



08, 09, 10 e 11 de novembro de 2022
ISSN 2177-3866

CORRELAÇÃO ENTRE OS FATORES DO ENSINO BÁSICO E INDICADORES SOCIAIS NOS MUNICÍPIOS BRASILEIROS.

RODRIGO CIRINO SILVA

UNIVERSIDADE ESTADUAL PAULISTA JÚLIO DE MESQUITA FILHO (UNESP)

FERNANDO FREIRE VASCONCELOS

FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE DA UNIVERSIDADE DE SÃO PAULO - FEA

CORRELAÇÃO ENTRE OS FATORES DO ENSINO BÁSICO E INDICADORES SOCIAIS NOS MUNICÍPIOS BRASILEIROS.

Introdução

No Brasil cerca de 50% da população vive abaixo da linha da pobreza multidimensional ou se encontra vulnerável à pobreza (Silva et al., 2020). Assim a atuação dos governantes é imprescindível para amenizar situações de risco, através de políticas públicas que tem influência direta na população (Carvalho et al., 2021), ligadas a áreas como saúde, educação, moradia, segurança, saneamento básico, entre outras (IPEA, 2018).

Os impactos dessas políticas podem ser medidos através de diferentes índices, pelos quais é possível observar a melhora ou não nos aspectos que levam ao desenvolvimento social e bem-estar (Oliveira e Santos, 2021).

Os índices mais amplamente utilizados são o Índice de desenvolvimento humano [IDH] e Índice de Gini (Dorsa e Constantino, 2018). No caso do IDH, trata-se de um indicador sintético - medida-resumo, onde o padrão de vida (renda) é medido pela Renda Nacional Bruta [RNB] per capita expressa em poder de paridade de compra [PPP] constante, em dólar (Bueno, 2009). Entretanto há mais fatores envolvidos no desenvolvimento humano, e o simples crescimento da renda per capita não garante melhora na qualidade de vida (Bueno, 2009).

No Índice de Gini, o parâmetro social avaliado é a concentração de renda apurada, evidenciando a discrepância entre os rendimentos da camada mais pobre e a mais rica da população (Campello e Gentili, 2017). Esse índice também pode ser calculado comparando-se outras variáveis sociais, como a concentração de terras, por exemplo (Dorsa e Constantino, 2018). Sendo assim, a queda na pobreza refletiria diretamente nesse índice (Campello e Gentili, 2017). Porém é necessário critério na interpretação dos seus resultados. Duas populações de mesmo coeficiente de Gini, porém diferentes rendas per capita podem significar diferentes cenários: Quanto maior a renda per capita, mais o índice estará relacionado com a qualidade de vida material. Quanto menor a renda per capita, ele também indicará o acesso à serviços básicos (Dorsa e Constantino, 2018).

Outro indicador social importante é o Produto Interno Bruto [PIB] e o PIB per capita (Silva et al., 2020). Segundo o Instituto Brasileiro de Geografia e Estatística [IBGE] (2021a), o PIB é a soma de todos os bens e serviços finais produzidos por um país, estados ou cidade, geralmente em um ano. Diferente do senso comum, o PIB não é apenas o total da riqueza de um país, mas um indicador do fluxo de caixa de bens e serviços finais produzidos durante um período (Oliveira e Santos, 2021).

Segundo Silva (2016), pode-se correlacionar a pobreza e a desigualdade social, qualitativa e quantitativamente ao déficit educacional. Essa relação é evidenciada principalmente em aspectos como acesso ao mercado de trabalho e desenvolvimento da capacidade crítica. Além disso, aspectos como número médio de anos de educação recebidos durante a vida a partir de 25 anos, e expectativa em anos de escolaridade para crianças na idade de iniciar a vida escolar influenciam diretamente o IDH de uma população (PNUD, 2021). Pode-se também evidenciar a correlação entre analfabetismo e renda familiar, conforme o IBGE (2020). No Brasil, municípios ou regiões com maior taxa de analfabetismo são também os que possuem situação de pobreza mais agravada (Silva, 2016).

Portanto, o conhecimento de como efetivamente os indicadores relativos à educação estão correlacionados com índices de disparidade social torna-se uma ferramenta importante para a tomada decisões públicas (Medeiros et al., 2019). Uma maneira de realizar esse tipo estudo é através de métodos de análise multivariada devido à complexidade dos dados envolvida (Cestero e Caballero, 2018), como a análise de componentes principais [PCA] e métodos de clusterização.

