012, restrito ao quinto e sexto ano.

Tabela 8 – Estimação do Probit para a permanência escolar utilizando as Habilidades Sociais

Pr(y_2017=1)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
mae_EM	0.006	0.006	0.003	-0.008	0.003	0.005	0.002	-0.005	-0.006	-0.007	-0.008
Branco	0.018	0.021	0.018	0.014	0.013	0.017	0.014	0.015	0.012	0.013	0.012
Homem	-0.028*	-0.009	-0.019	-0.005	-0.013	0.001	-0.010	0.007	-0.002	-0.001	-0.000
Matem	0.046***		0.031**	0.020*	0.040***		0.033***		0.027***	0.023*	0.026**
Port		0.041***	0.021*	0.012		0.032***	0.011	0.020**		0.006	0.002
Reprovada				-0.102***				-0.086***	-0.080***	-0.080***	-0.041*
Cooperatividade					0.025**	0.025**	0.024**	0.022*	0.022**	0.021*	0.020*
Assertividade					-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
Empatia					0.000	-0.001	-0.001	0.000	0.001	0.000	0.001
Auto conceito					-0.001	-0.000	0.000	-0.002	-0.002	-0.001	-0.003
Idade											-0.026***
Pr(y=1 x)	0.88	0.88	0.88	0.88	0.89	0.89	0.89	0.89	0.89	0.89	0.89
LR χ^2	52.31	46.29	57.27	107.12	60.98	49.61	61.75	86.32	92.24	92.79	110.45
p-valor	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
R ²	0.032	0.029	0.035	0.067	0.05	0.04	0.05	0.07	0.07	0.07	0.09
Observações	2235	2234	2231	2217	1979	1979	1976	1969	1969	1966	1966

Fonte: elaboração própria. ***significante a 1%. **significante a 5%. *significante a 10%. Valores indicam o efeito marginal da variável explicativa sobre a Pr(y=1). Matem: habilidade de matemática. Port: habilidade de língua portuguesa. mae_EM: mãe com Ensino Médio Completo. Modelo aplicado apenas nos alunos que participaram da coleta do primeiro semestre de 2012, restrito ao quinto e sexto ano.

As tabelas apresentadas nesta presente seção demonstram que habilidades não cognitivas podem trazer um poder preditivo maior aos modelos econométricos para avaliação do fluxo escolar no Brasil. Os resultados mostram que existem habilidades não cognitivas que são significantes mesmo quando controlado pela habilidade cognitiva e por fatores socioeconômicos dos estudantes, para a determinação do fluxo escolar. As formulações apresentadas não são exaustivas, sendo que os efeitos estimados dependem das outras covariadas utilizadas. Porém, as evidências parecem ser suficientes para destacar a importância da inclusão das habilidades não cognitivas em conjunto com as habilidades cognitivas na determinação de resultados de fluxo escolar.

Um exercício adicional foi realizado utilizando a metodologia conhecida como *heckprobit*, em que se realiza uma estimação controlada pela potencial seleção amostral para determinação da permanência escolar, natural de dados dinâmicos como os disponíveis em Sertãozinho - SP, a metodologia já foi utilizada anteriormente para investigação do fluxo escolar brasileiro, como no estudo de Leon e Menezes-Filho (2002). Os resultados deste exercício empírico estão apresentados na seção seguinte.

3.3 Contornando o viés de seleção amostral utilizando dados de 2008

A dinâmica dos dados de Sertãozinho - SP podem conduzir o leitor a se questionar se não existe um viés de seleção amostral nas estimações, natural deste tipo de problema. A metodologia apresentada pode indicar que, dos alunos presentes em 2008, apenas os mais hábeis são justamente aqueles que permaneceram estudando, sendo necessário realizar as estimações corrigindo esse potencial viés, seguindo essa coorte de alunos (e, portanto, descartando os "novos" alunos que entraram nas coletas de 2012 e 2017). Para essa correção, foi utilizada a metodologia *heckprobit*, apresentada na seção metodológica, que segue a ideia da correção de seleção amostral de Heckman (1977), aplicada para modelos não lineares.

