

Por que o Brasil vai Mal no PISA? Uma Análise dos Determinantes do Desempenho no Exame

Alex Hayato Sasaki, Giovanni Di Pietra,
Naercio Menezes Filho, Bruno Komatsu

Por que o Brasil vai Mal no PISA?

Uma Análise dos Determinantes do Desempenho no Exame

Alex Hayato Sasaki

Giovanni Avila Cardoso Di Pietra

Naercio Aquino Menezes Filho

Bruno Kawaoka Komatsu

Alex Hayato Sasaki
Universidade de São Paulo
Faculdade de Economia, Administração e Contabilidade
Avenida Professor Luciano Gualberto, 908
05508-010 - São Paulo/SP -Brasil
asasaki@usp.br

Giovanni Avila Cardoso Di Pietra
Insper Instituto de Ensino e Pesquisa
Centro de Políticas Públicas (CPP)
Rua Quatá, nº300
04546-042 - São Paulo, SP - Brasil
giovanniacp@insper.edu.br

Naercio A. Menezes Filho
Insper Instituto de Ensino e Pesquisa
Centro de Políticas Públicas (CPP)
Rua Quatá, nº300
04546-042 - São Paulo, SP - Brasil
naercioamf@insper.edu.br

Bruno Kawaoka Komatsu
Insper Instituto de Ensino e Pesquisa
Centro de Políticas Públicas (CPP)
Rua Quatá, nº300
04546-042 - São Paulo, SP - Brasil
brunokk@insper.edu.br

Copyright Insper. Todos os direitos reservados.

É proibida a reprodução parcial ou integral do conteúdo deste documento por qualquer meio de distribuição, digital ou impresso, sem a expressa autorização do Insper ou de seu autor.

A reprodução para fins didáticos é permitida observando-se

a citação completa do documento.

Por que o Brasil vai Mal no PISA?

Uma Análise dos Determinantes do Desempenho no Exame

Alex Hayato Sasaki

Giovanni Avila Cardoso Di Pietra

Naercio Aquino Menezes Filho

Bruno Kawaoka Komatsu

Centro de Políticas Públicas do Insper e USP

Resumo

Este artigo procura explicar o desempenho comparativamente baixo do Brasil no PISA de 2015. Em primeiro lugar, nós mostramos que os alunos brasileiros têm uma forte queda de desempenho ao longo da prova, assim como outros países latino-americanos. Além disso, nossos resultados mostram que o decaimento dos alunos brasileiros se deve principalmente à falta de conhecimento e habilidades para responder a prova, uma vez que eles gastam muito tempo nas primeiras questões e não alcançam as últimas. Na segunda parte da prova, depois de um breve intervalo, o desempenho melhora um pouco, o que indica que parte do mal desempenho do Brasil parece ser também devido à falta de habilidades para fazer provas como a do PISA. Assim, a posição ruim do Brasil é explicada essencialmente pela dificuldade em resolver as questões iniciais.

Palavras-chave: Habilidades socioemocionais, habilidades cognitivas, PISA, Desigualdade

Abstract

This article seeks to explain Brazil's comparatively low performance in the 2015 PISA. First, we show that Brazilian students have a strong performance decline throughout the exam, as well as other Latin American countries. In addition, our results show that the decline of Brazilian students is mainly due to the lack of knowledge and skills to do the exam, since they spend a lot of time on the first questions and do not reach the last ones. In the second part of the exam, after a brief break, the performance improves a little, which indicates that part of the poor performance of Brazil also seems to be due to the lack of skills to do exams such as PISA. Thus, Brazil's bad position is explained mainly by the difficulty in solving the initial questions.

Keywords: Socioemotional skills, Cognitive skills, PISA, Inequality

Área ANPEC: Área 8 – Microeconomia, Métodos Quantitativos e Finanças

JEL: O15, J24

I. 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 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, 2016).

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 e Kimko, 2000; Hanushek e Woessman, 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 daquelas avaliações padronizadas (Duckworth et al., 2009; Eklöf, 2007; Segal, 2012; Borghans et al., 2008; Almlund et al., 2011) e, de forma mais geral, para o desempenho econômico e outros *outcomes* individuais (Heckman, Stixrud e Urzua, 2006; Heckman, Pinto e Savelyev, 2013; Balart et al., 2018).

Nesse artigo, nós exploramos inovações na a organização da prova do PISA de 2015 para determinar quais aspectos daquela 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 forma que os alunos obrigatoriamente responderam à questões de forma sequencial, de forma que, uma vez finalizada a questão, eles não poderiam retornar a ela. Dessa forma, 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 Borghans e Schils (2012), Balart et al. (2018) e Marchioni (2017), nós exploramos uma variação exógena da posição de cada questão para identificar os efeitos de habilidades socioemocionais que se refletem no decaimento de rendimento que ocorre ao longo da prova. Dada a alocação aleatória da posição de cada questão, as diferenças na queda do desempenho ao longo de uma prova refletem habilidades socioemocionais, tais como persistência, concentração e determinação na população dos países. Medidas como essa, baseadas em resultados objetivos de uma avaliação, superam o caráter subjetivo de indicadores construídos a partir de informações auto relatadas, que limita a comparabilidade entre indivíduos e regiões. Nós também utilizamos uma medida de habilidades cognitivas, a partir da probabilidade de acerto da primeira questão do exame.