A PCA possibilita que se reduza a dimensionalidade dos dados, facilitando seu manejo e permitindo sua visualização em nos eixos mais representativos, as componentes principais (Cestero

e Caballero, 2018). Isso permite que se estabeleça a correlação, se existente, entre diversos fatores (Moita-Neto e Moita, 1998). Nesse caso em específico, evidenciando como a discrepância educacional pode afetar os indicadores sociais e econômicos dos Estados ou vice-versa. Já a análise de agrupamentos, clusterização, permite que, a partir do conjunto de dados apresentado por cada ponto amostral (nesse caso os estados brasileiros), e avalie-se quais deles possuem maior similaridade entre si (Moita-Neto e Moita, 1998). Neste estudo escolheu-se utilizar indicadores socioeconômicos disponibilizados pelo IBGE (2020) e compará-los com os indicadores educacionais utilizados pelo do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira [INEP] (2022a, 2022b).

Julnes (1999) já recomendava esse tipo de abordagem na avaliação de dados para a criação de políticas públicas, devido seu poder de integrar uma maior quantidade de dados que regressões lineares simples. Hoje, com a grande velocidade e quantidade em que os dados são produzidos, é imprescindível o uso da ciência de dados para extrair a informação existente neles (Saldanha et al., 2021).

Por fim, espera-se apontar como os indicadores sociais, econômicos e educacionais se correlacionam entre si, e uma vez conhecidas as relações mais importantes, priorizá-las em estudos futuros dentro dessa temática. Além disso, também se cria margem para comparação entre as políticas públicas das diferentes regiões do país e impactos que elas causam em seu cenário educacional e socioeconômico.

Implementação de Algoritmo de Machine Learning

Descrição dos datasets de indicadores socioeconômicos e educacionais

Neste trabalho foram utilizadas apenas fontes de dados abertas ao público, sem quaisquer restrições no uso e acesso. Os dados foram divididos em duas categorias: Indicadores socioeconômicos e educacionais. O quadro 1 abaixo descreve esses indicadores bem como o banco de dados utilizado.

Quadro 1 - Indicadores utilizados no estudo

Indicador	Tipo	Descrição	Ano	Fonte
IDHM	Socioeconômico	Índice de desenvolvimento humano municipal	2017	PNUD et al. (2017)
IDHM Renda	Socioeconômico	Índice de desenvolvimento humano municipal focado na renda média	2017	
IDHM Educação	Socioeconômico	Índice de desenvolvimento humano municipal focado no acesso à educação	2017	
IDHM Longevidade	Socioeconômico	Índice de desenvolvimento humano municipal focado na longevidade	2017	
Índice de Gini	Socioeconômico	Concentração de renda na população	2019	IBGE (2019a, 2021b)
PIBe	Socioeconômico	Produto Interno Bruto estadual	2019	
Média Hora-aula diária	Educacional	Média de Hora aula diária por instituição de ensino	2020	INEP (2022a,b)
Adequação do corpo docente	Educacional	% de professores com formação superior no ensino básico e médio	2020	
Resultados SAEB	Educacional	Nota média no Sistema de Avaliação do Ensino Básico [SAEB]	2020	

Resultados ENEM	Educacional	Nota média no Exame Nacional do Ensino Médio [ENEM]	2020	
--------------------	-------------	--	------	--

Fonte: Dados originais da pesquisa

O período temporal dos dados disponibilizados variou entre as diferentes bases de dados, variando entre 2017 até 2020. Portanto foi decidido utilizar a versão mais recente disponível de cada um dos indicadores de acordo com o descrito na Tabela 1.

Preparo dos dados para análise e compilação inicial

Uma vez obtida, a massa de dados foi carregada no software de análise POWER BI. Uma vez que as tabelas de microdados do ENEM excederam o número de linhas permitidas pelo MS Excel, não foi possível utilizá-lo. Além disso, o POWER BI permitiu criar tabelas e visibilidades essenciais para correlacionar os diversos indicadores estudados.

Depois de carregados os dados foram submetidos a uma compilação inicial, também no POWER BI, para a geração de uma tabela de resumo, contendo os valores de cada indicador separados por estado (Unidade da Federação ou UF). A tabela agrupou os dados de todos os indicadores descritos na Tabela 1. Uma vez feita essa etapa, os dados estão prontos para serem analisados nos algoritmos escolhidos.

