

Moreira, Ajax Reynaldo Bello

Working Paper

Escolas públicas: Perfil do desempenho e seus condicionantes

Texto para Discussão, No. 2426

Provided in Cooperation with:

Institute of Applied Economic Research (ipea), Brasília

Suggested Citation: Moreira, Ajax Reynaldo Bello (2018) : Escolas públicas: Perfil do desempenho e seus condicionantes, Texto para Discussão, No. 2426, Instituto de Pesquisa Econômica Aplicada (IPEA), Brasília

This Version is available at:

<https://hdl.handle.net/10419/211376>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

TEXTO PARA **DISCUSSÃO**

2426

**ESCOLAS PÚBLICAS: PERFIL
DO DESEMPENHO E SEUS
CONDICIONANTES**

Ajax Moreira



ESCOLAS PÚBLICAS: PERFIL DO DESEMPENHO E SEUS CONDICIONANTES

Ajax Moreira¹

1. Coordenador de estudos da conjuntura e flutuação cíclica na Diretoria de Estudos e Políticas Macroeconômicas (Dimac) do Ipea. Correio eletrônico: <ajax.moreira@ipea.gov.br>.

Governo Federal

Ministério do Planejamento, Desenvolvimento e Gestão

Ministro Esteves Pedro Colnago Junior

ipea

**Instituto de Pesquisa
Econômica Aplicada**

Fundação pública vinculada ao Ministério do Planejamento, Desenvolvimento e Gestão, o Ipea fornece suporte técnico e institucional às ações governamentais – possibilitando a formulação de inúmeras políticas públicas e programas de desenvolvimento brasileiros – e disponibiliza, para a sociedade, pesquisas e estudos realizados por seus técnicos.

Presidente

Ernesto Lozardo

Diretor de Desenvolvimento Institucional

Rogério Boueri Miranda

Diretor de Estudos e Políticas do Estado, das Instituições e da Democracia

Alexandre de Ávila Gomide

Diretor de Estudos e Políticas Macroeconômicas

José Ronaldo de Castro Souza Júnior

Diretor de Estudos e Políticas Regionais, Urbanas e Ambientais

Alexandre Xavier Ywata de Carvalho

Diretor de Estudos e Políticas Setoriais de Inovação e Infraestrutura

Fabiano Mezadre Pompermayer

Diretora de Estudos e Políticas Sociais

Lenita Maria Turchi

Diretor de Estudos e Relações Econômicas e Políticas Internacionais

Ivan Tiago Machado Oliveira

Assessora-chefe de Imprensa e Comunicação

Mylena Pinheiro Fiori

Ouvidoria: <http://www.ipea.gov.br/ouvidoria>

URL: <http://www.ipea.gov.br>

Texto para Discussão

Publicação seriada que divulga resultados de estudos e pesquisas em desenvolvimento pelo Ipea com o objetivo de fomentar o debate e oferecer subsídios à formulação e avaliação de políticas públicas.

© Instituto de Pesquisa Econômica Aplicada – **ipea** 2018

Texto para discussão / Instituto de Pesquisa Econômica Aplicada.- Brasília : Rio de Janeiro : Ipea , 1990-

ISSN 1415-4765

1. Brasil. 2. Aspectos Econômicos. 3. Aspectos Sociais.
I. Instituto de Pesquisa Econômica Aplicada.

CDD 330.908

As publicações do Ipea estão disponíveis para *download* gratuito nos formatos PDF (todas) e EPUB (livros e periódicos).
Acesse: <http://www.ipea.gov.br/portal/publicacoes>

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade dos autores, não exprimindo, necessariamente, o ponto de vista do Instituto de Pesquisa Econômica Aplicada ou do Ministério do Planejamento, Desenvolvimento e Gestão.

É permitida a reprodução deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais são proibidas.

JEL: I21; I28; C38; C45.

SUMÁRIO

SINOPSE

ABSTRACT

1 INTRODUÇÃO7

2 REVISÃO DA LITERATURA9

3 METODOLOGIA.....10

4 IMPLEMENTANDO A METODOLOGIA PROPOSTA16

5 CONCLUSÃO30

REFERÊNCIAS31

APÊNDICE34

SINOPSE

Neste estudo, propomos um instrumento de gestão para identificar grupos de escolas públicas que tenham padrões de desempenho similar e por isso tenham necessidade de apoio e orientação similar.

O resultado do aluno é medido a partir de suas notas na Prova Brasil (PB), nas disciplinas de matemática e língua portuguesa, e também pelo indicador da condição de estar realizando a PB com uma idade superior à esperada. O grau de inclusão é mensurado por meio das taxas de reprovação e abandono da escola. Definimos como produto da escola a parcela do resultado dos alunos que não pode ser explicada a partir de seus ambientes familiar e escolar, e identificamos uma partição em grupos de escolas com padrões dissimilares, mas compostos de escolas similares.

Essa metodologia foi implementada com as informações da Prova Brasil e do Censo Escolar de 2013. Os principais resultados empíricos demonstram que: *i*) é possível definir uma partição do conjunto das escolas públicas em nove grupos segundo o aprendizado e o grau de inclusão dos alunos; *ii*) estes grupos são linearmente separáveis; e *iii*) esta partição é parcialmente explicada com variáveis não observadas, comuns a todas as escolas de um município.

Palavras-chave: desempenho das escolas; classificação; análise de *cluster*.

ABSTRACT

We propose a management tool to identify groups of public schools that have similar standards of performance and therefore need similar support and guidance.

The student's result is measured with the "Prova Brasil" (PB) for mathematics and for Portuguese language, and with an indicator of the condition of being performing (PB) with a higher than expected age. The degree of inclusion measured by the rates of disapproval and dropping out of school. We defined as the school product the portion of the students' results that is not explained with their family and school environment, and we identified a partition in groups of schools with dissimilar patterns, but composed of similar schools.

We implemented this methodology with information from the “Prova Brasil” and the School Census of the year 2013. The main empirical results are: *i*) it is possible to define a partition of the set of public schools in 9 groups according to the learning and the degree of inclusion of the students; *ii*) these groups are linearly separable; and *iii*) this partition is partially explained with unobserved variables common to all schools in a municipality.

Keywords: performance of schools; ranking; cluster analysis

1 INTRODUÇÃO

As escolas públicas de ensino fundamental estão submetidas a um sistema de gestão complexo, descentralizado e monitorado centralmente. A direção do colégio tem autonomia, mas depende de orientação e recursos da secretaria municipal de educação, que por sua vez tem o mesmo tipo de vínculo com a secretaria estadual. Todo esse sistema recebe recursos do orçamento federal, que é monitorado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). O sistema favorece a diversidade de comportamento, a experimentação e a análise de resultados. O Brasil não está em uma posição favorável quando comparado com outros países. A Organização para a Cooperação e o Desenvolvimento Econômico (OECD) apresenta resultados da prova do Programa Internacional de Avaliação de Estudantes (Pisa) para 65 países. Esta avaliação mostra que a nota do Brasil em matemática é superior apenas quando comparada a oito desses países (OECD, 2012). Bruns, Evans e Luque (2012) analisam a educação no Brasil e afirmam que, embora a qualidade da educação tenha melhorado, ainda há muito o que ser feito.

Segundo Benhabib e Spiegel (2005), Ranis e Stewart (2005) e Pradhan e Hall (2014), o crescimento econômico futuro é condicionado pelo capital humano, e a formação deste capital é discutida em Hanushek (1986), Hanushek e Woessmann (2012) e Hanushek (2015) e em outros artigos resenhados em Tabosa (2017). Algumas pesquisas se limitam a identificar correlações entre práticas pedagógicas, características da escola ou incentivos e o desempenho da escola, mas outros, como Fryer Junior (2016), adotam abordagem experimental para analisar a questão e estabelecer nexos causais.

Neste estudo, a questão é distinta. Propomos um instrumento de gestão para identificar grupos de escolas públicas que tenham padrões de desempenho similar e assim tenham necessidade de apoio e orientação similar. A escola pública deve prover ensino de qualidade de forma inclusiva, ou seja, promover o aprendizado do conteúdo requerido das matérias, evitando o atraso escolar, a reprovação e o abandono da escola. O resultado do aluno é medido a partir de suas notas na Prova Brasil, especificamente nas disciplinas de matemática e língua portuguesa, e também mediante um indicador da condição de estar realizando a PB com uma idade superior à esperada. O grau de inclusão é mensurado a partir das taxas de reprovação e abandono da escola.

Desde Hanushek (1986), reconhece-se que o aprendizado do aluno é o resultado da atividade pedagógica e da orientação da escola, bem como produto dos ambientes familiar e escolar. O aluno chega à escola com interesses, disciplina e capacidade de convívio social formados no ambiente familiar, além do apoio e estímulo para o estudo. Naturalmente, alunos de famílias mais qualificadas e que estudam em escolas localizadas em vizinhança com famílias também mais qualificadas tendem a ter vantagem sobre os demais. Portanto, o produto da escola é uma quantidade não observada, pois a quantidade observada – desempenho dos alunos da escola – depende das famílias dos alunos. Por isto definimos como produto da escola a parcela do desempenho dos alunos que não pode ser explicada a partir de seus ambientes familiar e escolar.

As escolas podem estar distribuídas de forma homogênea no espaço, conforme as medidas do seu desempenho, ou se concentrar em grupos de escolas similares, ou seja, o conjunto das escolas pode estar distribuído em grupos distintos, mas composto de escolas similares. Nossa hipótese neste estudo é que o desempenho da escola revela características observadas e não observadas do seu funcionamento, e que os padrões de desempenho revelam o funcionamento ou as características das escolas. Por exemplo, as escolas que apresentam ensino de melhor qualidade podem ser as mesmas que, sendo muito exigentes, tendem a expelir os piores alunos, assim como as escolas inclusivas podem ser lenientes e estar abrindo mão da qualidade do ensino. Estes dois grupos de escola têm características muito diferentes e requerem apoio e orientação distintos, e podem ser identificados considerando conjuntamente as diversas dimensões do desempenho.

Estabelecida uma partição das escolas, cabe perguntar se esses grupos são separáveis e se o perfil dos professores e diretores da escola estão relacionados com o grupo em que a escola foi classificada. Estas duas questões foram abordadas com uma função discriminante linear. A seguir, a seção metodológica discute como mitigar o efeito da família no desempenho do aluno, como particionar as escolas em grupos similares e como avaliar a capacidade de um conjunto de informações explicar uma partição.