Nossos resultados mostram que, em comparação com alguns dos países que possuem as maiores médias do PISA, o Brasil possui baixa probabilidade de acerto da primeira questão, o que indica habilidades cognitivas também menores. Além disso, a probabilidade de responder corretamente às questões vai se reduzindo ao longo das duas fases da prova. Entre essas duas fases, há um intervalo, em que o Brasil também demonstra uma recuperação acentuada.

Quais são as implicações desse padrão para a interpretação dos indicadores de habilidades socioemocionais? Para responder essa questão procuramos investigar mais 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. Os resultados sugerem que o mal desempenho do Brasil se deve especialmente 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. O mal desempenho brasileiro seria mais relacionado ao conhecimento e aos níveis de habilidades cognitivas e, nesse sentido, a interpretação do decaimento como medida puramente de habilidades socioemocionais não parece muito clara.

Nosso segundo resultado 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 provas como a 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.

Além dessa introdução, a segunda seção traz uma revisão da bibliografia relevante para o desenvolvimento deste trabalho na segunda seção. A terceira seção descreve a estrutura do PISA 2015 e os dados que utilizaremos e a quarta seção descreve a metodologia para a estimação dos indicadores de habilidades socioemocionais usando os dados do PISA 2015. A quinta seção apresenta e discute os resultados a partir de tabelas e gráficos. A sexta seção apresenta uma discussão sobre os fatores associados ao decaimento e a particularidade do caso brasileiro, concluindo na sétima seção e apresentando 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 e Kimko, 2000; Hanushek e Woessman, 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 (Almlund et al., 2011), apesar de alguns estudos não encontrarem uma relação clara entre estas duas variáveis (Eklöf, 2007).

Em segundo lugar, nosso artigo se relaciona com a literatura psicométrica que foca em características dos exames, como o posicionamento dos itens, e no contexto do exame. 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 todas 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) com base no PISA de 2012, com o objetivo de avaliar as habilidades cognitivas e não-cognitivas dos países latino-americanos em comparação com os outros países que participam do exame, assim como 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 públicas 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. Dados

Utilizaremos os microdados do PISA 2015. O PISA é uma avaliação educacional realizada em nível internacional a cada três anos a 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 ou demais 35, países parceiros. No Brasil, o exame abrangeu 23.141 estudantes de 841 escolas.

O exame consiste em responder questões de ciências, literatura e matemática, além da resolução colaborativa de problemas, durante duas horas. Ele é estruturado de forma que há uma variedade de conjuntos de questões (chamados de *clusters*) de cada um dos assuntos. Há diversos formulários de prova e cada um é composto por quatro desses *clusters*, e a posição de cada *cluster* dentro da prova é fixada de acordo com o código do formulário e o assunto a que diz respeito. 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 (Codebook 2015) das questões apresenta o ordenamento das questões dentro de cada um dos *clusters*. O formulário de prova que um aluno recebe é aleatório.¹ 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 nível 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.² Após os dois blocos de provas e um outro intervalo maior, os alunos também deveriam responder ao final do exame a questões referentes aos seus dados socioeconômicos, familiares e 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 socioeconômicas individuais. Para esses testes, regredimos as *dummies* de cada

¹ A Tabela 3.1 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.

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

característica nas *dummies* de número do formulário, no número aleatório de alocação dos exames de ciência, e nas 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 de que os coeficientes são conjuntamente iguais a zero, ou seja, não há uma associação clara entre os números de formulário, o número de alocação dos exames de ciências e as características individuais. Dessa forma, temos maior confiança de que podemos considerar a alocação de cada questão como aleatória para cada estudante.

Tabela 1 - Teste de Aleatorização

| Variável Dependente | (1) Gênero | (2) Série | (3) Mês de Nasc. | (4) Possui. Escrivainha | (5) Quarto indiv. | (6) Lugar de estudos em casa | (7) Comp. para estudar | (8) Maior escolar. – Pai | (9) Maior escolar. – Mãe | (10) Qtd. de livros em casa |
|---------------------|---------------|--------------|---------------------|----------------------------|----------------------|---------------------------------|---------------------------|-----------------------------|-----------------------------|--------------------------------|
| F-Stat* | 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 | 22745 | 22745 | 22745 | 20638 | 19859 | 20195 | 20290 | 19882 | 20631 | 20875 |
| R-sq. | 0.017 | 0.018 | 0.015 | 0.020 | 0.021 | 0.020 | 0.017 | 0.018 | 0.021 | 0.018 |
| Obs. | 23141 | 23141 | 23141 | 21034 | 20255 | 20591 | 20686 | 20278 | 21027 | 21271 |

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. * 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. ** DF = Graus de Liberdade, tal que $F\text{-Stat} \sim F(\text{DF1}, \text{DF2})$.

