

ESCASSEZ DE ÁGUA E PRIMEIRA INFÂNCIA: UM ESTUDO SOBRE O DESEMPENHO EDUCACIONAL NO SEMIÁRIDO BRASILEIRO

Lauro Nogueira – Universidade Federal Rural do Semi-Árido – DECEN/UFERSA

Erik Figueiredo – Universidade Federal da Paraíba – PPGE/UFPB

Wallace Patrick Santos de Farias Souza – Universidade Federal da Paraíba – PPGE/UFPB

RESUMO

O principal objetivo deste artigo consiste em investigar as implicações da escassez de água e períodos de seca nas etapas iniciais da vida, em especial, a fase intrauterina, sobre os respectivos resultados educacionais das crianças do semiárido brasileiro observados na Prova Brasil nos exames de 2013 a 2017. Para tanto, é assumida a Teoria do Ciclo da Vida, proposta por Cunha e Heckman (2007), e aplicado dois métodos distintos de estimação associando as duas variáveis climáticas (escassez de chuvas e seca) semelhante a abordagem de Rocha e Soares (2015). Os principais resultados indicam a existência de efeitos negativos da escassez de água e altas temperaturas sobre os resultados educacionais dos alunos. Outras características socioeconômicas individuais e coletivas utilizadas como controle também se mostraram significativas na composição dos resultados dos exames. Especificamente, os resultados revelam que menos de 4% do diferencial do desempenho educacional na região são explicados por características pertencentes a escolas e/ou municípios. Por fim, destaca-se que esses achados evidenciam os fatos estilizados estabelecidos na Teoria do Ciclo da Vida, sobretudo, a não existência de substitutos perfeitos na composição das habilidades cognitivas e não cognitivas. Isto é, investimentos em fases posteriores não recompõem os agravos referentes às vulnerabilidades enfrentadas nas fases iniciais, notadamente, na primeira infância.

Palavras-chave: Teoria do Ciclo da Vida; Escassez de Água; Resultados Educacionais.

ABSTRACT

The main objective of this article is to investigate as a result of water scarcity and periods of drought in the early stages of life, in particular, an intrauterine phase, on the respective educational results of children from the Brazilian semiarid observed in Prova Brasil in the 2013 exams a 2017. Therefore, the Life Cycle Theory, proposed by Cunha and Heckman (2007), is adopted and two different estimation methods are applied associating the two climatic variables (scarcity of rain and drought) similar to the Rocha and Soares (2015). The main results indicate the existence of factors that affect water scarcity and the high educational results of students. Other individual and collective socioeconomic characteristics used as controls are also differentiated in the composition of the test results. Specifically, the results reveal that less than 4% of the educational performance differential in the region is explained by characteristics belonging to schools and / or municipalities. Finally, it is highlighted that these findings highlight the stylized facts added in the Life Cycle Theory, above all, the lack of perfect substitutes in the composition of cognitive and non-cognitive skills. That is, investments in later phases do not recombine the aggravations related to the vulnerabilities faced in the initial phases, especially in early childhood.

Keywords: Life Cycle Theory; Water Shortage; Educational Results.

Área 3 – Economia Regional e Urbana

JEL: I0; I1; I2

1. INTRODUÇÃO

A escassez de água e/ou períodos de seca extrema impactam diretamente na redução da produção agrícola, basicamente, dos pequenos produtores, como também, aceleram e elevam o processo de migração forçada. E, quando associado a condições de pobreza, origina diversos outros problemas, por exemplo, difícil acesso à água potável, mal hábitos de higiene pessoal, diminuição da renda, proliferação de doenças, fome e desnutrição, desequilíbrio ambiental, entre outros. Em outras palavras, em ambientes econômicos vulneráveis, a escassez de água decorrente dos choques climáticos pode diminuir o bem-estar social afetando negativamente setores essenciais como: saúde, educação, renda e segurança pública. Particularmente, crianças e mulheres grávidas estão mais susceptíveis a serem atingidas pelo problema (FAO, 2019).

De acordo com Giordano, Barron e Ünver (2019) há duas formas gerais de escassez de água: i) a escassez física: conjuntura onde os recursos hídricos naturais são insuficientes para atender à demanda de uma determinada região; ii) a escassez econômica: decorrente da má gestão dos recursos hídricos. Em ambos os casos, surgem consideráveis implicações negativas sobre a população da região afetada. É plausível destacar que a incidência recorrente da escassez de água e/ou períodos de secas certamente obstruem também níveis de escolaridade, uma vez que, crianças não saudáveis tendem a frequentar a escola de maneira mais irregular, especialmente em áreas rurais.

Segundo relatório do observatório de escassez hídrica vinculado a Agência Nacional de Águas (ANA, 2018), a região semiárida brasileira é formada por uma área de 1,03 milhão de quilômetros quadrados, equivalente a 12% do território nacional. Atualmente, 1.262 municípios pertencentes a dez estados compõem o semiárido: Alagoas, Bahia, Ceará, Maranhão, Minas Gerais, Paraíba, Pernambuco, Piauí, Rio Grande do Norte e Sergipe. Aproximadamente 27 milhões de pessoas residem na região, caracterizada pelo clima seco, altas temperaturas, períodos de chuvas instáveis e/ou insuficientes, rios intermitentes e alta evapotranspiração.

Em estudo relacionado, Rocha e Soares (2015) analisam as implicações da escassez de água e secas durante o período gestacional na saúde das crianças ao nascer no semiárido brasileiro. Os principais resultados sugerem que níveis de chuvas abaixo da média histórica e períodos de secas são fortemente correlacionados com maiores índices de mortalidade infantil, menor peso ao nascer e períodos mais curtos de gestação. Além disso, os achados indicam que esses efeitos são bem maiores durante o período fetal.

Pesquisas recentes e sob óticas distintas (desigualdade de oportunidades) convergem nesse caminho. Isto é, as condições iniciais desfavoráveis causam maiores danos quando ocorrem em períodos intrauterinos e/ou na denominada primeira infância. Pois comprometem a formação das habilidades cognitivas e não cognitivas dos indivíduos. Em síntese, as restrições naturais e sociais impostas nessas etapas da vida são determinantes nas decisões escolares, e, conseqüentemente, afetam salários, empregos, criminalidade, lucratividade das empresas, entre outros fatores não menos importantes (HECKMAN, STIXRUD E URZUA, 2006, HECKMAN, 2007; CUNHA e HECKMAN, 2007; MACCINI e YANG, 2009; HECKMAN E KARAPAKULA, 2019; BAKER, GRUBER E MILLIGAN, 2019).

Posto isso, este estudo tem como principal objetivo estimar o impacto da escassez de água e períodos de seca extrema na fase intrauterina sobre o desempenho educacional observado das crianças do semiárido brasileiro ao final do primeiro ciclo do ensino fundamental, especificamente, nas edições da Prova Brasil: 2013, 2015 e 2017. Para tanto, adota-se uma Função de Produção Educacional (FPE) semelhante a Hanushek (1979), acrescida por algumas especificidades da literatura. Em especial, utiliza-se duas variáveis independentes de interesse – desvios das chuvas em relação à média histórica e uma dummy

caracterizando períodos de seca extrema – semelhantes a abordagem adotada em Rocha e Soares (2015). Em síntese, pretende-se validar a Teoria do Ciclo de Vida (TCV) advogada por Heckman e Cunha (2007).

O conjunto de dados analisados possuem informações mensais sobre precipitação pluviométrica e temperatura média dos municípios que possuem estações meteorológicas físicas referentes ao período de janeiro de 1961 a dezembro de 2018. Esses subsídios são combinados com dados de nascimento e informações socioeconômicas individuais dos alunos do 5º ano que prestaram o exame da Prova Brasil no período analisado. Além disso, associou-se na base de dados atributos das escolas e dos respectivos municípios. A estratégia de identificação assume que os desvios das chuvas observadas e períodos de secas em relação à média histórica para cada município, condicionadas aos efeitos fixos por municípios, escolas e espaço temporal (mês), não estão correlacionadas com outras possíveis determinantes do desempenho educacional (habilidades cognitivas e não cognitivas) durante o período intrauterino, assim como, imediatamente anterior e posterior a gestação.

Admitindo essa hipótese, dois métodos de estimação são estabelecidos a fim de capturar o efeito das variáveis de interesse sobre os respectivos resultados educacionais obtidos: i) seguindo Raudenbush et al. (2004), optou-se por modelos multinível de regressão para dados em painel, a fim de captar a influência da escassez de água sobre os respectivos resultados educacionais, uma vez que, em estudos sobre resultados educacionais, as informações são aninhadas hierarquicamente de forma natural; ii) Quanto as consequências da escassez extrema de água (seca) aplicou-se um método de efeito tratamento análogo ao usado em Heckman e Karapakula (2019), atendendo algumas distinções exigidas.

Os resultados apontam que os choques negativos das chuvas, escassez de água e/ou períodos de seca, estão fortemente correlacionados com os baixos desempenhos educacionais observados no semiárido brasileiro. É importante destacar que esses efeitos somente são válidos estatisticamente durante a fase intrauterina, isto é, em períodos anteriores e posteriores a gestação, os coeficientes não se mostraram significativos. Por outro lado, a temperatura compensada média parece também afetar negativamente o desempenho dos alunos. As demais variáveis de controle (socioeconômicas individuais, escolares e municipais) se mostraram significativas e com efeito esperado. No geral, há indícios que validam a TCV, embora a abordagem somente capte efeitos de curto prazo (resultados educacionais no final do primeiro ciclo do ensino fundamental) no desenvolvimento cognitivo e não cognitivo das crianças.