Tabela 9 – Regressão da permanência escolar pelo método Heckprobit com variáveis socioemocionais e Locus de Controle

Efeitos Marginais	(1)	(2)	(3)	(4)	(5)	(6)
Abertura	0.010	0.010	0.010			
Conscienciosidade	-0.014	-0.006	-0.007			
Extroversão	-0.020*	-0.019*	-0.021**			
Neuroticismo	-0.002	-0.002	-0.002			
Amabilidade	0.027**	0.026**	0.025**			
Locus Interno				0.021**	0.022**	0.020*
Homem	-0.022	-0.018	-0.019	-0.025	-0.022	-0.021
Branco	-0.039*	-0.038	-0.040*	-0.038*	-0.037*	-0.039*
Matemática	0.01		-0.001	-0.019		0.001
Língua Portuguesa	-0.01	-0.001		0.012	-0.009	
Mãe EM	0.015	0.017	0.015	0.016	0.015	0.015
Pr(y=1 X)	0.88	0.88	0.88	0.88	0.88	0.88
Equação de Seleção	SIM	SIM	SIM	SIM	SIM	SIM
ρ	-0.52	-0.52	-0.52	-0.53	-0.59	-0.43
Teste LR (ρ) - pvalor	0.50	0.58	0.50	0.51	0.40	0.61
Observações	998	999	998	1004	1005	1005

Fonte: elaboração própria. ***significante a 1%. **significante a 5%. *significante a 10%. Valores indicam o efeito marginal da variável explicativa sobre a Pr(y=1). A equação de seleção inclui a nota da prova de 2008 (na mesma disciplina utilizada na equação principal), *dummy* se o indivíduo é branco, *dummy* se o indivíduo é homem e a *dummy* de escolaridade da mãe.

Os resultados das estimações tabelas 9 e 10, que totalizam nove variações diferentes do modelo, trazem inovações frente aos da seção anterior, dando uma importância para a Amabilidade e Extroversão na determinação do fluxo escolar. Porém, a extroversão apresentou sinal de seu efeito marginal negativo sobre a probabilidade do estudante permanecer na escola estudando.

A Extroversão, de fato, aparece na literatura com resultados confluentes sobre os resultados escolares. Poropat (2009) cita duas evidências presentes na literatura que são conflitantes a respeito desta característica: Raad e Schouwenburg (1996) indicam que estudantes extrover-

Tabela 10 – Regressão da permanência escolar pelo método Heckprobit com Habilidades Sociais

Efeitos Marginais	(1)	(2)	(3)
Cooperatividade	0.006	0.006	0.006
Assertividade	0.009	0.009	0.010
Empatia	-0.007	-0.006	-0.008
Auto-Estima	-0.001	-0.001	-0.001
Homem	-0.026	-0.022	-0.022
Branco	-0.038	-0.037*	-0.039*
Matemática	0.016		0.005
Língua Portuguesa	-0.015	-0.002	
Mãe EM	0.015	0.016	0.015
Pr(y=1 X)	0.89	0.89	0.89
Equação de Seleção	SIM	SIM	SIM
ρ	-0.74	-0.73	-0.73
Teste LR (ρ) - pvalor	0.45	0.52	0.52
Observações	1004	1005	1005

Fonte: elaboração própria. ***significante a 1%. **significante a 5%. *significante a 10%. Valores indicam o efeito marginal da variável explicativa sobre a $Pr(y=1)$. A equação de seleção inclui a nota da prova de 2008 (na mesma disciplina utilizada na equação principal), *dummy* se o indivíduo é branco, *dummy* se o indivíduo é home e a *dummy* de escolaridade da mãe.

tidos são mais enérgicos, o que facilitaria o aprendizado; por outro lado, (EYSENCK, 1992) argumenta que essa energia poderia ser utilizada em distrações ao invés de maior uso para estudo. Em uma vasta análise dos *Big Five*, Poporat (2009) sugere que esse efeito pode depender do estágio escolar. De acordo com o estudo, a extroversão, no Ensino Médio, passaria a afetar negativamente os resultados escolares, explicando o sinal encontrado neste presente trabalho.

