

# O que Explica o Desempenho do Brasil no PISA 2015?

GIOVANNI AVILA CARDOSO DI PIETRA\*

ALEX HAYATO SASSAKI†

BRUNO KAWAOKA KOMATSU‡

NAERCIO AQUINO MENEZES FILHO§

## Sumário

1. Introdução .....	167
2. Revisão da Literatura .....	170
3. Descrição do PISA 2015 .....	171
4. Estratégia empírica de decomposição do desempenho ..	173
5. Decaimento .....	176
6. Recuperação .....	188
7. Conclusão e desenvolvimentos futuros .....	192
Apêndice .....	196

## Palavras-chave

Habilidades socioemocionais, habilidades cognitivas, PISA, desigualdade

## JEL Codes

O15, J24

## Resumo · Abstract

Este artigo procura explicar o fraco desempenho dos alunos brasileiros no PISA 2015 explorando o fato de que nesse ano as provas foram feitas em computadores. Mostramos que os brasileiros apresentam um forte declínio de desempenho ao longo da prova e propomos uma nova medida desse decaimento usando a descontinuidade de acertos antes e depois do intervalo que existe durante a aplicação da prova. Mostramos que esse decaimento ocorre porque os alunos brasileiros gastam muito tempo nas questões iniciais e não alcançam as últimas questões em cada bloco. Apesar do desempenho ruim nesse indicador não cognitivo, o fraco desempenho brasileiro é explicado essencialmente pela dificuldade nas questões iniciais, salientando uma desvantagem nas habilidades cognitivas e/ou na experiência em fazer esse tipo de prova em computador com relação aos alunos dos outros países.

## 1. Introdução

A qualidade da educação no Brasil em comparações internacionais é ruim. Em 2015, o Brasil ficou na posição 59 a 66, dependendo da disciplina, de 73 regiões e países no Programa Internacional de Avaliação de Alunos (*Programme for International Student Assessment* – PISA), com médias de notas em matemática (401 pontos), leitura (407 pontos) e ciências (377 pontos) abaixo das médias dos alunos da OCDE (de respectivamente 493, 493 e 490 pontos). Essas médias não representaram melhora

\* Fundação Getúlio Vargas, Escola de Economia de São Paulo (FGV/EESP). Rua Itapeva, 474, Bela Vista, São Paulo, SP, CEP 01332-000, Brasil.  0000-0003-1192-8501

† Insper, Cátedra Ruth Cardoso. Rua Quatá, 300, Vila Olímpia, São Paulo, SP, CEP 04546-042, Brasil.  0000-0002-2825-8711

‡ Insper, Cátedra Ruth Cardoso. Rua Quatá, 300, Vila Olímpia, São Paulo, SP, CEP 04546-042, Brasil.  0000-0001-6045-9794

§ Insper, Cátedra Ruth Cardoso; e Universidade de São Paulo, Faculdade de Economia, Administração e Contabilidade (USP/FEA). Av. Prof. Luciano Gualberto, 908, Butantã, São Paulo, SP, CEP 05508-010, Brasil.  0000-0003-1241-6188

 di.pietra.giovanni@gmail.com

 asassaki@usp.br

 brunokk@insper.edu.br

 naercioamf@insper.edu.br

em relação aos últimos anos. A média brasileira de ciências tem se mantido estável desde 2006, e a de leitura, desde 2000. A média de matemática apresentou crescimento significativo de 21 pontos desde 2003, porém diminuiu 11 pontos entre 2012 e 2015 (OCDE, 2017).

Habilidades cognitivas e socioemocionais são determinantes para o desenvolvimento econômico de países ou regiões e para resultados socioeconômicos individuais. Medidas precisas de habilidades cognitivas baseadas em notas de exames internacionais padronizados, como as do PISA, têm mostrado que países com maiores níveis de habilidades cognitivas na população apresentam crescimento econômico mais acelerado (Hanushek & Kimko, 2000; Hanushek & Woessmann, 2012). Por outro lado, as diferenças de notas de exames entre os países representam não somente a diversidade de habilidades cognitivas, porém também a heterogeneidade de habilidades socioemocionais. As habilidades socioemocionais vem ganhando relevância na literatura econômica, uma vez que possuem papel importante dentro de avaliações padronizadas (Duckworth et al., 2009; Eklöf, 2007; Segal, 2012; Borghans, Duckworth, Heckman, & Ter Weel, 2008; Almund, Duckworth, Heckman, & Kautz, 2011) e, de forma mais geral, para o desempenho econômico e outros *outcomes* individuais (Heckman, Stixrud, & Urzua, 2006; Heckman, Pinto, & Savelyev, 2013; Balart, Oosterveen, & Webbink, 2018).

Nesse artigo, nós exploramos inovações na organização da prova do PISA de 2015 para examinarmos quais aspectos dessa avaliação são relacionados aos diferentes tipos de habilidades e quais deles determinam o desempenho relativamente ruim do Brasil. Em 2015, as provas realizadas em computador foram organizadas de modo que os alunos obrigatoriamente responderam à questões de forma sequencial: uma vez finalizada a questão, eles não poderiam retornar a ela. Dessa maneira, os alunos não puderam escolher responder em primeiro lugar somente as questões que sabiam responder.

Nós utilizamos os microdados do PISA 2015 para determinar medidas de habilidades cognitivas e socioemocionais. Em primeiro lugar, seguindo o procedimento em Borghans e Schils (2012), Balart et al. (2018) e Marchioni (2017), nós exploramos a variação exógena da posição de cada questão para identificar os efeitos de habilidades socioemocionais que se refletem no declínio de rendimento ao longo da prova, decompondo os resultados do PISA em duas dimensões: no desempenho inicial, que representa as habilidades cognitivas, e em variações no desempenho durante a prova, que representam habilidades socioemocionais. Dada a alocação aleatória da posição de cada questão, as diferenças na queda do desempenho ao longo de uma prova refletem aspectos individuais tais como persistência, concentração e determinação na população dos países.

Medidas baseadas em resultados objetivos de uma avaliação, como as aqui apresentadas, superam o caráter subjetivo de indicadores construídos a partir de informações auto relatadas, que limitam a comparabilidade entre indivíduos e regiões. Uma inovação metodológica proposta por este trabalho é levar em conta a influência do intervalo entre os blocos de provas para a estimação do decaimento do acerto ao longo da prova, obtendo resultados mais acurados para a probabilidade de acerto no início e no final do exame.

A medida tradicional de decaimento consiste em estimar a diferença entre a probabilidade de acerto na primeira questão e a probabilidade de acerto na última questão. Propomos neste trabalho uma medida alternativa, que explora a descontinuidade da probabilidade de acerto das questões ao redor do intervalo entre os dois blocos de provas, estimada com uma abordagem de RDD. Em geral, os alunos apresentam uma probabilidade de acerto muito baixa pouco antes do intervalo e uma alta probabilidade de acerto após o intervalo, semelhante à probabilidade de acerto no início da prova.

Nossos resultados mostram que, em comparação com alguns dos países que possuem as maiores médias do PISA, os alunos brasileiros possuem baixa probabilidade de acerto da primeira questão, o que pode indicar que as habilidades cognitivas também são menores. Além disso, a probabilidade de responder corretamente às questões vai se reduzindo ao longo dos dois blocos da prova. Entre esses dois blocos há um intervalo em que o Brasil também mostra uma recuperação acentuada.

Quais são as implicações desse padrão para a interpretação dos indicadores de habilidades socioemocionais? Para responder essa questão procuramos investigar com mais detalhes os determinantes do decaimento, com foco especial sobre a taxa de respondentes que chegam à questão e sobre o tempo de resposta médio utilizado em cada questão respondida. Promovemos uma discussão que sugere que o mal desempenho do Brasil se deve ao fato de que grande parte dos respondentes não conseguem chegar ao fim da prova. Esse comportamento pode ocorrer pelo fato de que os jovens demoram para entender o enunciado da questão e para desenvolver o raciocínio sobre a resposta. Também pode ser que a tradução da prova seja mal feita, que os alunos brasileiros não estejam acostumados a fazer provas no estilo da prova do PISA ou ainda, que eles não estejam acostumados a fazer provas em computadores. Nesse sentido, a interpretação do decaimento como medida puramente de habilidades socioemocionais não parece muito clara, motivando a busca por uma outra medida que possa melhor capturar a influência destas habilidades.

Nossa segunda discussão apresentada se relaciona ao padrão de decaimento da probabilidade de acerto ao longo da prova como um todo. Parte do mal desempenho do Brasil parece ser devido à falta de habilidade em fazer uma prova no formato do PISA. Nós mostramos que o tempo médio de resposta das primeiras questões do bloco após o intervalo são consideravelmente menores do que aqueles das primeiras questões do primeiro bloco. Além disso, a probabilidade de chegar ao final do segundo bloco é consideravelmente maior do que a probabilidade de chegar ao final do primeiro bloco. Uma vez que a dificuldade das questões do primeiro bloco é igual àquela das questões do segundo bloco, os jovens brasileiros parecem aprender a fazer a prova ao longo da prova, o que não ocorre de forma tão acentuada em países com as maiores notas.

Apresentamos neste trabalho uma medida alternativa ao decaimento para capturar a influência das habilidades socioemocionais ao longo de uma prova, a descontinuidade de desempenho entre os dois blocos da prova. Nesta métrica, o Brasil apresenta a maior descontinuidade, em parte por um possível aprendizado ao longo da prova e outros fatores elencados acima. A interpretação desta medida como algo estritamente negativo e sua aplicabilidade para realizar comparações entre os países ainda não é clara. No entanto, ao replicarmos a metodologia de crescimento em [Balart et al. \(2018\)](#) e [Hanushek e Woessmann \(2012\)](#), apresentamos uma série de exercícios empíricos

sustentando o argumento de que a recuperação entre os blocos pode ser uma medida superior ao decaimento na representação das habilidades socioemocionais.

Além dessa introdução, a segunda seção traz uma revisão da bibliografia relevante para o desenvolvimento deste trabalho. A terceira seção descreve a estrutura do PISA 2015 e os dados utilizados e a quarta seção descreve a metodologia para estimação dos indicadores de habilidades socioemocionais a partir da decomposição do desempenho no PISA 2015. A quinta seção apresenta os resultados para o decaimento e uma discussão sobre os fatores associados ao declínio de desempenho e a particularidade do caso brasileiro. Já a sexta seção apresenta as estimações para a descontinuidade e replica metodologias de outros trabalhos para validar a medida de recuperação. Concluímos na sétima seção e apresentamos os desenvolvimentos futuros propostos.

## 2. Revisão da Literatura

Esse artigo se relaciona com duas linhas principais da literatura. Em primeiro lugar, nosso artigo contribui com os estudos que investigam medidas de habilidades cognitivas e não cognitivas. Alguns trabalhos que relacionam habilidades cognitivas com crescimento econômico constroem medidas daquelas habilidades utilizando notas de avaliações internacionais (Hanushek & Kimko, 2000; Hanushek & Woessmann, 2012). Há, no entanto, um ramo da literatura que verifica a influência de traços de personalidade e motivação sobre o resultado de testes de inteligência (Duckworth et al., 2009; Segal, 2012; Borghans et al., 2008), e a influência do contexto e dos incentivos envolvidos nesses testes (Almund et al., 2011), apesar de alguns estudos não encontrarem uma relação clara entre estas duas variáveis (Eklöf, 2007).

Mais recentemente, Akyol, Krishna, e Wang (2018) decompõem o papel da motivação no desempenho em testes com baixas expectativas. Eles usam o PISA 2015 para identificar alunos que fazem a prova de maneira “não séria”, segundo critérios de tempo e padrões de resposta, sendo que 25% dos alunos de todos os países faz a prova dessa maneira. Para o Brasil, 67% dos alunos podem ser considerados “não sérios”. No entanto, da mesma forma que neste trabalho, ao considerar todos os alunos fazendo a prova de forma séria o Brasil não ganha muitas posições no ranking de probabilidade de acertar as questões na prova, devido ao baixo conhecimento dos alunos “não sérios”.

Em segundo lugar, nosso artigo se relaciona com a literatura psicométrica que foca no contexto do exame e em suas características, como o posicionamento dos itens. Davis e Ferdous (2005), constatam que a literatura tende a associar o crescimento da dificuldade dos itens como um possível efeito de “fadiga”, quando o desempenho diminui e a dificuldade do item, portanto, aumenta, enquanto os alunos avaliados passam a decair em aspectos relacionadas ao desempenho, como motivação, concentração e nível de energia. Albano (2013) verifica que há viés em estimações pela Teoria de Resposta ao Item (TRI) de exames que mantêm a mesma posição das questões para todos as provas, uma vez que existe um grau diferenciado de dificuldade associado a cada item de acordo com sua posição e não exclusivamente ao seu conteúdo.

Além disso, há dois trabalhos dessa literatura que se relacionam mais diretamente com nosso artigo. Borghans e Schils (2012) buscam identificar a influência das habilidades socioemocionais no decorrer de um exame com baixa expectativa. Os principais resultados apresentados demonstram uma grande queda do desempenho dos alunos

durante as provas, porém essa queda não está correlacionada com o resultado final da prova. Além disso, a medida de habilidades socioemocionais a partir do decaimento nas provas é correlacionada com a medida auto reportada e esses indicadores são bons preditores de resultados como renda e hábitos de saúde no futuro.