A metodologia proposta neste estudo foi implementada com as informações da Prova Brasil e do Censo Escolar de 2013. Os principais resultados empíricos demonstram o seguinte.

- 1) As escolas apresentam desempenho heterogêneo, sendo possível definir uma partição do conjunto das escolas públicas em nove grupos, segundo o aprendizado e o grau de inclusão dos alunos, sendo identificados os seguintes grupos: *i)* nota alta e inclusão média; *ii)* nota e exclusão média; *iii)* nota baixa e atraso médio; *iv)* nota média com alta taxa de reprovação; *v)* atraso escolar alto; *vi)* taxa de abandono alta; *vii)* taxa de reprovação alta; *viii)* nota e inclusão média; e *ix)* outros.
- 2) Essa partição é parcialmente explicada por variáveis não observadas comuns a todas as escolas de um município, o que sugere a importância das secretarias municipais/estaduais de educação.
- 3) Os perfis do professor, do diretor e da escola são obtidos por meio de questionário realizado em conjunto com a Prova Brasil, que constituem cerca de trezentas variáveis, e estas não apresentam capacidade de explicar a partição.

A seção 2 faz um breve resumo da literatura e a seção 3 descreve a metodologia para a construção da partição das escolas em grupos, identificando os padrões de funcionamento. Na seção 4, os modelos adotados explicam a participação das escolas nos grupos identificados. A seção 5 apresenta os resultados para 2013 e a conclusão.

2 REVISÃO DA LITERATURA

A relação entre educação e produtividade econômica motiva uma quantidade enorme de estudos que enfatizam diferentes aspectos desta relação e dos condicionantes do processo educacional. A questão do efeito da desigualdade na condição social dos alunos sobre o aprendizado é estudado desde o conhecido relatório Coleman. Hanushek (1986; 2015) e Hanushek e Woessmann (2012), em diversos artigos, desde a década de 1970, e em particular em 2015, enfatizam que a qualidade da educação – aprendizado de um tema – é mais importante do que a quantidade de educação – escolaridade – como medida da formação do capital humano e suas consequências sobre a produtividade e o crescimento econômico. Em uma outra vertente, Fryer Junior (2016) faz uma meta-análise de cerca de trezentos experimentos realizados nos Estados Unidos que atendem a padrões mínimos de qualidade metodológica para validar práticas pedagógicas e acompanhar a implementação de programas de apoio em escolas selecionadas. Desse modo, ele avaliou a eficiência econômica de diferentes tipos de intervenção, em particular o monitoramento dos professores, o aumento do tempo de ensino, o apoio aos alunos em desvantagem e, mais importante, a orientação para os diretores aperfeiçoarem a gestão da escola.

No Brasil, Bof (2016) e Simões (2016) fazem uma avaliação do Plano Nacional de Educação; Soares e Sátyro (2008) e Silva Filho (2016) avaliam o impacto de recursos sobre a aprendizagem e o atraso escolar; e Medeiros e Oliveira (2014) medem o efeito da desigualdade regional da distribuição de fatores sobre o resultado das escolas. Todos fazem avaliações que admitem implicitamente a homogeneidade das escolas e não enfatizam as diferenças de gestão. As escolas precisam formar as habilidades cognitivas e também socioemocionais dos seus alunos. São dimensões que requerem esforços específicos, mas que não necessariamente competem por recursos. Polo e Santos (2016) avaliam se as escolas enfrentam este dilema.

Em outra vertente e utilizando a análise de conglomerados, Hattie (2002) identifica padrões (tipos) das escolas da Nova Zelândia para avaliar a discriminação étnica. Fortin (2006) analisa os padrões dos alunos para avaliar o seu potencial de evasão. Estes artigos discutem o efeito da educação no crescimento e da adoção de práticas utilizando experimentos com escolas e, mediante registros administrativos, buscam avaliar o impacto das condições de infraestrutura e outras características das escolas sobre o aprendizado e o atraso escolar. A maioria destes textos não enfatizam o papel da gestão e a identificação de grupos de escolas com desempenhos similares no grupo e dissimilares entre os grupos, e, neste sentido, estamos, até onde temos informação, lidando com um tema inovador.

3 METODOLOGIA

3.1 Efeito do desempenho da escola

Por meio do Censo Escolar, o Inep compila anualmente informações sobre os alunos matriculados, reprovados e os que abandonaram a escola. A cada dois anos, este órgão realiza a Prova Brasil/Saeb para todos os alunos das escolas públicas com mais de vinte alunos no final de cada etapa do ensino fundamental. As provas testam o aprendizado em matemática e em língua portuguesa, e o resultado é comparável para todos os alunos. Juntamente com a prova, os alunos respondem a um questionário sobre as características de suas famílias. Os professores e os diretores também respondem a um questionário sobre suas características e as da escola. Então, seja $R(k, s, e, a)$ o resultado do aluno (a) que cursa a etapa $s = 1$ (2) do ensino fundamental na escola (e), relativa a três medidas de resultado: ($k = 1$) nota em matemática, ($k = 2$) nota em língua por-

tuguesa e ($k = 3$) o indicador de atraso escolar.¹ Seja $F(s, e, a)$ o vetor de características das famílias apresentado no apêndice² A que foram sintetizadas em 23 variáveis. O resultado da escola é a parcela do resultado do aluno $R(\cdot)$, que não é explicada com o perfil da sua família $F(\cdot)$,³ ou seja, é o efeito fixo $RX(\cdot)$ no modelo a seguir.

$$E(R(k, s, e, a)) = RX(k, s, e) + \beta(k, s) F(s, e, a) \quad k = 1, 2, 3 \quad s = 1, 2 \quad (1)$$

Seja $RX(k, s, e)$ e $k = 1, 2, 3$ as parcelas dos resultados do aluno que não são explicadas com o perfil da sua família, que é estimada no modelo 1; e $RX(k, s, e)$ e $k = 4, 5$, respectivamente, correspondem às taxas de reprovação e abandono da escola informadas no Censo Escolar para cada etapa do ensino fundamental.

O desempenho da escola é definido como a parcela de (RX) que não é explicada com o ambiente escolar, e este é medido com o perfil médio das famílias da escola: $F(e)$.

$$E(RX(k, s, e)) = \delta(k, s) + \gamma(k, s) F(e) \quad k = 1..5, \quad s = 1, 2 \quad (2)$$

Estimado o modelo, o desempenho da escola (e) na dimensão (k) da etapa (s) do ensino fundamental é definido como:

$$r(k, s, e) = RX(k, s, e) - (\delta(k, s) + \gamma(k, s) F(e)) \quad k = 1..5, \quad s = 1, 2 \quad (3)$$

3.2 Particionando o conjunto das escolas

A tabela 1 mostra que, depois de combinar as informações do Censo Escolar com as da Prova Brasil para 2013, obtivemos informação sobre 54 mil escolas, das quais apenas 27% tem as duas etapas do ensino fundamental. Isto nos motivou a analisar separadamente estas duas etapas. Pretendemos identificar grupos de escolas que sejam similares, considerando conjuntamente todas as cinco dimensões ($r(e, k) \quad k = 1..5$) do desempenho de cada uma das etapas do ensino fundamental.

1. Esse indicador assume o valor 1 quando o aluno tem mais que dois anos da idade esperada para a conclusão da etapa do ensino fundamental que está cursando.

2. Nele estão incluídas perguntas que caracterizam o nível de renda da família – eletrodomésticos, condições da moradia – educação dos pais e comportamento do aluno (tipo de escola em que começou a estudar, se faz dever de casa, se os pais ajudam/motivam o estudo).

3. O perfil da família está documentado no apêndice A, que apresenta 23 variáveis construídas a partir das 51 questões do questionário respondido pelo aluno.

Nosso objetivo é particionar as escolas em grupos $(G_1, \dots, G_m, \dots, G_g)$ que sejam distintos, mas compostos de escolas similares, ou seja, encontrar os pontos – centroides do grupo – $(r(m), m = 1 \dots g)$ tais que a soma da distância de cada escola do grupo G_m a estes pontos $d(e, m)$ seja a menor possível.

Mais formalmente, trata-se de encontrar $G(g) = (G_1, \dots, G_m, \dots, G_g)$, tal que $FC = \text{Min} \sum_m VI(m)$, em que $VI(m) = \sum_{e \in G_m} d(e, m)$, em que $d(e, m) = \sum_k (r(e, k) - r(m, k))^2$, e $r(m, k) = \sum_{e \in G_m} r(e, k) / \#(G_m)$, $r(m) = (r(m, k) \ k = 1 \dots 5)$.

A solução exata desse problema consiste em comparar todas as possíveis partições e escolher a que minimize (FC) . O número de partições possíveis é (2^n) , em que n é o número de escolas, o que torna esta abordagem inviável para conjuntos do tamanho do que temos.

No caso de n moderados (< 15 mil), os algoritmos hierárquicos constroem uma sequência de partições indexadas que encontram partições ótimas para um dado número de grupos da partição. A sequência de partições começa comparando as escolas isoladamente e as agrupa segundo as de maior similaridade em sucessivas hierarquias, com graus crescentes de aglomeração, até chegar ao conjunto de todas as escolas.

O conjunto das escolas tem entre 30 mil e 38 mil escolas, por isto adotamos uma abordagem em duas etapas para efetuar a partição das escolas.⁴ Na primeira etapa, utilizamos um algoritmo hierárquico⁵ para uma amostra aleatória de cerca de 15 mil escolas. Na segunda etapa, utilizamos o centroide obtido na primeira etapa, ponto de partida para o algoritmo (*kmeans*) que realiza a partição para o conjunto de todas as escolas.

É possível mostrar⁶ que:

- $VI(m) = \sum_k VI(m, k)$, em que $VI(m, k)$ é a variância intra da dimensão k no grupo m , mostrando de forma direta que o resultado depende da variância de cada dimensão;

4. A partição realizada em duas etapas é uma solução aproximada para a partição de minimiza $VI(G(g))$. Esta aproximação é tão boa quanto a amostra selecionada para o algoritmo hierárquico for representativa. Consideramos como representativa a amostra aleatória de 40% das 38 mil escolas que utilizamos.

