

<https://doi.org/10.18222/eae.v34.10119>

# A DESIGUALDADE DE OPORTUNIDADE NA AVALIAÇÃO NACIONAL DA ALFABETIZAÇÃO DE 2016

 RODRIGO PERES DE ÁVILA<sup>I</sup>

 RENATA SPERRHAKE<sup>II</sup>

<sup>I</sup> Universidade Federal do Rio Grande (FURG), Rio Grande-RS, Brasil; [rodrigopavila@gmail.com](mailto:rodrigopavila@gmail.com)

<sup>II</sup> Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre-RS, Brasil;  
[renata.sperrhake@gmail.com](mailto:renata.sperrhake@gmail.com)

## RESUMO

Este artigo investiga a desigualdade de oportunidade no acesso das crianças a níveis adequados de leitura, escrita e matemática. Para tanto, calcularam-se os Indicadores de Oportunidade Humana para os resultados da Avaliação Nacional da Alfabetização de 2016. Os resultados indicam problemas consideráveis nas três dimensões analisadas, agravados ao adotar um padrão mais exigente do que seria um nível adequado de aprendizado. Os indicadores estimados variaram entre 0.41 e 0.61 para o critério menos exigente, e entre 0.07 e 0.23 para o critério mais exigente de aprendizado. A decomposição de Shapley mostra que a desigualdade de oportunidade no acesso a uma boa educação está vinculada majoritariamente (90%) ao nível socioeconômico das escolas e sua localização geográfica.

**PALAVRAS-CHAVE** DESIGUALDADE DE OPORTUNIDADES • ALFABETIZAÇÃO • AVALIAÇÃO.

## COMO CITAR:

Ávila, R. P. de, & Sperrhake, R. (2023). A desigualdade de oportunidade na Avaliação Nacional da Alfabetização de 2016. *Estudos em Avaliação Educacional*, 34, Artigo e10119.  
<https://doi.org/10.18222/eae.v34.10119>

# LA DESIGUALDAD DE OPORTUNIDAD EN LA EVALUACIÓN NACIONAL DE ALFABETIZACIÓN DE 2016

## RESUMEN

Este artículo investiga la desigualdad de oportunidad en el acceso de los niños a niveles adecuados de lectura, escritura y matemáticas. Para ello, se calcularon los Indicadores de Oportunidad Humana para los resultados de la Avaliação Nacional da Alfabetização [Evaluación Nacional de Alfabetización] de 2016. Los resultados indican problemas considerables en las tres dimensiones analizadas, agravados al adoptar un estándar más exigente de lo que sería un nivel adecuado de aprendizaje. Los indicadores estimados variaron entre 0.41 y 0.61 para el criterio menos exigente, y entre 0.07 y 0.23 para el criterio más exigente de aprendizaje. La descomposición de Shapley muestra que la desigualdad de oportunidad en el acceso a una buena educación está vinculada de modo mayoritario (90%) al nivel socioeconómico de las escuelas y su localización geográfica.

**PALABRAS CLAVE** DESIGUALDAD DE OPORTUNIDADES • ALFABETIZACIÓN • EVALUACIÓN.

# INEQUALITY OF OPPORTUNITY IN THE 2016 NATIONAL LITERACY ASSESSMENT IN 2016

## ABSTRACT

The present article investigates the inequality of opportunity in children's access to appropriate levels of reading, writing and mathematics. To this end, Human Opportunity Indicators were used to calculate the results of the 2016 Avaliação Nacional da Alfabetização [National Literacy Assessment]. The results indicate considerable problems in the three dimensions analyzed, aggravated by adopting a more demanding standard of what would be an appropriate level of learning. The estimated indicators ranged between 0.41 and 0.61 for the least demanding criterion, and between 0.07 and 0.23 for the most demanding learning criterion. Shapley's decomposition shows that inequality of opportunity in access to a good education is linked mostly (90%) to the socio-economic level of schools and their geographic location.

**KEYWORDS** INEQUALITY OF OPPORTUNITY • LITERACY • ASSESSMENT.

Recebido em: 1 MARÇO 2023

Aprovado para publicação em: 13 NOVEMBRO 2023



Este é um artigo de acesso aberto distribuído nos termos da licença Creative Commons do tipo BY-NC.

## INTRODUÇÃO

A literatura de desigualdade de oportunidade, baseada nos trabalhos centrais de Roemer (1993, 1996, 1998), tem discutido a injustiça gerada pelo acesso desigual a resultados socioeconômicos – caracterizados como vantagens – sendo a injustiça vinculada à parcela dessa desigualdade associada a fatores de não responsabilidade dos indivíduos. É frequente o argumento de que as circunstâncias desfavoráveis para um agente podem impedi-lo de transformar seus esforços em resultados, e que, se uma sociedade tem essa característica, ela é indesejável sob o conceito de véu da ignorância de Rawls (1971). O conceito implica que todos os recursos de uma sociedade sejam distribuídos por indivíduos cobertos por um véu da ignorância, o que incentivaria alocações mais igualitárias, visto que ninguém saberia a sua probabilidade de acesso aos resultados. Simplificadamente, uma sociedade justa sob o véu da ignorância é aquela em que é desejável fazer parte sem saber a própria posição.

O modelo proposto por Dworkin (1981) inclui dois parâmetros para a construção da igualdade de oportunidades: a separação entre fatores de responsabilidade e fatores arbitrários. Tal ideia é abordada, com variações, por Arneson (1989), Cohen (1989) e, finalmente, Roemer (1998), sintetizando a noção de que para alcançar a justiça uma sociedade deve compensar os fatores de não responsabilidade dos indivíduos, entendidos como circunstâncias (background familiar, cor, gênero, dentre outros). Sem essa compensação, o prêmio ao esforço não é o mesmo entre os indivíduos, o que implica em mais desigualdade econômica futura.

A literatura empírica sobre desigualdade de oportunidades estuda diversas vantagens além da renda. De igual maneira, o subconjunto dessa literatura que calcula os Indicadores de Oportunidade Humana – representado pela sigla HOI, *human opportunity index*, em inglês – aborda dimensões não consideradas no trabalho original de Barros et al. (2009). A ideia de Sen (2000) de que um maior espaço informacional para definir justiça é um mecanismo importante para decisões sociais justas é largamente utilizada. A desigualdade de acesso a uma boa educação no Brasil, tema do presente artigo, é abordada por diversos trabalhos recentes, dentre os quais pode-se citar Procópio et al. (2015), Figueirêdo et al. (2014), Foguel e Veloso (2014), Carvalho e Waltenberg (2015), Dill e Gonçalves (2012, 2013), Souza et al. (2017), Ney et al. (2010), Gamboa e Waltenberg (2015), Ferreira e Gignoux (2011), Diaz (2012) e Araújo et al. (2019). Todos os estudos, em maior ou menor grau, atestam que os resultados escolares dependem das circunstâncias dos alunos, especialmente *background* familiar.

A literatura aponta o Brasil como um país que enfrenta notáveis dificuldades no campo da educação, em praticamente todos os níveis. Embora a escolaridade média da população tenha aumentado nas décadas recentes, a qualidade da educação – bem como sua justa distribuição – continua sendo um desafio difícil

de ser vencido, especialmente no que diz respeito ao ensino público. São diversos os resultados que corroboram a ideia de que a educação é uma dimensão crítica de nosso desenvolvimento como sociedade. Os resultados do Brasil no Programa de Internacional de Avaliação de Estudantes (PISA) mostram que metade dos estudantes brasileiros não têm nível básico em Leitura, sendo este o mínimo exigido para exercício da cidadania. Além disso, os dados de 2018 mostram uma estagnação dos resultados brasileiros desde 2009. O PISA mostra também que esse quadro está presente especialmente nas escolas públicas, municipais e estaduais, visto que os resultados das escolas federais e particulares superam a média da OCDE. Outro fato relevante sobre os problemas educacionais brasileiros é que apenas 12% dos jovens em idade universitária frequentam o ensino superior no Brasil, valor 5 vezes menor ao de países desenvolvidos, de acordo com Carvalho e Waltenberg (2015).

Roemer (1998) argumenta que focalizar nas vantagens/resultados educacionais é importante para que as crianças possam obter produtividade econômica futuramente. Pensando na forma intertemporal, o acesso a uma educação de qualidade pode ser entendido como uma vantagem no presente e uma circunstância favorável no futuro, quando a escolaridade de uma criança se transformar na renda de um adulto e conseqüentemente *background* familiar para outras crianças. Além disso, a alfabetização das crianças – tema específico do artigo – é um fator determinante no sucesso ao longo de toda a escolarização.

Existem diferentes estratégias possíveis para modelar e estimar as desigualdades de oportunidades. A mensuração do esforço, entendido como fator de responsabilidade, é usualmente considerada um desafio metodológico, visto que muitas vezes é uma variável não observável. Outra questão relevante diz respeito à dependência ou independência entre esforço e circunstâncias, o que basicamente implica aceitar ou não o Axioma de Identificação de Roemer (AIR). Uma discussão dessa questão no cenário brasileiro pode ser encontrada no trabalho de Figueirêdo et al. (2014).