Marchioni (2017) se utiliza da metodologia de Borghans e Schils (2012) e os dados do PISA de 2012 para avaliar as habilidades cognitivas e não-cognitivas dos países latino-americanos em comparação com os outros países que participam do exame, assim como para avaliar diferenças entre gênero e entre escolas públicas e privadas. Os principais resultados deste trabalho sugerem que os países latino-americanos, em geral, têm níveis menores de habilidades cognitivas e não-cognitivas em relação aos outros grupos de países do exame, que o gênero feminino está melhor no que tange às habilidades não-cognitivas e que não foram encontradas, entre países, diferenças gerais significativas entre os alunos de escolas pública e de escolas privadas. No entanto, para certos países essa diferença é significativa. É possível que os padrões de segregação socioeconômica por tipo de escola sejam o que orienta a diversidade de resultados encontrados, considerando o tipo de escola como uma *proxy* para o status socioeconômico de um aluno. O resultado é intuitivo nos casos em que o decaimento é menor para os alunos da escola particular, já que se pode pensar que o status socioeconômico da família é um importante fator na formação de habilidades não-cognitivas.

### 3. Descrição do PISA 2015

Utilizaremos os microdados do PISA 2015. O PISA é uma avaliação educacional realizada em nível internacional com o objetivo de avaliar o sistema educacional dos países participantes. O exame é feito a cada três anos por uma amostra de alunos com entre 15 anos e 2 meses e 16 anos e 3 meses de idade, matriculados em alguma instituição educacional. Em 2015, participaram 70 países, sendo 35 deles participantes da Organização para Cooperação e Desenvolvimento Econômico (OCDE) e os demais 35, países parceiros. No Brasil, o exame abrangiu 23.141 estudantes de 841 escolas.

O exame consiste em questões de ciências, literatura e matemática ou de resolução colaborativa de problemas, durante duas horas. Os alunos não fazem todas as mesmas provas e nem as mesmas questões. Cada aluno recebe uma combinação aleatória de questões, alternando as matérias e ordem de aparecimento de cada questão entre os alunos. Os alunos também não fazem provas de todas as matérias. O exame é estruturado de forma que há uma variedade de conjuntos de questões (chamados de *clusters*) de cada uma das matérias e a cada aluno são designados apenas quatro dos *clusters* possíveis. Há inúmeras combinações de *clusters*, essas combinações são identificadas por um número e dão origem a um formulário de prova, cada formulário é composto por quatro desses *clusters*. O formulário de prova que um aluno recebe é aleatório.

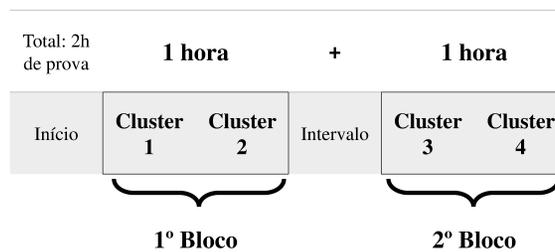
Para o ano de 2015, havia 30 formulários de provas do tipo PBA (*Paper-Based Assessment*) e 396 formulários de prova do tipo CBA (*Computer-Based Assessment*) no desenho geral do exame. Dos 70 países que participaram do PISA 2015, apenas 15 países receberam provas exclusivamente do tipo PBA, e nenhum dos países recebeu

simultaneamente provas do tipo CBA e PBA. O Brasil foi um dos países que, conforme a maioria, aplicou somente as provas do tipo CBA.

A documentação do PISA apresenta quais são e em que posição estão os *clusters* em cada um dos formulários de prova, assim como o livro de códigos das questões apresenta o ordenamento delas dentro de cada um dos *clusters* (OCDE, s. d.).

Como vimos, o formulário de prova que um aluno recebe é aleatório.<sup>1</sup> A Figura 1 ilustra o formato das provas que os alunos recebem. Devido a essa montagem do exame, a posição de cada uma das questões é aleatória e configura uma fonte de variação exógena. A intenção dos organizadores do PISA é de que os *clusters* possuam níveis de dificuldade homogêneos entre si e o tempo esperado para a resolução de cada um deles é de 30 minutos. Dessa forma, o exame tem o tempo máximo de duas horas, sendo dividido em dois blocos de uma hora cada e separados por uma pequena pausa entre os dois blocos, em geral de 5 minutos.<sup>2</sup> O exame de conhecimentos acaba no segundo bloco. Após os dois blocos de provas, os alunos respondem a questionários não avaliativos referentes aos seus dados socioeconômicos, familiares e a perguntas relacionadas ao funcionamento das escolas, atividades pedagógicas e sobre suas opiniões quanto ao estudo das matérias que são avaliadas no PISA.

A Tabela 1 mostra os testes de aleatorização da ordem das questões para características individuais e socioeconômicas, sendo essas: gênero, série que frequenta, mês de nascimento, possuir uma mesa de estudos em casa, possuir um quarto individual, possuir um lugar específico para estudar em casa, possuir um computador para estudar, o maior grau de escolaridade dos pais e possuir uma quantidade mínima de livros em casa. Para esses testes, regredimos as *dummies* de cada característica nas *dummies* dos códigos das provas e suas interações, e reportamos as estatísticas F de significância conjunta do modelo, assim como os p-valores associados. Os resultados mostram que não podemos rejeitar a hipótese nula de que os coeficientes são conjuntamente iguais a zero, ou seja, não há uma associação clara entre os códigos de provas e as características individuais. Dessa forma, temos maior confiança de que podemos considerar a alocação de cada questão como aleatória para cada estudante.



Fonte: Documentação do PISA 2015. Elaboração própria.

**Figura 1.** Formato da Prova do PISA 2015.

<sup>1</sup>A Tabela 1 mostra os resultados para regressões de variáveis *dummy* para cada um dos grupos de alunos que responderam a cada combinação de CBA nas características individuais dos alunos.

<sup>2</sup>Não há uma regra fixa para a duração do intervalo. No geral, ele é de 5 minutos, porém no Brasil, chega a aproximadamente 15 minutos.

**Tabela 1.** Teste de Aleatorização

Variável Dependente	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Gênero	Série	Mês de Nasc.	Mesa	Quarto	Lugar de estudos	Computador	Escolaridade Pai	Escolaridade Mãe	Livros
F-Stat <sup>a</sup>	1,025	1,034	0,884	1,057	1,058	1,020	0,869	0,907	1,112	0,969
P-val	0,355	0,309	0,952	0,210	0,207	0,381	0,971	0,905	0,0639	0,663
DF 1	395	395	395	395	395	395	395	395	395	395
DF 2	22.745	22.745	22.745	20.638	19.859	20.195	20.290	19.882	20.631	20.875
R <sup>2</sup>	0,017	0,018	0,015	0,020	0,021	0,020	0,017	0,018	0,021	0,018
Obs.	23.141	23.141	23.141	21.034	20.255	20.591	20.686	20.278	21.027	21.271

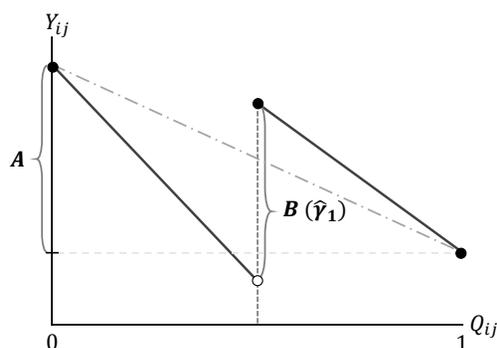
Notas: Regressões de *dummies* de Bookid, Science Cluster Random Number e suas respectivas interações nas características individuais. Os coeficientes de cada *dummy* não são apresentados nos resultados. Por suposição, todos os coeficientes estimados associados a cada grupo devem ser iguais a zero. *DF* = Graus de Liberdade, tal que  $F\text{-Stat} \sim F(DF1, DF2)$ . <sup>a</sup> F-Stat é a estatística do teste F para a hipótese de que todos os parâmetros de *Booklet*, *Random Science Cluster Number* e suas interações são conjuntamente iguais a zero.

#### 4. Estratégia empírica de decomposição do desempenho

Nessa seção descrevermos o procedimento de estimação do decaimento dos alunos ao longo da prova formulado em [Borghans e Schils \(2012\)](#), e também utilizado em [Marchionni \(2017\)](#) e [Balart et al. \(2018\)](#). Nós também proporemos um segundo procedimento, para estimar uma medida alternativa ao decaimento, que chamaremos de recuperação durante a prova. A medida de decaimento consiste em estimar a diferença entre a probabilidade de acerto na primeira questão e a probabilidade de acerto na última questão. Já a medida de recuperação explora a descontinuidade da probabilidade de acerto das questões no entorno do intervalo das provas, sendo elaborada a partir de um modelo polinomial com um desenho de RDD (*Regression Discontinuity Design*), tendo a distância ao intervalo como referência para o modelo. Em geral, os alunos apresentam uma probabilidade de acerto muito baixa pouco antes do intervalo e uma alta probabilidade de acerto após o intervalo, semelhante à probabilidade de acerto no início da prova.

Uma inovação metodológica proposta por este trabalho é levar em conta, mesmo para a estimação do decaimento, a influência do intervalo entre as provas. Obtendo a resultados mais acurados para a probabilidade de acerto no início e no final da prova. Tanto o decaimento como a recuperação são apresentados como uma tentativa de capturar a influência das habilidades não cognitivas durante uma prova.

Para dar mais clareza aos efeitos estimados, a [Figura 2](#) demonstra uma interpretação gráfica das duas medidas de habilidades não cognitivas que podem ser construídas, nela temos a probabilidade média de acerto no eixo das ordenadas e a posição das questões, variando de 0 a 1, no eixo das abscissas e trajetória da probabilidade de acerto representada em linhas sólidas. O ponto inicial representa a probabilidade de acertar a primeira questão, que pode ser considerada uma medida de habilidades cognitivas. O ponto final do gráfico representa a probabilidade de acerto da última questão da prova. O decaimento é a medida dessa diferença (Efeito A). Já a recuperação é representada pelo Efeito B, estimado a partir da descontinuidade da probabilidade de acerto no ponto mediano da prova.



**Figura 2.** Ilustração de resultados dos modelos.

O motivo da queda de desempenho ao longo do exame pode ser devido a vários fatores, mas a medida de decaimento em si, apesar de representar a influência das habilidades não cognitivas, é linearmente dependente da probabilidade de acerto da primeira questão, ou seja, das habilidades cognitivas. Ao considerarmos a descontinuidade de desempenho ao redor do intervalo, propomos uma outra medida que seja, ao menos, linearmente independente do ponto inicial, portanto, capaz de representar melhor fatores não cognitivos.

Borghans e Schils (2012) desenvolve uma abordagem para se decompor as notas do PISA em dois elementos: o desempenho inicial e o decaimento no desempenho durante o teste. Para estimar o decaimento de desempenho consideramos que a probabilidade de acertar uma questão durante um exame depende da habilidade cognitiva dos alunos, assim como depende também de suas habilidades socioemocionais. Outro fator que influencia essa probabilidade são as condições e o contexto em que cada questão é apresentada. Pelas características intrínsecas a cada um desses fatores, há grande dificuldade em discernir os seus efeitos separadamente, e qualquer estratégia de identificação possível depende de hipóteses sobre o tipo de influência de cada um desses sobre a probabilidade de acertar uma questão.

Utilizamos a posição relativa de cada questão no exame a fim de identificar o efeito das habilidades socioemocionais ou não cognitivas. Para isso, nos valem da hipótese de que as mesmas não influem na probabilidade de acertar uma questão logo no início da prova, i.e., no início de uma prova apenas a habilidade cognitiva influencia a probabilidade de acertar uma questão. Ao longo da prova, no entanto, traços da personalidade como atenção, motivação ou persistência passam a ter maior relevância na probabilidade de acerto das questões. Para que as estimativas desse efeito não sejam enviesadas, é preciso que a posição de cada questão seja não correlacionada com outras características que influenciem em sua dificuldade, como seu conteúdo e disciplina. Também é necessário que a posição de cada questão seja não correlacionada com a habilidade cognitiva ou com características individuais (posses individuais, contexto familiar, educação dos pais, etc.) que podem influenciar nas habilidades cognitivas de cada aluno realizando os exames. Essas duas últimas condições são garantidas pela aleatorização dos *clusters* de questões, confirmada nos resultados da Tabela 1.

A estratégia empírica aqui adotada para identificar o decaimento é a mesma adotada em Borghans e Schils (2012), Marchioni (2017) e em Balart et al. (2018), mas

incluindo também uma variável para capturar o efeito do pequeno intervalo entre as duas horas de prova. Para tanto, estimamos o seguinte modelo:

$$Z_{ij} = \beta_0 + \beta_1 I_1 Q_{ij} + \beta_2 Q_{ij} + \beta_3 I_1 + \sum_{j=2}^J \mu_j + \varepsilon_{ij}, \quad (1)$$

onde a variável dependente  $Z_{ij}$  é a *dummy* de acerto da questão  $j$  pelo indivíduo  $i$ . Como variáveis explicativas temos:  $Q_{ij}$  é a posição normalizada entre 0 e 1 da questão, sendo que para a primeira questão o valor atribuído a  $Q_{ij}$  é 0, para a última questão antes do intervalo o valor é de 0,5 e para última questão, é igual 1;  $I_1$  é uma variável *dummy* para as questões após o intervalo; o termo  $\mu_j$  é o efeito fixo de cada uma das questões, sua utilização se dá pela intenção de se controlar pela dificuldade específica de cada uma das questões. Com a especificação linear incluindo a variável para capturar o efeito do intervalo, a interpretação não é direta, o decaimento é obtido a partir da equação (2), com a diferença entre a probabilidade de acertar a primeira e últimas questões:

$$\begin{aligned} \text{Decaimento} &= \widehat{Z}_1 - \widehat{Z}_0, \\ \text{Decaimento} &= \widehat{\beta}_1 + \widehat{\beta}_2 + \widehat{\beta}_3. \end{aligned} \quad (2)$$

Como mencionado acima, estimaremos também uma medida de recuperação da probabilidade de acertos entre o intervalo, nos valendo das mesmas características de aleatorização para estimar o decaimento. Notamos que há dois momentos distintos em que o aluno perde a capacidade de acerto das questões, durante o primeiro bloco de provas e durante o segundo bloco de provas.