5. *Complete linkage*, disponível no programa *stata*.

6. Se $d(e, m)^2 = \sum_k (r(e, k) - r(m, k))^2$ e $VI(m) = \sum_e d(e, m)^2 = \sum_e \sum_k (r(e, k) - r(m, k))^2 = \sum_k \sum_e (r(e, k) - r(m, k))^2 = \sum_k VI(m, k)$ em que $VI(m, k) = \sum_e (r(e, k) - r(m, k))^2$ é a variância intra da dimensão k no grupo m .

- o algoritmo *kmeans*⁷ garante que os conjuntos $G(m)$ são convexos;⁸ e
- $\forall e \in G_m$ então $\{d(e,m) < d(e,j) \forall j \neq m\}$, portanto qualquer par de grupos (m, j) são disjuntos.

Como não temos um critério para avaliar a importância relativa de cada uma das dimensões de resultado, vamos considerar que todas elas sejam igualmente importantes, por isto foram padronizadas, fazendo que tenham média 0 e desvio-padrão unitário para o conjunto das escolas. Esta padronização garante que a unidade de medida do resultado não tenha efeito sobre a partição das escolas e facilita a interpretação da característica de cada um dos grupos identificados por meio dos seus centroides.

O critério para a construção dos grupos garante que conjuntos $(G(m) \ m = 1 \dots g)$ são convexos e disjuntos, ou seja, $G(i) \cap G(j)$ é vazia $\forall i, j$. Estas duas condições garantem a existência de um separador linear para cada par de grupos, questão discutida na seção 3.4.

3.3 Escolha do número de grupos da partição

O algoritmo de agrupamento obtém resultados condicionais ao número de grupos escolhido. O aumento do número de grupos aumenta a homogeneidade intragrupo, mas aumenta a complexidade do resultado. Mais complexidade significa mais imprecisão na estimativa dos parâmetros e maior dificuldade no uso da informação dos padrões identificados. Escolher o número de grupos é minimizar uma função de perda que considere estas duas dimensões, homogeneidade e complexidade.

Não dispomos de medidas para o custo da complexidade, e o efeito sobre a imprecisão é irrelevante para o grau de fragmentação de interesse, ou seja, não temos uma função de perda, e, portanto, não há um critério objetivo para a escolha do número de grupos. Por isto, apresentamos medidas descritivas de partições indexadas com o seu

7. Também disponível no programa *stata*, que dado os centroides $C(i)$ aloca cada escola (e) ao grupo que tem o centroide mais próximo, e recalcula o centroide do grupo. Repete-se este procedimento até que todas as escolas tenham sido alocadas.

8. Dado $e1, e2 \in G(i)$ e $y(k, \alpha) = (\alpha r(k, e1) + r(k, e2)(1 - \alpha)) \ \alpha \in [0, 1]$, $d(y, i) < d(ek, i) \ k = 1, 2, e$, portanto $y \in G(i)$ e $d(y, i) < d(y, j) \ \forall j$.

número de grupos, e, após uma avaliação subjetiva das partições obtidas, escolhemos arbitrariamente uma das alternativas.⁹

3.4 Critério de separação dos grupos

O algoritmo de partição adotado garante que os grupos de uma partição $G(g, r)$ definidos no espaço de resultados das escolas (r) são conjuntos convexos e disjuntos, portanto separáveis linearmente. Dado um par de grupos (i, j) e $I(i, e) = 1 \Leftrightarrow e \in G(i, r)$, então existe uma função linear $I(i, e) | \beta_r(e)$ que separa os grupos (i) e (j) no espaço de resultados (r), que foi o utilizado para determinar a partição.

A partição das escolas $G(g)$ revela padrões de funcionamento que podem ser o resultado de características observadas e não observadas da escola. Na Prova Brasil, as escolas estão caracterizadas com o seu perfil definido por cerca de trezentas unidades observadas $x = (PE, PD, PF)$, que caracterizam, respectivamente, a escola, o diretor (PD) e o professor (PF). Uma questão relevante é avaliar se a variabilidade do perfil (x) é capaz de distinguir uma escola do grupo “bom desempenho” de uma escola do grupo “mal desempenho”.

Isso pode ser avaliado definindo no espaço de (x) a mesma partição de escolas agora no espaço (x) , $G(g, x)$. Neste espaço, os grupos da partição não são necessariamente convexos nem disjuntos, mas a função discriminante linear continua sendo o separador ótimo, ou seja, dado um par de grupos (i, j) e $I(i, e) = 1 \Leftrightarrow e \in G(i, x)$ a função $I(i, e) | \beta_x(e)$ determina a menor probabilidade de erro de classificação.

O desempenho de uma escola é o resultado de características observadas da escola (x) e de características não observadas, das quais podemos estimar o efeito daquelas que são comuns a todas as escolas de um certo município, representando, por exemplo, a gestão da secretaria municipal de educação ou a do prefeito. A função discriminante implementada foi (4), que explica a classificação com as quantidades observadas $x(e)$ e com um efeito fixo municipal que representa todas as quantidades comuns de um município que distinguem a classificação da escola.

9. De fato, escolhemos uma partição e apresentamos os resultados para uma segunda partição no apêndice A.

$$E(J(e)|x(e)) = \alpha(m(e)) + \beta x(e) \quad \forall e \in G(i) \cup G(j), \text{ onde } J(e) = 1 \Leftrightarrow e \in G_i, 0 \Leftrightarrow e \in G_j \quad (4)$$

O erro de classificação consiste em identificar uma escola como do grupo (*i*) quando é do grupo (*j*) e vice-versa. A proporção de escolas que foram classificadas em cada uma destas condições constitui a probabilidade de erro de classificação.

Quando o interesse é avaliar a capacidade da variabilidade de (*x*) explicar a classificação, considere o indicador (5)

$$J(e, x, z) = 1 \Leftrightarrow E(J|x) = \hat{\beta}x > z, \quad p1(x, z) = P(j(e, x, z) = 0 | J(e) = 1), \quad p2(x, z) = P(j(e, x, z) = 1 | J(e) = 0) \quad (5)$$

Em que *p1* e *p2* são as probabilidades dos erros de classificação do tipo 1 e 2. Dependendo do problema de classificação, pode-se dar mais importância a um dos dois erros; portanto, não existe um critério único para a escola de *z*.

Escolhemos o (*z*), que minimiza o maior dos dois erros de classificação, e a nossa medida de capacidade de *x* distinguir entre os grupos (*i, j*)¹⁰ é $p(x, i, j) = \min_z (\max(p1(x, z) + p2(x, z)))$.

Quando o interesse for avaliar a capacidade de (*x*) e do efeito fixo municipal explicar a separação entre (*i, j*), temos

$$J(e, x, z) = 1 \Leftrightarrow E(J|x) = \alpha(\widehat{m}) + \hat{\beta}x > z, \quad q1(x, z) = P(j(e, x, z) = 0 | J(e) = 1), \quad q2(x, z) = P(j(e, x, z) = 1 | J(e) = 0) \quad (6)$$

A metodologia para o cálculo de $p(x, i, j)$ e $q(x, i, j)$ é empírica,¹¹ e o seu grau de precisão foi avaliado estimando a função discriminante, considerando como variável explicativa o vetor de desempenho das escolas, $E(J(e) | r(e)) = \beta r(e)$. Neste caso, o valor esperado de $p(r, i, j)$ é zero, e espera-se que o valor estimado seja próximo de zero.

A função discriminante pode ser estimada para distinguir o grupo *G* (*i*) dos demais grupos $\Omega(i) = \cup_{k \neq i} Gk$. Neste caso, a equação (4) é estimada $\forall e \in \Omega$, e utilizar a mesma metodologia das equações (5) e (6) para calcular as funções $p(x, i)$ e $q(x, i)$.

10. A separação também pode ser avaliada entre cada grupo e os demais grupos $g(k) \times UG_i \neq k$.

11. Essa função é calculada apenas a cada 5% da distribuição de *z*, o que coloca alguma imprecisão nos resultados.

Resumindo, a metodologia proposta consiste em realizar, em cada etapa do ensino fundamental, os seguintes passos:

- estimar os modelos 1 e 2, obtendo: $r(e, k)$ $k = 1, 2, 3, 4, 5$;
- padronizar o resultado: $r^*(e, k) = (r(e, k) - m(r(k, .))) / sd(r(k, .))$;
- utilizar um algoritmo hierárquico para efetuar a partição das escolas segundo a medida $r^*(e)$ para uma amostra de escolas e obter os centroides $C(g) = (r(m, k), m = 1...g, k = 1...5)$ para cada uma das alternativas $g = 1, \dots$;
- determinar partição $G(g)$ com o algoritmo *kmeans* dado o valor inicial $C(g)$ obtido na etapa anterior;
- calcular as estatísticas de resultado associado a $G(g)$;
- repetir as etapas 4 e 5 para diversos $g = 2, \dots$ e escolher um g^* ; e
- estimar a função discriminante (5) e as probabilidades de erro de classificação $p(r, i, j), p(x, i, j) q(x, i, j) \forall i, j$ (4).

4 IMPLEMENTANDO A METODOLOGIA PROPOSTA

A metodologia foi implementada utilizando a Prova Brasil e o Censo Escolar relativos ao ano de 2013. Estas duas fontes de informação foram combinadas, e a tabela 1 resume o resultado.

O Censo Escolar considera todas as escolas, mas a Prova Brasil¹² examinou apenas aquelas com mais de vinte alunos por turma. A tabela 1 mostra que no Censo existem 21.602 escolas com mais de cem alunos,¹³ porém não foram localizadas na Prova Brasil.¹⁴ Encontramos 54.835 escolas que possuem a informação completa. Destas, 38.784 têm turmas da primeira etapa (F1) e 30.906 têm turmas da segunda etapa (F2), das quais 14.855 têm as duas etapas do ensino fundamental.

12. Na Prova Brasil, são incluídos resultados de escolas codificadas de forma diferenciada (começando com 62) relativas a municípios que não existem e não foram localizadas no censo.

13. A Prova Brasil considera alunos no ano da prova, e nós computamos o número total de alunos em cada etapa do ensino fundamental.