As evidências encontradas apontam para prováveis consequências de médio e longo prazo já identificados em estudos semelhantes. Por exemplo: Almond e Currie, (2010); Maccini e Yang (2009); Deschenes e Moretti (2009). O restante do artigo obedece ao seguinte planejamento: a seção 2 retrata as principais características do semiárido brasileiro e traz uma síntese teórica sobre a formação das habilidades individuais; a seção 3 destina-se a explicar a identificação empírica, a metodologia, a descrição das variáveis de interesse, como também, o vetor de informações de controle; a seção seguinte é designada a apresentar e discutir os resultados; por fim, a seção 5 é dedicada as considerações finais.

2. ÁREA DE ESTUDO E FORMAÇÃO DAS HABILIDADES

Apresenta-se agora um breve resumo da literatura que trata da formação das habilidades individuais sob a ótica de quatro linhas científicas: psicologia, neurociência cognitiva, economia e epidemiologia. É importante destacar a admissão de dois pontos-chaves: i) o desenvolvimento teórico da formação cognitiva dos indivíduos onde a multidisciplinariedade permite estabelecer um modelo mais realista; ii) anular desigualdades existentes em famílias menos favorecidas é infinitamente mais eficiente quando a intervenção

ocorre nos anos iniciais de vida. No mais, a princípio descreve-se determinadas particularidades socioeconômicas da região estudada.

2.1 O Semiárido Brasileiro

O semiárido é fundamentalmente caracterizado pelas condições climáticas predominantes de semiaridez dado os baixos níveis de precipitação pluviométrica que originam um sistema hidrográfico bastante carente. O período de chuvas é marcado pela ausência de regularidade em toda região ao longo do tempo. Segundo Holanda (2013) os períodos de seca ocorrem no mínimo oito vezes a cada cem anos e permanecem entre três a cinco anos. Além disso, para cada dez anos a região somente apresenta chuvas satisfatórias e regulares em quatro anos. De acordo com relatório apresentado pela Superintendência de Desenvolvimento do Nordeste (SUDENE, 2019) a região retrata uma insolação média de 2.800 h/ano e as temperaturas médias anuais observadas estão entre 23 a 27 C°.

Figura 1 – Semiárido Brasileiro – Nova Delimitação



Fonte: IBGE (2019).

O semiárido ocupa aproximadamente 12% do território nacional (1,03 milhão de km²) com 1.262 municípios classificados nesse perfil. A maior parcela dos municípios é pertencente aos estados nordestinos, aonde o Ceará apresenta a maior incidência do seu território com essas características. O restante da região se estende pela parte setentrional de Minas Gerais (o Norte mineiro e o Vale do Jequitinhonha). Os critérios para delimitação do Semiárido aprovados pelas Resoluções do Conselho Deliberativo da Superintendência do Desenvolvimento do Nordeste – SUDENE – nº 107, de 27/07/2017 e de nº 115, de 23/11/2017 são, respectivamente:

- Precipitação pluviométrica média anual igual ou inferior a 800 mm;
- Índice de Aridez de Thornthwaite ¹ igual ou inferior a 0,50;
- Percentual diário de déficit hídrico igual ou superior a 60% dia/ano.

¹ Esse índice tem sido um dos principais indicadores de deficiência hídrica natural observada em uma determinada região. É calculado com base na razão entre a precipitação média em um determinado tempo pela evapotranspiração potencial médio (HOLZMAN E THORNTHWAITE, 1941).

Segundo o Atlas do Desenvolvimento 2010, a região semiárida brasileira tinha em 60,1% de seus municípios, um Índice de Desenvolvimento Humano Municipal (IDHM) variando entre muito baixo (0 - 0,499) e baixo (0,500 - 0,599), com todos eles registrando um IDHM inferior ao observado no Brasil (0,727). Para outros, 39,2% dos municípios, o IDHM oscilava entre médio (0,600 - 0,699) e somente 0,7% dos municípios apresentava IDHM considerado alto (0,700 - 0,799). Ao analisar os municípios por unidades federativas evidenciou-se que apenas no semiárido cearense, mineiro e potiguar a maior parte dos municípios possuem IDHM médio. Nos demais estados o número dos municípios com baixo IDHM prevalece significativamente.

Observa-se também, que em aproximadamente 64% dos municípios da região o Produto Interno Bruto (PIB) per capita variava entre R\$ 2.783 a R\$ 5.000. E em apenas oito municípios da região registrou-se um PIB per capita superior a R\$ 20.000. Essas informações confirmam a fragilidade econômica e social vivenciada na região.

2.2 Formações das Habilidades

O sucesso econômico individual é usualmente discutido na literatura relacionando-o as características individuais, familiares e sociais. Esses aspectos apresentam-se como essenciais em distintas dimensões de bem-estar social e cada vez mais desperta o interesse em compreender em que medida diferenças observadas de comportamento tem origem no desenvolvimento das habilidades individuais. O grande desafio, contudo, é identificar a formação dessas características e obter uma descrição do comportamento das habilidades.

Na psicologia, Piaget e Cook (1952) destaca dois elementos básicos ao desenvolvimento cognitivo dos indivíduos: i) fatores invariantes: conhecido como gênese, estabelece que ao nascer, o indivíduo recebe como herança uma série de estruturas biológicas (sensoriais e neurológicas) que permanecem constantes ao longo da vida, também conhecida como herdabilidade genética; ii) fatores variantes: advoga que o desenvolvimento cognitivo individual ocorre em função da interação entre o indivíduo e o meio em que ele vive, isto é, o denominado efeito do ambiente compartilhado.

Historicamente, guiados pelas possibilidades de medidas de inteligência como: escala Binet-Simon², (o teste de QI) e Spearman³ (o fator *g*), a literatura psicológica colabora neste rol de discussões ao impetrar que: i) há um processo cumulativo de conhecimento que varia com a idade dos indivíduos; ii) o conhecimento atual é insumo para o aprendizado posterior; e iii) a aprendizagem é captada de forma distinta em diferentes fases da vida, no qual pode haver períodos em que este aprendizado pode ser facilitado.

Por outro lado, a partir do refutado debate denominado “*Nature versus Nurture*” surgem às contribuições da neurociência. Estudos do cérebro e sua relação com a mente, à consciência e a cognição, apontam que o ciclo da vida possui períodos críticos e quanto mais os indivíduos acumulam novos conhecimentos e exercem novas atividades, mais circuitos neurais são criados e desenvolvidos, e de acordo com o ambiente, circunstâncias e experiências a que são expostos, maiores serão suas diferenças comportamentais (FIORI, 2008). Através da literatura epigenética, a neurociência cognitiva defende que a magnitude do estoque das habilidades individuais é susceptível de influências genéticas, mas também ambientais, inclusive as originadas nos períodos intrauterinos.

Ainda que o desenvolvimento cognitivo individual estivesse largamente discutido na literatura psicológica e na neurociência cognitiva, sua evidente relação com os ganhos de produtividade originados pelo “fator humano” na produção somente despertou notório

² Teste moderno de inteligência: escala Binet-Simon (1905).

³ Spearman (1904) propôs um método estatístico (Teoria do Fator *g*) provando que havia nos indivíduos um atributo capaz de captar a estrutura da inteligência.

interesse da literatura econômica a partir da Teoria do Capital Humano, doravante, TCH (MINCER, 1958; SCHULTZ, 1961; BECKER, 1964, entre outros estudos).

Desde então, a TCH passou a ser amplamente utilizada em trabalhos, tanto de ordem microeconômica, quanto em estudos macroeconômicos. Por exemplo, na microeconomia buscou-se identificar os determinantes e papel do capital humano na promoção da própria existência econômica dos indivíduos (HANUSHEK, 1979; BARROS et. al., 2001; FERREIRA e VELOSO, 2003). Por sua vez, estudos macroeconômicos passaram a investigar os ganhos de produtividade e seus efeitos acumulativos sobre a riqueza e bem-estar das nações a partir da TCH (ROMER, 1986; BARRO e LEE, 2013). É consensual que aperfeiçoar a qualidade da força de trabalho é um dos principais fatores para melhoria do ambiente econômico.

Na economia, a TCH adentra em diversas outras áreas de estudo, sobretudo aquelas que enfatizam o comportamento decisório dos agentes econômicos. Em resumo, a formação e determinantes do comportamento individual e/ou o desenvolvimento cognitivo dos indivíduos produz heterogeneidade entre os mesmos, explicando os diferenciais observados de acumulação de capital humano e refletindo consideráveis implicações em termos de políticas públicas (CARNEIRO e HECKMAN, 2003).

Embora haja uma crescente evolução da literatura econômica no que se refere ao desenvolvimento cognitivo dos indivíduos, especialmente, quando se fala em desigualdade de oportunidades⁴, é admissível identificar ao menos duas grandes lacunas: primeiro, de acordo com Cunha e Heckman (2007), a não observância de períodos distintos pressupõe que os insumos na composição das habilidades individuais são substitutos perfeitos. Isto é, ao considerar que a tecnologia de produção das habilidades é idêntica e independente da idade dos indivíduos, assume-se que a formação do capital humano é um processo estático.