4. Metodologia para a Análise de Decaimento do Desempenho dos Alunos

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.

Seguindo Borghans e Schils (2012), utilizamos a posição relativa de cada questão no exame a fim de identificar o efeito das habilidades socioemocionais ou não cognitivas.

Para isso, utilizamos a 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 a questão. Ao longo da prova, no entanto, traços da personalidade como atenção, motivação ou persistência passam a ser 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 Marchioni (2017) e em Borghans e Schils (2012). Estimamos o modelo:

$$P[Y_{ij} = 1] = \beta_0 + \beta_1 Q_{ij} + \sum_{j=2}^J \mu_j + \varepsilon_{ij} \quad (1)$$

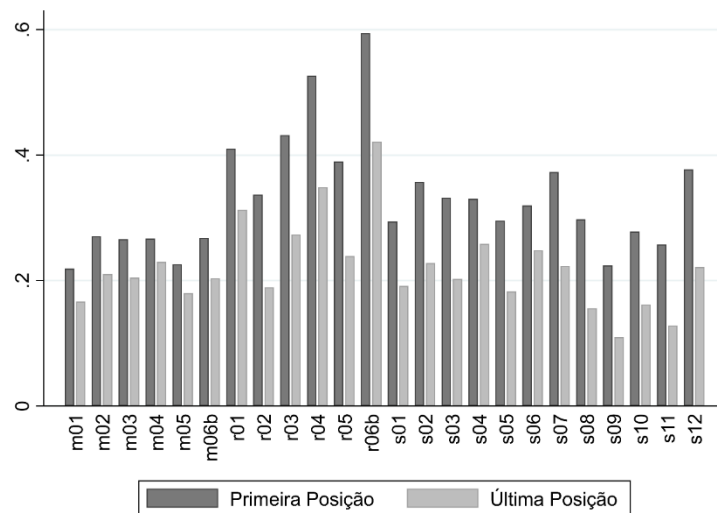
Onde Q_{ij} é a posição normalizada entre 0 e 1 da questão, sendo que para a primeira questão o valor atribuído a Q_{ij} é 0, e para última questão é igual a 1. O termo μ_j é o efeito fixo de cada uma das questões, sua utilização se dá pela intenção de se controlar pela dificuldade específica de cada uma das questões. Com a especificação linear, a interpretação é direta: o coeficiente β_1 estimado no modelo de regressão linear é o decaimento de desempenho.

5. Resultados

Em primeiro lugar, verificamos que há decaimento ao longo da prova. Como podemos ver na Figura 1, esse decaimento ocorre ao longo da prova, mesmo se comparamos o *cluster* de questões, entre a primeira e a última posições na prova. A Figura 2 mostra que mesmo se dividirmos as provas por disciplina, o padrão 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 ao fato de que, foi imposta aos participantes do exame uma

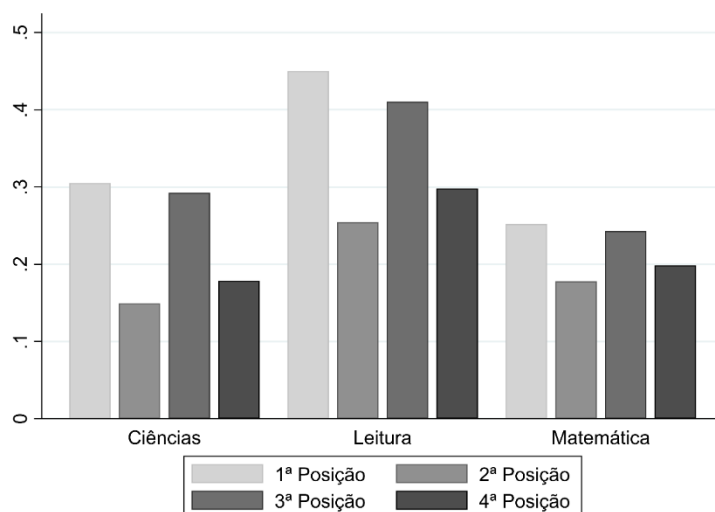
pequena pausa de no máximo 5 minutos antes de começar o bloco de testes dos *clusters* três e quatro, e o tempo de resolução para cada grupo de dois *clusters* foi limitado a apenas uma hora. Uma particularidade do exame em 2015 é de que os participantes que realizaram os testes em computadores (provas 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 que de fato há influência das habilidades socioemocionais no decorrer da prova.

Figura 1 – Percentual de Acertos Dentro dos Clusters, por Posição na Prova



Fonte: PISA 2015. Elaboração própria. Percentual de acertos utilizando a amostra de todos os países.

Figura 2 – Percentual Médio de Acertos dos *Clusters* de cada Disciplina, por Posição na Prova



Fonte: PISA 2015. Elaboração própria. Percentual de acertos utilizando a amostra de todos os países.