Algoritmos de análise

Foi utilizado um algoritmo de análise de fatores principais em código em R (Malik e Tuckfield, 2019), criado a partir do software R STUDIO. Primeiramente foi criada uma matriz de correlações para geração do mapa de calor para entender o inter-relacionamento das variáveis. O código completo dessa etapa está disponível no Apêndice 1.

De acordo com as correlações obtidas, optou-se por analisar os resultados em diferentes grupos de parâmetros para discussão. Quando necessário, os grupos passaram posteriormente pela aplicação de um algoritmo de clusterização (Malik e Tuckfield, 2019) também dentro do ambiente do RStudio, de forma a estudar os impactos de políticas semelhantes nos resultados educacionais de cada estado. Conforme os códigos descritos no Apêndice 2. Com isso foram gerados gráficos de agrupamento para posterior avaliação dos resultados.

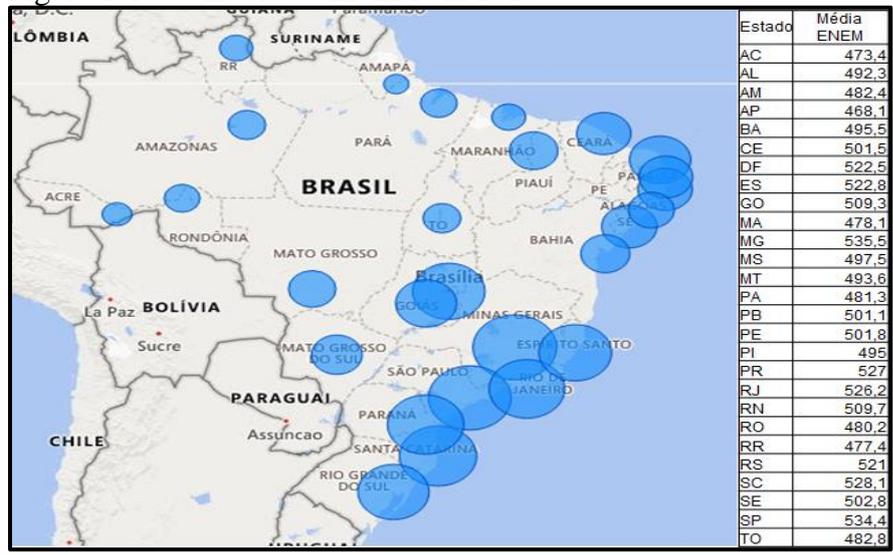
Resultados e Discussão

Compilação Inicial e Matriz de Correlações

A compilação inicial dos dados permitiu observar como os diferentes índices educacionais se distribuem dentro dos estados brasileiros. Começando pela média hora-aula, observou-se que esse valor foi maior nas escolas do nordeste (em especial PB). Já quando observada a quantidade de docentes no ensino fundamental que não possuem superior, observou-se uma grande disparidade entre as regiões norte e nordeste se comparadas com o sudeste, sul e centro oeste. Mostrando que apesar do maior volume hora aula na região nordeste, os professores não estão tão capacitados quanto em outras regiões. O mesmo padrão se repete quando observados os professores do ensino médio sem formação superior, com uma média maior nos estados do Norte e Nordeste. Entretanto, as médias observadas foram muito menores se comparadas com o ensino básico, indicando a maior necessidade de especialização para a docência durante esse ciclo específico.

Uma vez que há uma maior lacuna na formação de professores, já era esperado que isso refletisse nos resultados do ENEM, como observado na Figura 1 abaixo. Percebe-se que os estados com maior número de professores sem formação, também apresentaram as menores médias gerais no exame, enquanto os estados do Sul e Sudeste as maiores.