O trabalho de Barros et al. (2009) fornece uma solução simples e elegante para os problemas descritos no parágrafo anterior: a criação de um índice de oportunidade humana para o acesso de crianças aos resultados socioeconômicos, ou vantagens. A premissa é que, quanto mais jovens as crianças, menor é a diferença possível de esforço entre elas, ou dito de outra forma, crianças costumam empregar o mesmo empenho em seus estudos. Isso quer dizer que toda a dissimilaridade de um resultado entre crianças vem das suas circunstâncias, fatores de não responsabilidade. Na mesma linha, Kanbur e Wagstaff (2016) argumentam que o fato de crianças e adolescentes não se responsabilizarem totalmente por suas decisões faz com que seus resultados sejam predominantemente gerados pelas suas circunstâncias. Hufe et al. (2015) vão mais longe e argumentam que a totalidade das realizações das crianças deve ser vista como consequência de suas circunstâncias.

Nesse sentido, o propósito do presente artigo é aplicar o instrumental desenvolvido por Barros et al. (2009), o Indicador de Oportunidades Humanas (HOI), para a Avaliação Nacional da Alfabetização (ANA) de 2016, prova realizada por crianças do terceiro ano do ensino fundamental. Considera-se, portanto, que o bom desempenho em leitura, escrita e matemática é uma vantagem, e que toda a desigualdade observada nos resultados não vem do fato de que as crianças se esforçaram mais ou menos, mas das suas circunstâncias. Dentre os trabalhos recentes focados em educação básica no Brasil temos Procópio et al. (2015), que analisa os resultados da prova Geres para crianças nos primeiros quatro anos de escolarização, e Araújo et al. (2019), que estuda a desigualdade para o quinto ano do ensino fundamental através de dados do SAEB e do Censo Escolar de 2015. Uma aplicação específica do Índice de Oportunidade Humana já pode ser vista em Carvalho et al. (2015), que estima em 0.28 o HOI do acesso ao ensino superior, além de Foguel e Veloso (2014), que propõem modificações no índice proposto por Barros et al. (2009) e avaliam a desigualdade no âmbito de creches e pré-escolas.

Uma questão bastante atual na agenda de pesquisa empírica é a busca pela estabilidade dos resultados obtidos, ou, dito de outro modo, a investigação se determinado resultado não se deve também a critérios arbitrários ou até mesmo ao acaso. Um exemplo dessa preocupação no campo da desigualdade de oportunidades é o trabalho de Yalonetzky (2012). Assim, além de fornecer resultados inéditos sobre a ANA e contribuir com as estimativas de desigualdade de oportunidade no Brasil, este artigo tenta avançar no sentido de discutir a estabilidade dos resultados a diferentes critérios do que é um nível adequado de leitura, escrita e matemática. Para tanto, explora-se a literatura mais especializada sobre alfabetização, como Soares (2016, 2020), Morais e Leal (2020) e Spinillo (2013). Considerando que a escolha de uma nota de corte, embora alicerçada na literatura, implica em algum grau de arbitrariedade, é desejável que mudanças pequenas na linha de corte não alterem significativamente os resultados e as conclusões.

Por fim, utiliza-se algumas funções de distribuição acumulada e a decomposição de Shapley para entender quais circunstâncias são mais relevantes para a desigualdade de oportunidades observada. As funções de distribuição acumulada foram aplicadas à educação no Brasil por Figueirêdo et al. (2014), num estudo sobre desigualdade no acesso a uma boa nota no ENEM. A decomposição de Shapley (1953) foi proposta originalmente para aplicações de teoria dos jogos cooperativos, mas desde o trabalho de Shorrocks (1999) é amplamente utilizada para análise distributiva e, também, especificamente nos trabalhos que calculam o HOI. De modo simplificado, todo HOI tem um índice de dissimilaridade  $D$ , que reflete o tamanho da desigualdade no acesso a determinada vantagem. A decomposição de Shapley indica o peso percentual de cada circunstância considerada no cálculo do HOI no índice de

dissimilaridade. O trabalho de Carvalho et al. (2015), por exemplo, a partir do valor de Shapley, mostra que aproximadamente 70% da dissimilaridade do HOI de acesso ao ensino superior vem da instrução e da renda dos pais dos alunos.

Além dessa introdução, o artigo é composto por três seções e uma conclusão. Na primeira seção, apresenta-se mais detalhadamente a ANA e discute-se sobre os critérios suficientes para que uma criança de oito anos seja considerada adequadamente alfabetizada, além de uma breve descrição da base de dados e variáveis utilizadas. Na segunda, mostra-se resumidamente a teoria da igualdade de oportunidades e formaliza-se a estratégia metodológica para medi-la neste artigo. Na terceira, mostra-se os resultados e discute-se sua relação com os artigos empíricos publicados na área.

## **ALFABETIZAÇÃO E ALFABETIZAÇÃO MATEMÁTICA**

A já anunciada “História inacabada do analfabetismo no Brasil” (Ferraro, 2009), ganhou novas nuances conforme as sociedades grafocêntricas foram demandando dos sujeitos a participação em práticas de leitura e escrita mais sofisticadas. Uma das formas de visibilizar essa progressiva demanda mais qualificada do ler e do escrever está nas definições e nas formas de medir a alfabetização. Se, por um período, assinar o próprio nome era considerado critério para julgar se alguém era ou não alfabetizado, passa-se, posteriormente, à definição de alfabetização como a habilidade de ler e escrever um bilhete simples no idioma que conhece (IBGE, 2009). Mais contemporaneamente, especialmente após emergência das discussões sobre letramento, considera-se que não basta compreender o funcionamento do sistema de escrita alfabética para que alguém seja alfabetizado, é preciso também que este sujeito seja capaz de fazer uso da leitura e da escrita nas diferentes situações sociais e culturais em que estas se fazem presentes. Dessa forma, também as maneiras de medir a alfabetização passam a considerar critérios mais amplos e exigentes do que apenas o próprio nome ou a leitura de um bilhete. Também, é claro, considera-se o avanço nas metodologias psicométricas como uma forma de qualificar os processos e os resultados das medições na área da Educação.

Soares (2016) afirma que a alfabetização é um fenômeno multifacetado que circunscreve a compreensão da escrita como um sistema de representação no qual fonemas e grafemas estão relacionados a partir de certas propriedades e normas ortográficas, ou seja, um fenômeno que envolve a apropriação de uma tecnologia que permite ler e escrever a escrita alfabética-ortográfica, no nosso caso, do português brasileiro. Além disso, atrelado ao conceito de alfabetização, mas respondendo a processos cognitivos e linguísticos distintos (Soares, 2020), as habilidades vinculadas ao letramento também compõem as facetas do que hoje compreende-se

como alfabetização em um sentido mais abrangente. Assim, os usos e as funções que a escrita assume nas sociedades em diferentes tempos históricos, bem como a capacidade de inserir-se em práticas sociais nas quais a língua escrita está inclusa, também compõem o fenômeno aqui analisado.

A autora sintetiza a discussão das múltiplas facetas da alfabetização organizando-as, para fins didáticos, em três: 1) a faceta linguística, 2) a faceta interativa e 3) a faceta sociocultural (Soares, 2016).<sup>1</sup> A primeira abarca mais diretamente as relações letra-som (grafema-fonema) para ler ou escrever, a compreensão de que a escrita é um sistema de representação – alfabético-ortográfico –, que segue certas normas, bem como habilidades ligadas aos modos de ler e escrever. A segunda faceta contempla a relação com os textos, tanto com as habilidades de compreensão quanto com as habilidades de produção de textos orais e, sobretudo, escritos. Entende-se que toda a interação ocorre por meio de textos, por isso eles são centrais na aprendizagem e no ensino da língua escrita. Por fim, a terceira faceta abarca as dimensões sociais e culturais dos usos e funções da língua escrita em diferentes contextos nos quais ela é parte integrante das interações entre sujeitos.

Diferente do que se pode pensar de antemão, do ponto de vista do ensino e da aprendizagem, apesar desses processos serem de naturezas diferentes, eles ocorrem de modo simultâneos e são interdependentes (Soares, 2020). Nesse sentido, a aquisição da escrita como uma tecnologia não é pré-requisito para que uma criança se envolva em práticas sociais (e escolares) de leitura e produção de textos.

Como resultado do processo de ensino e aprendizagem da leitura e da escrita, a alfabetização pode ser considerada um dos alicerces para outras aprendizagens, especialmente as escolares, e, nesse sentido, um possível indicador de sucesso na escolarização. É a partir deste argumento que se torna importante e necessário a aferição do domínio das habilidades vinculadas ao ler e escrever das crianças.