5.1. Decaimento

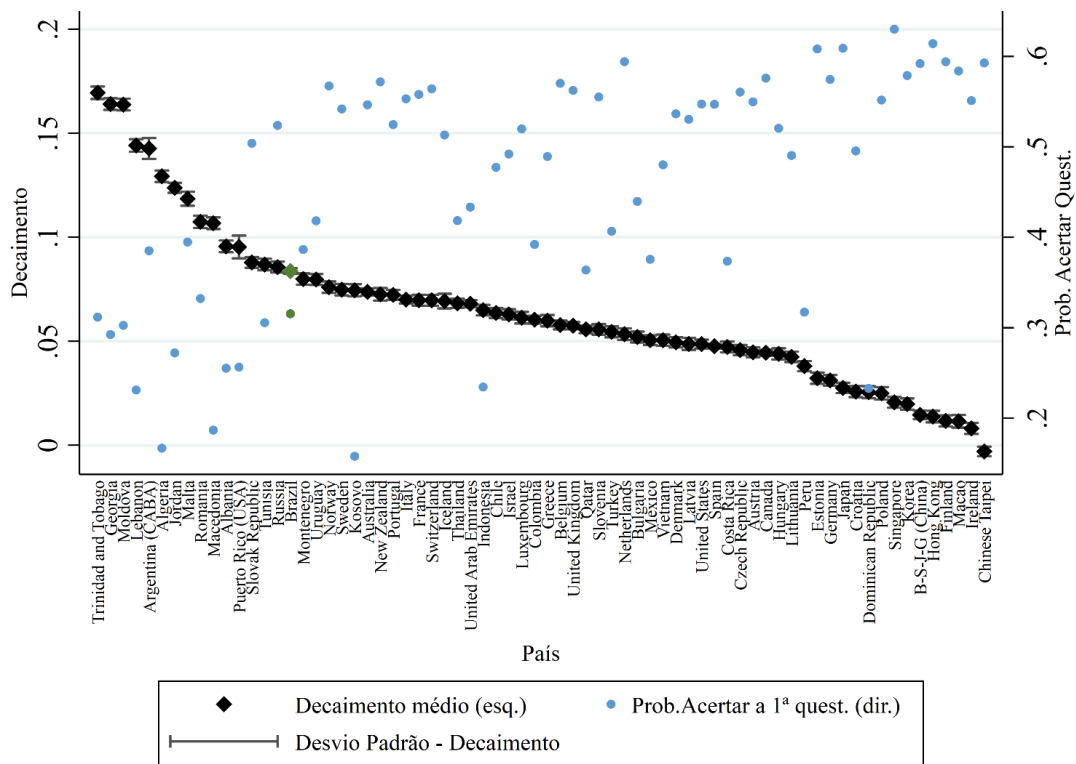
A Figura 3 mostra que a queda de desempenho estimada ao longo de uma prova extensa como o PISA difere significativamente entre países, assim como difere a probabilidade média dos alunos de cada país acertarem a primeira questão de uma mesma prova. O Brasil aparece com um dos menores níveis de probabilidade de acertar a primeira questão, o que indica baixo nível de habilidades cognitivas (posição 57 de 70)³. O nível de decaimento brasileiro é também um dos maiores (16º maior), o que representa, na interpretação de Borghans e Schils (2012) 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. Somente Taiwan (*Chinese Taipei*) possui um decaimento negativo, o que representa de fato uma melhora de desempenho ao longo da prova, porém muito próximo de zero.

Ilustramos o decaimento dos alunos brasileiro ao longo da prova, em comparação com países com bom desempenho na Figura 4, 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

³ Consideramos 70 regiões e países, uma vez que unificamos os territórios da Espanha e Estados Unidos, mantendo o território de Porto Rico como um país separado.

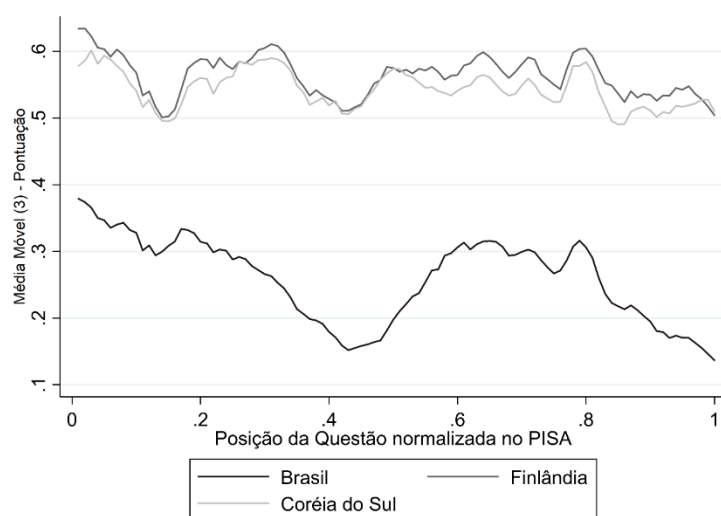
questão é igual a 1, a queda na taxa de acertos é muito mais acentuada para o Brasil do que para Finlândia e Coréia do Sul. O desempenho ao longo da prova destes últimos parece ser mais constante e menos afetado por fatores que não podem ser atribuídos ao “conhecimento” dos alunos, como o tempo total de prova ou o cansaço e a distração decorrente de uma prova mais longa.