Os resultados precisam ser analisados com cuidado, visto que dependem de algumas hipóteses, que constituem uma limitação do presente trabalho. Como as variáveis socioemocionais estão disponíveis apenas em 2012, não foram inseridas características não-cognitivas na equação de seleção do modelo *heckprobit*. O viés de variável omitida, em modelos probit, tende a subestimar os outros parâmetros (em valor absoluto), o que, apesar de ser um viés, reforça a significância dos parâmetros encontrados. Além disso, um argumento que favorece a omissão das características não-cognitivas na equação de seleção dos estudantes é a seguinte: quanto mais próximo da idade adulta está o indivíduo, maior o peso dele para a decisão de seguir na escola, frente a decisão dos pais, e assim, melhor seria a importância do socioemocional para tal escolha. No caso de 2008, os indivíduos, ainda no segundo ano do Ensino Fundamental, teriam suas escolhas educacionais quase que totalmente decidida pelos pais, sem sofrer efeito causal direto de suas habilidades não-cognitiva, ou seja, parece razoável assumir que os parâmetros

dos socioemocionais na equação de seleção são iguais a zero. Reforçando isso, a presença da escolaridade da mãe também está relacionada com as características emocionais das crianças, o que pode atenuar eventuais vieses que ainda restarem. Se o modelo escolhido fosse um *Logit*, com distribuição Logística, não é certo que omissão de variáveis tende a subestimar os parâmetros pelo modelo. Nesse caso, pouco poderia-se inferir acerca dos possíveis vieses no *heckprobit*.

A despeito das limitações metodológicas existentes, os resultados parecem ser robustos no sentido de que, em diversas formulações, encontra-se algum poder preditivo para habilidades não-cognitivas. Em especial, as duas variáveis, Extroversão e Amabilidade, estão relacionadas ao quanto o estudante está integrado ao seu grupo de amigos e ao ambiente escolar. Essa evidência estatística corrobora com a ideia do estudo clássico de Tinto (1975), citado na revisão da literatura, em que a integração acadêmica e social do estudante é relacionada a evasão escolar. Portanto, o presente trabalho traz a contribuição de confirmar empiricamente essa ideia do Modelo de Tinto para o Brasil, o que ainda se mostrava como uma lacuna na literatura nacional, mesmo sendo um tema de grande relevância no contexto apresentado.

4 CONCLUSÃO

O alto retorno da educação escolar, somado ao abrupto aumento dos gastos educacionais no Brasil, faz despertar uma curiosidade na literatura pelo fato de que, no país, a evasão no Ensino Médio permanece em patamares elevados. Era de se esperar que com o aumento de gastos educacionais ocorrido nos anos 2000 a escolaridade respondesse significativamente, dado que as restrições orçamentárias das famílias foram aliviadas, o que possibilita mais investimentos educacionais como também um menor abandono de jovens migrando para o mercado de trabalho precocemente.

Nesse cenário, as habilidades não-cognitivas ganharam a atenção da literatura internacional por se mostrarem preditores importantes dos resultados futuros dos alunos, além de serem positivamente associados a melhores desempenhos escolares, provendo um possível canal para explicação da estagnação dos índices de fluxo escolar brasileiros. Uma lacuna para o Brasil se dá em mostrar como as habilidades sociais/socioemocionais desempenham um papel na evasão escolar. A fundamentação para isso está no fato de que a educação pode apresentar retornos heterogêneos e tal heterogeneidade pode depender dos construtos de personalidade, sendo necessária uma exploração empírica para se levantar evidências sobre tais fatos, aproveitando uma oportunidade única no país de utilizar coleta de dados a nível de alunos, realizada na cidade de Sertãozinho em 2008, 2012 e 2017, permitindo produzir evidências sobre medidas de fluxo escolar, relacionando-as com habilidades não cognitivas. Devido a essa possibilidade, o presente trabalho buscou investigar o papel de um conjunto de habilidades não cognitivas (socioemocionais, habilidades sociais e Locus de Controle) sobre o fluxo escolar, escolhendo um conjunto de alunos que idealmente deveriam estar no Ensino Médio.