Morais e Leal (2020) apontam que as avaliações externas aplicadas desde 1990 já vem indicando o baixo desempenho dos estudantes nas habilidades de compreensão de textos escritos. Mesmo antes da existência de avaliações externas com foco na alfabetização, o fracasso escolar logo no início do processo de escolarização era aferido pelos altos índices de reprovação e evasão nas chamadas classes de alfabetização. Somado a esses fatores, o Plano Nacional de Educação estabelece, na meta 5, alfabetizar todas as crianças, no máximo, até o final do 3º (terceiro) ano do ensino fundamental, o que demanda formas de monitoramento e acompanhamento, sendo a avaliação externa uma das maneiras utilizadas para isso.

1 Magda Soares salienta que, embora use tal nomenclatura, todas as facetas têm algum componente linguístico.

O termo alfabetização é comumente associado às habilidades de ler e escrever palavras e textos escritos na língua materna. No entanto, quando complementado pelo termo “matemática”, ele relaciona-se à aprendizagem inicial da leitura e da escrita da linguagem matemática de modo a compreender seus conteúdos elementares (geometria, lógica, operações, etc.). O conceito de alfabetização matemática comparece na literatura acadêmica há ao menos duas décadas e meia. Salienta-se o trabalho de Danyluk (1998) como um que discutiu a importância da aprendizagem das formas de organização e compreensão da linguagem matemática como um conhecimento historicamente construído.

A alfabetização matemática ganha destaque com o Pacto Nacional Pela Alfabetização na Idade Certa, política pública de formação de professores no interior da qual a ANA estava inscrita. Segundo o material formativo do programa:

A Alfabetização Matemática que se propõe, por se preocupar com as diversificadas práticas de leitura e escrita que envolvem as crianças e com as quais as crianças se envolvem – no contexto escolar e fora dele –, refere-se ao trabalho pedagógico que contempla as relações com o espaço e as formas, processos de medição, registro e uso das medidas, bem como estratégias de produção, reunião, organização, registro, divulgação, leitura e análise de informações, mobilizando procedimentos de identificação e isolamento de atributos, comparação, classificação e ordenação. (Ministério da Educação, 2014, p. 31).

Nesse sentido, assim como a alfabetização inicial na língua materna, a alfabetização matemática também cumpre importante função como alicerce para aprendizagens escolares posteriores.

A delimitação clara do que se espera de uma criança alfabetizada é um fator ainda em discussão, o que é uma das motivações para a realização deste estudo. No momento da aplicação da Avaliação Nacional da Alfabetização, a definição da expectativa em relação à criança alfabetizada não estava explícita. Foi posterior à realização de algumas edições da avaliação que foram definidos os níveis que seriam considerados “suficientes” e “insuficientes”. Cabe destacar que, recentemente, o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) realizou uma pesquisa chamada “Alfabetiza Brasil” com 251 professoras alfabetizadoras de 206 municípios das diferentes regiões do Brasil com o objetivo de “determinar o ponto de corte que indica a alfabetização de uma criança ao final do 2º ano do ensino fundamental” (Inep, 2023). A partir disso foi possível definir “o conjunto de habilidades mínimas que devem caracterizar a criança alfabetizada, ao fim do 2º ano do ensino fundamental” (Inep, 2023). Tal movimento era inédito nas políticas de avaliação da alfabetização até então.



## A AVALIAÇÃO NACIONAL DA ALFABETIZAÇÃO (ANA)

A Avaliação Nacional da Alfabetização foi a primeira iniciativa brasileira que teve como objetivo avaliar a alfabetização de crianças em larga escala em nível nacional.<sup>2</sup> Criada em 2012, com aplicações em 2013, 2014 e 2016, a Avaliação Nacional da Alfabetização integrava as ações do Pacto Nacional pela Alfabetização na Idade Certa (PNAIC) e se configurava como um dos seus eixos de atuação. Foi uma avaliação censitária que teve como público-alvo os alunos do 3º ano do ensino fundamental – ano este que era o último do chamado Ciclo de Alfabetização – das escolas urbanas e rurais das redes públicas de ensino que tenham mais de 10 alunos matriculados. O PNAIC se constituiu como um compromisso assumido pelos governos federal, estaduais e municipais com vistas a alfabetizar as crianças até o final do Ciclo de Alfabetização. Dentre as ações previstas no PNAIC, estava a aferição anual da alfabetização em língua portuguesa e matemática por meio de exame periódico realizado pelo Inep. A ANA, após três aplicações, foi descontinuada pelo Ministério da Educação, em 2018, para se adequar à Base Nacional Comum Curricular (BNCC), que estabelece o 2º ano do ensino fundamental como momento de finalização do processo de alfabetização. Com as mudanças no Sistema Nacional de Avaliação da Educação Básica (Saeb), anunciadas também em 2018, a avaliação da alfabetização das crianças passou ao escopo desse sistema, com nova nomenclatura e com prova específica aplicada ao final do 2º ano do ensino fundamental.

A ANA era composta por duas provas objetivas, uma de leitura e outra de matemática, ambas contendo 20 itens de múltipla escolha, com quatro alternativas cada; e pela prova de resposta construída, contendo dois itens que avaliam a escrita de palavras e a escrita de um texto. A metodologia utilizada na construção das provas é a Teoria de Resposta ao Item (TRI) e os itens de cada um dos testes objetivam avaliar uma das habilidades descritas nas Matrizes de Referência, uma para Língua Portuguesa e uma para Matemática. O modelo básico da Teoria da Resposta ao Item considera itens categóricos (acerto ou erro) e uma função logística para modelar a Curva Característica do Item. De acordo com Andrade et al. (2000) podemos formalizar deste modo:

$$P(U_{ij} = 1|\theta_j) = c_i + \frac{(1-c_i)}{1+e^{-Da_i(\theta_j-b_j)}} \quad (1)$$

onde,  $P(U_{ij} = 1|\theta_j)$  representa a probabilidade da resposta do aluno  $j$  ao item  $i$  estar correta dada a sua proficiência (habilidade, traço latente)  $\theta_j$ ; note que  $U_{ij}$  é uma variável dicotômica, enquanto  $a_i$  representa o parâmetro de discriminação, com valor

2 A "Provinha Brasil", criada em 2008 e que também avaliava alfabetizandos (do 2º ano do ensino fundamental) tinha características diferentes da ANA.

proporcional à inclinação da Curva Característica do Item;  $b_i$  o parâmetro de dificuldade, medido na mesma escala da habilidade;  $c_i$  o parâmetro de acerto casual; finalmente,  $D$  é um fator de escala, constante e usualmente unitário.

Na ANA, as proficiências eram padronizadas com a média arbitrária de 500 e desvio padrão de 100. Os resultados de cada prova eram organizados em três escalas de proficiência independentes entre si e não comparáveis: uma para leitura, uma para escrita e uma para matemática. As escalas de leitura e de matemática são compostas “por quatro níveis progressivos e cumulativos. Isso significa uma organização da menor para a maior proficiência” (Ministério da Educação, 2012, p. 1). Já a escala de escrita é composta por cinco níveis.

O Inep estabelece os níveis 1 e 2 da escala de Leitura como insuficientes, e os níveis 3 e 4 como suficientes. A partir do nível 3 é que são solicitadas habilidades de leitura consideradas mais sofisticadas, dentre as quais a principal é a inferência. Segundo Spinillo (2013, p. 139), “inferir é derivar uma nova proposição a partir de outras proposições fornecidas pelo texto ou a partir dele na relação que o leitor estabelece entre o texto e seu conhecimento de mundo, possibilitando, assim, que as lacunas deixadas pelo escritor sejam preenchidas”. Estudos na área da leitura apontam que as dificuldades de compreensão de textos residem na dificuldade de estabelecer inferências autorizadas pelo texto.

Nesse sentido, e considerando que os níveis das escalas de proficiência são cumulativos, o desejável seria que, após três anos de escolarização no ensino fundamental, uma criança alcançasse as habilidades de leitura descritas no nível 4, visto que este abrange também as habilidades dos níveis anteriores. Entretanto, para ser considerado um aluno com alfabetização suficiente pelo Inep, não é necessário atingir o nível 4, ou seja, não é preciso ter consolidadas as habilidades de leitura que levariam o estudante a, por exemplo, inferir a relação de causa e consequência em um texto contendo humor ou identificar referentes de pronomes possessivos em textos não narrativos. Além disso, no nível 4 os textos apresentam maior diversificação em relação ao gênero, não se restringindo aos gêneros textuais de tipo narrativo, mais comuns nas salas de aula de alfabetização.

No que se refere às habilidades de escrita, divididas em escrita de palavras e de textos, o Inep considera os níveis 1, 2 e 3 insuficientes, e os níveis 4 e 5 suficientes. Com esse critério, as crianças que ainda não têm consolidadas habilidades relacionadas à estrutura do texto solicitado na prova – a saber: uma narrativa contendo enredo, personagens, narrador, tempo e espaço –, assim como habilidades relacionadas ao uso de recursos coesivos que não prejudiquem a compreensão do texto, são consideradas suficientemente alfabetizadas. No nível 4, as crianças também não usam sinais de pontuação, ou os usam de modo inadequado, além de apresentarem equívocos na segmentação de palavras.