Figura 3 – Queda de desempenho estimada ao longo da prova dos países selecionados



Fonte: PISA 2015. Elaboração própria. Decaimento estimado conforme a equação 1 e probabilidade de acertar a primeira questão, para cada uma das 70 regiões. Destacamos o Brasil em verde.

Figura 4 – Queda de desempenho observada ao longo da prova para o Brasil, Finlândia e Coréia do Sul



Fonte: PISA 2015. Elaboração própria. Médias móveis de 3 termos defasados da probabilidade de acertos 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.

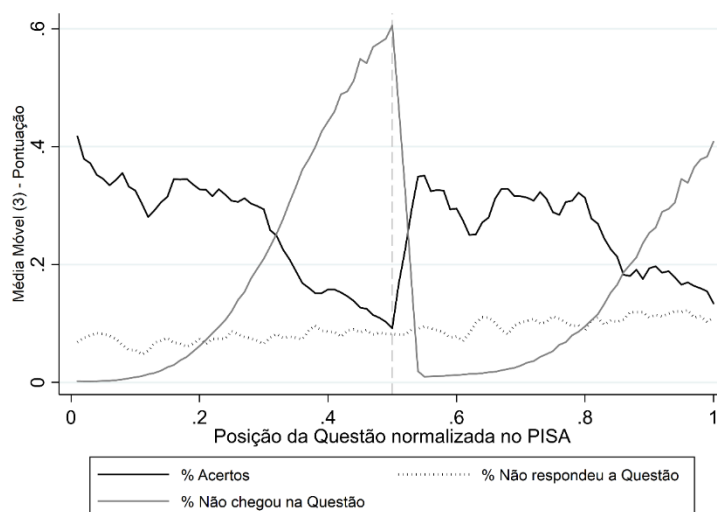
5.2. Determinantes do Decaimento

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

A Figura 5 mostra a proporção de acertos, a proporção de alunos que não responderam à questão e a proporção daqueles que não chegaram ao item, 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*. A taxa de não resposta é relativamente constante ao longo da prova. 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.⁴

Figura 5 – Queda de desempenho observada e porcentagem de alunos brasileiros que não respondeu às questões no PISA 2015



Fonte: PISA 2015. Elaboração própria. 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.

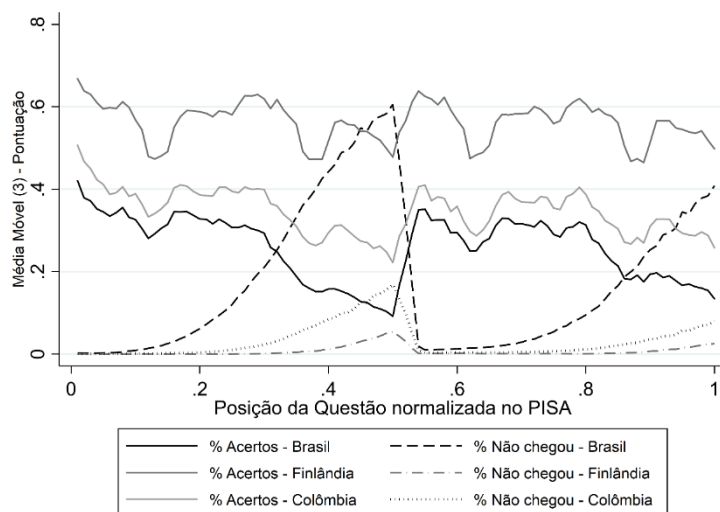
Em segundo lugar, a Figura 5 também deixa clara a diferença entre os dois blocos. A proporção dos alunos que não chegam à questão é claramente menor no segundo bloco, especialmente no final do bloco, em que chega a 40%, em comparação ao pico da primeira parte. Além disso, no segundo bloco, a taxa de 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 6, 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

⁴ 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 tem a ver com a 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.

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.

Figura 6 – Queda de desempenho e porcentagem de alunos que não chegaram às questões, Brasil em comparação a outros países no PISA 2015



Fonte: PISA 2015. Elaboração própria. 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.