Podemos supor que os alunos já muito prejudicados por fatores que atrapalham o seu desempenho ideal no final da primeira prova, ao serem expostos a uma pequena pausa, reiniciam a prova no segundo bloco mais concentrados, motivados ou ainda capazes de ler e interpretar melhor as questões pela experiência prévia na parte anterior. Começando o segundo bloco de questões quase tão bem dispostos quanto no início do primeiro bloco.

Desta forma, alunos que não são prejudicados por estes fatores não cognitivos não deveriam apresentar recuperação nenhuma, uma vez que o seu desempenho é resultado somente de suas habilidades cognitivas. Já alunos muito prejudicados no primeiro bloco devem apresentar uma grande recuperação entre os dois blocos, uma vez que quase toda a sua capacidade de responder as questões é afetada por fatores não cognitivos. Assim, a medida aqui proposta não é inteiramente substituta ou mesmo independente do decaimento, mas apenas se distancia dos extremos da prova para capturar a influência das habilidades socioemocionais.

Para estimarmos com maior precisão a descontinuidade do desempenho dos alunos no intervalo, apresentamos os resultados de uma forma polinomial cúbica da equação (1), apresentada na equação (3). Seguiremos a estrutura e o desenho de uma regressão descontínua, de modo a atenuar um possível viés gerado pela estimação de coeficientes lineares de uma variável contínua em torno da descontinuidade:

$$Z_{ij} = \gamma_0 + \gamma_1 I_1 + \gamma_2 I_1 X_{ij} + \gamma_3 X_{ij} + \gamma_4 I_1 X_{ij}^2 + \gamma_5 X_{ij}^2 + \gamma_6 I_1 X_{ij}^3 + \gamma_7 X_{ij}^3 + \sum_{j=2}^J \mu_j + \varepsilon_{ij}. \quad (3)$$

Nesta equação,  $X_{ij}$  é uma variável derivada da posição normalizada  $Q_{ij}$ , que mede a distância relativa de cada ponto da prova ao intervalo, variando de 0,5 a 0,5. A construção da variável deste modo faz com que todas as questões anteriores ao intervalo respondam por 50% da prova e todas as questões após o intervalo respondam pelos outros 50% de prova. Neste modelo, o coeficiente de interesse é  $\gamma_1$ , associado à variável *dummy* que denota as questões após o intervalo ( $I_1$ ), e exprime a dimensão da descontinuidade em relação à variável dependente do modelo. Ao utilizarmos um modelo de probabilidade linear, essa dimensão está limitada entre 0 e 1. Desta forma, a descontinuidade estimada para cada país será expressa pelo coeficiente  $\hat{\gamma}_1$  e interpretada como a recuperação, entre os dois blocos de prova, da capacidade de acerto das questões.

## 5. Decaimento

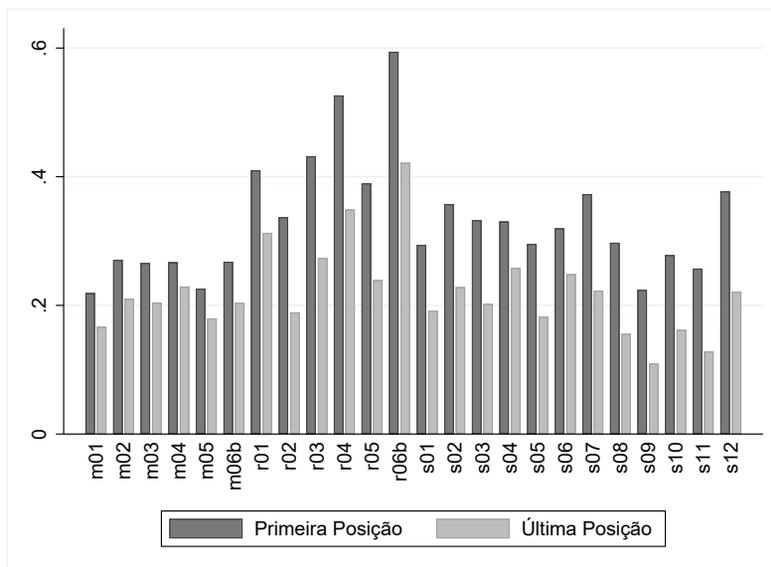
Em primeiro lugar, verificamos que há decaimento do desempenho dos alunos. Como podemos ver na [Figura 3](#), há declínio no desempenho mesmo se comparamos os acertos médio nos próprios *clusters* de questões, quando aparecem na primeira ou última posição da prova. A [Figura 4](#) mostra que mesmo se dividirmos as provas por disciplina, o padrão de queda do desempenho se mantém. Adicionalmente, para alguns *clusters* há um comportamento a princípio não esperado, pois quando apresentados na terceira posição há uma maior probabilidade de acerto do que quando apresentados na segunda, mas ainda assim a probabilidade é menor do que quando apresentados na primeira posição. Isso se deve, novamente, ao fato de que, foi imposta aos participantes do exame uma pequena pausa de no máximo 5 minutos antes de começar o bloco de testes dos *clusters* três e quatro, e o tempo de resolução para cada grupo de dois *clusters* foi limitado a apenas uma hora. Uma particularidade do exame em 2015 é de que os participantes que realizaram os testes em computadores (provas do tipo CBA) tiveram de resolver as questões do teste de uma forma fixa e sequencial, de modo que não podiam voltar para perguntas anteriores e revisar suas respostas depois de chegar ao final de cada teste. Mesmo assim, para nenhum *cluster* há probabilidade maior de acertar as questões em qualquer outra posição que não a primeira. Isso é um indício de que de fato há influência das habilidades socioemocionais no decorrer da prova.

### 5.1 Estimativas de decaimento

A [Tabela 2](#) mostra que a queda de desempenho estimado ao longo de uma prova extensa como o PISA difere significativamente entre países, assim como difere a probabilidade média dos alunos de cada país acertarem a primeira questão de uma mesma prova. O Brasil aparece com um dos menores níveis de probabilidade de acertar a primeira questão, o que pode indicar um baixo nível de habilidades cognitivas (posição 48 de 56).<sup>3</sup> O decaimento brasileiro é o maior, o que representa, na interpretação de [Borghans e Schils \(2012\)](#), um baixo nível de habilidades socioemocionais. Podemos notar uma correlação negativa entre as duas medidas, o que significa que os países que mais acertam a primeira questão são também aqueles que tem menor decaimento.

---

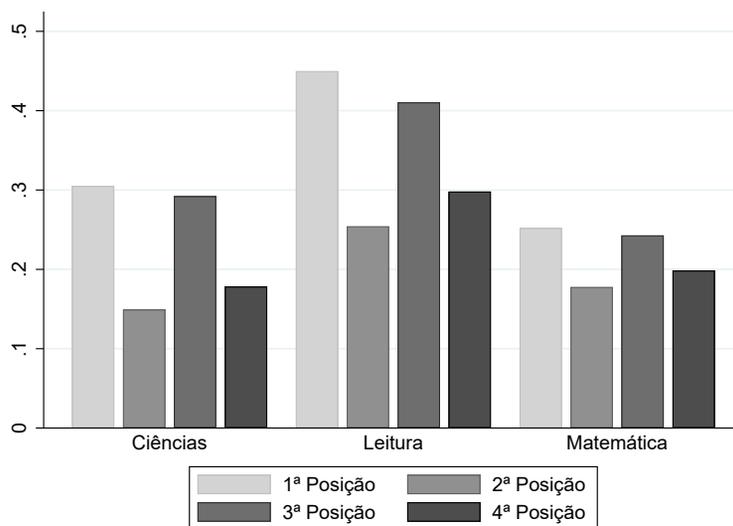
<sup>3</sup>Consideramos 56 regiões e países que fizeram a prova em CBA, excluindo regiões específicas dos EUA que fazem a prova de modo separado, excluindo também o território de Porto Rico, que não fez a prova em formato CBA.



Nota: Percentual de acertos utilizando a amostra de todos os países. Para cada cluster, a letra inicial indica sua disciplina ("m" para matemática, "r" para leitura e "s" para ciências).

Fonte: PISA 2015. Elaboração própria.

**Figura 3.** Percentual de Acertos Dentro dos *clusters*, por Posição na Prova.



Nota: Percentual de acertos utilizando a amostra de todos os países.

Fonte: PISA 2015. Elaboração própria.

**Figura 4.** Percentual Médio de Acertos em cada Disciplina por Posição na Prova.

**Tabela 2.** Queda de desempenho estimada para países que fizeram a prova em CBA.

País	Prob. Acertar a 1ª Questão	Prob. Acertar última antes do Break <sup>a</sup>	Prob. Acertar a 1ª após o Break <sup>b</sup>	Prob. Acertar a Última Questão	Decaimento (Total)	Decaimento Antes do Break <sup>c</sup>	Decaimento Após o Break <sup>b</sup>
Brazil	0,403	0,130	0,360	0,176	0,227	0,273	0,184
Uruguay	0,453	0,267	0,411	0,287	0,167	0,186	0,125
Tunisia	0,340	0,168	0,283	0,179	0,161	0,172	0,104
Russian Federation	0,541	0,369	0,508	0,383	0,158	0,172	0,125
Portugal	0,577	0,395	0,529	0,428	0,149	0,182	0,102
Sweden	0,564	0,402	0,523	0,420	0,143	0,162	0,102
Mexico	0,411	0,223	0,373	0,270	0,141	0,189	0,104
Peru	0,370	0,165	0,339	0,231	0,139	0,205	0,108
Colombia	0,413	0,243	0,371	0,276	0,137	0,170	0,095
Montenegro	0,397	0,269	0,360	0,261	0,136	0,128	0,099
Slovak Republic	0,503	0,383	0,459	0,369	0,134	0,120	0,090
New Zealand	0,581	0,446	0,545	0,449	0,132	0,135	0,096
Thailand	0,401	0,270	0,360	0,277	0,124	0,131	0,083
Italy	0,535	0,416	0,497	0,413	0,123	0,119	0,085
Israel	0,501	0,366	0,459	0,381	0,121	0,135	0,078
France	0,555	0,440	0,517	0,435	0,121	0,115	0,083
Norway	0,571	0,456	0,530	0,450	0,120	0,115	0,080
Bulgaria	0,463	0,318	0,427	0,348	0,115	0,146	0,079
Costa Rica	0,400	0,265	0,371	0,285	0,115	0,134	0,086
Chile	0,457	0,326	0,423	0,344	0,113	0,131	0,079
Greece	0,482	0,358	0,446	0,370	0,113	0,124	0,077
Luxembourg	0,527	0,405	0,489	0,417	0,111	0,122	0,072
Australia	0,563	0,472	0,534	0,455	0,108	0,091	0,080
Hungary	0,501	0,420	0,475	0,394	0,107	0,081	0,081
United Arab Emirates	0,436	0,353	0,395	0,331	0,105	0,083	0,064
Belgium	0,564	0,465	0,535	0,461	0,103	0,099	0,074
Spain	0,543	0,426	0,512	0,441	0,102	0,117	0,071
Spain (Regions)	0,546	0,434	0,513	0,445	0,101	0,112	0,068
Switzerland	0,571	0,468	0,535	0,470	0,101	0,103	0,066
Iceland	0,506	0,411	0,469	0,405	0,101	0,096	0,064
Dominican Republic	0,266	0,125	0,247	0,167	0,100	0,142	0,080
Qatar	0,369	0,267	0,334	0,274	0,095	0,102	0,060
United States	0,538	0,435	0,504	0,446	0,092	0,103	0,058
Austria	0,535	0,477	0,503	0,445	0,091	0,058	0,058
Slovenia	0,568	0,523	0,523	0,479	0,089	0,044	0,044
Japan	0,602	0,526	0,598	0,522	0,080	0,076	0,076
Czech Republic	0,536	0,456	0,514	0,456	0,080	0,080	0,057
Denmark	0,563	0,474	0,535	0,483	0,079	0,089	0,052
Latvia	0,512	0,452	0,493	0,433	0,079	0,060	0,060
United Kingdom	0,555	0,503	0,529	0,477	0,078	0,052	0,052
Canada	0,596	0,521	0,566	0,520	0,076	0,075	0,046
Turkey	0,399	0,341	0,382	0,324	0,075	0,058	0,058
Germany	0,564	0,488	0,550	0,496	0,068	0,076	0,054
Estonia	0,596	0,533	0,582	0,532	0,065	0,064	0,050
Lithuania	0,488	0,433	0,478	0,424	0,064	0,054	0,054
B-S-J-G (China)	0,571	0,493	0,572	0,510	0,061	0,078	0,063
Macao	0,591	0,495	0,577	0,536	0,055	0,095	0,042
Netherlands	0,569	0,530	0,554	0,515	0,054	0,039	0,039
Poland	0,540	0,486	0,546	0,492	0,048	0,054	0,054
Finland	0,581	0,538	0,582	0,539	0,043	0,043	0,043
Singapore	0,635	0,589	0,622	0,592	0,043	0,046	0,030
Hong Kong	0,597	0,558	0,595	0,557	0,040	0,038	0,038
Croatia	0,474	0,435	0,483	0,445	0,029	0,039	0,039
Korea	0,559	0,537	0,554	0,533	0,026	0,021	0,021
Ireland	0,539	0,507	0,546	0,515	0,024	0,031	0,031
Chinese Taipei	0,576	0,562	0,579	0,566	0,010	0,013	0,013

Nota: Decaimento estimado conforme a equação (1) e probabilidade de acertar a primeira questão, para cada sessão da prova, levando em conta o intervalo e os pesos amostrais do PISA 2015. Todas as estimativas são significativas ao menos a 5%.