14. Isso revela a falha na combinação dos dados do censo e da Prova Brasil, a qual precisa ainda ser entendida.

TABELA 1
Número de escolas segundo categorias

Prova Brasil	Localizada na Prova Brasil			Cruzando F1 e F2			
	Sim	Não	Total	Não F1	F1	Total	
Número de alunos < 100	2.696	74.021	76.717	Não F2	0	23.929	23.929
Número de alunos > 100	54.835	21.602	76.437	F2	16.051	14.855	30.906
Total	57.531	95.623	153.154	Total	16.051	38.784	54.835

Elaboração do autor.

A tabela 2¹⁵ apresenta os percentis da distribuição de cada dimensão do desempenho das escolas e mostra o quanto as escolas são heterogêneas. Basta comparar os resultados do percentil inferior com o superior das medidas de desempenho. Para a etapa 1, cerca de 25% das escolas apresentam resultados nulos para as medidas de não inclusão (atraso, reprovação e abandono) e 25%, valores elevados destas medidas. A mesma dispersão vale para as demais medidas e etapas.

TABELA 2
Distribuição do desempenho das escolas

Percentil	Fundamental 1					Fundamental 2				
	mt	lp	at	ab	rep	mt	lp	at	ab	rep
p5	153	145	0	0	0	208	204	0,12	0	0,5
p25	183	170	0	2	0	228	224	0,22	0,5	4,5
p50	206	190	0,3	5,6	3,5	242	238	0,3	2,3	9,9
p75	226	208	1,7	10,7	10	258	252	0,41	5,3	17,4
p95	253	233	5,9	20,2	23,9	284	275	0,64	12,3	30,2

Elaboração do autor.

A tabela 3 apresenta os resultados do ajuste dos modelos 1 e 2, que estimam as medidas de desempenho da escola que não podem ser atribuídas a família dos alunos. A tabela apresenta ainda o tamanho da amostra e a proporção da variância que é explicada com o perfil da família do aluno (modelo 1), assim como o perfil das famílias da escola (modelo 2). Os resultados mostram que as famílias explicam uma fração importante da variação dos resultados do aluno e da escola.

15. Apresentamos os resultado do desempenho com a seguinte notação: *mt*: nota de matemática; *lp*: nota de língua portuguesa; *ia*: indicador de atraso escolar; *rp*: taxa de reprovação; *ab*: taxa de abandono (*ab*).

TABELA 3
Sumário dos resultados dos modelos (1) e (2)

	Modelo 1				Modelo 2			
	R2-intra		#aluno ¹		R2		#escola ¹	
	F1	F2	F1	F2	F1	F2	F1	F2
Mt	0,106	0,081	2074	2040	0,52	0,43	41	33
Lp	0,124	0,11	2074	2040	0,529	0,394	41	33
la	0,571	0,714	2524	2720	0,367	0,426	41	33
Rp	-	-	-	-	0,197	0,157	39	31
Ab	-	-	-	-	0,218	0,249	39	31

Elaboração do autor.

Nota: ¹ Contagem em milhares.

Na falta de uma função de perda que pondere homogeneidade *versus* complexidade, apresentamos na tabela 4 os resultados para diversos indicadores calculados para partições com número de grupos crescentes, $ng = 2...15$ grupos. Consideramos os seguintes indicadores: *i*) BIC;¹⁶ *ii*) razão entre a variância intragrupos e a variância total (vi/vt); *iii*) para as escolas da primeira etapa do ensino fundamental, o índice proposto por Dunn¹⁷ (dun); e *iv*) índice proposto por Davies-Bouldin¹⁸ (db).

O BIC penaliza a complexidade considerando apenas o efeito da imprecisão na estimativa dos parâmetros, subestimando o custo da complexidade. A razão (vi/vt) é um indicador de fragmentação desnecessária, pois particionar um conjunto de dados que tem distribuição uniforme não reduz a quantidade (vi/vt). As outras medidas, (dun) e (db), foram calculadas apenas para as escolas da etapa F1. A estatística (db) mede o grau de proximidade relativa entre os grupos por meio da razão $k(i,j) = (d(i) + d(j)) / d(i,j)$, em que $d(i)$ é a distância média das escolas do grupo (i), e ($d(i,j)$) a distância entre os centroides dos grupos. A estatística (db) é a média do maior $k(i,j)$, e, portanto, quanto menor o índice db , mais os grupos estão dispersos. A estatística (du) é uma medida de dispersão para o pior caso.

A tabela 4 mostra, para as duas etapas do ensino fundamental, a redução progressiva do BIC e do vi/vt , o que indica que a fragmentação aumenta o grau de similaridade

16. $BIC = \ln(RSS/n) + k \ln(n)/n$ onde $k = 5 * ng$ é o número de parâmetros.

17. $Dun = \min_j \{d(i,j)\} / \max_i \{d(i)\}$.

18. $db = \sum_j \max_i \{(d(i) + d(j)) / d(i,j)\}$.

das escolas no grupo.¹⁹ No caso da etapa 1 do ensino fundamental, temos um resultado oscilante para o índice de Dunn e um valor mínimo para o índice *db*. Uma análise preliminar de algumas destas partições indicou que as partições com nove e com seis grupos forneciam resultados interessantes. Apresentamos os resultados para nove grupos a seguir, e no apêndice A estão os resultados para seis grupos.

TABELA 4
Características da partição segundo número de grupos

Ng	Fundamental 1				Fundamental 2	
	BIC	vi/vt	dun	Db	BIC	vi/vt
3	1,02	0,62	0,96	1,65	1,11	0,65
4	0,93	0,56	0,93	1,56	1,01	0,59
5	0,82	0,51	0,86	1,53	0,9	0,53
6	0,73	0,46	0,77	1,49	0,81	0,48
7	0,66	0,43	0,7	1,44	0,76	0,46
8	0,61	0,41	0,76	1,47	0,71	0,43
9	0,58	0,4	0,72	1,51	0,65	0,41
10	0,53	0,38	0,64	1,5	0,62	0,39
12	0,45	0,35	0,61	1,53	0,55	0,37
15	0,37	0,32	0,67	1,49	0,46	0,34

Elaboração do autor.

A tabela 5 mostra a média de cada uma das medidas de desempenho – centroide – de cada grupo, o número de escolas no grupo e uma medida do desvio-padrão da distância das escolas ao centroide de cada grupo que mede o “volume” do grupo. Nas colunas seguintes, apresentamos os valores do centroide utilizando uma notação que ajuda a interpretar e denominar cada grupo, e finalmente um “nome” que sumariza o perfil deste grupo, assim como um mnemônico para identificá-lo. Lembramos que as medidas de desempenho estão padronizadas, e, portanto, a média de cada grupo é o desvio da média geral em uma distribuição de variância unitária, e, portanto, uma indicação da significância estatística do desvio com a média do conjunto das escolas.

19. Fragmentar um conjunto que tenha distribuição uniforme não reduz a relação (*vi/vt*).

TABELA 5
Fundamental 1: centroides da partição com nove grupos

	mt	lp	id	rp	ab	#esc	%esc	DP	mt	lp	id	rp	ab	Nome
g5	1,88	1,82	-0,28	-0,68	-0,47	2985	0,08	1,615	A	A	b	b	b	pa
g2	0,44	0,46	-0,95	-0,84	-0,5	5562	0,14	1,152	a	a	b	b	b	pm
g9	-1,43	-1,44	0,06	0,14	-0,01	3982	0,1	1,439	B	B	0	a	0	pb
g4	0,57	0,59	0,1	0,15	-0,07	7209	0,19	1,104	a	a	0	a	0	sa
g3	-0,12	-0,11	1,78	-0,07	-0,12	2871	0,07	1,612	b	b	A	0	b	at
g6	-0,19	-0,19	-0,15	-0,09	1,89	1560	0,04	1,62	b	b	b	0	A	ab
g8	-0,34	-0,37	0,52	1,83	-0,02	3675	0,1	1,559	b	b	a	A	0	rp
g7	-0,73	-0,79	1,21	1,3	2,59	1017	0,03	2,08	b	b	A	A	A	ex
g1	-0,43	-0,43	-0,36	-0,3	-0,21	9541	0,25	1,049	b	b	b	b	b	ou

Elaboração do autor.

Notação: B (<-1);baixa: b (-1,-1); < média: a (-1,1); > média A (>1); alta.

Os resultados demonstram o seguinte.

- 1) O tamanho dos grupos varia entre 3% e 25% do total das escolas. Isto indica que a partição não identificou grupos raros. Lembramos que, se todos os grupos tivessem o mesmo tamanho, seriam cerca de 11% da amostra.
- 2) Os resultados do aprendizado – notas de matemática e língua portuguesa – são similares em todos os grupos, sugerindo que as escolas tendem a não ter especialidade destes conteúdos.
- 3) O grupo g_5 “nota alta” é constituído das melhores escolas, apresentam nota e inclusão alta, ou seja, valores abaixo da média para reprovação, abandono e atraso escolar.
- 4) O grupo g_7 “exclui” é composto pelas piores escolas, baixa *performance* e inclusão baixa nas três medidas.
- 5) Os grupos g_2 (g_9) “nota média (baixa)” é formado pelas escolas intermediárias, nota baixa (e muito baixa), e medidas de inclusão baixas ou nulas.
- 6) O grupo g_4 “seleciona” apresenta um padrão curioso, nota alta e inclusão baixa. São escolas que utilizam algum mecanismo para selecionar os melhores alunos, e não manter os alunos com pior aprendizado.
- 7) Os grupos g_3 , g_6 e g_8 têm baixo aprendizado e se especializam em uma das formas de exclusão, atraso ou reprovação ou abandono elevado.
- 8) Finalmente, o grupo g_1 é o grupo residual.

Notável nesse resultado é a especialização dos grupos. Os grupos g_5 , g_2 , g_9 se distinguem por seu aprendizado; os grupos g_3 , g_6 e g_8 se diferenciam por apresentarem uma alta taxa de um dos tipos de exclusão: atraso, reprovação e abandono. O grupo g_7 pode ser interpretado como o das piores escolas, pois apresenta uma alta taxa de exclusão nas três medidas. Finalmente, o grupo g_4 parece ter um padrão de não garantir que os alunos com pior aprendizado terminem a etapa do ensino fundamental.