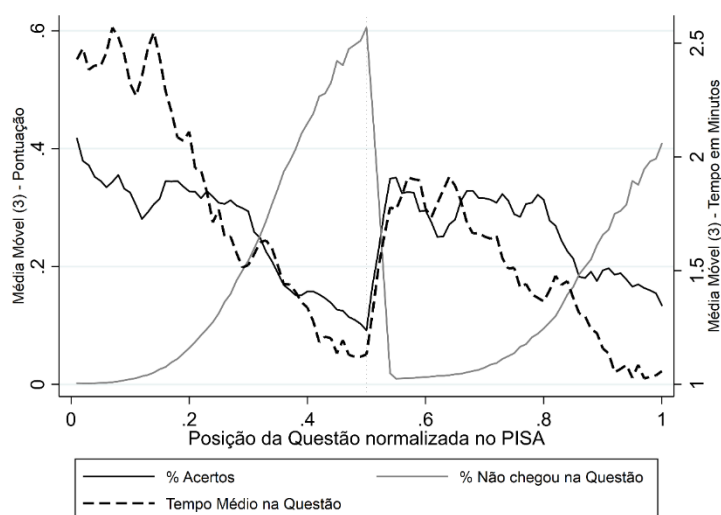
Os dados de aplicação das provas do tipo CBA no PISA 2015 apresentam também o tempo que os alunos levaram para responder cada questão de seus respectivos formulários de prova. Na Figura 7, reportamos o tempo médio de resposta ao longo da prova. A figura mostra que os alunos brasileiros gastam muito tempo 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 maior 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 8, podemos notar que este é muito menor para países exemplares no PISA como Coreia do Sul e Finlândia, e com

variância menor. Curiosamente, ao compararmos os resultados 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 6).

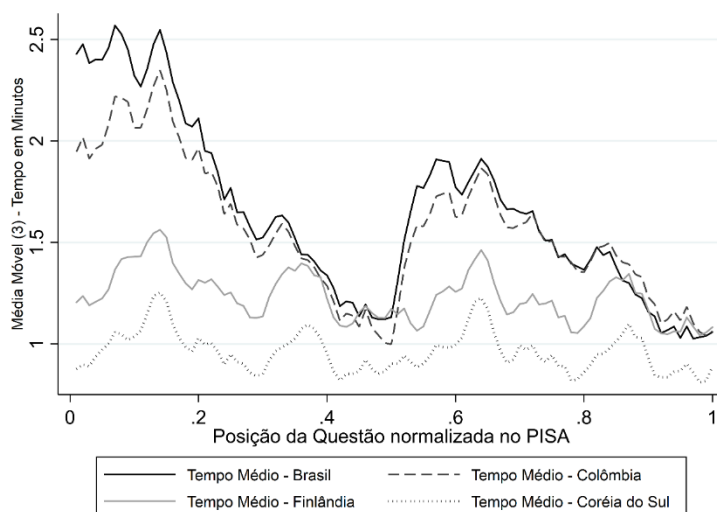
A redução do tempo médio ao longo de um bloco pode refletir dois movimentos. O primeiro deles, é um efeito composição, em que aqueles que gastam mais tempo nas primeiras questões acabam não chegando às questões mais para o final do bloco, de modo que aqueles que respondem a essas questões (e que entram 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.

Figura 7 – Queda de desempenho, porcentagem de alunos brasileiros que não conseguiu responder e tempo médio gasto nas questões no PISA 2015



Fonte: PISA 2015. Elaboração própria. 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.

Figura 8 – Tempo médio gasto nas questões, Brasil em comparação a outros países no PISA 2015



Fonte: PISA 2015. Elaboração própria. 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.

A Tabela 2 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.

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 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.

Tabela 3 – Relação entre Tempo Médio, Pontuação e participação ao longo do PISA 2015 para escolas públicas e privadas no Brasil

| Escolas | 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 |

Fonte: PISA 2015. Elaboração própria. 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.

Quando separamos os alunos brasileiros de escolas públicas e daqueles de escolas privadas, na Tabela 3, 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

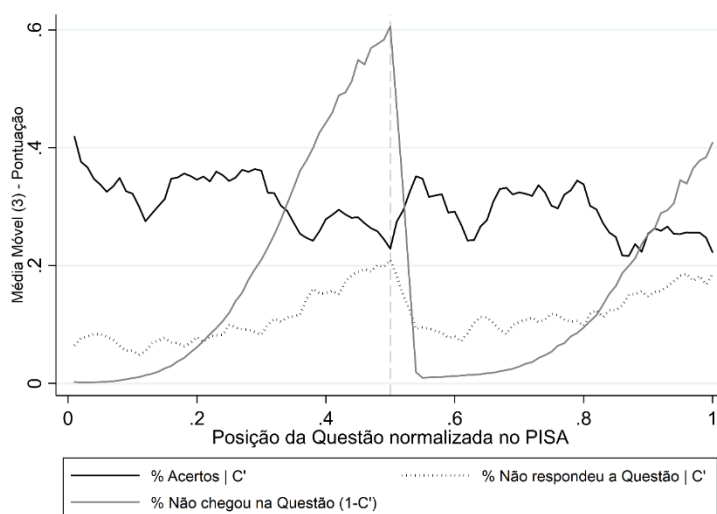
e significativo crescimento na probabilidade de chegar ao final do segundo bloco. No entanto, somente entre os alunos de escolas públicas a probabilidade de acertos parece aumentar de forma relevante, do segundo ao quarto grupos de alunos. Entre os alunos das escolas privadas, a probabilidade de acerto 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 9 mostra qual seria a pontuação condicionada ao percentual de alunos brasileiros que alcançam as 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 as 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.