Os resultados produzidos por diferentes metodologias indicam que Conscienciosidade, a Amabilidade e a Extroversão possuem um papel preditivo tanto sobre a probabilidade do estudante atingir o Ensino Médio quanto sobre a probabilidade de permanência na escola. Entre as formulações utilizadas foi elaborado um modelo de escolha discreta (Probit) com correção para a potencial seleção amostral nos dados, formulação permitida pela dinâmica presente nos dados, que permitiu testar se esse atrito produzia algum tipo de viés de seleção amostral. O resultado indica que não há uma seleção amostral, o que também reforça os resultados produzidos pelo Probit simples.

Uma potencial explicação para a associação das habilidades não cognitivas sobre o fluxo escolar se dá através do fato de que essas habilidades indicam a integração dos estudantes, o sentimento de pertencimento ao grupo, tanto ao ambiente escolar quanto ao grupo de colegas, sendo que essa integração é um determinante da evasão escolar, segundo o Modelo de Tinto. Outra explicação que se mostrou empiricamente significativa seria de que a conscienciosidade, que mostra perseverança e responsabilidade, determinaria a permanência na escola, pois pode mostrar a capacidade dos estudantes de superar obstáculos, o que permite o estudante seguir

estudando mesmo na brusca mudança que se mostra a transição para o Ensino Médio.

A importância desse resultado para a literatura é reforçado pelo fato de esses resultados persistirem mesmo quando controlado pela escolaridade da mãe e outras características demográficas, que funcionam como uma *proxy* da renda da familiar, e controlado pela habilidade cognitiva dos estudantes, tendo nos socioemocionais um poder preditivo maior do que das variáveis cognitivas.

REFERÊNCIAS

- BARROS, Ricardo Paes de et al. Determinantes do desempenho educacional no brasil. Instituto de Pesquisa Econômica Aplicada (Ipea), 2001.
- BECKER, Gary S. Human capital: A theoretical and empirical analysis, with special reference to schooling. *NY: National Bureau of Economic Research*, 1964.
- BECKER, Gary S; CHISWICK, Barry R. Education and the distribution of earnings. *The American Economic Review*, JSTOR, v. 56, n. 1/2, p. 358–369, 1966.
- BEN-PORATH, Yoram. The production of human capital and the life cycle of earnings. *Journal of political economy*, The University of Chicago Press, v. 75, n. 4, Part 1, p. 352–365, 1967.
- CARD, David. *Earnings, schooling, and ability revisited*. [S.l.], 1994.
- _____. Earnings, ability and schooling revisited. *Research in Labour Economics*, JAI Press: Greenwich, CT, v. 14, p. 23, 1995.
- CARNEIRO, Pedro. *Heterogeneity in the returns to schooling: implications for policy evaluation*. Tese (Doutorado) — University of Chicago, Department of Economics August 2003., 2003.
- CHIKITANI, Maurício. *Peer effects on locus of control*. Tese (Doutorado), 2015.
- COMAY, Yochanan; MELNIK, Arie; POLLATSCHEK, Moshe A. The option value of education and the optimal path for investment in human capital. *International Economic Review*, JSTOR, p. 421–435, 1973.
- CUNHA, Flavio; HECKMAN, James J; SCHENNACH, Susanne M. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, Wiley Online Library, v. 78, n. 3, p. 883–931, 2010.
- DUNCAN, Greg J; MAGNUSON, Katherine et al. The nature and impact of early achievement skills, attention skills, and behavior problems. *Whither opportunity*, p. 47–70, 2011.
- EYSENCK, Hans J. Four ways five factors are not basic. *Personality and individual differences*, Elsevier, v. 13, n. 6, p. 667–673, 1992.
- FERREIRA, Francisco HG; BARROS, Ricardo Paes de. The slippery slope: Explaining the increase in extreme poverty in urban brazil, 1976-96. 1999.
- FERREIRA, Francisco HG; LITCHFIELD, Julie A. Education or inflation?: the micro and macroeconomics of the brazilian income distribution during 1981-1995. *Cuadernos de economía*, SciELO Chile, v. 38, n. 114, p. 209–238, 2001.
- FINDLEY, Maureen J; COOPER, Harris M. Locus of control and academic achievement: A literature review. *Journal of personality and social psychology*, American Psychological Association, v. 44, n. 2, p. 419, 1983.
- GOLDBERG, Lewis R. The structure of phenotypic personality traits. *American psychologist*, American Psychological Association, v. 48, n. 1, p. 26, 1993.