A segunda problemática é consequência da primeira, uma vez que, investimentos tardios na formação das habilidades individuais não necessariamente são suficientes para suprir a ausência/carência de investimentos nas fases iniciais, dado que, há um *trade-off* entre equidade e eficiência para investimentos tardios. Estes problemas têm fortes implicações nas políticas públicas e pode gerar resultados não satisfatórios especialmente em ambientes que prevalecem à desigualdade de oportunidades.

Com a finalidade de sintetizar as principais evidências e conclusões relativas às literaturas da psicologia, neurociências e economia, Cunha e Heckman (2007) propõem um quadro teórico admitindo que as habilidades individuais sejam responsáveis por uma parcela significativa da variação observada nos resultados socioeconômicos dos indivíduos. Seguindo fielmente o estudo, o *mix* dessas literaturas permite estabelecer um modelo robusto para explicar seis fatos estilizados:

- i. As diferenças das habilidades individuais e entre os grupos socioeconômicos originam ambientes para políticas de intervenção em idades iniciais da vida. Isto implica que a qualidade das escolas e dos recursos educacionais exercem efeitos relativamente pequenos sobre os déficits das habilidades e apresenta pouco efeito sobre os resultados dos testes das crianças de diferentes grupos socioeconômicos;
- ii. Há períodos sensíveis e críticos no desenvolvimento da criança. Isto é, algumas habilidades são mais facilmente adquiridas em certas fases da infância do que outras características;

⁴ Em acordo com a literatura da Teoria da Justiça (Rawls, 2009) e o reconhecido Axioma de Identificação de Roemer (RIA), (Roemer, 1998) admite-se o sucesso econômico individual são decorrentes de dois conjuntos de fatores: i) variáveis de circunstância (não responsabilidade) – também defendidas na TCH: sexo, raça, renda, etc.; ii) variáveis de esforço (responsabilidade individual), por exemplo, decisão de migrar, anos de estudos, horas trabalhadas, entre outras.

- iii. Apesar dos baixos retornos para intervenções direcionadas em jovens adolescentes carentes, a literatura empírica mostra altos retornos econômicos para investimentos corretivos em jovens e crianças;
- iv. Se os investimentos nas fases iniciais das crianças (primeira infância) desfavorecidas não são acompanhados por investimentos em fase posteriores, o seu efeito em idades futuras será ainda menor;
- v. Efeitos de restrição de crédito sobre os resultados de adultos dependem da idade em que ele surgiu na família da criança. Isto é, renda familiar durante os anos de ensino superior desempenha um papel menor na determinação da participação do indivíduo.
- vi. Habilidades socioemocionais (não cognitivas) promovem habilidades cognitivas e são um importante produto de famílias afortunadas e intervenções governamentais bem sucedidas em famílias desfavorecidas.

Essa dependência desempenha intensas implicações na condução de políticas públicas, especialmente em economias em desenvolvimento. Por exemplo, no Brasil, a maior parte dos investimentos em educação é destinada a programas de incentivo ao acesso do ensino superior ou aqueles programas de formação profissional pública e de alfabetização de adultos. Em geral, há ausência, insignificância e/ou má gestão dos investimentos públicos em educação infantil.

Um dos pontos polêmicos que gera bastante controvérsia entre estudiosos e classes políticas compreende o primeiro fato, uma vez que, não atenta contra a importância da qualidade das escolas ou recursos educacionais, apenas adverte que investimentos educacionais nos primeiros anos de vida apresentam retornos mais expressivos (eficiência e efetividade) na condução de uma política pública que objetive igualar as oportunidades individuais. Existem outros pontos extremamente importantes na discussão proposta por Cunha e Heckman (2007), e que serão consideradas no desenvolvimento desse trabalho.

3. ASPECTOS METODOLÓGICOS

3.1 Identificação Empírica

A hipótese central assumida nesse trabalho é que os problemas relacionados à escassez de água na região do semiárido brasileiro, especificamente, na fase intrauterina podem afetar o desenvolvimento das capacidades/habilidades, e, portanto, o desempenho educacional individual. Assim sendo, a estratégia empírica adotada está baseada na Teoria do Ciclo da Vida (TCV) de formação das habilidades individuais proposta por Cunha e Heckman (2007), a qual defende que o desenvolvimento dessas habilidades satisfaz um processo dinâmico de múltiplos estágios/períodos. Onde cada período é formado por distintas taxas de retornos.

Em síntese, as habilidades cognitivas e não cognitivas individuais estão sujeitas a um complexo conjunto de características genéticas, sociais e econômicas. Formalmente, identificada na Equação (1).

$$\theta_{t+1}^k = f_t^k(\theta_t^C, \theta_t^N, I_t^k, h^C, h^N), \quad k \in \{C, N\} \quad (1)$$

Onde θ_{t+1}^k representa os resultados imputados a habilidades observados no período t . Sendo $\theta_t = \theta_t^C, \theta_t^N$ respectivamente as habilidades cognitivas e não cognitivas.⁵ Por sua vez, $I_t = I_t^C, I_t^N$ simula o vetor de investimentos cognitivos, não cognitivos e capacidades de saúde. Permita $h = h^C, h^N$ expressar as habilidades parentais cognitivas e não cognitivas. De

⁵ Spearman (1904) relata que a habilidade cognitiva geral demonstra a capacidade que os indivíduos adquirem competências e interagem em seu ambiente. Por exemplo, dominar linguagens, compreender fenômenos, enfrentar situações e problemas, construir argumentações, elaborar propostas, entre outras aptidões individuais.

modo que, a equação (1) relata que para cada período t , pode-se determinar uma tecnologia recursiva para habilidades cognitivas ($k = C$) e habilidades não cognitivas, ($k = N$).

Assume-se também que os indivíduos possuem um vetor de capacidades em cada período da sua vida, usualmente, representado pela idade. Inicialmente, já a partir da concepção, a criança herda qualidades iniciais genéticas e vivencia as condições ambientais, θ_1 (PLOMIN e SPINATH, 2004). Além disso, para cada investimento familiar, I_t , gera-se um vetor das habilidades, em que as capacidades parentais, h (por exemplo: QI, genes, educação, renda, entre outras), originam *outputs* de investimentos de genes e de seus próprios pais. Sendo a tecnologia expressa em (1), na forma recursiva, quando o indivíduo possui t anos de idade:

$$\theta_{t+1} = f_t(h, \theta_t, I_t), \quad (2)$$

Sendo $t = 1, 2, 3, \dots, T$. Para tanto, admite-se f_t uma função crescente e estritamente côncava em I_t , e duas vezes continuamente diferenciáveis para todos os argumentos. Assim, trocando em (2) por $\theta_t, \theta_{t-1}, \dots$; repetidamente, pode-se obter o estoque de recursos na fase $t + 1, \theta_{t+1}$, em função de todos os investimentos do passado:

$$\theta_{t+1} = m_t(h, \theta_1, I_1, \dots, I_t), \quad t = 1, 2, 3, \dots, T \quad (3)$$

Além do mais, postula-se:

- i. Autoprodutividade: habilidades adquiridas em períodos anteriores incrementam as habilidades posteriormente desenvolvidas: $\frac{\partial f_t(h, \theta_t, I_t)}{\partial \theta_t} > 0$.
- ii. Complementariedade Dinâmica: habilidades desenvolvidas em um período aumentam a produtividades em períodos posteriores: $\frac{\partial^2 f_t(h, \theta_t, I_t)}{\partial \theta_t \partial I_t} > 0$.

Excepcionalmente, segundo o estudo, esse processo tecnológico permite observar períodos sensíveis e críticos na formação das habilidades. Agora, admita t^* como período sensível a θ_{t+1} , de maneira que;

$$\frac{\partial \theta_{t+1}}{\partial I_s} = \frac{m_t(h, \theta_1, I_1, \dots, I_t)}{\partial I_s} \equiv 0 \quad \forall \quad h, \theta_1, I_1, \dots, I_s, \quad s \neq t^* \quad (4)$$

Todavia,

$$\frac{\partial \theta_{t+1}}{\partial I_{t^*}} = \frac{m_t(h, \theta_1, I_1, \dots, I_t)}{\partial I_{t^*}} > 0 \quad \text{para alguns } h, \theta_1, I_1, \dots, I_t \quad (5)$$

As equações (4) e (5) respectivamente informam que o período t^* é sensível em relação ao período s se, ao mesmo nível de insumos, o investimento no estágio t^* tem maior retorno produtivo em relação a outro estágio, $s \neq t^*$. A ideia central proposta é que exista considerável estabilidade no período sensível em função da influência do ambiente. Em outras palavras, crianças nascidas em ambientes restritos tendem a desenvolver habilidades menos produtivas, e, mesmo que se invista em períodos posteriores, conhecidos como períodos críticos, nunca constituíram um substituto perfeito dos períodos iniciais. Assim, o estoque de capacidades $h' = (\theta_{T+1})$ das crianças analisadas é uma função das características dos pais, das condições iniciais e dos investimentos ocorridos durante a primeira infância I_1 e I_2 :

$$h' = m_2(h, \theta_1, I_1, I_2), \quad (6)$$

Adota-se aqui, uma versão simplificada do modelo proposto por Cunha e Heckman (2007), para dois períodos/estágios ($t = 2$) seguido por um período φ após o termino da primeira infância $\{t \in (T = 1, T + 2, T + \varphi)\}$. Os resultados educacionais obtidos pelas crianças, $y_{T+\varphi}$ são provenientes de um *mix* das habilidades distintas, no estágio inicial do período $T = 1$ e $T + 2$:

$$y_{T+\varphi} = f_t(\theta_t, \theta_p, I_t, \eta_{k,t}), \quad (7)$$

Onde o estoque das habilidades evolui da seguinte forma: os indivíduos nascem com condições iniciais θ_1 . Sendo, o primeiro período $T = 1$ corresponde à fase intrauterina dos

indivíduos, a qual é influenciada pelas condições iniciais (log dos desvios das chuvas nos 12 meses que antecedem o nascimento) e fatores genéticos (características como sexo e raça), θ_t .