Em relação às habilidades de matemática, ao adotar o mesmo critério que em relação aos resultados em leitura, ou seja, considerar os níveis 3 e 4 como sendo suficientes, o Inep desconsidera como necessárias as habilidades do nível 4, tais como inferir uma resposta que não está explícita na questão e resolver problemas numéricos e algébricos mais complexos sem auxílio de imagens, visto que esse tipo de questão predomina nesse nível.

Apesar de Soares e Bergmann (2020, p. 92) destacarem que o Inep prefere não “dar etiquetas valorativas” para os níveis em que se divide a proficiência dos estudantes, os autores apontam que os diferentes níveis das escalas podem orientar o trabalho pedagógico sugerindo, por exemplo, que estudantes com proficiência nível 1 estariam abaixo do básico, tendo aprendido apenas habilidades elementares e que precisariam de intervenções para recuperação; enquanto isso, estudantes com proficiência nível 4 seriam considerados avançados e poderiam ser colocados diante de situações didáticas que os incentivassem a ir além do esperado. No entanto, convém destacar que o Inep, ao apresentar os resultados da ANA, organizava um agrupamento dos níveis em “suficiente” e “insuficiente”, tal como apresentamos anteriormente.

Assim, de forma simplificada, são esses conjuntos de habilidades em leitura, escrita e matemática que diferenciam o nível mais alto de proficiência do nível imediatamente anterior. Por considerarmos que tais habilidades são importantes e coerentes com o desejável para o terceiro ano de escolarização, na seção seguinte, ao calcular os indicadores de oportunidades humanas, adotaremos um critério alternativo, mais exigente, em que só será considerado alfabetizado de forma suficiente o aluno que atingir o nível 4 nas provas de Leitura e Matemática e o nível 5 na prova de Escrita.

Note que quanto mais alunos atingirem o nível mais alto (mesmo o nível anterior sendo considerado suficiente para o Inep), menor será a diferença entre os dois critérios, o que se refletirá em HOIs mais parecidos. Se as diferenças forem grandes, há um sinal de alerta possível, visto que não é desejável que uma linha métrica seja muito sensível a esse tipo de variação.

Outro fator que fortalece o argumento de maior exigência é que uma avaliação em larga escala da leitura e da escrita tem um potencial reduzido para avaliar um construto multifacetado como é a alfabetização, pois esta abrange não apenas a compreensão do sistema alfabético-ortográfico de escrita (aferido pela leitura e escrita de palavras, por exemplo), mas também as habilidades de leitura e escrita de textos de variados gêneros textuais, além dos usos e funções desses gêneros nas diversas situações comunicativas em que eles são demandados, tal como apontado em seção anterior neste texto. O recorte, necessário, realizado no construto “alfabetização” para que seja possível algum tipo de mensuração por uma avaliação em

larga escala já limita as habilidades a serem avaliadas, o que impõe, de antemão, um limitador aos resultados da avaliação, visto que eles visibilizam apenas algumas das habilidades que tornam um sujeito efetivamente proficiente em língua portuguesa e, portanto, leitor e produtor de textos.

### A base de dados

A base utilizada são os microdados da ANA por aluno.<sup>3</sup> Foram aplicadas 2,14 milhões de provas de português e 2,19 milhões de provas de matemática. Para formulação dos HOIs, utilizou-se os pesos de cada aluno dentro da amostra, procedimento padrão em dados censitários. Não são disponibilizados dados sobre gênero e raça dos alunos, os quais seriam variáveis clássicas de controle das circunstâncias. A Tabela 1, a seguir, contém a descrição das variáveis utilizadas.

**TABELA 1**  
**Descrição das variáveis utilizadas**

VARIÁVEL	DESCRIÇÃO
Nível de Leitura (1, 2, 3, e 4)	Quantidade de alunos que alcançou pontuação correspondente a cada um dos níveis da prova de Leitura
Nível de Escrita (1, 2, 3, 4 e 5)	Quantidade de alunos que alcançou pontuação correspondente a cada um dos níveis da prova de Escrita
Nível de Matemática (1, 2, 3 e 4)	Quantidade de alunos que alcançou pontuação correspondente a cada um dos níveis da prova de Matemática
Localização	Variável que indica se o aluno reside em área rural ou urbana
Região	Variável que indica em qual das cinco regiões do Brasil o aluno reside
Nível Socioeconômico (NSE)	Nível socioeconômico da escola, variando entre muito baixo, baixo, médio baixo, médio, médio alto, alto e muito alto
Dependência Administrativa da Escola	Variável que indica se a escola é federal, estadual ou municipal

Fonte: Elaboração dos autores.

Neste artigo, usa-se a dependência administrativa da escola como uma *proxy* de qualidade educacional, visto que as escolas federais costumam ter melhor infraestrutura e professores mais qualificados (os dados da Tabela 2 ilustram esse ponto). Tal iniciativa está de acordo com Albernaz et al. (2002), trabalho que aponta que os insumos escolares também exercem influência nas desigualdades educacionais. Por esse motivo criou-se uma *dummy* em que federal = 1 e estadual e municipal = 0. Em relação ao nível socioeconômico, criou-se uma *dummy* classificando como ricos os níveis médio alto, alto e muito alto, e como pobres as demais quatro categorias.

3 Exceto a variável NSE que originalmente era referente à média da escola. Fez-se uma agregação imputando aos alunos o nível médio da sua escola.

**TABELA 2**  
**Dependência administrativa como proxy de qualidade educacional**

Prova	PROFICIÊNCIA MÉDIA POR DEPENDÊNCIA ADMINISTRATIVA		
	Federal	Estadual	Municipal
Leitura	582	532	508
Escrita	565	523	501
Matemática	579	534	508

Fonte: Elaboração dos autores.

Por fim, a Tabela 3 contém algumas estatísticas descritivas básicas sobre os dados. Eles permitem ter uma ideia melhor da distribuição e do peso das circunstâncias consideradas no trabalho.

**TABELA 3**  
**Frequência e porcentagem acumulada das circunstâncias**

Variável	ESTATÍSTICAS DESCRITIVAS BÁSICAS		
	Frequência	Porcentagem	% Acumulada
Norte	328.521	12.13%	12.13%
Nordeste	750.016	27.70%	39.83%
Sudeste	1.042.711	38.51%	78.34%
Sul	373.803	13.81%	92.15%
Centro-Oeste	212.258	7.84%	100%
Rural	365.541	13.50%	13.50%
Urbano	2.341.768	86.50%	100%
NSE muito alto	4.378	0.18%	0.18%
NSE alto	303.534	12.59%	12.77%
NSE médio alto	953.125	39.53%	52.30%
NSE médio	597.093	24.76%	77.06%
NSE médio baixo	417.033	17.29%	94.35%
NSE baixo	121.390	5.03%	99.38%
NSE muito baixo	14.845	0.62%	100%
Escola federal	1.520	0.06%	0.06%
Escola estadual	478.526	17.68%	17.74%
Escola municipal	2.227.263	82.27%	100%

Fonte: Elaboração dos autores.

## IGUALDADE DE OPORTUNIDADES

Como já abordado, o conceito de igualdade de oportunidades está vinculado ao trabalho de Roemer (1993, 1996, 1998) e parte da ideia de que nem toda desigualdade

é indesejável, somente aquela que vem de fatores de não responsabilidade dos indivíduos. Diversos trabalhos apontam nessa direção, teórica e empiricamente, tais como Fleurbaey (2008), Ferreira e Gignoux (2011), Peragine (2011), Fleurbaey e Peragine (2013), Ferreira e Peragine (2015).

No debate sobre desigualdade anterior aos trabalhos de Roemer, não se fazia distinção entre a desigualdade desejável e indesejável, ou seja, toda a desigualdade era relevante. Baseado nas noções de justiça de Rawls (1971), Sen (1980), dentre outros, Roemer (1998) operacionalizou a ideia de que a desigualdade pode vir de dois grupos:

- i) Esforço (e): fator de responsabilidade, entendido como o conjunto de decisões que cada indivíduo toma e que o faz obter maior ou menor acesso a um resultado econômico: trabalhar mais ou menos horas, migrar ou não migrar, tomar mais ou menos educação formal, etc.
- ii) Circunstâncias (c): fatores de não responsabilidade, entendidas como as características exógenas ao indivíduo e que os fazem obter maior ou menor acesso a um resultado econômico: gênero, etnia, localização, herança, etc.

Note que nem sempre é trivial definir se uma determinada característica é exógena ou não a um indivíduo. Ter nascido em determinada localização é um fator exógeno, mas decidir migrar ou não é um fator de responsabilidade, da mesma forma que nascer em uma família rica é uma circunstância, mas decidir trabalhar mais horas e ganhar mais dinheiro é entendido como esforço. Um resultado econômico X qualquer, ou o acesso a uma determinada vantagem pode ser escrita formalmente como:

$$X = f(c, e) \quad (2)$$

Sendo a desigualdade de oportunidade a parte das diferenças de X causadas por c.