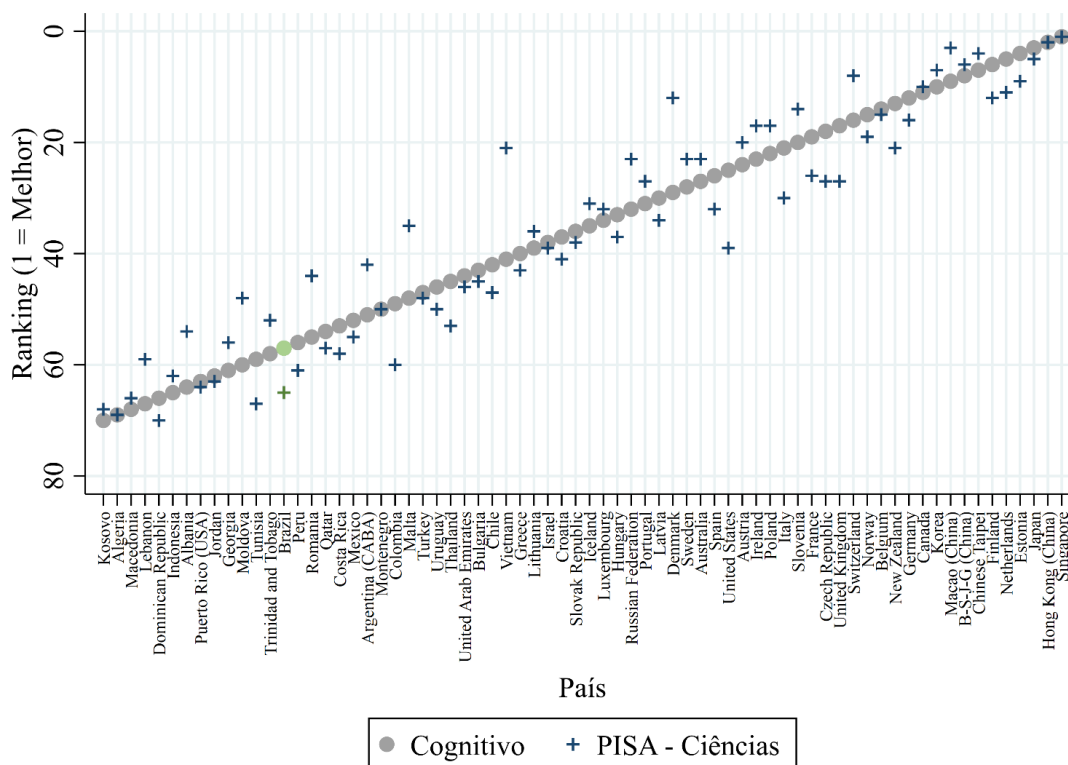
Figura 9 – Queda de desempenho observada e porcentagem de alunos brasileiros que não respondeu às questões no PISA 2015, condicionada ao percentual de alunos que alcançam as últimas questões



Fonte: PISA 2015. Elaboração própria. 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.

Por último, mostramos que apesar do decaimento do Brasil ser acentuado, ele não afeta sua 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 10 mostra os países ordenados de acordo com o ranking por habilidades cognitivas, além do ranking pela média de ciências. Entre as 70 regiões para as quais temos as notas e as estimativas de habilidades cognitivas, o Brasil aparece na posição 57 em relação a essas habilidade, e nas posições 65 em ciências, 58 em leitura e 63 em matemática. A Figura 10 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 pouco somente entre os últimos países do ranking. Esse resultado sugere que o decaimento não afeta de forma relevante a posição relativa do Brasil em relação aos demais países.

Figura 10 – Ranking de Habilidades Cognitivas e das Notas de Ciências



Fonte: PISA 2015. Elaboração própria. Ranking pela probabilidade de acertar a primeira questão e ranking pela nota de ciências, para cada uma das 70 regiões. Destacamos o Brasil em verde.

6. 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. Nós exploramos novas características da prova em 2015 que fazem os alunos responderem sequencialmente às questões para determinar quais aspectos daquela 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. Nossos resultados mostram que, o Brasil possui probabilidade comparativamente baixa de acerto da primeira questão, e uma das maiores medidas de decaimento ao longo da prova. No entanto, exercício contrafactual sugere que somente o nível das habilidades cognitivas parecem influenciar a posição relativa do Brasil entre os países.

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 mais relacionado 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 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 conseguissem 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 levado especialmente por alunos da rede pública.