Por sua vez, $h = \theta_p$ retrata as habilidades parentais cognitivas e não cognitivas (renda familiar, escolaridade dos pais e trabalho remunerado). No tocante à $T + 2$ assume-se ser o estágio compreendido entre o final da primeira infância (acima de 6 anos) e a prestação do exame da Prova Brasil (9 a 12 anos). E I_t origina um vetor das habilidades em $T + \varphi$, I_2 .

3.2 Parâmetros de Interesse

A construção das variáveis de clima para obtenção dos parâmetros de interesse é semelhante à estratégia seguida em Rocha e Soares (2015). Diferindo essencialmente em dois pontos: campo de estudo (todo o semiárido, incluindo municípios mineiros e obedecendo a nova delimitação 2017) e fonte de dados (dados históricos do Instituto Nacional de Meteorologia – INMET). As duas variáveis representativas da flutuação dos níveis de chuvas durante o período de gestação de um indivíduo seguem as seguintes equações:

- i. **Log dos desvios das chuvas dos últimos 12 meses:** a percentagem de desvio da precipitação média:

$$P_{i\tau} = \ln \left(\sum_{t=\tau-11}^{\tau} p_{im} \right) - \ln (\bar{p}_i) \quad (8)$$

Onde p_{im} representa a precipitação pluviométrica mensal no município i e mês m ; \bar{p}_i relata a precipitação anual média histórica, considerando o período mensal de 1961 a 2018, no município i ; τ indica o mês de nascimento do indivíduo.

- ii. **Seca dos últimos 12 meses:** dummy que assume valor 1 se houve seca no período. Em resumo procura-se capturar condições extremas de escassez de água:

$$S_{i\tau} = 1 \Leftrightarrow \sum_{t=\tau-11}^{\tau} p_{im} < (\bar{p}_i - p_i^\sigma) \quad (9)$$

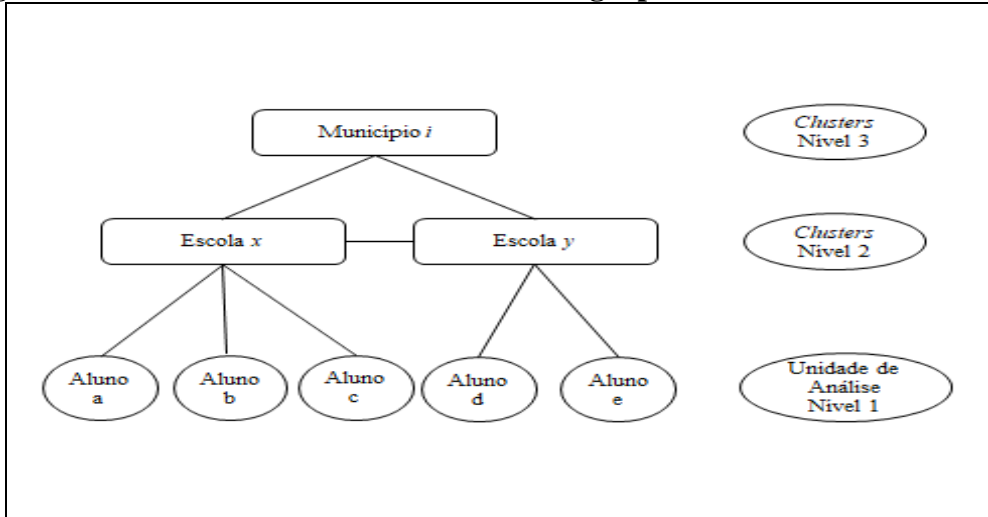
Sendo p_i^σ o desvio padrão histórico anual, considerando o período mensal analisado entre 1961 a 2018. Em síntese, a variável dummy seca ($S_{i\tau} = 1$) assumira valor igual a um, quando o volume total dos níveis de precipitações pluviométricas ao longo dos 12 meses que antecedem o nascimento de um indivíduo for mais do que um desvio padrão abaixo da média histórica para o município i .

3.2.1 Modelos Multinível Linear Hierárquico – Desvios das chuvas

Para apurar a relação entre o volume de precipitação de chuvas (desvios das chuvas) e o desempenho educacional observado durante e após o período intrauterino optou-se por empregar os Modelos Multinível Linear Hierárquico, denominados HLM. A escolha pelo HLM decorre da existência de um aninhamento natural na base de dados: proficiência e informações socioeconômicas dos alunos oriundos de escolas geridas pelos respectivos municípios pertencentes ao semiárido brasileiro. Em síntese, aplica-se um modelo linear hierárquico de três níveis com dados agrupados, dada as seguintes características: alunos (nível 1) agrupados em escolas (nível 2) localizadas em municípios do semiárido brasileiro (nível 3).

Conforme destaca Raudenbush et al. (2004), fenômenos que envolvem subsídios de desempenho e/ou questões sociais e comportamentais, habitualmente possuem estruturas aninhadas. E cada nível dessa arquitetura expressa uma relação com variações residuais distintas entre os diversos níveis existentes. A seguir a Figura 2 retrata esquematicamente o HLM.

Figura 2 – Modelo Multinível Linear Dados Agrupados em Três Níveis– HLM3



Fonte: Elaboração dos autores.

De forma semelhante, Soto e Morera (2005), relatam a existência hierárquica natural em conjuntos de dados de caráter social e considera a adoção dos HLM uma grande vantagem para investigações dessa natureza. Dado que, designam submodelos que retratam relações internas entre as variáveis dentro de cada nível e explica como as mesmas contribuem para as relações presentes em outros níveis.

Vale salientar que os períodos em que alunos, escolas e municípios são observados ocorrem de forma absoluta, isto é, cada aluno e escola estar associado somente a um grupo, isto é, a um único município. Especificamente, serão contemplados diversos cenários para um modelo hierárquico linear com dados agrupado em três níveis, denominado HLM3, seguindo literalmente West, Welch e Galecki (2014). Inicialmente, tem-se:

$$Y_{ijk} = \pi_{0jk} + \sum_{p=1}^P \pi_{pjk} \cdot Z_{pjk} + e_{ijk} \quad (10)$$

Onde o vetor de covariadas (variáveis explicativas) na expressão (10) $Z = z_1, \dots, z_p$ representam o nível 1 associado a observação $i = 1, \dots, I$ de nível 1. Em contrapartida, π_{pjk} , com $p = 0, \dots, P$ referem-se aos coeficientes de nível 1; Z_{pjk} é uma p -ésima variável explicativa de nível 1 para a observação i (aluno) na unidade de nível 2 j (escola) e na unidade de nível 3 k (município). Por sua vez, e_{ijk} representa o termo de erro nível 1 assumindo distribuição normal, média zero e variância amostral s^2 .

De forma análoga, em (11) as variáveis explicativas do vetor $X = x_1, \dots, x_q$ são alusivas às unidades de nível 2 $j = 1, \dots, J$.

$$\pi_{pjk} = b_{p0k} + \sum_{q=1}^{Q_p} b_{pqk} \cdot X_{qjk} + \varepsilon_{pjk} \quad (11)$$

b_{p0k} , sendo $q = 0, \dots, Q_p$, expressam os coeficientes de nível 2; X_{qjk} representa uma q -ésima variável explicativa de nível 2 para a unidade j na unidade de nível 3 k ; Por outro lado, ε_{ijk} exibem os efeitos aleatórios do nível 2, admitindo para cada unidade j , que o vetor $\varepsilon_{0jk}, \dots, \varepsilon_{pjk}$ possui distribuição normal multivariada, média zero e variância $\tau_{\varepsilon ppp}$.

Por fim, no terceiro nível (12), as variáveis explicativas $W = w_1, \dots, w_s$ referem-se as unidades $k = 1, \dots, K$.

$$b_{pqk} = \gamma_{pq0} + \sum_{s=1}^{S_{pq}} \gamma_{pqs} \cdot W_{sk} + u_{pqk} \quad (12)$$

De forma que, γ_{pqs} para $S = 0, \dots, S_{pq}$ estabelece os coeficientes de nível 3; W_{sk} denota uma s -ésima variável explicativa de nível 3 para a unidade k ; u_{pqk} expõem os efeitos aleatórios do nível 3, assumindo que para cada unidade k , o vetor composto pelos termos u_{pqk} conota distribuição normal multivariada e cada elemento apresenta média zero e variância $\tau_{u\pi pp}$, resultante na matriz de variância-covariância T_b com dimensão máxima igual a:

$$\text{Dimensão}_{\text{máx}} T_b = \sum_{p=0}^P (Q_p + 1) \cdot \sum_{p=0}^P (Q_p + 1) \quad (13)$$

Dependente da quantidade de coeficientes do nível 3 especificados com termos aleatórios.