- GREENE, William H. *Econometric analysis* (international edition). Pearson US Imports & PHIPES, 2000.
- GRESHAM, Frank M; ELLIOTT, Stephen N. *Social skills rating system (SSRS)*. [S.l.]: American Guidance Service, 1990.
- HECKMAN, J; CAMERON, Stephen. The dynamics of educational attainment for black, hispanic, and white males. *Journal of Political Economy*, v. 109, n. 3, p. 455–499, 2001.
- HECKMAN, James J. *Sample selection bias as a specification error (with an application to the estimation of labor supply functions)*. [S.l.]: National Bureau of Economic Research Cambridge, Mass., USA, 1977.
- HECKMAN, James J; LOCHNER, Lance J; TODD, Petra E. *Fifty years of Mincer earnings regressions*. [S.l.], 2003.
- HECKMAN, James J; RUBINSTEIN, Yona. The importance of noncognitive skills: Lessons from the ged testing program. *The American Economic Review*, JSTOR, v. 91, n. 2, p. 145–149, 2001.
- JOHN, Oliver P; SRIVASTAVA, Sanjay. The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, Guilford, v. 2, n. 1999, p. 102–138, 1999.
- LEON, Fernanda Leite Lopez de; MENEZES-FILHO, Naércio Aquino. *Reprovação, avanço e evasão escolar no brasil*. Instituto de Pesquisa Econômica Aplicada (Ipea), 2002.
- LLERAS, Christy. Do skills and behaviors in high school matter? the contribution of noncognitive factors in explaining differences in educational attainment and earnings. *Social Science Research*, Elsevier, v. 37, n. 3, p. 888–902, 2008.
- MILGRAM, Norman A; MILGRAM, Roberta M. Dimensions of locus of control in children. *Psychological Reports*, SAGE Publications Sage CA: Los Angeles, CA, v. 37, n. 2, p. 523–538, 1975.
- MINCER, Jacob. Investment in human capital and personal income distribution. *Journal of political economy*, The University of Chicago Press, v. 66, n. 4, p. 281–302, 1958.
- _____. *Schooling, experience, and earnings*. human behavior & social institutions no. 2. ERIC, 1974.
- MISCHEL, Walter; SHODA, Yuichi; RODRIGUEZ, Monica L. Delay of gratification in children. *Science*, JSTOR, v. 244, n. 4907, p. 933–938, 1989.
- MOURA, Rodrigo Leandro de. Testando as hipóteses do modelo de mincer para o brasil. *Revista brasileira de economia*, SciELO Brasil, v. 62, n. 4, p. 407–449, 2008.
- NERI, M et al. The effects of idiosyncratic shocks to father's income on child labor, school drop-outs and repetition rates in brazil. *Anais do XXII Encontro Brasileiro de Econometria, SBE*, 2000.
- NERI, Marcelo et al. *Motivos da evasão escolar*. Brasília: Fundação Getúlio Vargas, p. 1–34, 2009.
- _____. _____. 2015.

POROPAT, Arthur E. A meta-analysis of the five-factor model of personality and academic performance. *Psychological bulletin*, American Psychological Association, v. 135, n. 2, p. 322, 2009.