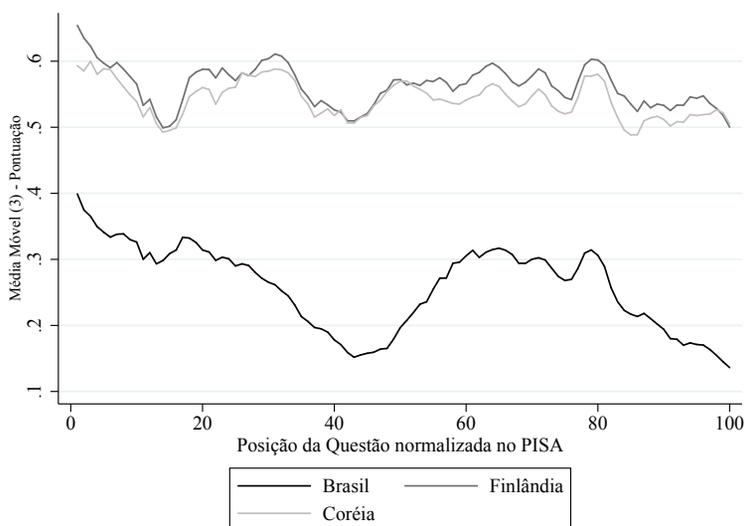
Fonte: PISA 2015. Elaboração própria.

Ainda nesta tabela, temos as probabilidades de acertos estimadas em pontos específicos da prova, como a última questão antes do intervalo (break) e a primeira questão após o intervalo, nota-se uma diferença razoável entre estas duas estimativas, sugerindo que a validade de alguma recuperação envolvendo o intervalo entre os blocos de prova. Mesmo com alguma recuperação providenciada pelo intervalo, a probabilidade de acerto da primeira questão após o intervalo dos alunos brasileiros é ainda menor em comparação aos outros países (posição 52 de 56).

Ilustramos o decaimento dos alunos brasileiros ao longo da prova, em comparação com países com bom desempenho na [Figura 5](#), onde comparamos a probabilidade de acertos de alunos do Brasil, Finlândia e Coréia do Sul ao longo da prova. Podemos notar que, à medida em que os alunos avançam para o final da prova, onde a posição normalizada da questão é igual a 1, a queda na taxa de acertos é muito mais acentuada para o Brasil do que para Finlândia e Coréia do Sul. O desempenho ao longo da prova destes últimos parece ser mais constante e menos afetado por fatores que não podem ser atribuídos ao “conhecimento” dos alunos, como o tempo total de prova ou o cansaço e a distração decorrente de uma prova mais longa.

## 5.2 Possíveis explicações para o decaimento

Por que o Brasil possui um decaimento tão grande em relação ao dos demais países? Isso se deve somente às habilidades socioemocionais, ou também há outros fatores envolvidos? Nessa seção, exploramos com mais detalhes a estrutura do exame e as variáveis de tempo de cada item, para tentar responder a essas questões. Na edição de 2015, os exames do tipo CBA tinham a limitação de uma estrutura sequencial das



*Nota:* Médias móveis de 3 termos defasados da probabilidade de acertos observados ao longo da prova para Brasil, Finlândia e Coréia do Sul, por posição normalizada da questão. O eixo horizontal mostra a posição normalizada das questões, colocando o intervalo na posição 0,5 e o final da prova na posição 1.

*Fonte:* PISA 2015. Elaboração própria.

**Figura 5.** Queda de desempenho observada ao longo da prova para o Brasil, Finlândia e Coréia do Sul.

questões, com tempo fixo por bloco de prova. Essas características evidenciam alguns pontos que nos ajudam a entender o desempenho do Brasil.

A [Tabela A-1](#), no [Apêndice](#), mostra a proporção de acertos, a proporção daqueles que não chegaram ao item e tempo médio em cada questão, ao longo da prova. Em primeiro lugar, como o esperado, a taxa de acertos decai ao longo da prova e dentro de cada bloco de dois *clusters*. Como também é esperado, a proporção daqueles que não chegam nem a responder à questão é crescente em cada bloco. No entanto, o aumento da proporção é muito acelerado e chega ao máximo de 60%. Esse padrão sugere que os alunos brasileiros não conseguem administrar bem o tempo de prova e que a pontuação dos alunos brasileiros poderia ser maior, caso estes conseguissem completar os blocos de questões.<sup>4</sup>

Em segundo lugar, a [Figura 6](#), que mostra apenas estas estatísticas para o Brasil, deixa clara a diferença entre os dois blocos. A proporção dos alunos que não chegam à questão é claramente menor no segundo bloco, especialmente no final do bloco, em que chega a 40%, em comparação ao pico da primeira parte. Além disso, no segundo bloco, a taxa de acertos parte de um nível menor e chega a um nível maior do que no primeiro bloco, sendo menos decrescente e relativamente mais estável. Essas diferenças sugerem que os alunos brasileiros aprendem a fazer a prova entre os dois blocos, distribuindo melhor o tempo de resolução das questões.

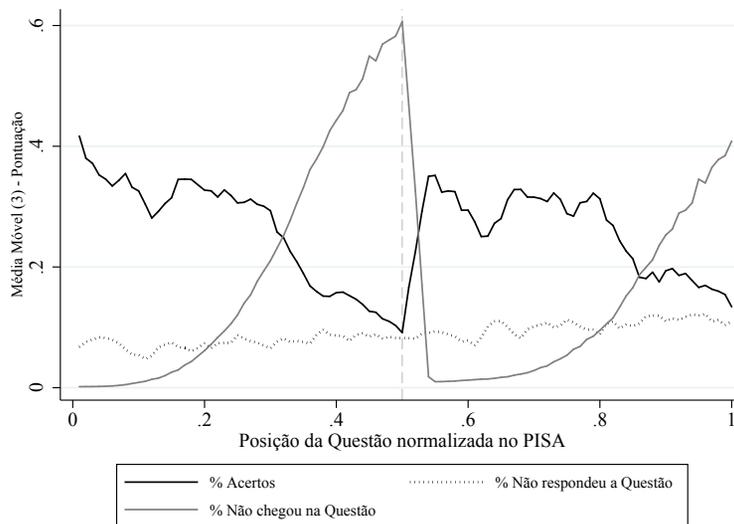
Na [Figura 7](#), comparamos a participação dos alunos em cada posição das questões na prova com a de outros países. Notamos que a porcentagem de questões não alcançadas pelos alunos brasileiros é muito maior em relação à mesma para países exemplares no PISA, como a Finlândia, ou mesmo para a Colômbia, cujo desempenho no PISA é comparável ao desempenho brasileiro. Enquanto a última questão do primeiro bloco de uma hora não é alcançada por, aproximadamente, 6% dos alunos finlandeses e 18% dos alunos colombianos, cerca de 61% dos alunos brasileiros não alcançam essa mesma questão.

Os dados de aplicação das provas do tipo CBA no PISA 2015 apresentam também o tempo que os alunos levaram para responder cada questão de seus respectivos formulários de prova. Na [Figura 8](#), reportamos o tempo médio de resposta ao longo da prova. A figura mostra que os alunos brasileiros gastam muito tempo para responder às questões iniciais da prova, fazendo com que poucos consigam chegar ao final das provas. É possível que os alunos que não chegam ao final da prova levem mais tempo para ler e entender o enunciado das questões e elaborar o raciocínio que leva à resposta. Nesse sentido, o mal desempenho dos alunos teria pouco a ver com habilidades socioemocionais e dependeriam mais das habilidades cognitivas.

Ao compararmos o tempo médio de resolução das questões ao longo da prova de outros países com o tempo médio dos alunos brasileiros, na [Figura 9](#), podemos notar que este é muito menor para países exemplares no PISA, como Coréia do Sul e Finlândia, e com variância menor. Curiosamente, ao compararmos os resultados

---

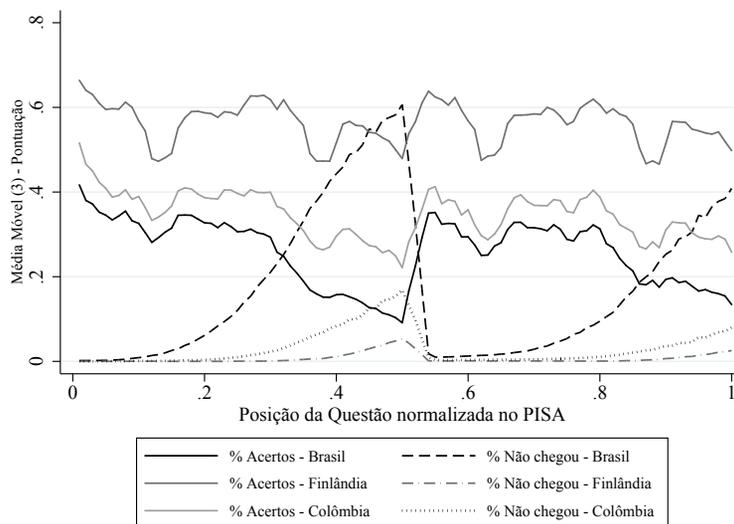
<sup>4</sup>Mais especificamente, o desempenho seria melhorado, caso o efeito marginal de cada unidade de tempo sobre a probabilidade de acerto do item fosse maior no final da prova (para uma questão adicional) do que no início da prova (para questões que já estão sendo resolvidas). Esse aspecto está associado à organização da prova, uma vez que, ao contrário das provas do tipo PBA, os alunos não podem escolher as questões que sabem responder e reflete mais a estratégia de resolução da prova, do que o conhecimento sobre os assuntos.



Nota: Proporção de acertos, de não respostas e de casos em que o aluno não chega na questão, por posição normalizada da questão. Os três percentuais somados, mais a proporção de respostas erradas, totaliza 100%. O eixo horizontal mostra a posição normalizada das questões, colocando o intervalo na posição 0,5 e o final da prova na posição 1.

Fonte: PISA 2015. Elaboração própria.

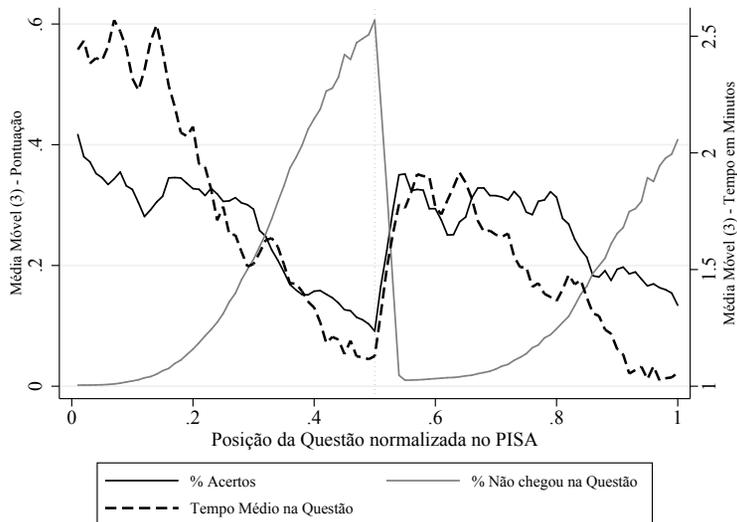
**Figura 6.** Queda de desempenho observada e porcentagem de alunos brasileiros que não respondeu às questões no PISA 2015.



Nota: Proporção de acertos e de casos em que o aluno não chega na questão, por posição normalizada da questão. Os dois percentuais somados, mais a proporção de respostas erradas e de não resposta, totaliza 100%. O eixo horizontal mostra a posição normalizada das questões, colocando o intervalo na posição 0,5 e o final da prova na posição 1.

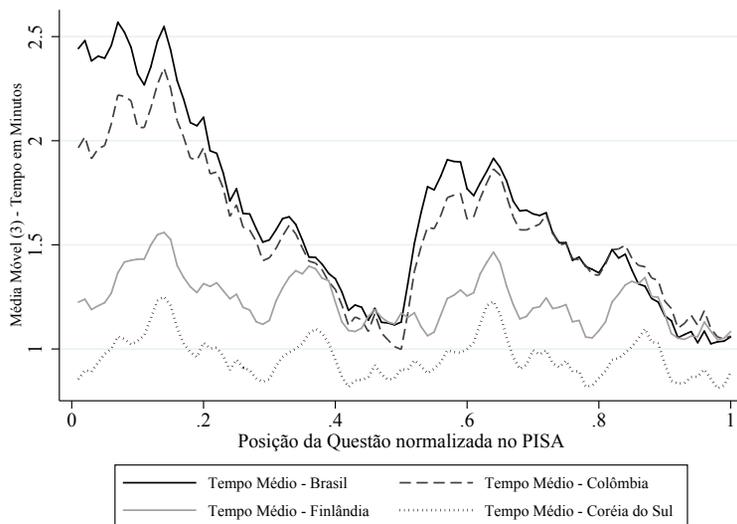
Fonte: PISA 2015. Elaboração própria.

**Figura 7.** Queda de desempenho e porcentagem de alunos que não chegaram às questões, Brasil em comparação a outros países no PISA 2015.



*Nota:* Proporção de acertos, de casos em que o aluno não chega na questão e tempo médio gasto na questão, por posição normalizada da questão. Os dois percentuais somados, mais a proporção de respostas erradas e de não resposta, totaliza 100%. O eixo horizontal mostra a posição normalizada das questões, colocando o intervalo na posição 0,5 e o final da prova na posição 1.  
*Fonte:* PISA 2015. Elaboração própria.

**Figura 8.** Queda de desempenho, porcentagem de alunos brasileiros que não conseguiu responder e tempo médio gasto nas questões no PISA 2015.



*Nota:* Tempo médio gasto na questão, por posição normalizada da questão. O eixo horizontal mostra a posição normalizada das questões, colocando o intervalo na posição 0,5 e o final da prova na posição 1.  
*Fonte:* PISA 2015. Elaboração própria.