Sinteticamente, os HLM possibilitam analisar o comportamento de determinada variável dependente Y , que expressa o parâmetro de interesse em relação a certas variáveis independentes (explicativas) sujeitas a variações, para dados agrupados, entre observações e entre grupos a que competem essas observações. Para tanto, é preciso haver variáveis que expressam elementos que se alteram entre indivíduos pertencentes a determinado nível, mas, permanecem inalteradas para determinados grupos de indivíduos que constituem um nível superior.

Outra vantagem expressiva em comparação a modelos tradicionais de regressão, por exemplo, modelos usuais de dados em painel (efeitos fixos ou aleatórios) refletem-se na adequação natural dos dados. Isto é, os HLM permitem identificar e analisar as heterogeneidades individuais e entre grupos a que compõem esses indivíduos tornando possível considerar partes aleatórias em cada nível da análise. Atendendo esse aspecto os HLM lineares também são denominados modelos de coeficientes aleatórios

3.2.2 Efeito Tratamento Quantílico – Seca

Para estimar a relação da segunda variável (seca) sobre o desempenho educacional segue-se literalmente a expressão (14) proposta por Heckman e Karapakula (2019). Destaca-se que a abordagem agora considerada reflete um exercício diferente com a aplicação de uma metodologia distinta. Isto é:

$$y_i(b, a) = \alpha + \beta D_i + \gamma \cdot Z_i + W_i(b, a) + \varepsilon_i(b, a) \quad (14)$$

Onde $y_i(b, a)$ denota o resultado de interesse na idade (a) para a criança (i) nascida no ano (b); Por sua vez, D_i expressa o status de tratamento da criança (i). Em outras Palavras, $D_i = 1$ se nos doze meses que antecedem o nascimento da criança (i) prevaleceram períodos de seca, $D_i = 0$ caso contrário; Z_i expressa o vetor das covariáveis que antecedem o período de tratamento (seca). Por exemplo, características municipais, precipitação 13 a 24 meses antes do nascimento, entre outras; W_i expressa o vetor de características da criança nascida no ano (b) na idade (a). Por fim, $\varepsilon_i(b, a) = v_i + u_i(b, a)$ onde v_i representa um efeito fixo para o i tratado e $u_i(b, a)$ é um termo de erro específico para o corte de nascimento e idade da criança. E,

$$y_i(b, a) = (1 - D_i)y_i^0(b, a) - D_i y_i^1(b, a) \quad (15)$$

O efeito do tratamento ω_i na idade a para uma criança nascida no ano b é:

$$\omega_i(b, a) = y_i^1(b, a) - y_i^0(b, a) \quad (16)$$

A opção por essa estratégia empírica baseia-se justamente no fato de que condições extremas negativas (secas) ocasionam externalidades que permeiam seus efeitos nocivos durante períodos ou fases posteriores da vida, sobretudo, em momentos logo após a primeira

infância. Nesse sentido, acredita-se haver uma exímia sintonia entre o objetivo proposto, a literatura da primeira infância *versus* resultados educacionais e a metodologia adotada.

3.3 Dados e Descrição das Variáveis

O conjunto de dados é referente aos alunos do quinto ano do ensino fundamental natural e residentes (não migrantes) dos 74 municípios do semiárido que possuem respectivamente estações físicas de coleta meteorológica e informações referentes aos relativos exames educacionais⁶. A variável dependente é a proficiência obtida na Prova Brasil (média geral, matemática e português). A base amostral é composta por 136.237 alunos na faixa etária entre 9 a 12 anos de idade pertencentes a praticamente todos os estados do Nordeste brasileiro, como também, a alguns municípios localizados ao norte do estado de Minas Gerais⁷. A Tabela 1 retrata a estruturação de acordo com os três níveis de variáveis consideradas.

Tabela 1 – Conjunto de Dados Agrupados (Aluno, Escola e Município)

Classificação	Variável	Descrição
Nível Individual	Prova Brasil/SAEB	Log da nota (média geral, português e matemática).
Nível 1 – Alunos (variáveis de controle)	Sexo	Dummy: 0 representa as mulheres e 1 homens.
	Raça	Dummy: 0 representa não brancos e 1 brancos.
	Idade	Log da idade dos alunos faixa etária: 9 a 12 anos.
	Semestre nascimento	Dummy: 0 nascidos no 2º semestre e 1 no 1º.
	Trabalho infantil	Dummy: 0 trabalha e 1 não trabalha.
	Pré-escola	Dummy: 0 não frequentou e 1 frequentou.
	Reprovação	Dummy: 0 possui reprovação e 1 não possui.
	Evasão	Dummy: 0 se já abandonou o ano e 1 caso contrário.
	Família Biparental	Dummy: 0 família monoparental e 1 biparental.
	Motivação Escolar	Dummy: 0 não motivado e 1 motivado.
	Educação Parental	Log da soma dos anos de estudos (mãe e pai).
	Hiato Educ. Parental	Diferença absoluta dos anos de estudos (mãe e pai).
	Renda (Nível)	Dummy: 0 menor que a média e 1 caso contrário.
	Dprtl_12_nasc.	Desvios das chuvas dos últimos 12 meses (antes).
Dprtl_13_24_nasc.	Desvios das chuvas: 13º ao 24º mês (antes).	
Dprtn_1_12_nasc.	Desvios das chuvas dos 12 meses (após).	
Seca_12_nasc.	Dummy: 1 período de seca e 0 caso contrário (antes).	
TCM_12_nasc.	Log da Temperatura compensada média (antes).	
Nível 2 - Escola	Dependência	Dummy: 0 pública e 1 privada.
	Localização	Dummy: 0 zona rural e 1 zona urbana.
	IDEB (escola)	Log da nota média proficiência SAEB da escola.
Nível 3 – Municípios Variáveis Municipais	IFDM	Índice Firjan de Desenvolvimento Municipal.
	IDEB (município)	Índice de Desenvolvimento da Educação Básica.
	PIB per capita	Log do Produto Interno Bruto por habitante.
	População	Log da população municipal.

Fonte: Elaboração dos autores.

As características socioeconômicas individuais (raça, renda familiar, educação parental, sexo, naturalidade, etc.) e atributos escolares (dependência administrativa, índice de

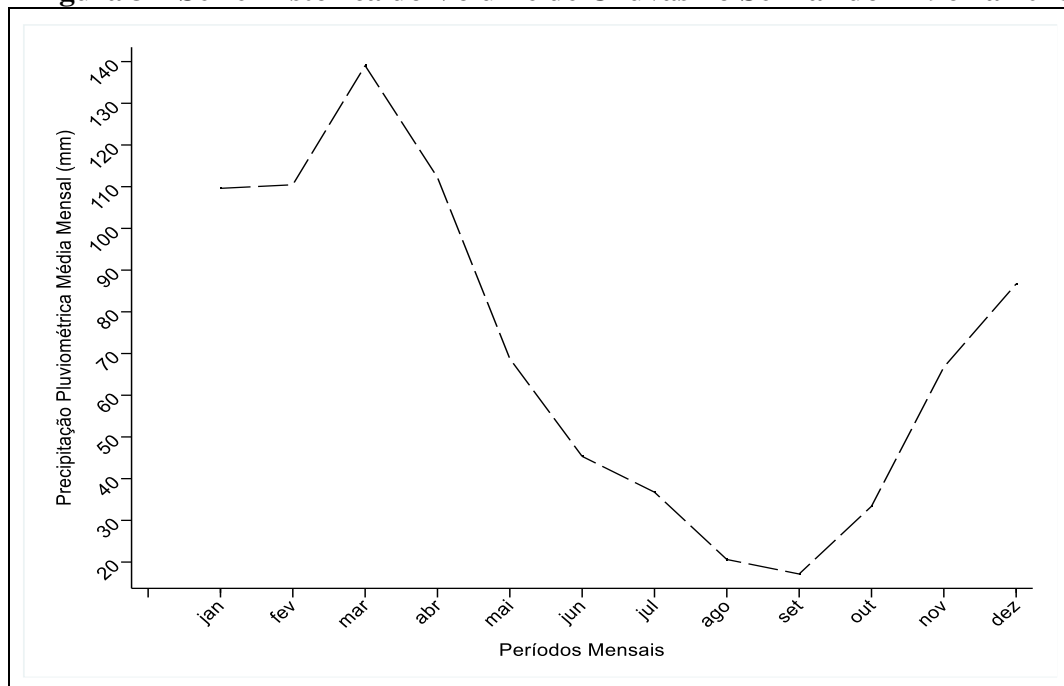
⁶ Dos 75 municípios do semiárido nacional que possuem estações físicas de coletas de informações meteorológicas somente o município de Iguatu/CE não continha informações disponíveis referentes à Prova Brasil 2015.

⁷ Exceto o estado do Maranhão, uma vez que, apenas dois municípios maranhenses (Araioses e Timon) pertencem à região semiárida brasileira e em ambos não há estações meteorológicas físicas.

desempenho da educação básica, etc.) foram coletadas do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) proveniente dos exames: Prova Brasil, Saeb e Censo Escolar (edições 2013, 2015 e 2017). Os dados referentes ao PIB per capita e população pertencem a base do Instituto Brasileiro de Geografia e Estatística (IBGE). Quanto as informações sobre o IFDM, foram obtidos na Federação das Indústrias do Rio de Janeiro (FIRJAN).