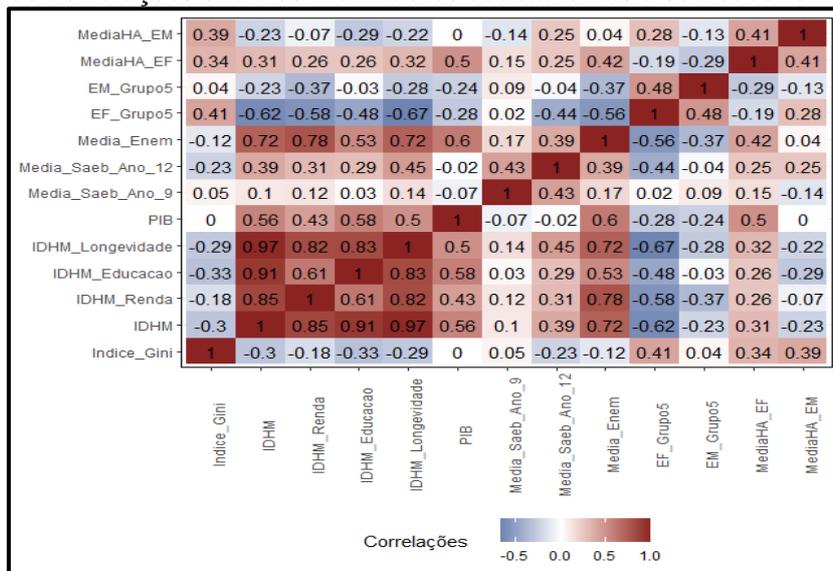
Figura 1. Nota média do ENEM em 2020



Fonte: INEP (2022b)

Buscando compreender melhor os resultados educacionais obtidos, outras dimensões sociais e econômicas foram incluídas visando encontrar correlações que possam trazer maiores esclarecimentos. Então, para entender melhor o porquê desse cenário, os dados educacionais foram correlacionados com os dados socioeconômicos dos estados, através da matriz de correlações, gerando o mapa de calor da Figura 2.

Figura 2 - Matriz de correlações entre os indicadores educacionais e socioeconômicos



Nota: Onde: EF_Grupo5 é a média de professores sem ensino superior no ensino fundamental, EM_Grupo5 é a média de professores sem ensino superior no ensino médio, MédiaHA_EF é a média

de horas-aula diária para escolas do ensino fundamental e MédiaHA_EM é a média de horas-aula diária para escolas do ensino médio

Fonte: Dados originais da pesquisa

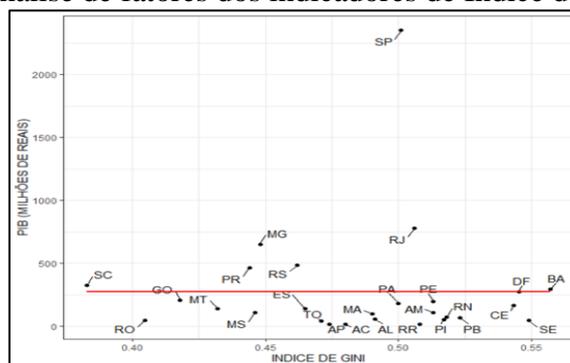
Observando-se a intensidade das correlações apresentadas, pode-se observar a formação de 3 grupos de variáveis relevantes para aprofundamento dos resultados e aplicação do algoritmo de clusterização:

- Grupo 1: Índice de Gini e PIB, devido à sua correlação aproximadamente igual a zero.
- Grupo 2: Média ENEM, IDHM (e suas componentes), pois o ENEM é um momento relevante para direcionamento dos jovens para o ensino superior e pode ser afetado por fatores além do próprio esforço individual.
- Grupo 3: Adequação do corpo docente, média de hora aula, média do ENEM e resultados do Saeb, para compreender como a preparação docente e volume de aulas pode afetar os resultados nos exames nacionais.

Comparação entre Índice de Gini e PIB

A primeira observação interessante acerca do Índice de Gini e do PIB é que a correlação entre eles é de aproximadamente zero (0,001), conforme a Figura 2. Essa correlação praticamente inexistente também foi observada em alguns cenários estudados por Pinheiro (2019) e foi exposta na análise de fatores na Figura 3.

Figura 3 - Análise de fatores dos indicadores de Índice de Gini x PIB



Fonte: Resultados originais da pesquisa

Nota-se que o estado de São Paulo detém o maior PIB e um elevado valor de concentração de renda (Índice de Gini \rightarrow 1). Rio de Janeiro e Minas Gerais, embora possuam PIBs semelhantes (e elevados se comparados aos demais estados, exceto São Paulo), diferem consideravelmente quanto à concentração de renda.

No outro extremo do PIB, Roraima e Acre possuem os menores valores, entretanto diferem com relação à concentração de renda, sendo que Roraima tem um Índice de Gini maior que o Acre.