**Figura 9.** Tempo médio gasto nas questões, Brasil em comparação a outros países no PISA 2015.

colombianos com os resultados brasileiros, vemos que, mesmo com um tempo médio pouco menor do que o tempo médio brasileiro, a parcela dos alunos colombianos que chega ao final das provas é consideravelmente maior do que a parcela de alunos brasileiros (Figura 7).

A redução do tempo médio ao longo de um bloco pode refletir dois movimentos. O primeiro deles, é um efeito composição, em que aqueles que gastam mais tempo nas primeiras questões acabam não chegando às questões ao final do bloco, de modo que aqueles que respondem a essas questões (e que entram na conta da média) são os alunos que já tinham tempo menor de resposta. O segundo efeito é de aprendizado, ou de estratégia de resolução da prova, em que o mesmo aluno pode reduzir o tempo de resolução da cada questão para tentar responder mais questões. A diferença entre o Brasil e a Colômbia pode ser explicada por um ou outro desses fatores ter maior peso. Enquanto na Colômbia o segundo efeito parece ser mais acentuado, no Brasil o efeito composição parece ser maior, de modo que os alunos que gastam mais tempo nas primeiras questões não passam a controlar o tempo de resposta por questão para chegar mais longe na prova.

A Tabela 3 descreve a influência do tempo médio gasto nos primeiros 5% de questões após o início de cada bloco de uma hora de provas. Vemos que, para os países com desempenho relativamente constante ao longo da prova (Finlândia e Coreia do Sul), a média de tempo por questão é menor e os tempos máximos dos quatro grupo de alunos são mais concentrados em torno desta média. Além disso, os alunos que tomam mais tempo nas primeiras questões de fato tem menor probabilidade de chegar ao final de cada bloco do exame, porém apresentam uma probabilidade de acertar maior do que a de seus pares que demoram menos tempo. Já para o Brasil e a Colômbia, vemos que o tempo médio por questão é maior, e os valores máximos dos grupo de alunos são mais dispersos. Além disso, apesar das médias de tempo suficientemente parecidas, há grandes diferenças em relação à probabilidade de chegar ao final do bloco. Entre os alunos brasileiros do grupo com menor tempo de resolução das questões, apenas 54% conseguem atingir os últimos 5% de questões no primeiro bloco de provas, o que é um resultado muito abaixo do esperado, mesmo se comparado ao grupo que gasta mais tempo nas primeiras questões na Colômbia, em que 77,5% atingem os últimos 5% de questões no primeiro bloco.

A Tabela 3 mostra, além disso, que o possível aprendizado dos alunos brasileiros em fazer a prova se reflete em uma redução significativa do tempo de resolução das primeiras questões, além de um aumento acentuado da probabilidade de chegar ao final do bloco. O grupo que gasta mais tempo no segundo bloco chega a uma probabilidade de alcançar o final do bloco superior àquela do grupo que gastou menos tempo no primeiro bloco. O grupo brasileiro que gasta menos tempo no segundo bloco chega a uma probabilidade superior ao grupo que gastou menos tempo na Colômbia no primeiro bloco.

Como as questões possuem ordem aleatória, a única coisa que se altera é a experiência de ter realizado um bloco da prova. Nesse sentido, é interessante notar que mesmo com o aumento da proporção de alunos que chegam ao final da prova aumentar no segundo bloco, a proporção de acertos também aumenta, do segundo ao quarto grupos de alunos. Esse resultado sugere que, entre os alunos que aprendem a fazer a

**Tabela 3.** Relação entre Tempo Médio, Pontuação e participação ao longo do PISA 2015 para o Brasil em comparação a outros países.

País	Tempo (5% Início)	Quartil (Tempo)	Tempo Max.	Prob. (>45%)	Acertos** (>45%)	Tempo (5% Intervalo)	Tempo Max.	Prob. (>95%)	Acertos** (>95%)	Prob. (>45%) e (>95%)
Brasil	2,48	1	1,24	0,547	0,512	1,82	0,91	0,881	0,460	0,404
		2	1,98	0,483	0,520		1,44	0,791	0,535	0,360
		3	3,12	0,389	0,487		2,26	0,711	0,544	0,290
		4*	9,94	0,285	0,453		7,56	0,569	0,523	0,205
Colômbia	2,07	1	0,97	0,845	0,612	1,64	0,79	0,968	0,577	0,684
		2	1,73	0,875	0,622		1,33	0,962	0,645	0,717
		3	2,76	0,824	0,583		2,13	0,950	0,640	0,674
		4*	7,89	0,775	0,539		6,38	0,920	0,613	0,617
Finlândia	1,26	1	0,66	0,946	0,831	1,11	0,57	0,986	0,787	0,769
		2	1,03	0,948	0,848		0,89	0,988	0,847	0,798
		3	1,58	0,929	0,836		1,40	0,983	0,863	0,784
		4*	4,88	0,886	0,859		4,44	0,969	0,862	0,748
Coreia do Sul	0,97	1	0,43	0,968	0,838	0,92	0,39	0,993	0,739	0,784
		2	0,75	0,961	0,862		0,70	0,994	0,862	0,796
		3	1,24	0,959	0,866		1,19	0,993	0,883	0,820
		4*	4,14	0,952	0,869		4,13	0,989	0,896	0,810

Notas: Tabela com tempo médio de resposta nas 5% primeiras questões de cada bloco para o total dos alunos por país, além do tempo máximo, probabilidade de chegar às 5% últimas questões do bloco e probabilidade de acerto dessas questões, condicional a atingi-las, para alunos divididos em 4 grupos de tamanho igual, segundo o tempo utilizado nas 5% primeiras questões. \*Valores extremos (acima de 99% da distribuição) de tempo por questão removidos. \*\*Probabilidade de acertar uma das últimas questões, dado que o aluno atingiu a questão.

Fonte: PISA 2015. Elaboração própria.

prova no segundo bloco, o efeito marginal do tempo sobre a probabilidade de acerto em novos itens no final da prova é maior do que o efeito sobre a probabilidade de acerto de itens que já seriam vistos de qualquer forma no início da prova. Em outras palavras, para aumentar a nota, vale mais a pena o aluno gastar tempo em novos itens, deixando de lado aqueles em que possivelmente tem maior dificuldade no início da prova. Esse resultado parece reforçar o argumento de que a diferença de decaimento do Brasil em relação aos demais países não se relaciona somente às habilidades socioemocionais.

Quando separamos os alunos brasileiros de escolas públicas daqueles de escolas privadas, na [Tabela 4](#), os primeiros apresentam maior tempo de resolução, proporções menores de alunos que chegam ao final do bloco e de acertos. Com essa desagregação, podemos observar, além disso, que ambos os grupos apresentam redução grande no tempo de resposta das primeiras questões do segundo bloco em comparação ao primeiro bloco e significativo crescimento na probabilidade de chegar ao final do segundo bloco. No entanto, somente entre os alunos de escolas públicas a probabilidade de acertos parece aumentar de forma relevante, do segundo ao quarto grupos de alunos. Entre os alunos das escolas privadas, a probabilidade de acerto diminui ou se mantém relativamente constante.

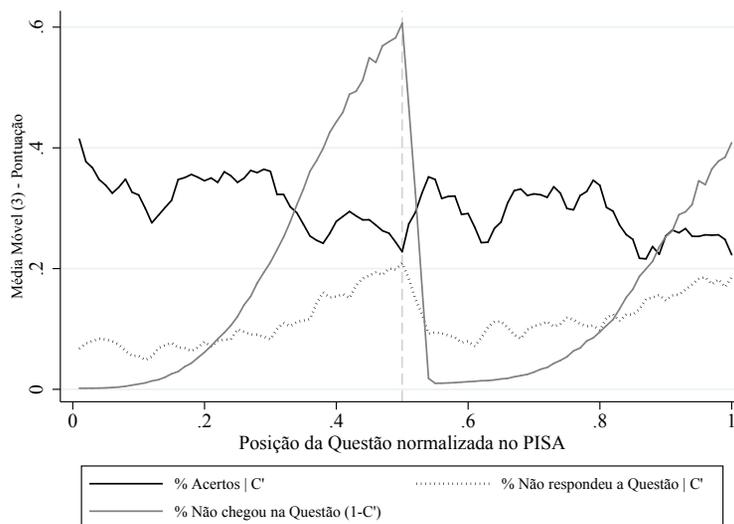
Dadas as comparações aqui apresentadas, parece intuitivo propor que um modelo para estimação do decaimento de desempenho entre diferentes grupos deva levar em conta a probabilidade de acerto condicional ao aluno ter chegado à questão e o tempo gasto em cada questão, que a princípio deve exercer um efeito positivo na probabilidade de acerto das questões. A [Figura 10](#) mostra qual seria a pontuação condicionada ao percentual de alunos brasileiros que alcançam as questões.

**Tabela 4.** Relação entre Tempo Médio, Pontuação e participação ao longo do PISA 2015 para escolas públicas e privadas no Brasil.

País	Tempo (5% Início)	Quartil Dist. de Tempo	Tempo Max.	Prob. (>45%)	Acertos** (>45%)	Tempo (5% Intervalo)	Tempo Max.	Prob. (>95%)	Acertos** (>95%)	Prob. (>45%) e (>95%)
Públicas	2,51	1	1,26	0,541	0,482	1,82	0,91	0,884	0,437	0,400
		2	2,01	0,476	0,498		1,44	0,792	0,510	0,351
		3	3,15	0,381	0,462		2,26	0,708	0,516	0,284
		4*	10,04	0,284	0,435		7,62	0,558	0,489	0,204
Privadas	2,10	1	1,08	0,591	0,695	1,70	0,90	0,872	0,627	0,441
		2	1,69	0,574	0,718		1,39	0,816	0,705	0,430
		3	2,65	0,481	0,691		2,14	0,758	0,696	0,357
		4*	7,95	0,401	0,659		6,28	0,665	0,672	0,290

Notas: Tabela com tempo médio de resposta nas 5% primeiras questões de cada bloco para o total dos alunos por país, além do tempo máximo, probabilidade de chegar às 5% últimas questões do bloco e probabilidade de acerto dessas questões, condicional a atingi-las, para alunos divididos em 4 grupos de tamanho igual, segundo o tempo utilizado nas 5% primeiras questões. \*Valores extremos (acima de 99% da distribuição) de tempo por questão removidos. \*\*Probabilidade de acertar uma das últimas questões, dado que o aluno atingiu a questão.

Fonte: PISA 2015. Elaboração própria.



Nota: Proporção de acertos e de não resposta, dado que o aluno chegou à questão, por posição normalizada da questão, e proporção de casos em que o aluno não chega na questão por posição normalizada da questão. Os dois primeiros percentuais somados, mais a proporção de respostas erradas, totaliza 100%. O eixo horizontal mostra a posição normalizada das questões, colocando o intervalo na posição 0,5 e o final da prova na posição 1.

Fonte: PISA 2015. Elaboração própria.

**Figura 10.** Queda de desempenho observada e porcentagem de alunos brasileiros que não respondeu às questões no PISA 2015, condicionada ao percentual de alunos que alcançam as últimas questões.

Nessa nova configuração, o decaimento da proporção de acertos se torna menos acentuada em cada bloco. As consequências da escolha de utilizar somente as questões atingidas para a estimação do decaimento são brevemente avaliadas em [Borghans e Schils \(2012\)](#) utilizando dados mais antigos do PISA 2009. Segundo os autores, a estimação utilizando as questões não atingidas como erradas, num modelo binário, leva à superestimação do decaimento. No entanto, atentando ao fato de que, em média, os alunos que são capazes de chegar ao final da prova têm aproveitamento de tempo melhor, esta formulação levaria a uma correlação estrita entre os fatores não observados e a posição da questão, uma vez que só os melhores alunos são capazes de responder às últimas questões, implicando em um decaimento estimado viesado. Portanto, seria necessário testar a validade desta formulação para implementá-la, o que não pode ser feito ainda neste trabalho.

Além disso, a proporção de questões não respondidas aumenta ao longo da prova. Esse aumento pode novamente ser explicado por efeitos de composição e de tempo de prova. Conforme avançamos ao longo das questões da prova, os alunos que gastaram mais tempo nas primeiras questões deixam de responder as últimas questões e não entram no cálculo da taxa de não resposta. É possível que alunos com taxas maiores de não resposta avancem mais ao longo da prova, alcançando às questões finais. Além disso, o segundo efeito é de que, o comportamento dos alunos que chegam ao final da prova pode mudar, com maiores taxas de não resposta devido ao menor tempo disponível.

Por último, mostramos que apesar do decaimento do Brasil ser acentuado, ele não é único determinante de sua baixa posição relativa de notas entre os países participantes. No cenário contrafactual em que não há decaimento ao longo dos blocos, a variável relevante para comparar os países é a média da probabilidade estimada de acerto da primeira questão, que utilizamos como medida de habilidades cognitivas.