Por fim, os dados climáticos observacionais foram extraídos da base do Instituto Nacional de Meteorologia (INMET) referente a períodos mensais de janeiro de 1961 a dezembro de 2018. A escolha por essa fonte de dados decorre essencialmente da maior precisão nas informações. Conforme observa-se na Figura 3 há um período crítico de baixos índices de chuvas na região semiárida entre os meses de maio a novembro. Os picos de chuvas (volume médio histórico) constatados ocorrem entre janeiro a abril.

Figura 3 – Série Histórica do Volume de Chuvas no Semiárido – 1961 a 2018



Fonte: Elaboração dos autores.

4. RESULTADOS E DISCUSSÕES

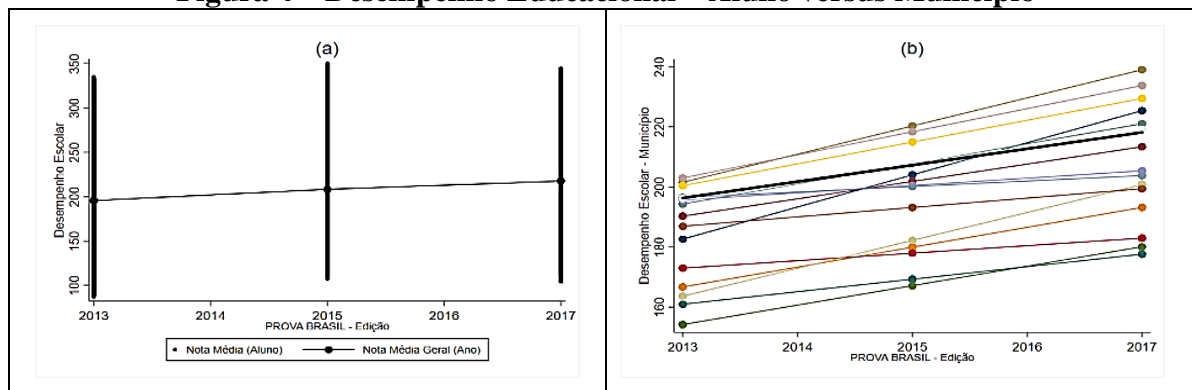
Há ao menos duas consideráveis inovações na proposta: i) relacionar problemas de escassez de água e períodos de seca vivenciada nos doze meses que antecedem o nascimento de um indivíduo a seus resultados educacionais; ii) correlacionar situações desfavoráveis no período intrauterino com a realização educacional observada posteriormente a primeira infância. Nesse intuito, os resultados foram divididos em dois blocos, a subseção 4.1 destina-se a identificar os efeitos referentes à escassez de água sobre os resultados educacionais. Já a 4.2 reserva-se a analisar a relação entre seca e desempenho educacional.

4.1 Desvios de Chuvas e Resultados Educacionais

De acordo com West, Welch e Galecki (2014), sistemas educacionais são exemplos característicos de estrutura agregada hierárquica de três níveis, onde os alunos estão aninhados em turmas pertencentes a escolas agrupadas em um determinado distrito educacional e/ou município. A Figura 4 contém algumas características importantes da

variável dependente (desempenho educacional) que fortalecem o uso do HLM3. Em primeiro lugar, observa-se uma tendência de crescimento linear, embora varie por aluno, escola e município. Conforme mostrado em 4(b) existe diferentes interceptos e inclinações tanto a nível de escola quanto a nível de município.

Figura 4 – Desempenho Educacional – Aluno versus Município



Fonte: Elaboração dos autores.

A princípio estimou-se um modelo não condicional (modelo nulo) a fim de identificar a existência de variabilidade no desempenho educacional entre os estudantes de uma determinada escola, como também, pertencentes a escolas distintas, oriundo de um mesmo município e/ou de municípios diferentes. É importante destacar que o modelo nulo não está condicionado a nenhuma variável explicativa (a nenhum efeito fixo), além do intercepto, embora ainda seja condicional aos efeitos aleatórios. Em síntese, o modelo nulo inclui um intercepto geral fixo, efeitos aleatórios associados ao intercepto para escolas nos respectivos municípios e efeitos aleatórios associados ao intercepto para os municípios.

A Tabela 2 mostra a significância estatística de todos os coeficientes da decomposição amostral da variância para o modelo incondicional. Os resultados ocorrem convergentes com a literatura e informam que a média geral esperada dos alunos se situa em 193,26 pontos. Naturalmente, a maior parcela da variação total é explicada pelas próprias características dos discentes. Em números, quase 71% da variabilidade total no desempenho observado são determinadas por um conjunto de fatores individuais.

Tabela 2 – Decomposição da Variância – Modelo Nulo

Efeito Fixo	Coefficiente	Desvio Padrão	$p > z $
Média Geral (γ_{000}) log	5,264***	0,011	0,000
Efeito Aleatório		Componente da variância	
Varição entre Alunos		0,0356***	
Varição entre Escolas		0,0047***	
Varição entre Municípios		0,0098***	
Decomposição da Variância		% por Nível	
Nível 1 (Alunos)		70,96	
Nível 2 (Escolas)		9,48	
Nível 3 (Municípios)		19,56	

Fonte: Elaboração dos autores.

Notas: *** $p < 0,01$; ** $p < 0,05$.

Por outro lado, apenas 9,48% da variação dos desempenhos analisados são explicadas pelas diferenças entre as escolas de um mesmo município, contra 19,56% decorrentes de escolas de municípios diferentes. Esses achados são fundamentais para embasar a aplicação do modelo hierárquico de dados agrupados, especialmente, pela rejeição da hipótese nula que os interceptos aleatórios sejam iguais à zero. Para escolha do modelo mais ajustado adota-se a

metodologia “*step-up strategy*”. De acordo com Raudenbush (2004) esse método de identificação de modelos multiníveis lineares para dados agrupados é um artifício de expansão que considera múltiplos panoramas, significância estatística dos parâmetros nos diversos níveis existentes, como também, associa critérios estatísticos para tomada de decisão de qual modelo oferece o maior ajuste ao problema proposto.

Fundamentado nos resultados gerais do modelo HLM3 exibidos na Tabela 3, o desempenho médio esperado dos alunos na avaliação de matemática (115,24), língua portuguesa (75,49) e média geral (95,77) são bem abaixo do percentual de escala considerada satisfatória. Destaca-se que todos os coeficientes referentes às variáveis de controle do nível 1 (alunos) da análise mostraram-se estatisticamente válidos. Apenas as variáveis localização da escola (nível 2: escola), IDEB e PIB municipal (nível 3: municípios) apresentaram coeficientes de efeitos fixos não diferentes de zero estatisticamente.

Focando nas variáveis de interesse (parâmetros climáticos), em todos os cenários estudados há fortes indícios que a escassez de chuvas durante os doze meses que antecedem o nascimento influencia negativamente os resultados educacionais de curto prazo das crianças do semiárido brasileiro. Esses resultados são subsidiados pela Teoria do Ciclo da Vida proposta por Cunha e Heckman (2007). É importante ressaltar que esse problema é consideravelmente maior no desempenho de língua portuguesa.

Tabela 3 – Resultados do Modelo HLM3 – Interceptos Aleatórios

Parâmetros: Efeito Fixo	Coef.	DP	Z
β_0 Intercepto	4,747***	0,085	55,74
β_1 Precipitação: 12 meses antes	0,011**	0,005	2,03
β_2 Precipitação: 13 a 24 meses antes	-0,006	0,004	-1,42
β_3 Precipitação: 1 a 12 meses após	-0,011	0,006	-1,84
β_4 Temperatura: 1 a 12 meses antes	-0,046***	0,014	-3,30
Parâmetros: Efeito Aleatório	Coef.	DP	Z
σ^2 Município (1.87%)	0,0006576***	0,0001429	4,60
σ^2 Escola (1.43%)	0,0005006***	0,0000725	6,90
σ^2 Variância residual (96.7%)	0,0338344***	0,0002499	135,39
Parâmetros: Efeito Fixo	Coef.	DP	Z
β_0 Intercepto	4,324***	0,098	43,86
β_1 Precipitação: 12 meses antes	0,015***	0,006	2,63
β_2 Precipitação: 13 a 24 meses antes	-0,002	0,005	-0,47
β_3 Precipitação: 1 a 12 meses após	-0,008	0,006	-1,24
β_4 Temperatura: 1 a 12 meses antes	-0,045***	0,015	-2,93
Parâmetros: Efeito Aleatório	Coef.	DP	Z
σ^2 Município (2.97%)	0,0012077**	0,0002573	4,69
σ^2 Escola (1.23%)	0,0005003**	0,0000809	6,18
σ^2 Variância residual (95.80%)	0,0390235**	0,0002884	135,31
Parâmetros: Efeito Fixo	Coef.	DP	Z
β_0 Intercepto	4,562**	0,081	56,19
β_1 Precipitação 12 meses antes	0,013**	0,005	2,73
β_2 Precipitação 13 a 24 meses antes	-0,004	0,004	-0,99
β_3 Precipitação 1 a 12 meses após	-0,008	0,005	-1,61
β_4 Temperatura 1 a 12 meses antes	-0,043**	0,012	-3,38
Parâmetros: Efeito Aleatório	Coef.	DP	Z
σ^2 Município (2.52%)	0,0007295**	0,0001524	4,78
σ^2 Escola (1.46%)	0,0004221**	0,0000613	6,88
σ^2 Variância residual (96.02%)	0,0278072**	0,0002055	135,31

Fonte: Elaboração dos autores.