É possível reduzir a desigualdade de oportunidades de dois modos: maximizar os menores resultados dentro de cada tranche<sup>4</sup> ou reduzir a desigualdade entre tipos<sup>5</sup>. O argumento de Roemer (1998) pode ser sintetizado como permitir que os indivíduos transformem seus esforços em resultado, não sendo prejudicados por circunstâncias desfavoráveis. No mundo ideal, indivíduos que exercem o mesmo nível de esforço não deveriam obter resultados significativamente distintos. Como já salientado, essa ideia está diretamente ligada ao conceito de véu da ignorância proposto em Rawls (1971), em que uma sociedade é mais justa ou igualitária no acesso a bens primários se for desejável a um indivíduo ser parte dessa sociedade mesmo sem saber em qual dos grupos existentes.

4 Indivíduos com mesmo nível de esforço formam um tranche.

5 Indivíduos com as mesmas circunstâncias formam um tipo.

Como sintetiza Carvalho et al. (2015), o que Roemer propõe é um tipo de meritocracia condicional ou uma meritocracia entre iguais. Quando consideramos somente as notas em uma prova, segundo a teoria da igualdade de oportunidades, não estamos sendo justos porque estamos tratando igualmente pessoas diferentes, com características distintas. Esse é o princípio da compensação, segundo o qual a desigualdade de resultados derivada por diferentes dotações iniciais deve ser compensada pela política pública.

A literatura discute basicamente duas formas de mensuração da Equação (2). A forma direta, ou explícita, procura identificar o nível de desigualdade entre os conjuntos individuais de oportunidades. O entrave é que muitos conjuntos de oportunidades são não observáveis. A forma indireta é consequencialista, foca nos resultados e não nas oportunidades diretamente.

### **O Indicador de Oportunidades Humanas (HOI)**

O Indicador de Oportunidades Humanas (HOI), criado por Barros et al. (2009), caracteriza-se por medir o acesso a serviços essenciais ao desenvolvimento humano, entendidos nesse contexto como oportunidades. Uma característica metodológica importante do indicador é que deve ser calculado apenas para crianças, o que visa excluir o fator esforço na determinação do acesso às oportunidades. Presume-se assim que toda dissimilaridade estimada através de um HOI vem das circunstâncias, que são fatores de não responsabilidade, entendidos como injustos. Essa é uma contribuição importante do HOI à literatura visto que parte considerável da discussão de desigualdade de oportunidades trata de como estimar esforço adequadamente, ou como dissociar adequadamente circunstâncias e esforço.

Formalmente temos que  $0 \leq HOI \leq 1$ , onde valores maiores indicam mais oportunidades humanas. Mais especificamente O HOI é composto por (i) uma Taxa de Cobertura [C], entendida como a média das probabilidades individuais de acesso a uma oportunidade; e (ii) por um Índice de Dissimilaridade [D], que capta as diferenças das taxas de cobertura por tipos, sendo estes definidos como indivíduos com diferentes conjuntos de circunstâncias. A ideia geral é que a Dissimilaridade atue como uma penalização a taxas de coberturas mal distribuídas entre diferentes grupos de uma sociedade.

Note que se  $C=1$  não temos por definição um problema de Dissimilaridade, e quanto menor a taxa de Cobertura mais há espaço para que questões vinculadas à Dissimilaridade sejam relevantes. Note também que dois países podem ter a mesma taxa de Cobertura para determinada oportunidade, como por exemplo alfabetização, mas HOIs bastante distintos, quanto maiores forem as diferenças em seus índices de Dissimilaridade.<sup>6</sup>

6 País A:  $C = 0,8$ ,  $D = 0,1$ ,  $HOI = 0,72$ . País B:  $C = 0,8$ ,  $D = 0,5$ ,  $HOI = 0,4$ .

Podemos assim escrever que

$$HOI = C(1-D) \quad (3)$$

onde

$$C = \frac{1}{N} \sum_{i=1}^n P_i \quad (4)$$

A taxa de cobertura,  $0 \leq C \leq 1$ , reflete a média das probabilidades  $P$ , obtidas via *logit*, de cada  $i$  dos  $n$  indivíduos terem acesso ( $A=1$ ) a uma dada oportunidade, de forma condicionada ao seu vetor de circunstâncias  $X_i$ :

$$P_i (A=1|X_i) \quad (5)$$

Para obter tal probabilidade  $P_i$  formalmente são estimados os  $\beta$ s da seguinte equação:

$$\frac{P(A=1|x_{1i}...x_{mi})}{(1-P)(A=1|x_{1i}...x_{mi})} = e^{\beta_0 + \sum_{k=1}^m x_{ki}\beta_k} \quad (6)$$

Dado que o procedimento de uma regressão logística é feito por uma função de máxima verossimilhança, as probabilidades obtidas com a estimação dos  $\beta$ s ainda não são as de interesse. Para que a probabilidade individual  $P_i$  seja obtida, é necessária a seguinte relativização:

$$P_i = \left[ \frac{e^{\beta_0 + \sum_{k=1}^m x_{ki}\beta_k}}{1 + e^{\beta_0 + \sum_{k=1}^m x_{ki}\beta_k}} \right] \quad (7)$$

Já o índice de Dissimilaridade,  $0 \leq D \leq 1$ , mede as diferenças das taxas de acesso por tipos (grupos com circunstâncias distintas) em relação à cobertura média. Se  $D=0$  temos que a taxa de Cobertura é perfeitamente igualitária, de modo que  $(1-D)$  pode ser entendida como a proporção de oportunidades igualitárias.

$$D = 1/2C \sum_{i=1}^n \frac{1}{N} (P_i - C) \quad (8)$$

O HOI é, portanto, a diferença entre a Taxa de Cobertura total  $C$ , obtida via probabilidades individuais de acesso, e a penalidade  $CD$  relativa à cobertura desigual dessa oportunidade entre tipos. Não obstante, o trabalho de Hoyos e Narayan (2012), destaca algumas propriedades importantes de um HOI: homogeneidade de grau  $n = 1$ , sensibilidade a mudanças Pareto-eficientes e sensibilidade à redistribuição.



## A Decomposição de Shapley

A decomposição de Shapley, proposta originalmente no contexto da teoria dos jogos, nos permite entender como uma medida qualquer está distribuída entre diferentes grupos. No contexto da desigualdade de oportunidades, temos interesse em estimar a contribuição de cada circunstância no índice de Dissimilaridade de acesso a um resultado. Isso implica dizer que seja qual for o valor de  $D$ , podemos examinar o peso de cada circunstância na determinação desse valor, de forma que a soma desses pesos é sempre igual a um (toda a Dissimilaridade observada).

Uma propriedade importante do HOI e seus componentes é que o acréscimo de uma circunstância sempre aumenta a Dissimilaridade, reduzindo o HOI. Formalmente, se ao conjunto de circunstâncias  $A$  for acrescida a circunstância  $G$ , e se não houver intersecção entre  $A$  e  $G$ , temos que  $D(A, G) \geq D(A)$ , o que implica que  $HOI(A, G) \leq HOI(A)$ . Deste modo, o impacto de adicionar a circunstância  $G$  a um conjunto de circunstâncias qualquer é dado por:

$$D_G = \sum_{S \cup N \setminus \{G\}} \frac{s!(n-s-1)!}{n!} [D(S \cup \{G\}) - D(S)] \quad (9)$$

Onde  $N$  é o conjunto de todas as circunstâncias, que inclui o total de  $n$  circunstâncias e  $S$  é um subconjunto de  $N$  (contendo  $s$  circunstâncias) que não contém a circunstância particular  $G$ . Note então que  $[D(S \cup \{G\}) - D(S)]$  é a diferença entre a Dissimilaridade gerada antes e depois da inclusão de  $G$ . Por consequência podemos escrever que:

$$M_G = \frac{D_G}{D(N)} \text{ (contribuição \% da circunstância } G \text{ na Dissimilaridade total)} \quad (10)$$

$$\text{Com } \sum_{i \in N} M_i = 1$$

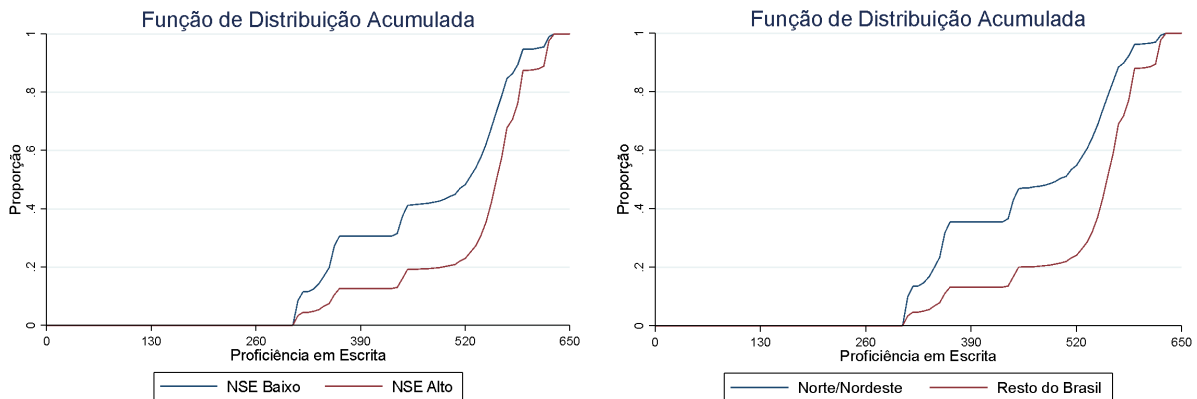
## RESULTADOS E DISCUSSÃO

Nesta seção são apresentados os resultados relacionados ao HOI e à decomposição de Shapley, tendo como objeto de pesquisa a Avaliação Nacional da Alfabetização (ANA) de 2016. Entende-se, portanto, o acesso a níveis adequados de leitura, escrita e conhecimento matemático como uma oportunidade essencial ao desenvolvimento humano. Como a prova é aplicada no terceiro ano do ensino fundamental, se está obedecendo a premissa de estimar o HOI para crianças, o que implica dizer que os diferentes resultados observados estão vinculados às diferentes circunstâncias as quais as crianças estão submetidas, sem que elas tenham responsabilidade sobre seu desempenho.