PREARO, Murilo Fernandez. *Impacto de realização de pré-escola sobre locus de controle*. Tese (PhD dissertation) — Escola de Economia de São Paulo/FGV, 2012.

PRIMI, R; SANTOS, D. Desenvolvimento socioemocional e aprendizado escolar: Uma proposta de mensuração para apoiar políticas públicas. resultado preliminares do projeto de competências socioemocionais no rio de janeiro. *Instituto Ayrton Senna. São Paulo*, 2014.

RAAD, Boele De; SCHOUWENBURG, Henri C. Personality in learning and education: A review. *European Journal of personality*, Wiley Online Library, v. 10, n. 5, p. 303–336, 1996.

ROTTER, Julian B. Generalized expectancies for internal versus external control of reinforcement. *Psychological monographs: General and applied*, American Psychological Association, v. 80, n. 1, p. 1, 1966.

SACHSIDA, Adolfo; LOUREIRO, Paulo Roberto Amorim; MENDONÇA, Mário Jorge Cardoso de. Um estudo sobre retorno em escolaridade no brasil. *Revista Brasileira de Economia*, SciELO Brasil, v. 58, n. 2, p. 249–265, 2004.

SPENCE, Michael. Job market signaling. *The quarterly journal of Economics*, MIT Press, v. 87, n. 3, p. 355–374, 1973.

TEIXEIRA, Wladimir Machado; MENEZES-FILHO, Naércio Aquino. Estimando o retorno à educação do brasil considerando a legislação educacional brasileira como um instrumento. *Revista de Economia Política*, SciELO Brasil, v. 32, n. 3, p. 479–496, 2012.

TINTO, Vincent. Dropout from higher education: A theoretical synthesis of recent research. *Review of educational research*, Sage Publications Sage CA: Thousand Oaks, CA, v. 45, n. 1, p. 89–125, 1975.

WEISBROD, Burton A. Education and investment in human capital. *Journal of Political Economy*, The University of Chicago Press, v. 70, n. 5, Part 2, p. 106–123, 1962.

APÊNDICE A – ESTIMAÇÕES DO PROBIT EM SUB-AMOSTRAS

Tabela 11 – Estimação do Probit para probabilidade de se atingir o Ensino Médio na sub-amostra do quinto ano do Ensino Fundamental de 2012

Pr(y_EM=1)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
mae_EM	-0.005	-0.004	-0.012	-0.030	-0.017	-0.017	-0.022	-0.036	-0.034	-0.038	-0.039
Branco	0.012	0.019	0.011	0.014	0.009	0.017	0.010	0.019	0.011	0.013	0.010
Homem	-0.140***	-0.073**	-0.116***	-0.097***	-0.132***	-0.069**	-0.117***	-0.057*	-0.110***	-0.101***	-0.100***
Matem	0.145***		0.106***	0.091***	0.135***		0.108***		0.113***	0.094***	0.095***
Port		0.137***	0.057**	0.039		0.122***	0.043*	0.096***		0.031	0.027
Reprovada				-0.140***				-0.144***	-0.125***	-0.122***	-0.089**
Am					-0.010	-0.011	-0.014	-0.012	-0.012	-0.014	-0.014
Cc					0.044**	0.046**	0.043**	0.043**	0.041**	0.040**	0.041**
Ex					0.013	0.020	0.012	0.020	0.014	0.013	0.013
Nr					-0.011	-0.006	-0.011	-0.002	-0.006	-0.006	-0.005
Op					-0.007	-0.009	-0.008	-0.014	-0.013	-0.013	-0.014
Idade											-0.025
Pr(y=1 x)	0.713	0.714	0.714	0.715	0.738	0.738	0.738	0.740	0.739	0.740	0.740
LR χ^2	134.538	112.264	143.053	167.805	130.319	104.315	134.792	130.176	150.080	153.347	155.805
pvalor	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R ²	0.099	0.083	0.106	0.125	0.111	0.089	0.115	0.112	0.129	0.132	0.134
Observações	1131	1129	1129	1121	1019	1017	1017	1012	1014	1012	1012