TABELA 6
Fundamental 2: centroides da partição com nove grupos

	mt	lp	at	rp	ab	#esc	%esc	dp	mt	lp	at	rp	ab	pa	Nota alta
g3	1,61	1,53	-0,1	-0,45	-0,37	2921	0,1	1,525	A	A	0	b	b	b	Nota alta
g7	-0,7	-0,69	-0,41	-0,57	-0,25	6015	0,2	1,152	b	b	b	b	b	b	Nota média
g9	-1,65	-1,76	0,23	0,11	0,09	2114	0,07	1,568	B	B	a	a	0	pb	Nota baixa
g5	0,76	0,78	0,19	1,49	-0,08	2572	0,08	1,522	a	a	a	A	0	sa	seleção
g6	-0,1	-0,11	2,09	0,09	-0,02	1818	0,06	1,754	0	0	A	0	0	at	atrasa
g8	-0,23	-0,23	0,45	0,19	2,32	1829	0,06	1,871	b	b	a	a	A	ab	abandono
g4	-0,56	-0,61	0	1,38	-0,09	3425	0,11	1,436	b	b	0	A	0	rp	reprova
g2	0,33	0,39	-1,04	-0,64	-0,64	3735	0,12	1,289	a	a	B	b	b	in	inclui
g1	0,24	0,3	0,11	-0,33	0,07	6308	0,21	1,058	a	a	a	b	0	ou	Outros

Elaboração do autor.

Notação: B (<-1); baixa: b(-1,-1); < média: 0(-1,-1); = média: a(-1,1); > média: A(>1); alta.

Os resultados para a segunda etapa do ensino fundamental seguem em geral o mesmo padrão dos resultados para a etapa 1. A principal diferença é que no F2 não é identificado o padrão das escolas com baixa inclusão, por exemplo, grupo “excluí” do F1, mas é identificado o padrão das escolas que “inclui”, pois têm notas maiores e baixa exclusão, em particular o indicador de atraso.

A tabela 7 apresenta, para as escolas com as duas etapas do ensino fundamental, o cruzamento das partições das escolas segundo cada uma das etapas. Apresenta também na coluna $\hat{n}F2$ a proporção de escolas no grupo que não tem a segunda etapa e nas linhas $F1\&F2$, $\hat{n}F1$ e total, a mesma proporção, respectivamente, para o conjunto das escolas que têm as duas etapas, para as escolas que só têm a etapa F2 e, finalmente, para o conjunto das escolas.

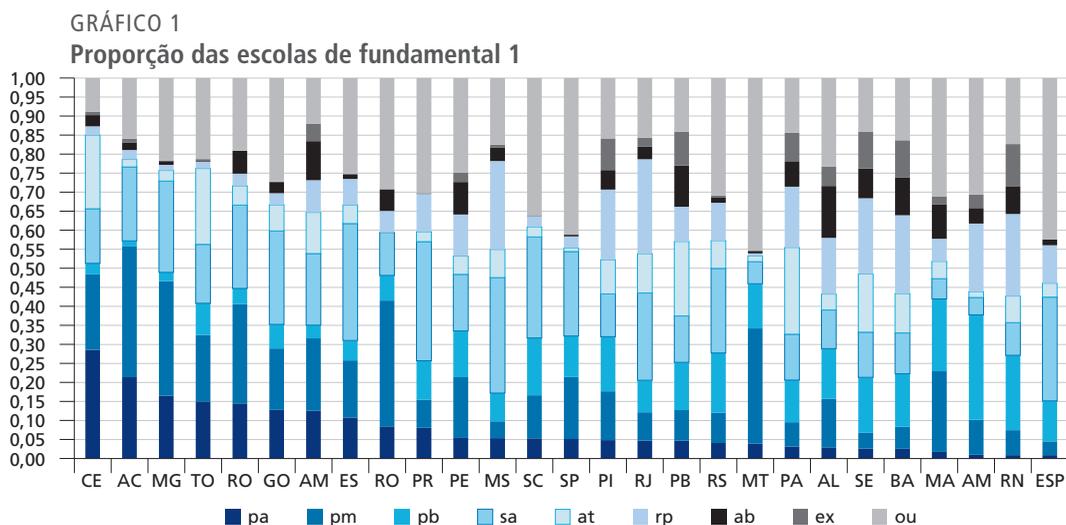
Os resultados mostram que, para a maioria dos cruzamentos, as escolas não permanecem no grupo de mesma denominação nas duas etapas. Notável exceção é o caso das boas escolas no fundamental 1 (pa). Neste caso, 33% destas escolas são identificadas como boas escolas na segunda etapa. Nos demais casos, ocorre troca de denominação mais frequente ou os resultados estão diluídos. O aumento do número de grupos torna as fronteiras dos grupos mais ambíguas, o que favorece a ocorrência desse resultado. Vale mencionar que no apêndice A estão apresentados os resultados para a partição com seis grupos. Neste caso, as escolas tendem a permanecer nos grupos de mesma denominação nas duas etapas do ensino fundamental.

A partição das escolas podem ter múltiplas finalidades para o gestor que avalia o conjunto das escolas. O desempenho das secretarias estaduais de educação pode ser avaliado medindo a incidência destes grupos em cada estado, por etapa do ensino fundamental. O gráfico 1 apresenta, por estado, a proporção das escolas do fundamental 1, em cada um dos grupos da partição em nove grupos. Por conveniência, os estados estão ordenados segundo a incidência de escolas de aprendizado alto. Chama atenção que a incidência dos diversos grupos de escolas seja tão diferenciada.

TABELA 7
Proporção das escolas da partição segundo F2 por grupo do F1

F1	Fundamental 2												F2	rF2	
	g3	g7	g9	g5	g6	g4	g8	g2	g1	ou	g1	ou			
g5	pa	0,33	0,08	0,02	0,08	0,06	0,04	0,01	0,2	0,18	0,18	0,2	0,18	1	0,63
g2	pm	0,12	0,23	0,04	0,06	0,05	0,06	0,03	0,24	0,17	0,17	0,24	0,17	1	0,62
g9	pb	0,03	0,22	0,13	0,08	0,06	0,18	0,06	0,11	0,13	0,13	0,11	0,13	1	0,6
g4	sa	0,16	0,12	0,03	0,16	0,05	0,1	0,03	0,09	0,25	0,25	0,09	0,25	1	0,63
g3	at	0,11	0,12	0,07	0,14	0,13	0,11	0,05	0,09	0,18	0,18	0,09	0,18	1	0,61
g8	rp	0,05	0,09	0,06	0,24	0,08	0,23	0,08	0,06	0,12	0,12	0,06	0,12	1	0,6
g6	ab	0,04	0,13	0,08	0,08	0,06	0,14	0,22	0,09	0,16	0,16	0,09	0,16	1	0,61
g7	ex	0,03	0,07	0,06	0,14	0,13	0,17	0,27	0,02	0,11	0,11	0,02	0,11	1	0,63
g1	ou	0,06	0,25	0,08	0,07	0,04	0,12	0,03	0,15	0,21	0,21	0,15	0,21	1	0,63
F1&F2		0,11	0,17	0,06	0,11	0,06	0,12	0,05	0,13	0,18	0,18	0,13	0,18	1	0,62
rF1		0,08	0,22	0,07	0,06	0,06	0,1	0,07	0,11	0,23	0,23	0,11	0,23	1	0
Total		0,1	0,2	0,07	0,08	0,06	0,11	0,06	0,12	0,21	0,21	0,12	0,21	1	0,44

Elaboração do autor.

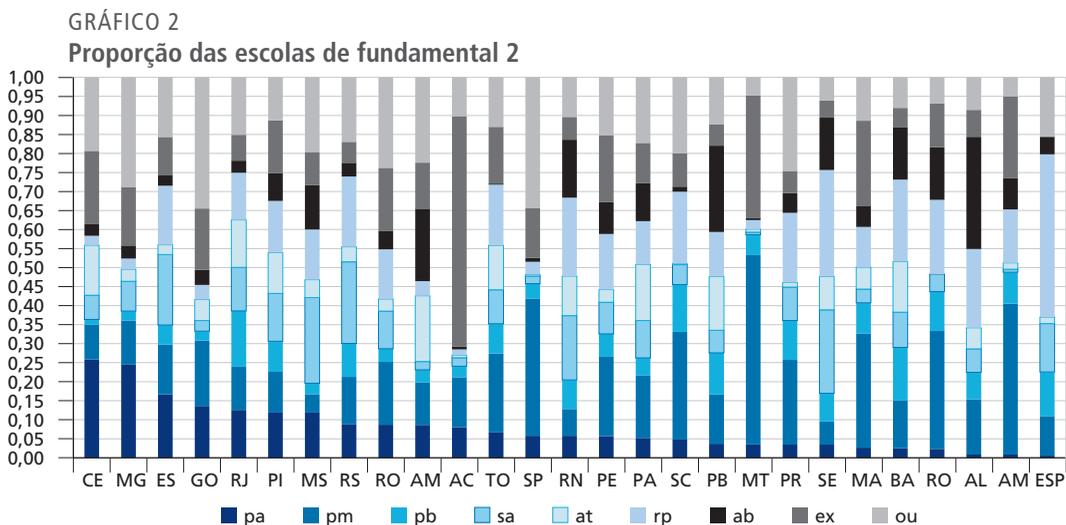


O desempenho do Ceará é notável, especialmente quando comparado com o estado de São Paulo. Aquele é o estado do Nordeste com a menor renda *per capita*, recursos e outras desvantagens comparativas, no entanto, para o fundamental 1, 25% das suas escolas são do grupo de nota alta e cerca de 20% está nos grupos de nota média e baixa. No caso de São Paulo, os números são, respectivamente, de 5% e 25%. O estado de Minas Gerais tem um desempenho comparável ao do estado do Ceará. Chama também atenção a alta incidência do grupo de escolas com nota acima da média e com alta taxa de exclusão – mais que 20% –, grupo este denominado por “selecionar”. Esse grupo está mais presente nos estados do Paraná, do Espírito Santo, de Mato Grosso do Sul e de Santa Catarina.

Os estados do Rio de Janeiro, de Mato Grosso do Sul, do Rio Grande do Norte, da Bahia e de Sergipe apresentam alta incidência de escolas no grupo com alta taxa de reprovação. O estado do Rio de Janeiro é singular, por ser um estado com alta renda *per capita* e ainda assim ter escolas públicas com alta incidência dos padrões considerados ruins.