Antes disso, usa-se de forma exploratória uma ferramenta útil para avaliar o impacto das circunstâncias no desempenho na ANA, as funções de

distribuição acumulada dos níveis de proficiência das três provas, entre grupos de interesse. Tal estratégia, já aplicada à desigualdade de oportunidades por Figueirêdo et al. (2014), é um complemento aos resultados principais do artigo. As figuras 1, 2 e 3, a seguir, mostram as curvas das funções para dois vetores de circunstâncias: nível socioeconômico e região das escolas.

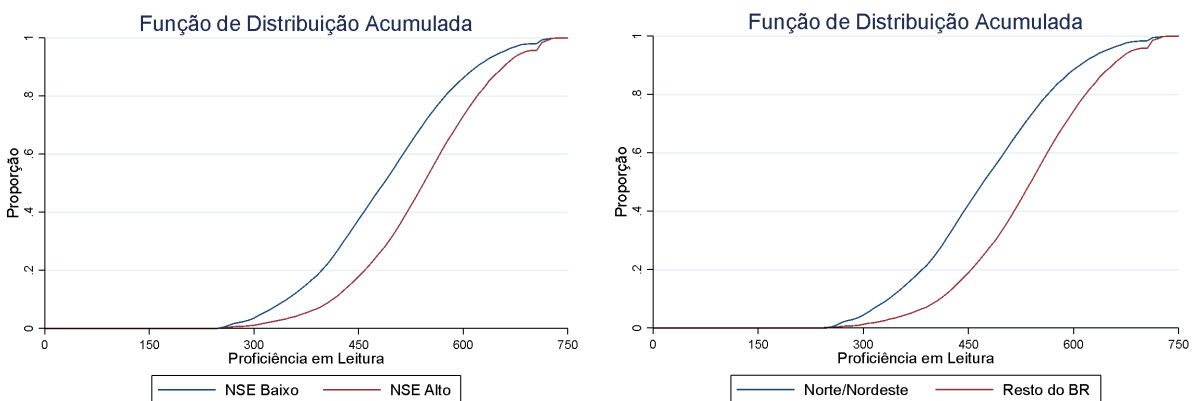
**FIGURA 1**  
**FDA da Proficiência em Escrita**



Fonte: Elaboração dos autores.

Para qualquer nível de proficiência na prova de Escrita (Figura 1), pertencer ao nível socioeconômico alto ou não pertencer às regiões Norte ou Nordeste representa estatisticamente um ganho. A probabilidade de um aluno do Norte ou Nordeste, escolhido ao acaso, atingir a nota necessária (500) para ser considerado com uma escrita adequada é de aproximadamente 50%, enquanto para o resto do Brasil é de quase 80%. As probabilidades são semelhantes na comparação entre os dois níveis socioeconômicos.

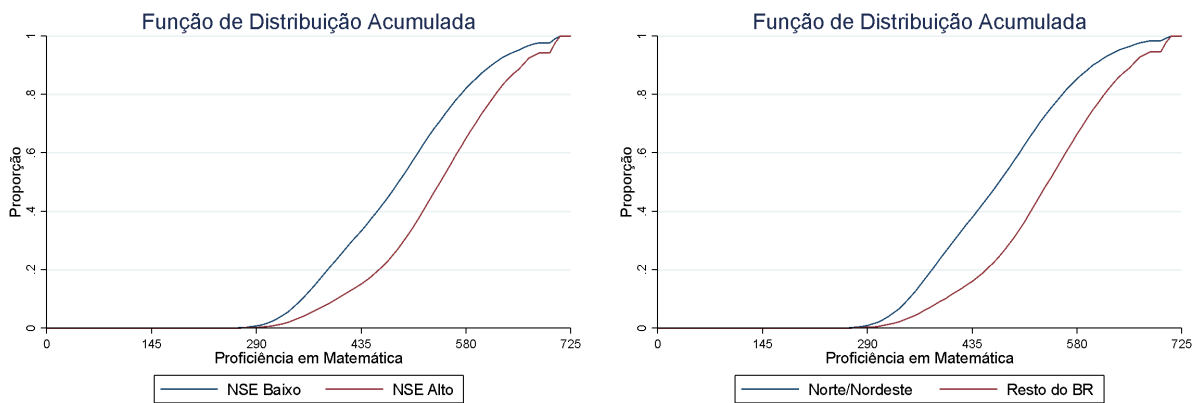
**FIGURA 2**  
**FDA da Proficiência em Leitura**



Fonte: Elaboração dos autores.

Em relação ao nível de leitura (Figura 2), as funções de distribuição acumulada também deixam claro o quanto é um ganho fazer parte do nível socioeconômico alto ou não pertencer às regiões Norte ou Nordeste. Enquanto a probabilidade de um aluno pertencente ao nível socioeconômico alto, escolhido ao acaso, atingir a nota necessária (525) para ser considerado com nível de leitura adequado é de 60%, para o nível socioeconômico baixo é de pouco menos de 40%. Tais diferenças estão de acordo com as encontradas por Figueirêdo et al. (2014) em relação ao acesso a boas notas no ENEM, bem como são coerentes com os resultados gerais da literatura empírica voltada ao ensino básico, em que o nível socioeconômico é usualmente apontado como um determinante da desigualdade educacional.

**FIGURA 3**  
**FDA da Proficiência em Matemática**



Fonte: Elaboração dos autores.

Nos resultados da prova de matemática (Figura 3) observa-se a mesma tendência de penalização ao fato do aluno pertencer à região Norte ou Nordeste ou ao nível socioeconômico baixo. A probabilidade de atingir a nota adequada (525) é aproximadamente a metade em relação aos grupos que representam melhores circunstâncias, com o valor caindo de aproximadamente 60% para pouco mais de 30%. Assim como Procópio et al. (2015), não se verifica até aqui diferenças significativas entre a desigualdade nos resultados de português e matemática.

Sobre os resultados principais, relativos ao HOI e à Decomposição de Shapley, com base na discussão já apresentada sobre os diferentes critérios possíveis para inferir se uma criança tem níveis de leitura, escrita e matemática adequados ao terceiro ano do ensino fundamental, optou-se por apresentar as estimativas em dois cenários: para o critério observado pelo Inep e para um critério mais exigente. Para fins de simplificação chamaremos de critérios fraco e forte.

**TABELA 4**  
**O Indicador de Oportunidades Humanas para a ANA 2016**

Oportunidade	Critério Fraco (Inep)			Critério Forte		
	Cobertura	Dissimilaridade	HOI	Cobertura	Dissimilaridade	HOI
Leitura	0,46	0,12	0,41	0,13	0,18	0,11
Escrita	0,67	0,10	0,61	0,09	0,22	0,07
Matemática	0,47	0,13	0,41	0,28	0,16	0,23

Fonte: Elaboração dos autores.

Os números expostos na Tabela 4 indicam inicialmente que, considerando o Critério Fraco, os resultados são intermediários. Note que, na aplicação de um HOI, não há um parâmetro de referência para categorizar o que são baixos ou altos níveis de Oportunidade Humana (como no IDH, por exemplo). A interpretação possível é que se fosse sorteada ao acaso uma criança que fez a prova, as probabilidades de ela ter conhecimento adequado de leitura, escrita e matemática seriam respectivamente de 46%, 67% e 47%. Não obstante, o fato de essas probabilidades serem afetadas por diferentes conjuntos de circunstâncias, como mostram os índices de Dissimilaridade, penalizam essa taxa de cobertura, gerando HOIs menores, entre 0,41 e 0,61.