Fonte: elaboração própria. ***significante a 1%. **significante a 5%. *significante a 10%. Valores indicam o efeito marginal da variável explicativa sobre a Pr(y=1). Am: Amabilidade. Cc: conscienciosidade. Ex: extroversão. Nr: Neuroticismo. Op: Abertura ao novo. Matem: habilidade de matemática. Port: habilidade de língua portuguesa. mae_EM: mãe com Ensino Médio Completo. Modelo aplicado apenas nos alunos que participaram da coleta do primeiro semestre de 2012, restrito ao quinto e sexto ano.

Tabela 12 – Estimação do Probit para a probabilidade de atingir o Ensino Médio na sub-amostra do sexto ano do Ensino Fundamental de 2012

Pr(y_EM)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Mae_EM	0.029	0.033	0.024	0.009	0.040	0.050	0.038	0.034	0.028	0.027	0.024
Branco	0.010	0.014	0.007	-0.007	0.003	0.010	0.002	-0.003	-0.007	-0.007	-0.003
Homem	-0.103***	-0.069**	-0.084***	-0.062**	-0.085***	-0.060*	-0.074**	-0.043	-0.063**	-0.054*	-0.052*
Matem	0.102***		0.073***	0.052**	0.095***		0.071***		0.070***	0.052**	0.045**
Port		0.077***	0.038**	0.027*		0.073***	0.035*	0.053***		0.028	0.025
Reprovada				-0.169***				-0.157***	-0.148***	-0.144***	-0.074*
Am					0.032*	0.023	0.026	0.020	0.027	0.023	0.024
Cc					0.014	0.018	0.014	0.018	0.015	0.015	0.012
Ex					-0.018	-0.019	-0.021	-0.017	-0.016	-0.018	-0.016
Nr					0.002	0.002	0.001	0.004	0.005	0.004	0.005
Op					-0.006	-0.002	-0.005	-0.005	-0.008	-0.008	-0.008
Idade											-0.046***
Predict	0.804	0.806	0.806	0.807	0.828	0.829	0.828	0.829	0.829	0.829	0.829
LR χ^2	89.983	75.892	95.168	147.612	89.593	76.892	95.198	120.642	127.797	131.392	145.985
pvalor	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R ²	0.082	0.070	0.088	0.137	0.103	0.088	0.109	0.139	0.148	0.152	0.169
Observações	1104	1105	1102	1096	949	952	949	947	944	944	944

Fonte: elaboração própria. ***significante a 1%. **significante a 5%. *significante a 10%. Valores indicam o efeito marginal da variável explicativa sobre a Pr(y=1). Am: Amabilidade. Cc: conscienciosidade. Ex: extroversão. Nr: Neuroticismo. Op: Abertura ao novo. Matem: habilidade de matemática. Port: habilidade de língua portuguesa. mae_EM: mãe com Ensino Médio Completo. Modelo aplicado apenas nos alunos que participaram da coleta do primeiro semestre de 2012, restrito ao quinto e sexto ano.

Tabela 13 – Estimação do Probit para a permanência escolar na sub-amostra do quinto ano do Ensino Fundamental de 2012