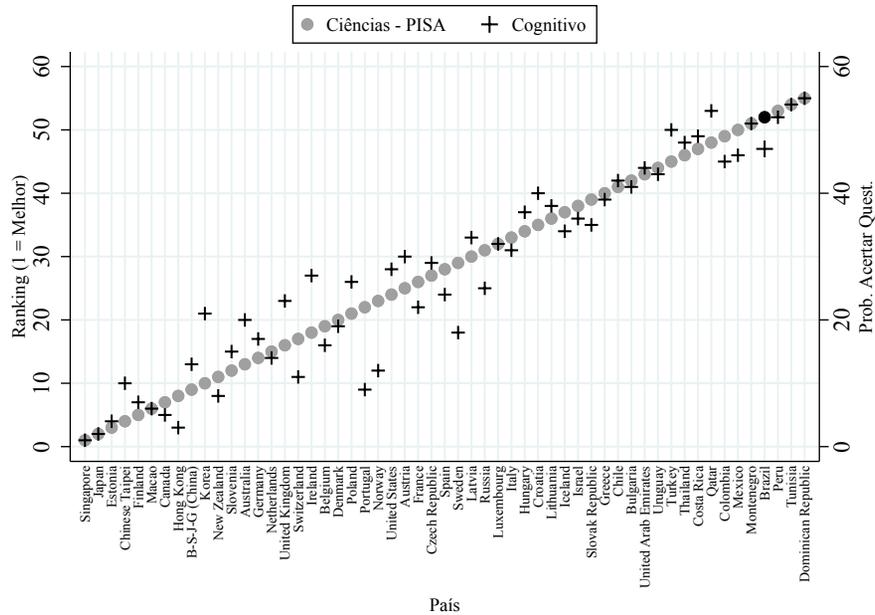
A [Figura 11](#) mostra os países ordenados de acordo com um ranking das notas de Ciências,<sup>5</sup> numa linha de 45°, e apresenta, em destaque, qual seria a posição de cada um desses países num ranking de acertar a primeira questão da prova. A distância vertical, para cada país, entre os dois pontos que representam estes dois rankings mostra os ganhos ou perdas de posições de acordo com a variável escolhida. Entre as 55 regiões<sup>6</sup> para as quais temos as notas e as estimativas de habilidades cognitivas, o Brasil aparece na posição 47 em relação a essas habilidades e na posição 52 em Ciências. Esse resultado mostra que se o Brasil e todos os demais países mantivessem o mesmo desempenho do início da prova sem serem afetados pelo decaimento, sua posição relativa melhoraria cinco posições entre os últimos países do ranking.

A [Figura 12](#) apresenta um gráfico somente para estes ganhos e perdas de posições para cada país nesses dois rankings comparados, ordenados da maior perda para o maior ganho. Dos 55 países, 24 perdem posições ao considerarmos apenas a probabilidade de acerto da primeira questão e outros 24 ganham posições. Ao observar os ganhos de outros países, podemos acreditar que um ganho de 5 posições não é desprezível, pois há apenas outros 5 países que ganham mais posições do que o Brasil.

---

<sup>5</sup>Ciências era a prova principal do exame de 2015.

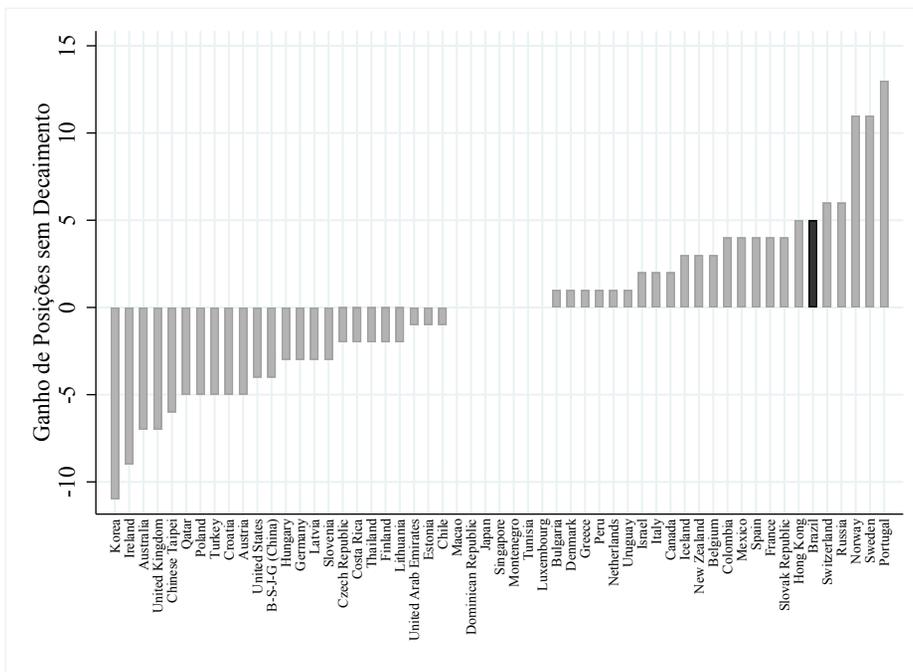
<sup>6</sup>Somente regiões ou países que fizeram provas do tipo CBA, mas com uma região a menos, excluindo as regiões da Espanha que fazem o PISA de modo separado.



Nota: Ranking pela probabilidade de acertar a primeira questão e ranking pela nota de ciências, para cada uma das 55 regiões. Destacamos o Brasil em preto.

Fonte: PISA 2015. Elaboração própria.

Figura 11. Ranking de Habilidades Cognitivas e das Notas de Ciências.



Nota: Ganho de posições num ranking entre a probabilidade de acertar a primeira questão e a nota de ciências, para cada uma das 55 regiões. Destacamos o Brasil em preto.

Fonte: PISA 2015. Elaboração própria.

Figura 12. Mudanças entre os Rankings de Habilidades Cognitivas e de Notas de Ciências.

Mesmo assim, o ganho de alguns países, como Noruega, Suécia e Portugal é muito mais expressivo do que o salto de posições brasileiro. Desta forma, há indícios de que o decaimento afeta de forma relevante a posição relativa do Brasil em comparação aos demais países, mas não pode ser apontado como o principal determinante do desempenho no exame.

## 6. Recuperação

### 6.1 Estimativas da descontinuidade

A medida de recuperação é o tamanho da descontinuidade apresentada no modelo apresentado na equação (3) deste trabalho e expressa pelo coeficiente  $\hat{\gamma}_1$ . A Tabela 5 ordena os países de acordo com a recuperação estimada, do maior para o menor. Vemos que neste ordenamento, em geral, o países que apresentam maior decaimento também apresentam maior recuperação de desempenho, sendo que o Brasil tem a maior recuperação entre os dois blocos. Peru, Portugal, México, Uruguai, Rússia, Colômbia e República Dominicana também estão entre os países com maiores descontinuidades de desempenho entre os blocos de prova e os maiores decaimentos de desempenho ao longo do exame. Este pode ser um indício de que tanto o decaimento quanto a medida de recuperação captam variações parecidas no comportamento dos alunos ao longo da prova.

Salientamos que apesar desta medida aludir a uma recuperação do desempenho, a descontinuidade não é uma dimensão positiva para as habilidades socioemocionais. Uma descontinuidade maior está associada a uma desestabilização maior por fatores externos ou motivacionais no primeiro bloco de provas, em relação aos alunos que não são prejudicados por estes fatores.

**Tabela 5.** Recuperação estimada para países que fizeram a prova em CBA.

País	$\hat{\gamma}_1$	País	$\hat{\gamma}_1$	País	$\hat{\gamma}_1$	País	$\hat{\gamma}_1$
Brazil	0,197	France	0,105	Hungary	0,08	Ireland	0,046
Peru	0,166	New Zealand	0,104	Denmark	0,079	Lithuania	0,046
Portugal	0,165	Israel	0,103	Switzerland	0,078	Austria	0,036
Uruguay	0,151	Norway	0,101	Thailand	0,076	Latvia	0,035
Sweden	0,14	Italy	0,1	Greece	0,076	United Kingdom	0,028
Mexico	0,138	Macao	0,1	Qatar	0,073	Slovenia	0,025 <sup>s5</sup>
Russian Federation	0,13	Germany	0,098	Estonia	0,069	Singapore	0,024
Costa Rica	0,12	Belgium	0,091	Finland	0,066	Chinese Taipei	0,021
Dominican Republic	0,119	Slovak Republic	0,09	Australia	0,066	United Arab Emirates	0,02
Colombia	0,119	United States	0,086	Czech Republic	0,06	Hong Kong	0,015 <sup>s10</sup>
Spain	0,118	B-S-J-G (China)	0,086	Iceland	0,058	Netherlands	0,009 <sup>ns</sup>
Bulgaria	0,11	Japan	0,084	Poland	0,058	Korea	0,009 <sup>ns</sup>
Chile	0,109	Montenegro	0,082	Canada	0,056	Turkey	0,002 <sup>ns</sup>
Luxembourg	0,109	Tunisia	0,08	Croatia	0,055		

Notas: Tabela com os valores estimados da recuperação de desempenho dos alunos entre os pontos antes e depois do intervalo, a partir da equação (2). Todos os coeficientes são significativos a menos de 1%, a não ser os valores marcados com: <sup>s5</sup> significativo somente a 5%; <sup>s10</sup> significativo somente a 10%; <sup>ns</sup> não significativo a 10%.

Fonte: PISA 2015. Elaboração própria.

Para entender quais habilidades dos alunos esta medida de recuperação é capaz de captar podemos examinar as correlações entre as medidas de habilidades cognitivas (probabilidade de acerto da primeira questão) e não-cognitivas (decaimento e recuperação), sem nenhuma pretensão de estimar relações causais entre elas. A [Tabela 6](#) mostra as regressões das medidas de habilidades cognitivas nas outras duas possíveis medidas de habilidades não cognitivas. Podemos verificar que a recuperação estimada não se mostra correlacionada com a nota em Ciências do PISA 2015 ou com a probabilidade de acertar a primeira questão da prova, mas fortemente correlacionada com a própria medida de decaimento. Assim, há evidências para que a recuperação também seja uma boa *proxy* para as habilidades socioemocionais ou outros fatores alternativos à habilidade cognitiva que influem no desempenho dos alunos durante a prova.

## 6.2 Recuperação e habilidades não cognitivas

Para esclarecer que tipo de habilidades a medida de recuperação entre os dois blocos pode capturar, replicamos as estimações principais apresentadas em [Hanushek e Woessmann \(2012\)](#) e [Balart et al. \(2018\)](#) de como as habilidades não cognitivas influenciam o crescimento de longo prazo dos países, a fim de promover um certo tipo de validação externa.

[Hanushek e Woessmann \(2012\)](#) desenvolvem uma nova métrica para a distribuição do desempenho educacional em países que pode descrever melhor a distribuição de habilidades cognitivas dentro dos países e ao longo tempo. Apresentando regressões de crescimento entre países com uma relação estreita entre desempenho educacional e crescimento do PIB. Os autores mostram que essa relação é estável através de análises de sensibilidade das especificações, períodos de tempo e amostras de países, utilizando diversas medidas de desempenho educacional, inclusive dados do PISA. Os autores propõem um construto (referenciado posteriormente como HW-Index) para capturar o desenvolvimento do capital humano nos países e mostram seu efeito sobre o crescimento de longo prazo dos mesmos. [Balart et al. \(2018\)](#) se valem da mesma metodologia para as regressões de longo prazo e incluem as estatísticas decompostas a partir da

**Tabela 6.** Regressões de medidas de habilidades cognitivas em habilidades socioemocionais.

	(1)	(2)	(3)
	Ciências PISA 2015	$P[Y_0 = 1]$	Decaimento
Decaimento	-709,3 *** (3,67)	-1,048 ** (2,89)	
Recuperação ( $\hat{\gamma}_1$ )	77,11 (0,41)	0,210 (0,60)	0,735 *** (8,60)
Constante	540,8 *** (40,82)	0,595 *** (23,89)	0,0380 *** (4,85)
N	55	55	55
R <sup>2</sup> Ajustado	0,318	0,187	0,575

Notas: Estatística *t* entre parênteses; \*  $p < 0,10$ ; \*\*  $p < 0,05$ ; \*\*\*  $p < 0,01$ .

Fonte: PISA 2015. Elaboração própria.

nota do PISA (desempenho inicial e decaimento) como representativos para o capital humano, sendo o decaimento uma medida adicional para as habilidades não cognitivas na formação de capital humano. Mostrando que o decaimento é uma medida tão boa quanto as medidas anteriores para as habilidades cognitivas e que o efeito estimado do decaimento no crescimento (habilidades não cognitivas) é aproximadamente igual ao efeito estimado do desempenho inicial (habilidades cognitivas). Os autores mostram também que o decaimento como medida de habilidades não cognitivas é constante ao longo do tempo e que as próprias habilidades não cognitivas também são constantes ao longo do tempo, utilizando os dados do PISA para os anos de 2003, 2006 e 2009.

A análise empírica do efeito das habilidades sobre crescimento econômico a ser replicada é uma regressão padrão de crescimento, incluindo a pontuação no PISA como uma *proxy* para o capital humano. Essa pontuação é decomposta em medidas que separam as habilidades cognitivas das não cognitivas: as pontuações do PISA no desempenho inicial ( $Z_{0p}$ ), expresso na probabilidade estimada de acerto da primeira questão, para cada país  $p$ , e o decaimento do desempenho durante o teste ( $D_p$ ). Incluiremos no modelo a recuperação estimada para cada país ( $\hat{\gamma}_{1p}$ ), como uma medida que pode ser complementar ou substituta do decaimento na identificação das habilidades não cognitivas.

Usamos os mesmos dados sobre o crescimento econômico (Hanushek & Woessmann, 2012), o mesmo crescimento de longo prazo (1960–2000), controles idênticos, mas utilizaremos o desempenho estimado em 2015 para a amostra de países presentes tanto em Balart et al. (2018) e nos dados do PISA 2015:

$$G_p = \delta_0 + \delta_1 Z_{0p} + \delta_2 D_p + \delta_3 \hat{\gamma}_{1p} + \delta_4 GDP1960_p + \sum_c \delta_c T_{cp} + \theta_p. \quad (4)$$

A equação (4) apresenta o modelo a ser estimado por MQO, apenas com o objetivo de validar a recuperação como uma medida de habilidades não cognitivas. Este modelo tem como variável dependente a taxa média de crescimento anual do PIB per capita de 1960–2000 ( $G_p$ ). Como variáveis explicativas incluímos também uma constante e o PIB per capita em 1960 ( $GDP1960_p$ ). Os controles sintetizados em  $\sum_c \delta_c T_{cp}$  são o grau abertura da economia, a proteção contra a expropriação, a fertilidade dos solos e uma *dummy* para a localização tropical descritos em Hanushek e Woessmann (2012). A Tabela 7 apresenta os resultados das regressões, incluindo somente o decaimento, a recuperação e incluindo ambas as medidas.