Observe, ainda sob o critério fraco, que as taxas de Dissimilaridade são semelhantes entre as dimensões do aprendizado, variando entre 10% e 13%. Isso indica que, sob a métrica oficial do Inep, as diferentes circunstâncias consideradas não exercem efeito especial em alguma habilidade de aprendizado específico, o que está de acordo com as funções de distribuição apresentadas anteriormente. Os resultados sobre dissimilaridade expostos na Tabela 4 também estão de acordo com os recentemente encontrados por Souza et al. (2017) para uma pesquisa com base no Saeb. De acordo com os autores, a desigualdade de oportunidade pode explicar entre 13% e 22% da desigualdade total verificada na educação dos alunos da 4ª série (5º ano) do ensino fundamental, no período entre 2005 e 2011, considerando-se os dados em nível nacional.

Por outro lado, a taxa de Cobertura da Escrita é sensivelmente maior do que a de Leitura e Matemática, o que se reflete diretamente no HOI da Escrita, que é 20 pontos maior do que o HOI das outras duas dimensões. Considerando que o HOI varia entre 0 e 1, os resultados em torno de 0,5 obtidos não podem ser vistos como ideais, tampouco sugerem que a qualidade educacional e sua distribuição sejam as piores possíveis. Para fins de comparação, o HOI referente ao acesso ao ensino superior calculado por Carvalho et al. (2015) é de 0,28, o que de certa forma nos levaria a divergir do diagnóstico consensual de que os problemas na educação brasileira são mais graves no ensino básico. Sob a métrica do HOI, haveria um agravamento do problema quando avançamos no ciclo de escolarização.

Um pesquisador, ao encerrar sua análise com base nesses parâmetros oficiais (Critério Fraco), pode concluir genericamente que: (i) de forma geral a aprendizagem medida pela prova ANA é de nível intermediário e (ii) a Escrita é a dimensão de melhor desempenho, ou mais próxima de um desempenho satisfatório. Uma consequência lógica razoável seria a recomendação de que é preciso intensificar esforços em termos de política educacional para melhorar especialmente os resultados de Leitura e Matemática.

Contudo, analisando os dados com um pouco mais de rigor, essa conclusão geral mostra-se um equívoco. Ao utilizarmos um critério mais exigente do que o Inep para o que deve refletir um nível de aprendizado adequado, os indicadores de Cobertura e Dissimilaridade pioram sensivelmente. Como consequência, observa-se que o HOI da Escrita, que era de 0,61, torna-se pouco mais do que um décimo disso (0,07), o que é um resultado alarmante. Note que o HOI da Leitura também cai sensivelmente, tornando-se a (pouco mais de 1/3 do valor original, assim como o HOI da Matemática, que cai praticamente pela metade. Os resultados passam a ser mais coerentes com o que Carvalho et al. (2015) estimou para o acesso ao ensino superior, permitindo que voltemos ao consenso de que os problemas na educação básica são tão ou mais graves do que os presentes nas demais fases da escolarização brasileira.

Não apenas os três indicadores de Cobertura pioram, visto que o critério de aprendizado satisfatório se tornou mais rígido, mas os indicadores de Dissimilaridade, obtidos com o mesmo conjunto de circunstâncias, também pioram de forma expressiva, especialmente os de Leitura (aumento de 50%) e Escrita (aumento de 120%). Com base nessas estimativas, parece seguro inferir que critérios menos rigorosos de aprendizado têm como externalidade negativa a subestimação da desigualdade no acesso a uma oportunidade educacional.

Em suma, as conclusões advindas da análise baseada no Critério Fraco não são consistentes com aquelas geradas pelo Critério Forte, o que por si só é um sinal de alerta aos formuladores de política pública. Se aceitarmos um nível de exigência mais rígido sobre o que constitui um aprendizado adequado, temos um problema extremamente grave nos dois componentes do HOI: baixa taxa de cobertura, ou seja, baixa probabilidade de uma criança ter boa leitura, escrita e habilidade matemática; bem como maior desigualdade no acesso a esses resultados – que já seriam ruins mesmo com dissimilaridade nula ( $HOI = C$ ).

A Tabela 5, a seguir, complementa a análise, decompondo a Dissimilaridade observada na tabela 4, de acordo com as circunstâncias observadas para o cálculo do HOI, quais sejam: (i) nível socioeconômico da escola onde o aluno estuda; (ii) se o aluno reside em área rural ou urbana; (iii) se estuda em escola Federal ou Estadual/Municipal; e (iv) se faz parte das regiões Norte e Nordeste ou não. Por

construção, a soma das porcentagens de uma decomposição de Shapley corresponde a 100% de cada Dissimilaridade calculada anteriormente. Tomando como exemplo a Dissimilaridade de 0,22 da Escrita pelo Critério Forte, a decomposição indica que 43,7% dessa Dissimilaridade de 0,22 vem da circunstância de estudar em uma escola de nível socioeconômico pobre.

**TABELA 5**  
**A Decomposição de Shapley para a ANA 2016**

D. Shapley	CRITÉRIO FRACO (INEP)				CRITÉRIO FORTE			
	Pobres	Rural	Est./Mun.	N/NE	Pobres	Rural	Est./Mun.	N/NE
Leitura	44,6	9,6	0,1	45,7	46,5	8,5	0,1	44,9
Escrita	41,0	8,4	0,1	50,5	43,7	7,8	0,1	48,4
Matemática	43,6	9,0	0,1	47,3	44,7	8,0	0,1	47,2

Fonte: Elaboração dos autores.

Um aspecto relevante a ser salientado da Tabela 5 é o fato de que a composição da desigualdade não varia essencialmente entre critérios de aprendizagem diferentes, tampouco entre as três dimensões de aprendizagem consideradas. As circunstâncias nível socioeconômico da escola e região do Brasil a qual o aluno pertence explicam, em todos os seis cenários, mais de 90% da dissimilaridade observada. Esse resultado, ainda que extremo, está de acordo com a literatura empírica, como Procópio et al. (2015), Carvalho et al. (2015) e Araújo et al. (2019), que apontam nível socioeconômico e localização dentre as dimensões mais relevantes da desigualdade educacional brasileira. Não obstante, a dependência administrativa da escola, utilizada como *proxy* de qualidade educacional, mostrou-se praticamente irrelevante em todos os casos. Note ainda que ao se considerar o critério forte, a importância do nível socioeconômico da escola aumenta marginalmente, enquanto o peso da localização diminui. Embora seja um resultado consistente para leitura, escrita e matemática, o que é um indicativo de robustez, o tamanho das diferenças obtidas não permite maior aprofundamento da discussão.

O resultado geral da Tabela 5 é especialmente interessante frente ao fato já salientado de que o tamanho da Dissimilaridade variou quando mudamos o critério de aprendizado e se mostrou diferente entre as dimensões consideradas. Há, portanto, sensibilidade do tamanho da Dissimilaridade entre os cenários abordados, bem como regularidade na sua composição, no que diz respeito ao aprendizado satisfatório de leitura, escrita e matemática, medidos pela ANA de 2016.

## CONCLUSÃO

O artigo se propôs a estudar a desigualdade de oportunidade no acesso à alfabetização, medida pela Avaliação Nacional de Alfabetização (ANA) de 2016. Por se tratar de crianças, o instrumental desenvolvido por Barros et al. (2009) é adequado, visto que essas não exercem esforço significativamente diferente. Com isso, toda a desigualdade observada no acesso à alfabetização é indesejável, pois é fruto de fatores de não responsabilidade das crianças. Complementarmente, utilizou-se funções de distribuição acumulada para medir tal desigualdade, o que levou a conclusões coerentes com o Indicador de Oportunidades Humanas.

A conclusão principal do artigo é que ainda restam muitos obstáculos para que se alcance a igualdade de oportunidades na alfabetização brasileira. Não apenas ainda existem problemas grandes em relação à taxa de Cobertura (probabilidade de uma criança escolhida ao acaso ser alfabetizada), como há questões relativas à dissimilaridade, o que foi demonstrado tanto com as funções de distribuição acumulada das proficiências como com a decomposição de Shapley das Dissimilaridades observadas. Pertencer aos níveis socioeconômicos mais altos e não pertencer às regiões Norte e Nordeste se mostrou consistentemente um benefício para as crianças que fizeram a prova, resultado que está de acordo com a literatura sobre o tema.

Partindo do princípio que o nível satisfatório de alfabetização, quando exposto na forma de uma métrica de corte, inclui algum nível de arbitrariedade, testou-se como uma pequena modificação na nota de corte afeta os resultados de leitura escrita e matemática. Considera-se, idealmente, que os resultados não deveriam ser excessivamente sensíveis a essa mudança, que foi fundamentada em um conceito menos restrito de alfabetização, tendo como referência as habilidades descritas nas escalas de proficiência da ANA.

Esse exercício de consideração de um critério mais exigente de aprendizado afetou severamente os Indicadores de Oportunidades Humanas, o que se traduz no mínimo como um sinal de alerta para os tomadores de decisão de política pública. Os resultados indicam que a situação da alfabetização brasileira, especialmente no que diz respeito à leitura e escrita, pode ser ainda pior do que o diagnosticado através das métricas oficiais. Além disso, o tamanho da dissimilaridade aumenta sob esse critério mais exigente, o que indica que a desigualdade educacional no Brasil pode estar sendo subdimensionada.