Notas: *** $p < 0,01$; ** $p < 0,05$.

Um resultado bem interessante refere-se à temperatura compensada média, isto é, quanto maior a temperatura observada nos doze meses que antecedem o nascimento das

crianças do semiárido brasileiro menor o desempenho médio alcançado na Prova Brasil. Esses achados associados à não significância estatística dos períodos anteriores aos doze meses que antecedem o nascimento, como também, aos doze meses posteriores ao nascimento permitem inferir que fatores climáticos como, escassez de água e altas temperaturas, somente exercem efeitos estatístico comprovado sobre o desempenho educacional das crianças na vida intrauterina. Em síntese, os resultados indicam haver importantes implicações negativas sobre os resultados educacionais das crianças associadas à escassez de água e altas temperaturas enfrentadas na fase intrauterina na região semiárida brasileira.

Esses achados podem ser considerados bastantes intrigantes, especialmente, quando questionados se esses efeitos são canais exercidos de forma direta e/ou indireta. Talvez, ao supor que esses problemas são de natureza indireta, especialmente, relacionados a problemas de saúde decorrentes da escassez de água e altas temperaturas recorrentes na região analisada seja uma tarefa mais trivial. Porém, deve-se atentar para algumas possíveis explicações de caráter direto, por exemplo, a escassez de água e altas temperaturas enfrentadas pelas mulheres gestantes na região diminuem o desenvolvimento cognitivo das crianças? Esses problemas afetam a herdabilidade genética? Enfim, ambientes desfavoráveis em períodos iniciais da vida podem afetar o desenvolvimento cognitivo, e, conseqüentemente, afetar o desempenho em períodos posteriores conforme defende amplamente a Teoria do Ciclo da Vida.

Ao analisar a decomposição da variância dos termos aleatórios da regressão verifica-se que aproximadamente 96% de toda variabilidade observada no desempenho educacional é resultante das diferenças proveniente das características individuais dos próprios alunos. Por exemplo, das condições de escassez de água e variação das temperaturas que cada um enfrentou no período intrauterino, como também, das particularidades socioeconômicas.

Por outro lado, menos de 4% das diferenças nas notas observadas são explicadas pelas especificidades referentes a escolas e/ou municípios. Uma provável explicação para essa pequena participação das escolas e municípios nos resultados alcançados seja a homogeneidade encontrada em ambos. Essas evidências fortalecem a aceitação dos fatos estilizados propostos por Heckman e Cunha (2007), em específico, o fato (1) onde assume-se que a qualidade das escolas e dos recursos educacionais desempenham papel relativamente baixo sobre os déficits das habilidades, e, portanto, cumpre influência mínima sobre os resultados dos testes realizados pelas crianças de diferentes grupos socioeconômicos. Em outras palavras, problemas enfrentados em fases/períodos anteriores não são totalmente recompostos por investimentos tardios.

4.2 Seca e Resultados Educacionais

Agora serão exibidos os resultados referentes a condições extremas de escassez de água (seca) sobre os resultados educacionais das crianças do semiárido ao final do primeiro ciclo do ensino fundamental. Isto é, busca-se captar qual o efeito dos períodos de seca nos doze meses que precedem o nascimento dessas crianças sobre os seus respectivos resultados obtidos na Prova Brasil. A priori é executado um teste simples de média populacional a fim de corroborar a existência de diferenças estatisticamente válidas entre os grupos analisados (tratamento e controle).

Tabela 4 – Teste das Médias Amostrais para Resultado Educacional

Grupo	Observações	μ	σ^2
Controle (0)	98.018	208,45	44,85
Tratamento (1)	38.219	205,30	47,70
Total	136.236	$\neq 3,15$	45,69
$(\neq) = \mu (0) - \mu (1)$		$t=11,45$	

H ₀ : diferença = 0		$\varphi = 136.235$
H _a : diferença < 0	H _a : diferença ≠ 0	H _a : diferença > 0

Fonte: Elaboração dos autores.

É importante ressaltar que o grupo considerado tratado se refere às crianças que enfrentaram períodos de seca nos doze meses que antecederam o seu nascimento. Os casos contrários fazem parte do grupo de controle. Em conformidade com os resultados do teste das médias amostrais, rejeita-se a hipótese nula de igualdade entre médias amostrais. Conclui-se, portanto, haver diferença estatística significativa entre os desempenhos observados nos respectivos grupos. Sinteticamente, em média, o grupo de tratamento (vida intrauterina com seca) possui rendimento escolar inferior ao grupo de controle.

Posto isso, a Tabela 5 apresenta o impacto das condições de extrema escassez de água (seca) sobre o desempenho educacional das crianças do semiárido brasileiro na Prova Brasil 2013 a 2017.

Tabela 5: Efeito (Seca) Tratamento Médio Quantílico – ETMQ

Desempenho Educacional na Prova Brasil			
Quantil	Matemática	Português	Média Geral
0.25	-0.03626***	-0.03860***	-0.03729***
0.50	-0.02901***	-0.03130***	-0.03054***
0.75	-0.01525***	-0.01723***	-0.01668***

Fonte: Elaboração dos autores.

Notas: *** $p < 0,01$; ** $p < 0,05$.

Em todos os cenários (estimações) considerados utilizou-se as variáveis de controle específicas dos indivíduos, escolas e municípios apresentadas anteriormente. Além disso, destaca-se que todos os coeficientes estimados se mostraram estatisticamente diferentes de zero. Em média, as crianças que enfrentaram períodos de seca na fase intrauterina obtiveram desempenho na Prova Brasil inferior as demais em todos os pontos da distribuição dos resultados.

Especificamente, esses efeitos são consideravelmente maiores nos extratos inferiores. Isto é, nos 25% dos alunos que obtiveram o pior desempenho esse efeito negativo (seca e resultado educacional) situa-se em 0,036, 0,038 e 0,037 respectivamente em matemática, português e média geral. Porém, à medida que o desempenho dos alunos avança essa diferença vai perdendo força. E quando olhamos para o extrato superior, ou seja, o 25% dos alunos que alcançaram as notas mais altas nos exames, observa-se que o efeito negativo dos períodos de seca representa menos da metade em relação ao extrato inferior e bem menos expressivo em relação ao extrato médio.

Esses resultados permitem supor que quanto mais tarde forem os investimentos tardios maiores serão os problemas da formação cognitiva e não cognitiva individual. Uma vez que, os efeitos negativos causados pelas condições extremas de escassez de água (seca) são expressivamente mais severos em crianças pertencentes ao grupo de alunos que obtiveram o desempenho mais fraco. Vale salientar que esse comportamento é visualizado em praticamente toda a distribuição. Em resumo, conforme o desempenho dos alunos avança o problema fica cada vez menor, e na parcela que corresponde aos 10% dos alunos com maiores médias observadas nos exames analisados não há efeito estatístico comprovado. Por sua vez, na parte inferior (10% dos alunos com piores desempenhos) verifica-se haver uma tendência crescente, embora pequena, em relação aos demais níveis da distribuição.

Todos esses efeitos são bem mais abrangentes quando se considera somente a nota em língua portuguesa, como já exposto na Tabela 5. Diante disso, agora é feito um exercício separando as diferenças observadas por sexo onde as meninas representam 51,5% contra 48,5% dos meninos. Os resultados expostos na Tabela 6 corroboram a literatura e

demonstram haver maiores diferenças entre os meninos nos respectivos grupos de tratamento e controle. O diferencial calculado em média é aproximadamente 10% superior em relação as meninas. Quando se analisa essa questão dentro do mesmo grupo observa-se um padrão similar, embora, a amplitude das diferenças dos meninos como já dito seja bem maior.

Tabela 6: Desempenho Educacional: Efeitos dos Períodos de Seca por Sexo

Quantil	Meninas			Meninos		
	Matemática	Português	Média Geral	Matemática	Português	Média Geral
0.25	-0,0336***	-0,0333***	-0,0340***	-0,0362***	-0,0386***	-0,0372***
0.50	-0,0246***	-0,0261***	-0,0261***	-0,0290***	-0,0313***	-0,0305***
0.75	-0,0100***	-0,0131***	-0,0117***	-0,0152***	-0,0172***	-0,0166***

Fonte: Elaboração dos autores.

Notas: *** $p < 0,01$; ** $p < 0,05$.

Em resumo, as meninas obtiveram um desempenho médio 1,67% superior aos meninos. Essa discrepância é três vezes maior quando se compara somente os resultados de língua portuguesa. Em contrapartida, os meninos obtiveram nota em matemática ligeiramente superior. Algumas explicações para esses aspectos são encontradas nos relatórios do Programa Internacional de Avaliação de Estudantes (PISA), por exemplo, a favor das meninas: em 2018, 24% dos meninos e 44% das meninas admitiram que a leitura é dos passatempos favoritos. Por sua vez, 60% dos meninos e 39% das meninas afirmaram que leem somente quando necessário. Outra possível explicação para as meninas superarem os meninos em língua portuguesa se refere ao fato de que elas em geral dedicam mais tempo as atividades escolares e são bem mais concentradas. Especificamente, Em todas as unidades federativas brasileiras, o desempenho em leitura das meninas superou o dos meninos em 2015.