Os resultados das regressões de cada medida de habilidades não cognitivas, os modelos 1 e 3, que não incluem todos controles, mostram que a medida de recuperação pode cumprir o mesmo papel que a medida de decaimento em explicar as diferenças de crescimento de longo prazo entre os países da amostra. Além de o efeito da recuperação ser robusto à introdução de controles (modelos 2 e 4) e à estimação do modelo contando com as duas medidas (modelo 5). Deste modo, há evidências convincentes de que a recuperação de desempenho entre os dois blocos captura variações nas habilidades não cognitivas dos alunos.

Hanushek e Woessmann (2012) apresentam uma medida consistente a níveis nacionais que seja capaz de sintetizar o desempenho de alunos em vários períodos e avaliações, chamado de HW-Index em Balart et al. (2018). Combinando os ajustes nos níveis de pontuações e o ajuste nas variações, calculam pontuações padronizadas de

**Tabela 7.** Regressões de crescimento em medidas de habilidades cognitivas e habilidades socioemocionais.

	(1)	(2) <sup>e</sup>	(3)	(4) <sup>e</sup>	(5) <sup>e</sup>
Cognitivo ( $Z_0$ )	10,65 *** (4,76)	11,04 * (2,71)	11,61 *** (6,77)	10,35 ** (3,23)	13,78 ** (3,60)
Decaimento	-7,460 * (2,23)	-2,692 (0,76)			8,129 (1,55)
Recuperação ( $\hat{y}_1$ )			-8,088 ** (3,32)	-5,237 * (2,19)	-9,946 * (2,59)
N	36	34	36	34	34
R <sup>2</sup> Ajustado	0,772	0,802	0,804	0,829	0,838

Notas: Variável dependente: taxa média de crescimento anual do PIB per capita de 1960–2000; Regressões incluem uma constante e PIB per capita em 1960. <sup>e</sup>Incluindo controles de abertura da economia, proteção contra a expropriação, fertilidade e localização tropical. Estatística *t* entre parênteses; \*  $p < 0,10$ ; \*\*  $p < 0,05$ ; \*\*\*  $p < 0,01$ .

Fonte: PISA 2015. Elaboração própria.

escores dos testes cognitivos para todos os países em todas as avaliações disponíveis no período de 1964 a 2003, agregando neste índice os escores ao longo do tempo, para diferentes idades e matérias. Desta forma, o HW-Index é representação mais completa dos níveis de capital humano que podem ser obtidos a partir do desempenho individual de alunos em exames padronizados.

Um terceiro exercício proposto para a validação da recuperação como uma medida representativa para as habilidades não cognitivas é o estudo das correlações entre o HW-Index e as medidas de habilidades cognitivas (probabilidade de acerto da primeira questão) e não-cognitivas (decaimento e recuperação). Mais uma vez, sem a pretensão de estimar relações causais entre essas variáveis.

A Tabela 8 mostra as regressões do HW-Index nas duas possíveis medidas de habilidades não cognitivas e na medida de habilidades cognitivas. Podemos verificar que quando incluídos separadamente, tanto o decaimento quanto a recuperação (respectivamente, modelos 1 e 2) são significativamente correlacionados com o índice, assim como a probabilidade de acerto da primeira questão. Ao contrário do que acontece com o decaimento, a correlação da recuperação com índice é robusta à estimação utilizando ambas as medidas simultaneamente (modelo 3). Com isso, vemos que o nível do capital humano pode ser separado em duas dimensões: cognitiva e não cognitiva, como propõe Balart et al. (2018). Vemos ainda que podemos supor que a recuperação estimada seja uma *proxy* melhor para as habilidades não cognitivas (ou outra segunda dimensão do capital humano) do que o decaimento.

Apesar de o período coberto pelos dados de crescimento ser desconexo com o período de extração das medidas de capital humano, as afirmações apresentadas em Balart et al. (2018) e Hanushek e Woessmann (2012) indicam a validade de se utilizar este tipo de dados, no nosso caso para o PISA 2015. Os autores argumentam que as variáveis de capital humano e as estatísticas decompostas do PISA são, de certa forma, constantes ao longo do tempo e das edições da prova. Uma vez que o objetivo desta sessão é apenas validar a medida de recuperação e sua relação com o capital humano e

**Tabela 8.** Regressões do HW-Index nas medidas decompostas da nota do PISA 2015.

	(1)	(2)	(3)
	HW-Index 2012	HW-Index 2012	HW-Index 2012
Cognitivo ( $Z_0$ )	5,232 *** (5,99)	5,457 *** (9,18)	5,689 *** (7,37)
Decaimento	-3,009 *** (3,78)		1,241 (0,71)
Recuperação ( $\hat{\gamma}_1$ )		-3,458 *** (4,32)	-4,309 * (2,62)
Constante	2,324 *** (4,38)	2,191 *** (6,35)	2,015 *** (4,19)
N	36	36	36
R <sup>2</sup> Ajustado	0,835	0,877	0,876

Notas: Estatística *t* entre parênteses; \*  $p < 0,10$ ; \*\*  $p < 0,05$ ; \*\*\*  $p < 0,01$ .

Fonte: PISA 2015 e Hanushek e Woessmann (2012).

não inferir causalidade desta para com o crescimento de longo prazo, os coeficientes apresentados por estas estimativas devem ser interpretados de forma cautelosa.

## 7. Conclusão e desenvolvimentos futuros

Nesse trabalho, nós examinamos os motivos que levam o Brasil a ter baixo desempenho no exame do PISA 2015 em comparação com outros países, além de propor uma medida alternativa para a mensuração objetiva de habilidades não cognitivas, a partir de uma descontinuidade no desempenho ao longo do exame. Nós exploramos novas características da prova em 2015 que fazem os alunos responderem sequencialmente às questões para determinar quais aspectos desta avaliação são relacionados aos diferentes tipos de habilidades e quais deles influenciam o desempenho relativamente ruim do Brasil.

Nós utilizamos a variação exógena da posição de cada questão para identificar efeitos de habilidades socioemocionais a partir de uma medida de decaimento na probabilidade de acerto dos itens ao longo da prova. Também calculamos uma medida de habilidades cognitivas, a partir da probabilidade de acerto da primeira questão do exame. A principal hipótese por trás dessas medidas é de que no início da prova somente as habilidades cognitivas afetam a probabilidade de acertar uma questão, e ao longo da prova, em contraste, a probabilidade de acertos declina sob a influência de personalidade, como motivação e empenho.

Mostramos que existe uma descontinuidade entre os blocos de prova e que esta pode ser entendida como uma recuperação de desempenho. Discutimos que a interpretação desta recuperação como uma medida de habilidades socioemocionais pode ser superior ao uso do decaimento para representar esta dimensão. No entanto, o decaimento apresenta uma abordagem mais prática do problema: é possível construir um ordenamento onde não existe decaimento a partir da decomposição do rendimento da prova entre desempenho inicial e queda de desempenho. Sendo que o cenário

contrafactual sem a influência de habilidades socioemocionais seria claramente a representação apenas do desempenho inicial. Para a recuperação, é difícil construir um cenário contrafactual com uma abordagem tão clara.

Nossos resultados indicam que o Brasil possui probabilidade comparativamente baixa de acerto da primeira questão, um dos maiores declínios de desempenho ao longo da prova e a maior descontinuidade de desempenho entre os dois blocos de provas. O exercício contrafactual sugere que não somente o nível das habilidades cognitivas influencia a posição relativa do Brasil num ranking de desempenho entre os países. Mesmo que o ganho de posições na ausência de decaimento não seja muito grande, ainda é um dos maiores observados entre os países que fizeram as provas do tipo CBA.

Nossa análise dos determinantes do decaimento mostra que o mal desempenho do Brasil se deve especialmente ao fato de que grande parte dos respondentes não conseguem chegar ao fim da prova, o que pode estar relacionado à demora para entender o enunciado da questão e para desenvolver o raciocínio sobre a resposta. Dessa forma, o decaimento parece estar associado também às habilidades cognitivas.

Adicionalmente, o desempenho ruim do Brasil pode estar ligado à dificuldade de se fazer exames como o PISA. Enquanto o tempo médio de resposta das primeiras questões do bloco após o intervalo são consideravelmente menores do que aqueles das primeiras questões do primeiro bloco, a probabilidade de chegar ao final do segundo bloco é consideravelmente maior do que a probabilidade de chegar ao final do primeiro bloco. Dada a aleatorização da ordem das questões, aspectos relacionados à dificuldade de se responder às questões se mantêm constantes ao longo da prova, de modo que os alunos brasileiros aprendem a fazer a prova entre os blocos. O mesmo não ocorre de forma tão acentuada em países com as maiores notas. As taxas de acerto dos alunos brasileiros que chegam ao final do segundo bloco são em geral maiores do que as daqueles que chegam ao final do primeiro bloco, o que sugere que o tempo é gasto de forma mais eficiente no segundo bloco. Dessa forma, é possível que o desempenho brasileiro aumentasse, caso os alunos gastassem menos tempo nas questões iniciais em que têm mais dificuldade, e utilizassem esse tempo para alcançar as questões finais da prova. Esse resultado é elevado especialmente para alunos da rede pública.

Assim como em [Akyol et al. \(2018\)](#), numa análise contrafactual de “equilíbrio geral”, onde todos os alunos não seriam prejudicados pelo decaimento, ou pela baixa motivação como os autores propõe, o Brasil não ganha tantas posições quanto se esperaria num ranking relativo ao desempenho dos países, apesar deste ganho não ser desprezível. Nossos resultados corroboram a conclusão de que isto se deve ao conhecimento dos alunos “não sérios” já ser muito baixo, estendendo a explicação para o fato de que poucos alunos brasileiros realizam o exame utilizando a melhor estratégia possível, em termos de tempo de prova e aproveitamento, para responderem a todas as questões do exame.

O PISA vem, ao longo das suas últimas edições, ampliando a coleta de dados socioeconômicos e socioemocionais auxiliares ao exame, com o uso destes dados –por exemplo a riqueza familiar ou atitude e motivação em relação ao esforço – seria possível aprofundar a análise das diferenças de decaimento e recuperação, validando as medidas objetivas aqui propostas a partir de outros construtos socioemocionais subjetivos reportados no exame, além de testar a hipótese levantada em [Marchioni](#)

(2017) de que as diferenças de decaimento entre escolas públicas e privadas seriam advindas da diferença de renda entre alunos e tipo de escola seria apenas uma *proxy* para esta condição.

Desenvolvimentos futuros sobre o tema devem, portanto, considerar de que forma o decaimento médio do desempenho dos alunos brasileiros pode ser mediado pela porcentagem de alunos que conseguem atingir as últimas questões da prova e em que extensão esta porcentagem ou o tempo médio de resolução de questões podem contribuir para uma medida de habilidades não cognitivas mais acurada entre os estados e, também, a extensão da aplicabilidade da recuperação entre os blocos como medida de habilidades socioemocionais, algumas destas questões já são adereçadas em Akyol et al. (2018).

As formulações apresentadas e também referenciadas neste trabalho (Borghans & Schils, 2012; Marchioni, 2017; Balart et al., 2018) estão limitadas, em termos práticos, à estimação de uma tendência linear da probabilidade de acerto entre a última e a primeira questão, não capturando bem a variação desta probabilidade em torno do intervalo e dificultando a introdução do tempo como variável explicativa. O intervalo parece intermediar alguma relação entre o tempo médio e a probabilidade de acerto das questões, como é demonstrado a partir da medida de recuperação. Com a estimação desta descontinuidade, podem ser propostos modelos que sejam mais adequadas à introdução do tempo como uma variável explicativa do desempenho.

Ao menos para o caso brasileiro, podemos notar que os alunos têm desempenho mais eficiente em termos de tempo por questão no segundo bloco de provas<sup>7</sup> e o decaimento observado é menor em relação ao decaimento do primeiro bloco. Esse resultado sugere que, ao longo do exame, o desempenho dos alunos brasileiros não é exclusivamente prejudicado pelo cansaço ou fadiga, mas que há algum aprendizado quanto à natureza da prova e um maior entendimento de como responder (ou não responder) às questões para ser capaz de chegar ao final do exame.

## Referências bibliográficas

- Akyol, Ş. P., Krishna, K., & Wang, J. (2018, August). *Taking PISA seriously: How accurate are low stakes exams?* (Working Paper N° 24930). Cambridge, MA: National Bureau of Economic Research (NBER). <http://dx.doi.org/10.3386/w24930>
- Albano, A. D. (2013). Multilevel modeling of item position effects. *Journal of Educational Measurement*, 50(4), 408–426. <http://dx.doi.org/10.1111/jedm.12026>
- Almund, M., Duckworth, A. L., Heckman, J., & Kautz, T. (2011). Personality psychology and economics. In E. Hanushek, S. Machin, & L. Woessmann (Orgs.), *Handbook of the economics of education* (Vol. 4, pp. 1–181). Amsterdam: Elsevier.
- Balart, P., Oosterveen, M., & Webbink, D. (2018). Test scores, noncognitive skills and economic growth. *Economics of Education Review*, 63, 134–153. <http://dx.doi.org/10.1016/j.econedurev.2017.12.004>
- Borghans, L., Duckworth, A. L., Heckman, J. J., & Ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of Human Resources*, 43(4), 972–1059. <http://dx.doi.org/10.3368/jhr.43.4.972>

---

<sup>7</sup>As tabelas 5 e 6, demonstram que em todos os casos observados a [Prob. > 95 %] é sempre maior do que [Prob. > 45 %].