## AGRADECIMENTOS

Agradecemos à Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) pelo apoio financeiro.

**REFERÊNCIAS**

- Albernaz, A., Ferreira, F., & Franco, C. (2002). Qualidade e equidade no ensino fundamental brasileiro. *Pesquisa e Planejamento Econômico*, 32(3), 453-76.
- Andrade, D. F., Tavares, H. R., & Valle, R. C. (2000). *Teoria da Resposta ao Item: Conceitos e aplicações*. Associação Brasileira de Estatística.
- Araújo, S. M. de, Farias Souza, W. P. S. de, & Malbouisson, C. S. (2019). Desigualdade de oportunidades na educação fundamental no Brasil: Analisando o papel das circunstâncias da infância. In *Anais do Encontro Nacional da Associação Brasileira de Estudos Regionais e Urbanos*. Universidade Federal Rural do Rio de Janeiro.
- Arneson, R. (1989). Equality and equal opportunity for Welfare. *Philosophical Studies*, 56(1), 77-93. <https://www.jstor.org/stable/4320032>
- Barros, R. P., Ferreira, F., Vega, J., & Chanduvi, J. (2009). *Measuring inequality of opportunities in Latin America and the Caribbean*. World Bank Publications e Plagrave McMillan.
- Carvalho, M. M., & Waltenberg, F. D. (2015). Desigualdade de oportunidades no acesso ao ensino superior no Brasil: Uma comparação entre 2003 e 2013. *Economia Aplicada*, 19(2), 369-396. <https://doi.org/10.1590/1413-8050/ea124777>
- Cohen, G. A. (1989). On the currency of egalitarian justice. *Ethics*, 99, 906-944. <https://www.jstor.org/stable/2381239>
- Danyluk, O. S. (1998). *Alfabetização Matemática: As primeiras manifestações da escrita infantil*. Sulina.
- Diaz, M. (2012). (Des)igualdades de oportunidades no ensino médio brasileiro: Escolas públicas e privadas. *Revista Economia*, 13(3), 553-568.
- Dill, H. C., & Gonçalves, F. (2012). Índice de oportunidade no Brasil entre 1999 e 2009: Estimação e decomposição através do valor de Shapley. *Pesquisa e Planejamento Econômico*, 42(2), 187-210.
- Dill, H. C., & Gonçalves, F. (2013). Igualdade de oportunidades entre os estados brasileiros: Uma análise microeconômica a partir dos dados da PNAD 2009. *Nova Economia*, 23(2), 307-328. <https://doi.org/10.1590/S0103-63512013000200003>
- Dworkin, R. (1981). What is equality? Part I: Equality of Welfare. *Philosophy & Public Affairs*, 10, 185-246. <https://www.jstor.org/stable/2264894>
- Ferraro, A. R. (2009). *História inacabada do analfabetismo no Brasil*. Cortez Editora.
- Ferreira, F. H. G., & Gignoux, J. (2011). The measurement of inequality of opportunity: Theory and application to Latin America. *Review of Income and Wealth*, 57(4), 622-657. <https://doi.org/10.1111/j.1475-4991.2011.00467.x>
- Ferreira, F. H. G., & Peragine, V. (2015). Equality of opportunity: Theory and evidence. *World Bank Policy Research Paper*, Article 7217. <https://doi.org/10.1596/1813-9450-7217>
- Figueirêdo, E., Nogueira, L., & Santana, F. L. (2014). Igualdade de oportunidades: Analisando o papel das circunstâncias no desempenho do Enem. *Revista Brasileira de Economia*, 68(3), 373-392. <https://doi.org/10.1590/S0034-71402014000300005>
- Fleurbaey, M. (2008). *Fairness, responsibility and welfare*. Oxford University Press.
- Fleurbaey, M., & Peragine, V. (2013). Ex ante versus ex post equality of opportunity. *Economica*, 80, 118-130. <https://doi.org/10.1111/j.1468-0335.2012.00941.x>
- Foguel, M., & Veloso, F. (2014). Inequality of opportunity in daycare and preschool services in Brazil. *Journal of Economic Inequality*, 12(2), 191-220. <https://doi.org/10.1007/s10888-013-9245-8>
- Gamboa, L., & Waltenberg, F. (2015). Inequality of opportunity for educational achievement in Latin America: Evidence from PISA 2006-2009. *Economics of Education Review*, 31(5), 694-708. <https://doi.org/10.1016/j.econedurev.2012.05.002>



- Hoyos, A., & Narayan, A. (2012). Inequality of opportunities among children: How much does gender matter? In *World Development Report 2012: Gender Equality and Development*. World Bank Publications.
- Hufe, P., Peichl, A., Roemer, J., & Ungerer, M. (2015). *Inequality of income acquisition: The role of childhood circumstances*. ZEW Discussion Papers.
- Instituto Brasileiro de Geografia e Estatística (IBGE). (2009). *Síntese de indicadores sociais: Uma análise das condições de vida da população brasileira*. IBGE.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). (2023). *Alfabetiza Brasil*. Inep. <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/alfabetiza-brasil/a-pesquisa>
- Kanbur, R., & Wagstaff, A. (2016). How useful is inequality of opportunity as a policy construct? In K. Basu, & J. Stiglitz (Eds.), *Inequality and growth: Patterns and policy* (pp. 131-150). Palgrave Macmillan. [https://doi.org/10.1057/9781137554543\\_4](https://doi.org/10.1057/9781137554543_4)
- Ministério da Educação. (2012). *Pacto Nacional pela Alfabetização na Idade Certa: Currículo na alfabetização – Concepções e princípios – Ano 1 – Unidade 1*. Ministério da Educação.
- Ministério da Educação. (2014). *Pacto Nacional pela Alfabetização na Idade Certa: Apresentação*. Ministério da Educação.
- Morais, A. G., & Leal, T. F. (2020). Alfabetização e sua avaliação no Brasil: Um balanço crítico a partir de diferentes pesquisas. *Em Aberto*, 33(108), 19-24. <https://doi.org/10.24109/emaberto.v33i108.4694>
- Ney, M. G., Souza, P. M. de, & Ponciano, N. J. (2015). Desigualdade de acesso à educação e evasão escolar entre ricos e pobres no Brasil rural e urbano. *InterSciencePlace*, 1(13), 33-55.
- Peragine, V. (2011). Review of “measuring inequality of opportunities in Latin America and the Caribbean”. *Journal of Economic Inequality*, 9, 137-143. <https://doi.org/10.1007/s10888-010-9151-2>
- Procópio, I. V., Freguglia, R. S., & Chein, F. (2015). Desigualdade de oportunidades na formação de habilidades: Uma análise com dados longitudinais. *Economia Aplicada*, 19(2), 326-348. <https://doi.org/10.1590/1413-8050/ea110460>
- Rawls, J. (1971). *A Theory of Justice*. Harvard University Press.
- Roemer, J. (1993). A pragmatic theory of responsibility for the egalitarian planner. *Philosophy & Public Affairs*, 22(2), 146-166.
- Roemer, J. (1996). *Theories of distributive justice*. Harvard University Press.
- Roemer, J. (1998). *Equality of opportunity*. Harvard University Press.
- Sen, A. (1980). Equality of what? In S. M. McMurrin (Ed.), *Tanner lectures on human values* (Vol. 1, pp. 197-220). Harvard University Press.
- Sen, A. (2000). *Desenvolvimento como liberdade*. Companhia das Letras.
- Shapley, L. (1953). A value for n-person games. In *Contributions to the Theory of Games* (pp. 307-317). Princeton University Press.
- Shorrocks, A. (1999). *Decomposition procedures for distributional analysis: A unified framework based on Shapley value*. Department of Economics, University of Essex.
- Soares, J. F., & Bergmann, L. (2020). Avaliação Nacional da Alfabetização: Síntese sobre os resultados das escolas associados a variáveis socioeducacionais. *Em Aberto*, 33(108), 83-99.
- Soares, M. (2016). *Alfabetização: A questão dos métodos*. Contexto.
- Soares, M. (2020). *Alfaletrar: Toda criança pode aprender a ler e escrever*. Contexto.

- Souza, W. P. S. F., Oliveira, V. R., & Anegues, A. C. (2017). Desigualdade de oportunidades na educação fundamental brasileira: Novas evidências a partir de um índice para o SAEB. *Ensaio FEE*, 38, 329-356.
- Spinillo, A. G. (2013). Alfabetização e consciência metalinguística: Da leitura da palavra à leitura do texto. In R. Maluf, & R. Cardoso-Martins (Orgs.), *Alfabetização no século XXI: Como se aprende a ler e a escrever* (pp. 138-154). Penso.
- Yalonetzky, G. A. (2012). Dissimilarity index of multidimensional inequality of opportunity. *The Journal of Economic Inequality*, 10(3), 343-373. <https://doi.org/10.1007/s10888-010-9162-z>