Por outro lado, o diferencial observado em matemática nos dados analisados a favor dos meninos pode ser explicado por questões apontadas no PISA em foco 2015. Segundo relatório, os meninos são mais predispostos a pensar como cientistas quando são submetidos a formular situações matemáticas ou explicar fenômenos científicos. Além disso, segundo o relatório, os pais são mais propensos a esperar (motivar) que os seus filhos, em vez de suas filhas, trabalhem em áreas que envolvem ciência, tecnologia, engenharia ou matemática, mesmo em situações que meninos e meninas apresentem o mesmo nível de desempenho em matemática.

5. CONSIDERAÇÕES FINAIS

Este artigo investigou as implicações das restrições pluviométricas e/ou períodos de escassez extrema de água (seca) nos períodos iniciais da vida, incluindo a fase intrauterina, de crianças (9 a 12 anos de idade) sobre os seus respectivos resultados educacionais nos exames da Prova Brasil no período de 2013 a 2017. É importante destacar que a fase/período de vida das crianças analisadas ocorre imediatamente após a denominada primeira infância. As informações analisadas referem-se a 74 municípios representantes de todos os estados pertencentes aos semiárido brasileiro que dispunham de estações meteorológicas físicas e dados dos respectivos exames.

Para tanto, utilizou-se dois métodos distintos de estimação associados a duas variáveis climáticas de interesse adotada em Rocha e Soares (2015). A priori foram utilizados modelos multiníveis de regressão para dados em painel e, em seguida, uma metodologia similar a proposta por Heckman e Karapakula (2019). Os principais resultados indicam a existência de efeitos negativos da escassez de água (baixos níveis pluviométricos ou períodos de seca),

como também, das altas temperaturas observadas na região semiárida brasileira sobre os respectivos resultados educacionais dos alunos.

Outras variáveis/características socioeconômicas individuais e coletivas utilizadas como controle também se mostraram significativas na composição dos resultados das provas, o que era esperado. Todavia, dando ênfase ao objetivo proposto, a pesquisa revela que menos de 4% do diferencial do resultado educacional observado na região são explicados por características pertencentes a escolas e/ou municípios. Esses resultados validam os fatos estilizados assumidos na proposta, especialmente, aqueles que pregam a minimização dos efeitos referentes a qualidade das escolas e recursos educacionais em momentos posteriores a primeira infância.

Por fim, é importante destacar a presença considerável de fatos importantes da Teoria do Ciclo da Vida nos resultados encontrados na pesquisa, a fim de promover uma proposta de desenho e formulação mais eficiente de políticas públicas que buscam combater as desigualdades de oportunidades educacionais. Essencialmente, pelo fato da não existência de substitutos perfeitos em fases/períodos posteriores da vida na formação das habilidades cognitivas e não cognitivas individuais, principalmente, na primeira infância.

REFERÊNCIAS

AGENCIA NACIONAL DE ÁGUAS (ANA, 2018). **Observatório de Escassez Hídrica no Semiárido**. Ministério do Desenvolvimento Regional, Brasília.

ALMOND, D.; CURRIE, J. Human capital development before age 5. **Handbook of Labor Economics Volume 4b**, Edited by Ashenfelter O, Card D, p. 1315-1487, 2010.

BAKER, M.; GRUBER, J.; MILLIGAN, K. The long-run impacts of a universal child care program. **American Economic Journal: Economic Policy**, v. 11, n. 3, p. 1-26, 2019.

BARROS, R. P. D.; MENDONÇA, R.; SANTOS, D. D. D.; QUINTAES, G. Determinantes do desempenho educacional no Brasil. **Pesquisa e Planejamento Econômico**, v. 31, n. 1, p. 1-42, 2001.

BARRO, R. J.; LEE, J. W. A new data set of educational attainment in the world, 1950–2010. **Journal of development economics**, v. 104, p. 184-198, 2013.

BECKER, G. S. Investment in human capital: A theoretical analysis. **The journal of political economy**, p. 9-49, 1964.

BINET, A.; SIMON, T. Métodos nuevos para el diagnóstico del nivel intelectual de lo subnormales. J. Gondra, **La psicología moderna: textos básicos para su génesis y desarrollo histórico**. (1992), 1905.

CARNEIRO, P. M.; HECKMAN J. J. Política de Capital Humano. **Bureau Nacional de Pesquisa Econômica**. NBER n. w9495. 2003.

CUNHA, F.; HECKMAN, J. J. The technology of skill formation. **American Economic Review**, v. 97, n. 2, p. 31-47, 2007.

DESCHENES, O.; MORETTI, E. Extreme weather events, mortality, and migration. **The Review of Economics and Statistics**, v. 91, n. 4, p. 659-681, 2009.

FAO, WWC. Escassez de Água na Agricultura. **1º Fórum Internacional do WASAG Sobre Escassez de Água na Agricultura**. Organização das Nações Unidas para Alimentação e Agricultura. Cabo Verde, 2019.

FERREIRA, S. G.; VELOSO, F. A. Mobilidade intergeracional de educação no Brasil. **Pesquisa e Planejamento Econômico**, v. 33, n. 3, p. 481-513. 2003.

FIORI, N. **As neurociências cognitivas**. Petrópolis, RJ: Vozes, 2008.

GIORDANO, M.; BARRON, J.; ÜNVER, O. **Water Scarcity and Challenges for Smallholder Agriculture**. In **Sustainable Food and Agriculture**. Academic Press, 2019.

HANUSHEK, E. A. Conceptual and empirical issues in the estimation of educational production functions. **Journal of human Resources**, v. 14, n. 3, p. 351-388. 1979.

HECKMAN, J. J. A economia, tecnologia e neurociência da formação da capacidade humana. **Anais da Academia Nacional de Ciências**, v. 104, n. 33, p. 13250-13255, 2007.

HECKMAN, J. J.; STIXRUD, J.; URZUA, S. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. **Journal of Labor economics**, v. 24, n. 3, p. 411-482, 2006.

HECKMAN, J. J.; KARAPAKULA, G. **Intergenerational and intragenerational externalities of the perry preschool project (No. w25889)**. National Bureau of Economic Research, 2019.

HOLANDA, N. O fenômeno da seca no Nordeste. In: Bancada Federal do Nordeste. **Seca: análises, pressupostos, diretrizes, projetos e metas para o planejamento de um novo Nordeste**. Brasília: Câmara dos Deputados. p. 10-19, 2013.

HOLZMAN, B.; THORNTHWAITTE, C. Evaporation and transpiration. **Climate and Man: Yearbook of Agriculture**, p. 545-550, 1941.

MACCINI, S.; YANG, D. Under the weather: Health, schooling, and economic consequences of early-life rainfall. **American Economic Review**, v. 99, n. 3, p. 1006-26, 2009.

MINCER, J. Investment in human capital and personal income distribution. **Journal of political economy**, v. 66, n. 4, p. 281-302, 1958.

ORGANIZAÇÃO PARA COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO (OCDE). **PISA em foco 2015**. Obtido em: <http://download.inep.gov.br/acoes-internacionais/pisa/resultados/2015/pisa-2015-brazil-prt.pdf>

ORGANIZAÇÃO PARA COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO (OCDE). **PISA 2018 results**. Obtido em <https://www.oecd.org/pisa/publications/pisa-2018-results.htm>.

PIAGET, J.; COOK, M. **The origins of intelligence in children**. New York: International Universities Press. V. 8, n. 5, p. 8, 1952.

PLOMIN, R; SPINATH, F. M. Intelligence: genetics, genes, and genomics. **Journal of personality and social psychology**, v. 86, n. 1, p. 112, 2004.

RAWLS, J. **A theory of justice**. Harvard university press, 2009.

RAUDENBUSH, S. W. **HLM 6: Hierarchical linear and nonlinear modeling**. Scientific Software International, 2004.

ROCHA, R.; SOARES, R. R. Water scarcity and birth outcomes in the Brazilian semiarid. **Journal of Development Economics**, v. 112, p. 72-91, 2015.

ROMER, P. M. Increasing returns and long-run growth. **Journal of political economy**, v. 94, n. 5, p. 1002-1037, 1986.

ROEMER, J. E. **Theories of distributive justice**. Harvard University Press, 1998.

SCHULTZ, T. W. Investment in human capital. **The American economic review**, p. 1-17, 1961.

SOTO, J. L. G.; MORERA, M. C. **Modelos jerárquicos lineales**. Editorial La Muralla, 2005.

SPEARMAN, C. General Intelligence, Objectively Determined and Measured. **The American Journal of Psychology**, v. 15, n. 2, p. 201-292, 1904.

SUPERINTENDÊNCIA DO DESENVOLVIMENTO DO NORDESTE (SUDENE, 2019). **Delimitação do Semiárido**. Ministério do Desenvolvimento Regional, Brasília.

WEST, Brady T.; WELCH, Kathleen B.; GALECKI, Andrzej T. **Linear mixed models: a practical guide using statistical software**. CRC Press, 2014.