Dados os desenvolvimentos alcançados entendemos ser interessante aprofundar a esta análise, estendendo-a para os anos anteriores da prova para compreender de que forma se dá a evolução desses indicadores cognitivos e não cognitivos. A mensuração dessa evolução juntamente com os dados de indicadores de desempenho econômico nos permitirá traçar um panorama de como as habilidades socioemocionais influenciam essas variáveis. O PISA vem, ao longo das suas últimas edições, ampliando a coleta de dados

socioeconômicos e sociomocionais auxiliares ao exame, com o uso destes dados – como por exemplo a riqueza familiar ou atitude e motivação em relação ao esforço – seria possível aprofundar a análise das diferenças de decaimento médio ao nível de indivíduo, principalmente, validando a medida objetiva aqui proposta a partir de outros construtos sociemocionais 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. As formulações apresentadas e também referenciadas neste trabalho (Borghans e Schils, 2012; Marchioni, 2017; Balart et. al, 2018) estão limitadas, em termos práticos, a uma estimação da 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 – uma vez que o intervalo parece intermediar alguma relação entre o tempo médio e a probabilidade de acerto das questões. 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 provas⁵ e o decaimento observado é menor em relação ao decaimento do primeiro bloco. Sugerindo 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.

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

Referências Bibliográficas

DAVIS, J.; FERDOUS, A.; Using item difficulty and item position to measure test fatigue. In: ANNUAL MEETING AMERICAN EDUCATIONAL RESEARCH ASSOCIATION, 2005, Montreal, Québec. Proceedings... Washington, DC: AMERICAN EDUCATIONAL RESEARCH ASSOCIATION, 2005. Disponível online em: <https://convention2.allacademic.com/one/aera/aera05/index.php?click_key=1&cmd=Multi+Search+Search+Load+Publication+For+Extra&publication_id=9091&PHPSESSID=817c3aamf6crecsve4ciu86g24>.

ALBANO, Anthony D. Multilevel modeling of item position effects. **Journal of Educational Measurement**, v. 50, n. 4, p. 408-426, 2013.

ALMUND, M.; DUCKWORTH, A. L., Heckman, J.; KAUTZ, T. Personality psychology and economics. In: HANUSHEK, E; MACHIN, S. and WOESSMANN, L. (Eds.) **Handbook of the Economics of Education**. Vol. 4. Amsterdam: Elsevier, 2011. pp. 1-181.

BALART, Pau; OOSTERVEEN, Matthijs; WEBBINK, Dinand. Test scores, noncognitive skills and economic growth. **Economics of Education Review**, v. 63, p. 134-153, 2018.

BORGHANS, L.; SCHILS, T. The leaning tower of Pisa: decomposing achievement test scores into cognitive and noncognitive components. Unpublished manuscript, 2012. Disponível online em:

<<http://www.sole-jole.org/13260.pdf>>.

BORGHANS, Lex; DUCKWORTH, Angela Lee; HECKMAN, James J.; TER WEEL, Bas. The economics and psychology of personality traits. **Journal of human Resources**, v. 43, n. 4, p. 972-1059, 2008.

DUCKWORTH, Angela Lee; QUINN, Patrick D.; LYNAM, Donald; LOEBER, Rolf; STOUTHAMER-LOEBER, Magda; MOFFITT, Terrie E.; CASPI, Avshalom. What intelligence tests test: Individual differences in test motivation and IQ. Manuscript submitted for publication, 2009. Disponível online em: <https://www.researchgate.net/publication/237626525_What_Intelligence_Tests_Test_Individual_Differences_in_Test_Motivation_and_IQ>.

EKLÖF, Hanna. Test-taking motivation and mathematics performance in TIMSS 2003. **International Journal of Testing**, v. 7, n. 3, p. 311-326, 2007.

HANUSHEK, Eric A.; KIMKO, Dennis D. Schooling, Labor-Force Quality, and the Growth of Nations. **The American Economic Review**, v. 90, n. 5, pp. 1184–1208, 2000. Disponível online em: <www.jstor.org/stable/2677847>.

HANUSHEK, E; WOESSMANN, L. Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. **Journal of Economic Growth**, v. 17, n. 4, 267-321, Dec. 2012. ISSN: 13814338.

HECKMAN, James J.; PINTO, Rodrigo; SAVELYEV, Peter. Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. **The American economic review**, v. 103, n. 6, p. 2052-2086, 2013.

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

MARCHIONI, Cynthia G. Habilidades no cognitivas en América Latina. Una medición desde pruebas estandarizadas. 2017. Dissertação (Mestrado em Economia) – Facultad de Ciencias Económicas, Universidad Nacional de La Plata, Buenos Aires. Disponível online em: <<http://www.depeco.econo.unlp.edu.ar/wp/wp-content/uploads/2017/09/145-tesis-marchioni.pdf>>.

OCDE. PISA 2015: Technical Report. 2017. Disponível em: <<http://www.oecd.org/pisa/data/2015-technical-report/>>. Acesso em: abr. 2017.

SEGAL, Carmit. Working when no one is watching: Motivation, test scores, and economic success. **Management Science**, v. 58, n. 8, p. 1438-1457, 2012.