Repetimos o exercício para as escolas do fundamental 2. Os resultados são similares. O estado do Ceará também apresenta alta incidência de escolas de alto aprendizado – quase 30% – e São Paulo apresenta a mesma incidência de 5%. Estes resultados, tal como outros já documentados, mostram que algo precisa ser aprendido

com as práticas da Secretaria de Educação do Ceará, bem como observar os estados com baixo desempenho e cujas práticas precisam ser evitadas.



Esses mesmos resultados estão tabulados na tabela A.1 do apêndice e podem motivar outras indagações que estão fora do escopo deste estudo. Quais são as práticas de gestão dos estados que apresentam melhor desempenho? Ou o que os estados fazem para ajudar as escolas da sua área que apresentam desempenho pior?

4.1 Explicando a partição das escolas segundo as variáveis observadas e o efeito municipal

A heterogeneidade do desempenho das escolas motiva particioná-las em grupos de escolas similares em cada grupo, e dissimilares entre os grupos. A partição selecionada agrupou as escolas segundo as dimensões de aprendizado e o grau de inclusão em nove categorias. Estas revelam os padrões de funcionamento como consequência de práticas pedagógicas e qualificação dos professores, disponibilidade de infraestrutura e gestão da escola, orientação e apoio da secretaria municipal de educação e das políticas estaduais na área da educação. Algumas destas características x (e) são observadas pelos questionários

respondidos junto com a Prova Brasil,²⁰ que caracterizam o perfil do diretor, da escola e dos seus professores.²¹ Outras características não são observadas, mas são comuns às escolas de um município que não são explicadas por $x(e)$. Esta quantidade tem o seu efeito medido com um efeito fixo por município $\alpha (m)$.

A seção 3.4 apresentou a função discriminante e os critérios para avaliar a capacidade de (x) e do efeito fixo municipal distinguirem as escolas do grupo G_i das escolas de um outro grupo G_j por meio das funções $p(x, i, j)$ e $q(x, i, j)$, ou, avaliando a capacidade de x distinguir as escolas do grupo G_i das demais escolas.

No primeiro caso, em que estamos avaliando a separação entre um par de grupos (i, j) , sabemos que estes são linearmente separáveis; no segundo caso, quando estamos testando se o grupo (i) é separável do grupo $NG(i) \cup_{k \neq i} G(k)$, a separabilidade não é garantida, pois o grupo $NG(i)$ não é necessariamente um conjunto convexo.

Vale comentar que as funções $p(x)$ e $q(x)$ são calculadas empiricamente e sujeitas a erro de medida. Por isto, estimamos para todas as situações a função discriminante utilizando o desempenho das escolas (r) . No caso em que estamos distinguindo apenas dois grupos, sabe-se que $p(r)$ é nulo. Isto não se verificou em muitos casos estudados, o que deve ser interpretado como uma imprecisão do método de cálculo das probabilidades de erro mencionadas. A função $p(x)$ mede a probabilidade de erro de classificação, e quanto maior este valor, menor a capacidade de x distinguir o grupo em análise.

Analisamos a partição com nove grupos, avaliando o efeito do desempenho das escolas (r) , e das 310 variáveis que caracterizam os professores, os diretores e a escola (x) , e o efeito adicional do efeito fixo municipal, calculando as funções $p(r)$, $p(x)$ e $q(x)$. Este exercício foi repetido para distinguir entre todos os pares $G(i)$, $G(j) \forall i, j$ e também para distinguir as escolas do grupo $G(i)$ das demais escolas. A tabela 9 apresenta os resultados de forma sintética, de modo que as probabilidades de erro de classificação para todos os pares de grupos (i, j) estão resumidas para cada grupo com o

20. São respondidas 74 perguntas sobre localização/infraestrutura da escola, 125 sobre a qualificação do diretor e suas práticas e 111 sobre a qualificação e as práticas dos professores, constituindo um total de 310 variáveis, em que as perguntas sobre os professores são convertidas em valores médios dos professores da escola e etapa do ensino fundamental.
21. Nesse caso, consideramos o professor típico da escola, medido com a média dos questionários respondidos pelos professores.

intervalo de variação de $p(r, i, j)$, $p(x, i, j)$ e $q(x, i, j)$, ou seja, é apresentado para cada grupo i o menor e o maior valor destas funções.

Como esperado, os resultados mostram que o erro de classificação entre pares de grupo ($p(r, i, j)$) é pequena ou nula, e o erro de classificação entre um grupo e os demais ($p(r, i)$) é substancialmente maior. As probabilidades de erro, $p(x, i)$ e $p(x, i, j)$, que utilizam a variabilidade das características observadas nas escolas – descritas com as 310 componentes do perfil da escola, do diretor e do professor – implicam uma probabilidade de erro muito elevada mesmo quando consideramos os erros de classificações entre pares de grupos. Este resultado indica que a variabilidade das características observadas tem pouca ou nenhuma capacidade de distinguir os padrões de desempenho das escolas. Finalmente, quando incorporamos o efeito fixo municipal no modelo de classificação, temos a função $q(x, i, j)$, que apresenta valores muito menores, indicando que os efeitos comuns a todas as escolas de um município são importantes e têm capacidade de explicar os padrões de desempenho. Na análise para a segunda etapa do ensino fundamental, os resultados são similares.

Vale lembrar que esse resultado não depende das características das famílias do município nem dos perfis dos professores, dos diretores e da escola. A partição dos grupos é explicada por características não observadas e comuns a todas as escolas do município, as quais podem ser atribuídas às práticas da secretaria de educação do município e do estado, o que enfatiza a importância da gestão municipal sobre o desempenho das escolas. A tabela A.1 do apêndice detalha os resultados da tabela 8, apresentando as probabilidades de erro de classificação para os 36 pares de grupos que podem ser formados com uma partição de nove grupos.²²

22. O emprego da mesma metodologia que utiliza o perfil da família para explicar a classificação mostrou, como esperado, que a probabilidade $p(F, i, j)$ apresenta valor elevado, que o perfil da família não explica o tipo da escola.

TABELA 8
Probabilidade de erro de classificação segundo grupo e variável explicativa

i	Fundamental 1									Fundamental 2								
	g1	g2	g3	g4	g5	g6	g7	g8	g9	g1	g2	g3	g4	g5	g6	g7	g8	g9
p (r,i)	0,28	0,18	0,14	0,29	0,03	0,06	0,08	0,12	0,11	0,34	0,12	0,12	0,15	0,13	0,06	0,17	0,04	0,11
maxj p (r, i, j)	0,07	0,06	0,07	0,07	0,06	0,07	0,07	0,06	0,05	0,06	0,05	0,06	0,06	0,06	0,07	0,07	0,07	0,06
minj p (r, i, j)	0	0	0,02	0,01	0,01	0	0	0	0,01	0,02	0,02	0,02	0,03	0	0,01	0	0,01	0,03
p (x,i)	0,45	0,45	0,46	0,43	0,41	0,45	0,34	0,41	0,45	0,43	0,45	0,4	0,44	0,4	0,42	0,44	0,35	0,43
maxj p (x, i, j)	0,42	0,43	0,47	0,42	0,4	0,42	0,38	0,42	0,47	0,41	0,41	0,4	0,41	0,41	0,44	0,41	0,41	0,44
minj p (x, i, j)	0,31	0,35	0,35	0,26	0,33	0,32	0,26	0,34	0,35	0,34	0,34	0,32	0,35	0,35	0,35	0,35	0,32	0,36
Q (x,i)	0,28	0,28	0,21	0,27	0,21	0,2	0,13	0,2	0,25	0,26	0,24	0,19	0,19	0,2	0,22	0,25	0,2	0,21
maxj q (x, i, j)	0,24	0,23	0,17	0,23	0,17	0,14	0,15	0,17	0,24	0,19	0,2	0,18	0,18	0,18	0,12	0,2	0,14	0,17
minj q (x, i, j)	0,09	0,07	0,1	0,08	0,06	0,08	0,06	0,08	0,08	0,11	0,1	0,05	0,07	0,09	0,09	0,1	0,07	0,05

Elaboração do autor.

5 CONCLUSÃO

Este estudo propôs uma metodologia para identificar grupos de escolas públicas do ensino fundamental que possuem características similares de aprendizado e de inclusão dos alunos. Admite-se que estes grupos revelam padrões de funcionamento e de características das escolas que podem não ser observáveis com o conjunto de informação disponível. A análise foi conduzida procurando mitigar o efeito das diferenças do perfil familiar entre os alunos e as escolas. Os padrões foram identificados utilizando a análise de agrupamento que constrói grupos de escolas similares segundo a sua medida de desempenho, e selecionada uma partição das escolas em nove grupos. Em seguida, foi avaliada a capacidade explicativa das características observadas das escolas/diretores/professores coletadas junto com a Prova Brasil.

Esse exercício foi implementado com os dados da Prova Brasil e do Censo Escolar para o ano de 2013. O grau de especialização das escolas nas duas etapas do ensino fundamental motivou a realização do exercício para cada uma das etapas. Os resultados para esse ano sugeriram a existência de nove grupos de escolas: *i)* nota alta e inclusão média; *ii)* nota e exclusão média; *iii)* nota baixa e atraso média; *iv)* nota média com alta taxa de reprovação; *v)* atraso escolar alto; *vi)* taxa de abandono alta; *vii)* taxa de reprovação alta; *viii)* nota média e inclusão média; e *ix)* outros. Naturalmente, estas características precisam ser avaliadas e validadas por especialistas, mas esta distinção das escolas é no mínimo curiosa, especialmente porque constituem grupos separáveis.

Outro resultado interessante é identificar as variáveis que explicam a partição das escolas. Verificamos que a localização da escola (município) é capaz de explicar a partição e o perfil da escola/diretor/professor não é capaz. Este resultado enfatiza a importância das secretarias municipais e estaduais de educação no resultado escolar.

O exercício deixa em aberto algumas indagações. As medidas de desempenho utilizadas são suficientemente completas no sentido de mensurar as principais características observáveis das escolas,²³ e também vale perguntar se a partição escolhida deixa de identificar algum grupo de escolas considerado relevante pelos especialistas.