Pr(y ₂₀₁₇₌₁)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
mae_EM	0.009	0.005	0.004	-0.007	0.000	-0.002	-0.003	-0.015	-0.012	-0.015	-0.015
Branco	0.016	0.017	0.016	0.017	0.027	0.031	0.029	0.033	0.030	0.033	0.030
Homem	-0.015	0.006	-0.000	0.013	-0.009	0.010	0.002	0.019	0.008	0.015	0.016
Matem	0.041***		0.016	0.006	0.038***		0.018		0.021	0.008	0.008
Port		0.047***	0.035*	0.023		0.044***	0.031*	0.027*		0.022	0.018
Reprovada				-0.096***				-0.093***	-0.093***	-0.091***	-0.061*
Am					0.007	0.003	0.003	0.001	0.004	0.001	0.001
Cc					-0.011	-0.011	-0.012	-0.011	-0.012	-0.012	-0.011
Ex					0.007	0.006	0.005	0.008	0.009	0.007	0.007
Nr					-0.001	-0.000	-0.001	0.001	0.001	0.001	0.001
Op					0.002	0.001	0.001	-0.003	-0.003	-0.003	-0.004
Idade											-0.023*
Pr(y=1 x)	0.88684	0.88753	0.88753	0.88670	0.90092	0.90172	0.90173	0.90122	0.90042	0.90122	0.90125
LR χ^2	20.175	24.611	25.937	47.892	21.532	23.914	25.544	45.152	42.427	45.441	50.336
p-valor	0.000	0.000	0.000	0.000	0.010	0.004	0.004	0.000	0.000	0.000	0.000
R ²	0.025	0.031	0.033	0.060	0.033	0.037	0.039	0.069	0.065	0.070	0.077
Observações	1131	1129	1129	1121	1019	1017	1017	1012	1014	1012	1012

Fonte: elaboração própria. ***significante a 1%. **significante a 5%. *significante a 10%. Valores indicam o efeito marginal da variável explicativa sobre a Pr(y=1). Am: Amabilidade. Cc: conscienciosidade. Ex: extroversão. Nr: Neuroticismo. Op: Abertura ao novo. Matem: habilidade de matemática. Port: habilidade de língua portuguesa. mae_EM: mãe com Ensino Médio Completo. Modelo aplicado apenas nos alunos que participaram da coleta do primeiro semestre de 2012, restrito ao quinto e sexto ano.

Tabela 14 – Estimação do Probit para a permanência escolar na sub-amostra do sexto ano do Ensino Fundamental de 2012

Pr(y ₂₀₁₇₌₁)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
mae_EM	-0.001	0.004	-0.002	-0.012	0.004	0.012	0.004	0.002	-0.003	-0.003	-0.005
Branco	0.019	0.024	0.018	0.010	0.012	0.018	0.012	0.011	0.006	0.006	0.010
Homem	-0.040*	-0.023	-0.035	-0.022	-0.030	-0.017	-0.030	-0.009	-0.020	-0.021	-0.021
Matem	0.060***		0.054***	0.041**	0.058***		0.057***		0.046***	0.047***	0.043**
Port		0.038***	0.009	0.002		0.033***	0.001	0.022*		-0.003	-0.005
Reprovada				-0.106***				-0.089***	-0.077***	-0.077***	-0.024
Am					0.031**	0.027*	0.031*	0.025*	0.028*	0.028*	0.029*
Cc					0.006	0.009	0.006	0.010	0.007	0.007	0.004
Ex					-0.020*	-0.019	-0.021*	-0.017	-0.019	-0.019	-0.016
Nr					0.004	0.005	0.004	0.006	0.005	0.005	0.006
Op					-0.010	-0.007	-0.010	-0.009	-0.011	-0.011	-0.011
Idade											-0.033***
Pr(y=1 x)	0.87678	0.87782	0.87745	0.87860	0.89467	0.89501	0.89466	0.89564	0.89529	0.89529	0.89567
LR χ^2	38.601	24.421	39.230	65.536	46.516	29.170	46.520	47.359	59.271	59.311	70.695
p-valor	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
R ²	0.047	0.030	0.048	0.081	0.073	0.046	0.073	0.075	0.094	0.094	0.112
Observations	1104	1105	1102	1096	949	952	949	947	944	944	944

Fonte: elaboração própria. ***significante a 1%. **significante a 5%. *significante a 10%. Valores indicam o efeito marginal da variável explicativa sobre a Pr(y=1). Am: Amabilidade. Cc: conscienciosidade. Ex: extroversão. Nr: Neuroticismo. Op: Abertura ao novo. Matem: habilidade de matemática. Port: habilidade de língua portuguesa. mae_EM: mãe com Ensino Médio Completo. Modelo aplicado apenas nos alunos que participaram da coleta do primeiro semestre de 2012, restrito ao quinto e sexto ano.