- Borghans, L., & Schils, T. (2012). *The leaning tower of Pisa: Decomposing achievement test scores into cognitive and noncognitive components*. <http://www.sole-jole.org/13260.pdf>
- Davis, J., & Ferdous, A. (2005, 12 de abril). Using item difficulty and item position to measure test fatigue. In M. Cochran-Smith & A.M. Villegas (Chair), *American Educational Research Association 2005 Annual Meeting*, Montreal, Québec. [https://convention2.allacademic.com/one/aera/aera05/index.php?click\\_key=1&cmd=Multi+Search+Search+Load+Publication&publication\\_id=9091&PHPSESSID=116qpqjsj53raqj7cpujk2rsul](https://convention2.allacademic.com/one/aera/aera05/index.php?click_key=1&cmd=Multi+Search+Search+Load+Publication&publication_id=9091&PHPSESSID=116qpqjsj53raqj7cpujk2rsul)
- Duckworth, A. L., Quinn, P. D., Lynam, D., Loeber, R., Stouthamer-Loeber, M., Moffitt, T. E., & Caspi, A. (2009). *What intelligence tests test: Individual differences in test motivation and IQ*. ResearchGate. [https://www.researchgate.net/publication/237626525\\_What\\_Intelligence\\_Tests\\_Test\\_Individual\\_Differences\\_in\\_Test\\_Motivation\\_and\\_IQ](https://www.researchgate.net/publication/237626525_What_Intelligence_Tests_Test_Individual_Differences_in_Test_Motivation_and_IQ)
- Eklöf, H. (2007). Test-taking motivation and mathematics performance in TIMSS 2003. *International Journal of Testing*, 7(3), 311–326. <http://dx.doi.org/10.1080/15305050701438074>
- Hanushek, E. A., & Kimko, D. D. (2000). Schooling, labor-force quality, and the growth of nations. *The American Economic Review*, 90(5), 1184–1208. <http://dx.doi.org/10.1257/aer.90.5.1184>
- Hanushek, E. A., & Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth*, 17(4), 267–321. <http://dx.doi.org/10.1007/s10887-012-9081-x>
- Heckman, J. J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *The American Economic Review*, 103(6), 2052–2086. <http://dx.doi.org/10.1257/aer.103.6.2052>
- Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3), 411–482. <http://dx.doi.org/10.1086/504455>
- Marchioni, C. G. (2017). *Habilidades no cognitivas en América Latina: Una medición desde pruebas estandarizadas* (Tesis de maestria, Universidad Nacional de La Plata, La Plata, Argentina). <http://dx.doi.org/0.35537/10915/70188>
- OCDE. (s. d.). *PISA 2015 database*. Organização para a Cooperação e Desenvolvimento Econômico: Programme for International Students Assessment. <https://www.oecd.org/pisa/data/2015database/>
- OCDE. (2017). *Pisa 2015: Technical report*. OCDE. Acessado em abril de 2017: <http://www.oecd.org/pisa/data/2015-technical-report/>
- Segal, C. (2012). Working when no one is watching: Motivation, test scores, and economic success. *Management Science*, 58(8), 1438–1457. <http://dx.doi.org/10.1287/mnsc.1110.1509>

## Apêndice.

Tabela A-1. Relação entre Pontuação, participação e Tempo Médio ao longo do PISA 2015.

País	Pontuação						Não chega na Questão						Tempo Médio (Min.)					
	(Percentual de Prova)						(Percentual de Prova)						(Percentual de Prova)					
	5	25	50	55	75	95	5	25	50	55	75	95	5	25	50	55	75	95
Korea	0,62	0,55	0,49	0,63	0,50	0,52	0,00	0,03	0,04	0,00	0,00	0,00	1,03	1,05	1,08	0,91	0,97	0,86
Chinese Taipei	0,60	0,60	0,54	0,63	0,55	0,54	0,00	0,03	0,05	0,00	0,02	0,00	1,14	1,45	1,32	1,06	1,35	0,97
Netherlands	0,59	0,56	0,50	0,58	0,53	0,50	0,00	0,00	0,03	0,00	0,00	0,00	1,25	1,22	1,10	1,03	1,02	0,94
Slovenia	0,57	0,49	0,43	0,56	0,46	0,48	0,00	0,01	0,06	0,00	0,00	0,01	1,28	1,27	1,23	1,02	1,08	0,95
Finland	0,61	0,60	0,49	0,63	0,54	0,58	0,00	0,01	0,08	0,00	0,00	0,02	1,28	1,35	1,31	1,14	1,19	1,07
United Kingdom	0,53	0,53	0,42	0,53	0,47	0,48	0,00	0,02	0,06	0,00	0,01	0,01	1,31	1,36	1,22	1,13	1,14	0,96
Croatia	0,50	0,50	0,37	0,55	0,44	0,39	0,00	0,02	0,12	0,00	0,00	0,02	1,31	1,37	1,34	1,15	1,20	1,10
Ireland	0,56	0,54	0,42	0,57	0,49	0,52	0,00	0,01	0,08	0,00	0,00	0,01	1,37	1,41	1,35	1,28	1,25	1,10
Austria	0,55	0,50	0,47	0,56	0,44	0,49	0,00	0,00	0,07	0,00	0,00	0,02	1,39	1,30	1,27	1,19	1,11	1,06
Canada	0,57	0,54	0,44	0,56	0,49	0,50	0,00	0,01	0,08	0,00	0,00	0,02	1,40	1,44	1,22	1,14	1,18	1,03
Hong Kong	0,61	0,64	0,52	0,65	0,57	0,58	0,00	0,04	0,08	0,00	0,02	0,01	1,40	1,70	1,37	1,35	1,50	0,94
Iceland	0,51	0,46	0,41	0,52	0,42	0,37	0,00	0,00	0,10	0,02	0,00	0,05	1,41	1,39	1,21	1,27	1,15	1,00
Poland	0,54	0,53	0,39	0,56	0,50	0,49	0,00	0,02	0,10	0,00	0,01	0,02	1,41	1,43	1,35	1,24	1,26	1,02
Singapore	0,62	0,58	0,50	0,63	0,54	0,60	0,00	0,02	0,08	0,00	0,00	0,01	1,42	1,54	1,35	1,16	1,31	1,04
Macao	0,63	0,54	0,51	0,60	0,57	0,53	0,00	0,00	0,11	0,01	0,00	0,05	1,44	1,60	1,41	1,42	1,46	1,13
Germany	0,60	0,55	0,42	0,63	0,51	0,52	0,00	0,04	0,14	0,00	0,02	0,03	1,47	1,48	1,34	1,23	1,31	1,11
Japan	0,62	0,55	0,44	0,64	0,52	0,56	0,00	0,03	0,14	0,00	0,01	0,03	1,47	1,57	1,55	1,33	1,43	1,06
Estonia	0,61	0,61	0,52	0,62	0,55	0,54	0,00	0,01	0,07	0,01	0,01	0,02	1,47	1,45	1,36	1,26	1,27	1,14
Lithuania	0,52	0,48	0,33	0,52	0,42	0,38	0,00	0,03	0,08	0,00	0,01	0,01	1,48	1,41	1,25	1,18	1,14	1,02
Belgium	0,58	0,55	0,39	0,59	0,49	0,46	0,00	0,03	0,16	0,00	0,02	0,07	1,48	1,51	1,36	1,22	1,33	1,11
Norway	0,58	0,54	0,38	0,58	0,46	0,43	0,00	0,03	0,11	0,00	0,02	0,04	1,48	1,46	1,34	1,23	1,27	1,08
Switzerland	0,58	0,53	0,39	0,57	0,47	0,47	0,00	0,02	0,14	0,00	0,01	0,03	1,49	1,42	1,24	1,19	1,17	0,98
Denmark	0,56	0,51	0,36	0,56	0,45	0,45	0,00	0,03	0,11	0,00	0,02	0,02	1,50	1,49	1,15	1,18	1,27	1,07
Hungary	0,51	0,51	0,37	0,54	0,43	0,43	0,00	0,04	0,14	0,00	0,01	0,04	1,50	1,41	1,35	1,20	1,31	1,02
Turkey	0,43	0,37	0,27	0,45	0,32	0,27	0,00	0,01	0,04	0,00	0,01	0,00	1,50	1,53	1,21	1,25	1,25	1,01
Czech Republic	0,58	0,51	0,42	0,56	0,47	0,50	0,00	0,01	0,10	0,00	0,00	0,03	1,51	1,42	1,50	1,23	1,21	1,10
Australia	0,54	0,51	0,37	0,52	0,44	0,42	0,00	0,04	0,14	0,01	0,02	0,06	1,51	1,55	1,23	1,25	1,32	1,02
Latvia	0,52	0,54	0,37	0,57	0,46	0,46	0,00	0,04	0,09	0,00	0,03	0,02	1,52	1,50	1,43	1,28	1,39	1,13
United States	0,55	0,52	0,36	0,55	0,46	0,46	0,00	0,03	0,16	0,00	0,02	0,04	1,53	1,60	1,28	1,20	1,35	1,03
B-S-J-G (China)	0,61	0,61	0,47	0,63	0,56	0,47	0,00	0,04	0,08	0,00	0,02	0,01	1,54	1,56	1,46	1,29	1,42	0,98
Spain	0,56	0,54	0,41	0,56	0,47	0,44	0,00	0,01	0,14	0,00	0,00	0,04	1,56	1,51	1,35	1,31	1,29	1,14
France	0,58	0,51	0,45	0,54	0,46	0,45	0,00	0,00	0,09	0,01	0,01	0,04	1,57	1,47	1,37	1,45	1,25	1,05
Luxembourg	0,54	0,48	0,33	0,53	0,43	0,42	0,00	0,02	0,21	0,00	0,01	0,04	1,58	1,54	1,35	1,25	1,25	1,06
Qatar	0,36	0,32	0,21	0,34	0,26	0,27	0,07	0,13	0,09	0,08	0,09	0,07	1,58	1,37	0,89	1,18	1,04	0,77
New Zealand	0,57	0,53	0,35	0,60	0,48	0,44	0,00	0,04	0,18	0,00	0,02	0,08	1,61	1,63	1,21	1,30	1,32	1,04
Israel	0,53	0,48	0,30	0,51	0,41	0,35	0,03	0,05	0,19	0,03	0,04	0,05	1,63	1,46	1,12	1,25	1,17	1,00
Sweden	0,56	0,52	0,31	0,57	0,44	0,42	0,00	0,04	0,21	0,00	0,02	0,07	1,64	1,56	1,24	1,30	1,33	1,04
Chile	0,53	0,46	0,28	0,50	0,38	0,37	0,00	0,02	0,18	0,00	0,00	0,06	1,67	1,58	1,26	1,38	1,39	1,11
Portugal	0,56	0,51	0,29	0,54	0,42	0,38	0,00	0,04	0,26	0,00	0,01	0,09	1,68	1,63	1,36	1,32	1,35	1,09
Italy	0,58	0,53	0,35	0,57	0,45	0,45	0,00	0,02	0,11	0,00	0,01	0,03	1,68	1,63	1,43	1,33	1,37	1,09
Slovak Republic	0,51	0,46	0,37	0,48	0,39	0,35	0,00	0,01	0,10	0,00	0,01	0,07	1,69	1,45	1,30	1,30	1,21	1,01
United Arab Emirates	0,45	0,42	0,29	0,46	0,32	0,35	0,00	0,04	0,06	0,01	0,04	0,03	1,70	1,58	1,31	1,31	1,40	0,98
Greece	0,52	0,47	0,36	0,50	0,41	0,38	0,00	0,00	0,07	0,00	0,00	0,02	1,73	1,61	1,34	1,37	1,37	1,06
Montenegro	0,41	0,32	0,24	0,41	0,27	0,24	0,00	0,02	0,14	0,01	0,01	0,07	1,73	1,36	1,24	1,27	1,16	0,93
Costa Rica	0,42	0,37	0,19	0,42	0,29	0,28	0,00	0,06	0,29	0,01	0,03	0,11	1,85	1,70	1,27	1,45	1,47	1,08
Bulgaria	0,48	0,42	0,31	0,45	0,35	0,32	0,00	0,01	0,16	0,00	0,00	0,05	1,90	1,49	1,26	1,33	1,31	1,01
Russian Federation	0,54	0,48	0,37	0,55	0,43	0,38	0,00	0,02	0,18	0,01	0,02	0,12	1,98	1,70	1,47	1,56	1,43	1,16
Uruguay	0,46	0,40	0,20	0,46	0,30	0,28	0,00	0,06	0,26	0,00	0,03	0,09	1,99	1,64	1,15	1,48	1,43	1,01
Thailand	0,42	0,40	0,29	0,41	0,32	0,28	0,00	0,01	0,18	0,01	0,01	0,11	2,00	1,71	1,59	1,59	1,48	1,14
Colombia	0,45	0,38	0,18	0,47	0,32	0,29	0,00	0,05	0,20	0,00	0,03	0,08	2,03	1,81	1,16	1,57	1,63	1,11
Mexico	0,42	0,33	0,20	0,40	0,29	0,28	0,00	0,05	0,33	0,00	0,02	0,14	2,10	1,75	1,49	1,58	1,50	1,10
Dominican Republic	0,32	0,22	0,08	0,34	0,17	0,20	0,00	0,04	0,08	0,00	0,03	0,02	2,12	1,62	0,61	1,64	1,32	0,82
Brazil	0,38	0,30	0,09	0,39	0,25	0,15	0,00	0,17	0,62	0,01	0,09	0,40	2,42	1,85	1,24	1,73	1,59	1,00
Peru	0,36	0,28	0,20	0,34	0,24	0,20	0,00	0,08	0,40	0,01	0,02	0,22	2,58	1,80	1,85	2,06	1,65	1,25
Tunisia	0,35	0,27	0,11	0,33	0,20	0,16	0,00	0,08	0,22	0,07	0,12	0,20	2,78	1,82	1,19	1,79	1,54	0,91

Fonte: PISA 2015. Elaboração própria.