23. Cabe perguntar se a forma como o perfil da família/diretor/professor/escola foi medido pode ser aperfeiçoada.

Uma extensão imediata desse exercício é avaliar a dinâmica dos padrões de funcionamento das escolas, construindo partições das escolas para os demais anos – 2007/2009/2011/2015 – em que existem dados disponíveis, e avaliando a probabilidade de transição intertemporal dos padrões identificados, em particular, discutindo o grau de estabilidade da escola em um certo padrão e os seus condicionantes.

REFERÊNCIAS

BENHABIB, J.; SPIEGEL, M. M. Human capital and technology diffusion. *In*: AGHION, P.; DURLAUF, S. (Eds.). **Handbook of economic growth**. 1 ed. v. 1. Elsevier, 2005. cap. 13. p. 935-966.

BOF, A. M. **A aprendizagem dos alunos e os desafios do PNE**. Brasília: Inep, 2016. (PNE em Movimento 5).

BRUNS, B.; EVANS, D.; LUQUE, J. **Achieving world-class education in Brazil: the next agenda**. Washington: The World Bank, 2012.

FORTIN, L. *et al.* Typology of students at risk of dropping out of school: description by personal family and school factors. **European Journal of Psychology of Education**, v. 21, n. 4, p. 363-383, 2006.

FRYER JUNIOR, R. G. **The production of human capital in developed countries: evidence from 196 randomized field experiment**. Cambridge: National Bureau of Economic Research, mar. 2016. (NBER Working Paper, n. 22130).

HANUSHEK, E. The economics of schooling: production and efficiency in public schools. **Journal of Economic Literature**, v. 24, n. 3, p. 1141-1177, set. 1986.

_____. Why standard measures of human capital are misleading. **Journal of Economic Policy**, v. 37, n. 2, p. 22-39, 2015.

HANUSHEK, E.; WOESSMANN, L. Schooling, educational achievement, and the Latin American growth puzzle. **Journal of Development Economics**, v. 99, n. 2, p. 497-512, 2012.

HATTIE, J. **Schools like mine: cluster analysis of New Zealand schools**. Auckland: University of Auckland, 2002. (Technical Report 14, project asTTle).

INEP – INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Panorama da educação: destaques do Education at Glance**. Brasília: Inep, 2016.

MEDEIROS, M.; OLIVEIRA, L. F. B. Desigualdades regionais em educação: potencial de convergência. **Revista Sociedade e Estado**, v. 29, n. 2, maio/ago. 2014.

OECD – ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **OECD, country note:** program for international student assessment (PISA) results from PISA 2012. 2012.

POLO, F.; SANTOS, D. D. **As escolas enfrentam trade-offs entre o desenvolvimento de habilidades cognitivas e socioemocionais?** Uma abordagem por análise envoltória de dados. 2016. Mimeografado.

PRADHAN R. P.; HALL, J. H. Causality between the pillars of development and economic growth: a panel VAR application. **International Journal of Services Economics and Management**, v. 6, n. 1, 2014.

RANIS, G.; STEWART, F. **Dynamic links between the economy and human development.** Nova Iorque: UN DESA, 2005. (DESA Working Paper, n. 8).

SILVA FILHO, G. A. **Higher salaries, more teaching, better performance?** The effects of the introduction of the minimum salary for teachers in Brazil. Tese (Doutorado) – Fundação Getúlio Vargas, 2016.

SIMÕES, A. A. **As metas de universalização da educação básica no Plano Nacional de Educação:** o desafio do acesso e a evasão dos jovens de baixa renda no Brasil. Brasília: Inep, 2016. (PNE em Movimento 4).

SOARES, S.; SÁTYRO, N. **O impacto da infraestrutura escolas na taxa de distorção idade-série das escolas brasileiras do ensino fundamental – 1998 a 2005.** Rio de Janeiro: Ipea, 2008. (Texto para Discussão, n. 1338).

TABOSA, J. S. **Impacto dos gastos com educação e saúde no crescimento econômico.** Ipea, 2017. Mimeografado.

BIBLIOGRAFIA COMPLEMENTAR

ALTINOK, N.; ANGRIST, N.; PATRINOS, H. A. **Global data set on education quality (1965-2015).** The World Bank, jan. 2018. (Policy Research Working Paper, n. 8314).

DIAZ, M. D. M. Qualidade do gasto público municipal em ensino fundamental no Brasil. **Revista de Economia Política**, v. 32, n. 1 (126), p. 128-141, jan./mar. 2012.

DOURADO, L. F. **Plano Nacional de Educação:** política de Estado para a educação brasileira. Brasília: Inep, 2016. (PNE em Movimento 1).

MENEZES FILHO, N.; PAZELLO, E. **Does money in schools matter?** Evaluating the effects of FUNDEF on wages and test scores in Brazil. 2004. Mimeografado.

PISA – PROGRAM FOR INTERNATIONAL STUDENT ASSESSMENT. **Brazil – country note: results from PISA 2012.** 2012.

RUGGIERO, J. Efficiency of educational production: an analysis of New York School Districts. **The Review of Economics and Statistics**, v. 78, n. 3, p. 499-509, ago. 1996.

SILVA FILHO, G. A.; CARVALHO, M. V. **O efeito da formação inicial do professor sobre o desempenho escolar em matemática nos anos iniciais do ensino fundamental.** Brasília: Inep, 2017. (Texto para Discussão, n. 43).

APÊNDICE A

TABELA A.1
Incidência de escolas por estado, por grupo e por etapa do ensino fundamental

	Fundamental 1									Fundamental 2								
	pa	pm	pb	sa	at	rp	ab	ex	ou	pa	pm	pb	sa	at	rp	ab	in	ou
RO	0,15	0,26	0,04	0,22	0,05	0,03	0,06	0	0,19	0,09	0,17	0,03	0,1	0,03	0,13	0,05	0,17	0,24
AC	0,21	0,34	0,01	0,19	0,02	0,02	0,02	0,01	0,16	0,08	0,13	0,03	0,02	0,01	0,01	0,01	0,61	0,1
AM	0,13	0,19	0,03	0,19	0,11	0,08	0,1	0,05	0,12	0,09	0,11	0,03	0,02	0,17	0,04	0,19	0,12	0,22
RO	0,08	0,33	0,07	0,11	0	0,06	0,06	0	0,29	0,02	0,31	0,1	0,05	0	0,2	0,14	0,11	0,07
PA	0,03	0,06	0,11	0,12	0,23	0,16	0,07	0,07	0,14	0,05	0,16	0,05	0,1	0,15	0,11	0,1	0,11	0,17
AM	0,01	0,09	0,28	0,05	0,02	0,18	0,04	0,04	0,31	0,01	0,4	0,08	0,01	0,02	0,14	0,08	0,21	0,05
TO	0,15	0,17	0,08	0,15	0,2	0,02	0	0	0,21	0,07	0,21	0,08	0,09	0,12	0,16	0	0,15	0,13
MA	0,02	0,21	0,19	0,05	0,05	0,06	0,09	0,02	0,31	0,03	0,3	0,08	0,04	0,06	0,11	0,05	0,23	0,11
PI	0,05	0,13	0,14	0,11	0,09	0,18	0,05	0,08	0,16	0,12	0,11	0,08	0,13	0,11	0,14	0,07	0,14	0,11
CE	0,29	0,2	0,03	0,14	0,19	0,02	0,03	0,01	0,09	0,26	0,09	0,01	0,06	0,13	0,02	0,03	0,19	0,19
RN	0,01	0,07	0,2	0,09	0,07	0,22	0,07	0,11	0,17	0,06	0,07	0,08	0,17	0,1	0,21	0,15	0,06	0,1
PB	0,05	0,08	0,13	0,12	0,2	0,09	0,11	0,09	0,14	0,04	0,13	0,11	0,06	0,14	0,12	0,23	0,06	0,12
PE	0,05	0,16	0,12	0,15	0,05	0,11	0,09	0,02	0,25	0,06	0,21	0,06	0,08	0,03	0,15	0,08	0,18	0,15
AL	0,03	0,13	0,13	0,1	0,04	0,15	0,14	0,05	0,23	0,01	0,15	0,07	0,06	0,06	0,21	0,29	0,07	0,09
SE	0,03	0,04	0,14	0,12	0,15	0,2	0,08	0,1	0,14	0,03	0,06	0,07	0,22	0,09	0,28	0,14	0,04	0,06
BA	0,03	0,06	0,14	0,11	0,1	0,21	0,1	0,1	0,16	0,03	0,13	0,14	0,09	0,13	0,22	0,14	0,05	0,08
MG	0,17	0,3	0,02	0,24	0,03	0,02	0,01	0	0,22	0,25	0,11	0,02	0,08	0,03	0,03	0,03	0,15	0,29
ES	0,11	0,15	0,05	0,31	0,05	0,07	0,01	0	0,25	0,17	0,13	0,05	0,19	0,03	0,16	0,03	0,1	0,16
RJ	0,05	0,07	0,08	0,23	0,1	0,25	0,03	0,02	0,16	0,13	0,11	0,15	0,11	0,12	0,12	0,03	0,07	0,15
SP	0,05	0,16	0,11	0,22	0,01	0,03	0	0	0,41	0,06	0,36	0,04	0,02	0	0,03	0,01	0,13	0,34
PR	0,08	0,07	0,1	0,31	0,03	0,1	0	0	0,3	0,03	0,22	0,1	0,09	0,01	0,18	0,05	0,06	0,25
SC	0,05	0,11	0,15	0,27	0,03	0,03	0	0	0,36	0,05	0,28	0,12	0,05	0	0,19	0,01	0,09	0,2
RS	0,04	0,08	0,16	0,22	0,07	0,1	0,01	0	0,31	0,09	0,12	0,09	0,21	0,04	0,18	0,04	0,06	0,17
MS	0,05	0,04	0,07	0,3	0,07	0,23	0,03	0,01	0,18	0,12	0,05	0,03	0,23	0,05	0,13	0,12	0,09	0,2
MT	0,04	0,3	0,12	0,06	0,01	0,01	0,01	0	0,45	0,03	0,5	0,05	0,01	0,01	0,02	0,01	0,32	0,05
GO	0,13	0,16	0,06	0,25	0,07	0,03	0,03	0	0,27	0,14	0,17	0,02	0,03	0,06	0,04	0,04	0,16	0,34
ES	0,01	0,04	0,11	0,27	0,04	0,1	0,02	0	0,42	0,01	0,1	0,12	0,13	0,02	0,43	0,05	0	0,16

Elaboração do autor.

TABELA A.2
Perfil da família: N (contagem) I (indicadora)

	Tipo	Descrição
1	N	Idade
2	N	Equipamentos (tv, rádio, dvd, geladeira, geladeira duplex, freezer, máquina de lavar)
3	I	Tem carro em casa
4	I	Tem computador
5	N	Quantas pessoas vivem na casa
6	I	Mora com a mãe
7	N	Escolaridade da mãe
8	I	Sua mãe sabe ler
9	I	Vê sua mãe lendo
10	I	Mora com o pai
11	N	Escolaridade do pai
12	I	Seu pai sabe ler
13	I	Vê o pai lendo
14	N	Frequência com que os pais vão à reunião de pais
15	N	Incentivo dos pais (estudar, dever de casa, ler, ir à escola, conversam sobre a escola)
16	N	Frequência que lê (jornal, livro, revista, quadrinhos, notícias, biblioteca, cinema, espetáculo)
17	N	Tempo que assiste tv ou trabalha em casa
18	I	Trabalha fora de casa
19	N	Idade em que entrou na escola
20	I	Faz o dever de casa de português
21	I	Faz o dever de casa de matemática
22	I	Lê livros da biblioteca
23	I	Pretende trabalhar

Elaboração do autor.

TABELA A.3

Fundamental 1: centroides da partição com seis grupos

	mt	lp	id	rp	ab	frq	DP	mt	lp	id	rp	ab		
g4	1,52	1,5	-0,41	-0,59	-0,41	5245	1,58	A	A	b	b	b	pa	Nota alta
g2	0,3	0,32	0,63	0,04	-0,05	8568	1,346	a	a	a	0	0	pm	Nota média
g6	-1,05	-1,06	-0,06	-0,08	-0,07	7990	1,379	B	B	0	0	0	pb	Nota baixa
g3	-0,44	-0,48	0,78	1,77	0,04	4306	1,75	b	b	a	A	0	rp	Reprova
g5	-0,46	-0,49	0,51	0,5	2,44	1963	2,017	b	b	a	a	A	ab	Abandono
g1	0,04	0,05	-0,74	-0,53	-0,33	10330	1,168	0	0	b	b	b	ou	Outros

Elaboração do autor.

Notação: B (< -1): baixa: b (-1,-.1): < média: 0: (-.1, .1): = média: a (.1,1): > média A (> 1): alta.

TABELA A.4

Fundamental 2: centroides da partição com seis grupos

	mt	lp	id	rp	ab	frq	Dp	mt	lp	id	rp	ab		
g2	1,27	1,25	-0,18	-0,24	-0,28	5644	1,566	A	A	b	b	b	Pa	Nota alta
g6	-1,21	-1,26	-0,14	-0,06	-0,08	5618	1,489	B	B	0	0	0	Pb	Nota baixa
g4	-0,02	-0,02	1,65	0	-0,05	2927	1,682	0	0	A	0	0	At	Atrasa
g3	-0,02	-0,04	0,08	1,58	-0,07	4466	1,603	0	0	0	A	0	Rp	Reprova
g5	-0,2	-0,19	0,4	0,17	2,1	2307	1,867	B	b	a	a	A	Ab	Abandono
g1	-0,06	-0,01	-0,48	-0,58	-0,3	9775	1,221	0	0	b	b	b	ou	Outros

Elaboração do autor.

Notação: B (< -1): baixa: b (-1, -.1): < média: 0: (-.1, .1): = média: a (.1, 1): > média A (> 1): alta.

TABELA A.5

Cruzando escolas do fundamental 1 e 2 para partição com seis grupos

	Fundamental 1								
	Pa	Pb	At	Rp	Ab	Ou			
F1							1		
pa		g4	0,43	0,07	0,1	0,08	0,03	0,29	1
pm		g2	0,27	0,13	0,12	0,18	0,05	0,25	1
pb		g6	0,1	0,27	0,08	0,17	0,06	0,33	1
rp		g3	0,12	0,14	0,13	0,37	0,1	0,13	1
ab		g5	0,1	0,15	0,11	0,2	0,31	0,14	1
ou		g1	0,19	0,19	0,07	0,11	0,04	0,41	1

Elaboração do autor.

TABELA A.6
Probabilidade de erro de classificação segundo pares de grupos da partição $g = 9$

GA	GB	p(r, i, j)		p(D, i, j)		q(D, i, j)		GA	GB	p(i, i, j)		p(D, i, j)		q(D, i, j)	
		f1	f2	f1	f2	f1	f2			f1	f2	f1	f2	f1	f2
1	2	0,05	0,04	0,41	0,38	0,23	0,18	3	7	0,05	0,05	0,35	0,39	0,1	0,1
1	3	0,07	0,04	0,41	0,39	0,14	0,18	3	8	0,05	0,02	0,41	0,32	0,17	0,07
1	4	0,05	0,06	0,42	0,37	0,23	0,14	3	9	0,04	0,04	0,47	0,36	0,14	0,05
1	5	0,06	0,02	0,4	0,35	0,13	0,15	4	5	0,01	0,05	0,38	0,41	0,17	0,18
1	6	0,07	0,05	0,42	0,35	0,13	0,12	4	6	0,03	0,06	0,32	0,4	0,12	0,12
1	7	0	0,03	0,31	0,41	0,09	0,19	4	7	0,07	0,04	0,26	0,38	0,08	0,14
1	8	0,03	0,06	0,38	0,34	0,14	0,11	4	8	0,06	0,06	0,34	0,41	0,16	0,14
1	9	0,01	0,05	0,41	0,36	0,24	0,12	4	9	0,05	0,03	0,36	0,41	0,15	0,17
2	3	0,06	0,05	0,41	0,4	0,13	0,15	5	6	0,06	0,03	0,33	0,41	0,08	0,1
2	4	0,06	0,04	0,38	0,37	0,19	0,11	5	7	0,05	0	0,33	0,35	0,06	0,1
2	5	0,06	0,02	0,4	0,36	0,17	0,11	5	8	0,05	0,04	0,35	0,36	0,08	0,1
2	6	0,05	0,05	0,41	0,41	0,11	0,11	5	9	0,05	0,06	0,39	0,37	0,08	0,09
2	7	0,05	0,03	0,35	0,4	0,07	0,2	6	7	0,01	0,07	0,38	0,39	0,12	0,1
2	8	0	0,05	0,37	0,34	0,08	0,1	6	8	0	0,01	0,42	0,37	0,14	0,1
2	9	0,04	0,05	0,43	0,39	0,15	0,11	6	9	0,03	0,05	0,42	0,44	0,13	0,09
3	4	0,02	0,06	0,37	0,35	0,17	0,07	7	8	0,05	0,07	0,36	0,35	0,15	0,1
3	5	0,02	0,04	0,38	0,36	0,11	0,13	7	9	0,01	0,05	0,35	0,41	0,09	0,17
3	6	0,06	0,03	0,42	0,38	0,14	0,1	8	9	0,04	0,05	0,41	0,4	0,17	0,1

Elaboração do autor.

TABELA A.7
Probabilidade de erro de classificação segundo pares de grupos da partição $g = 6$

A	B	Fundamental 1			
		$p(r, l, j)$	$p(F, l, j)$	$p(D, l, j)$	$q(D, l, j)$
1	2	0,09	0,45	0,46	0,27
1	3	0,01	0,6	0,39	0,14
1	4	0,1	0,49	0,46	0,22
1	5	0,05	0,63	0,38	0,08
1	6	0,08	0,47	0,47	0,29
2	3	0,1	0,57	0,38	0,2
2	4	0,03	0,45	0,44	0,2
2	5	0,02	0,49	0,35	0,17
2	6	0,03	0,49	0,44	0,24
3	4	0,09	0,56	0,34	0,09
3	5	0,04	0,43	0,46	0,21
3	6	0,08	0,43	0,39	0,2
4	5	0,04	0,67	0,32	0,07
4	6	0,01	0,46	0,39	0,17
5	6	0,01	0,31	0,43	0,15

Elaboração do autor.

Ipea – Instituto de Pesquisa Econômica Aplicada

Assessoria de Imprensa e Comunicação

EDITORIAL

Coordenação

Cláudio Passos de Oliveira

Supervisão

Everson da Silva Moura

Leonardo Moreira Vallejo

Revisão

Ana Clara Escórcio Xavier

Camilla de Miranda Mariath Gomes

Clícia Silveira Rodrigues

Idalina Barbara de Castro

Olavo Mesquita de Carvalho

Regina Marta de Aguiar

Reginaldo da Silva Domingos

Alice Souza Lopes (estagiária)

Amanda Ramos Marques (estagiária)

Isabela Monteiro de Oliveira (estagiária)

Isabella Silva Queiroz da Cunha (estagiária)

Lauane Campos Souza (estagiária)

Lynda Luanne Almeida Duarte (estagiária)

Polyanne Alves do Santos (estagiária)

Editoração

Aeromilson Trajano de Mesquita

Bernar José Vieira

Cristiano Ferreira de Araújo

Danilo Leite de Macedo Tavares

Herllyson da Silva Souza

Jeovah Herculano Szervinsk Junior

Leonardo Hideki Higa

Capa

Danielle de Oliveira Ayres

Flaviane Dias de Sant'ana

Projeto Gráfico

Renato Rodrigues Bueno

*The manuscripts in languages other than Portuguese
published herein have not been proofread.*

Livraria Ipea

SBS – Quadra 1 – Bloco J – Ed. BNDES, Térreo

70076-900 – Brasília – DF

Tel.: (61) 2026-5336

Correio eletrônico: livraria@ipea.gov.br

Composto em adobe garamond pro 12/16 (texto)
Frutiger 67 bold condensed (títulos, gráficos e tabelas)
Impresso em offset 90g/m² (miolo)
Cartão supremo 250g/m² (capa)
Brasília-DF

Missão do Ipea

Aprimorar as políticas públicas essenciais ao desenvolvimento brasileiro por meio da produção e disseminação de conhecimentos e da assessoria ao Estado nas suas decisões estratégicas.

ipea Instituto de Pesquisa
Econômica Aplicada

MINISTÉRIO DO
**PLANEJAMENTO,
DESENVOLVIMENTO E GESTÃO**



ISSN 1415-4765