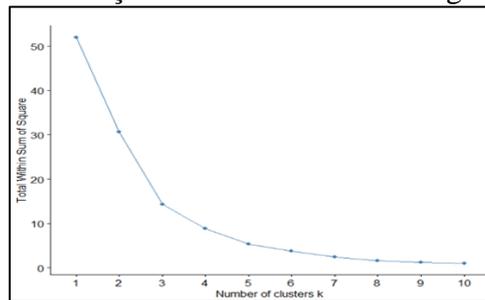
Nas extremidades do Índice de Gini, percebe-se que elas ocorrem em grandezas semelhantes de PIB, sendo que Santa Catarina apresenta a menor concentração de renda do Brasil, enquanto a Bahia apresenta a maior. Por fim, vale ressaltar que, por essa análise apenas, o Estado do Sergipe poderia ser foco de investimentos para melhor distribuir a sua renda, ao mesmo tempo em que aumente seu PIB, visto que o Estado estaria no pior quadrante desta comparação (menores PIBs e maior concentração de renda). Essa percepção coincide com o fenômeno estudado por Oliveira (2012), em que se cita que existe grande concentração de propriedade no Sergipe, fato somado

também ao fenômeno de migração campo-cidade. Além disso, ainda através do estudo de Oliveira (2012), o fechamento de escolas rurais isoladas e unidocentes (com apenas um professor) aumentou as disparidades sociais e agrava a concentração de renda.

Apesar destas considerações, ainda se conclui que a concentração de renda independe do PIB do Estado (Pinheiro, 2019).

Adicionalmente, pode-se agrupar os Estados por semelhanças para melhorar o entendimento dos cenários que levam a menores PIBEs e ou maiores concentrações de renda. Isso foi realizado através do algoritmo de clusterização por k-means. Com as variáveis definidas como PIB para o PIBE padronizado e “Índice_Gini” para o índice de Gini padronizado. Em seguida, o método Elbow auxilia na definição do número de grupos (figura 4).

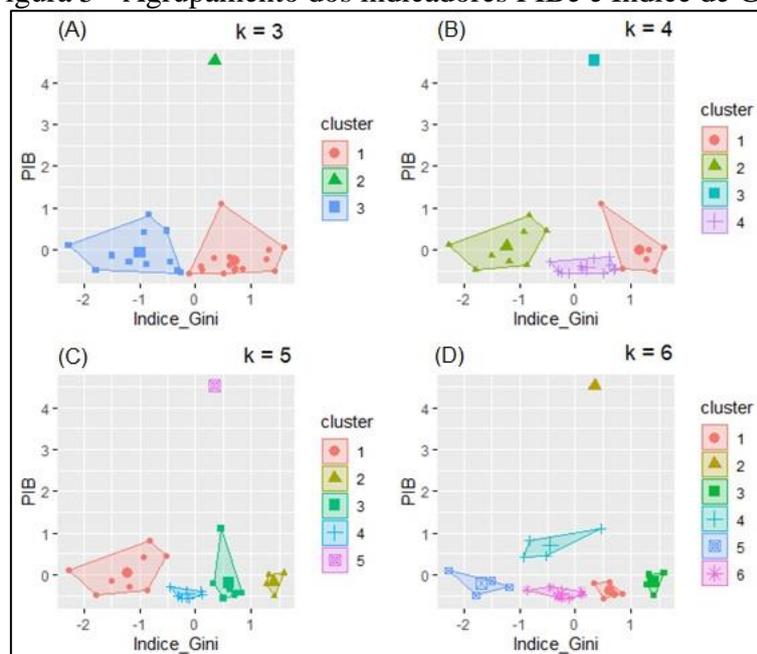
Figura 4 - Definição do número ótimo de agrupamentos



Fonte: Resultados originais da pesquisa

Analisando a Figura 4, nota-se que de 3 a 6 grupos se obtém os melhores resultados. Menos de 3 grupos os erros associados aos agrupamentos são elevados, enquanto para agrupamentos maiores que 6 grupos a melhoria no erro se torna pequena. Ainda assim, para decidir qual a melhor quantidade de grupos, optou-se por visualizar estas 4 opções, conforme Figura 5.

Figura 5 - Agrupamento dos indicadores PIBE e Índice de Gini

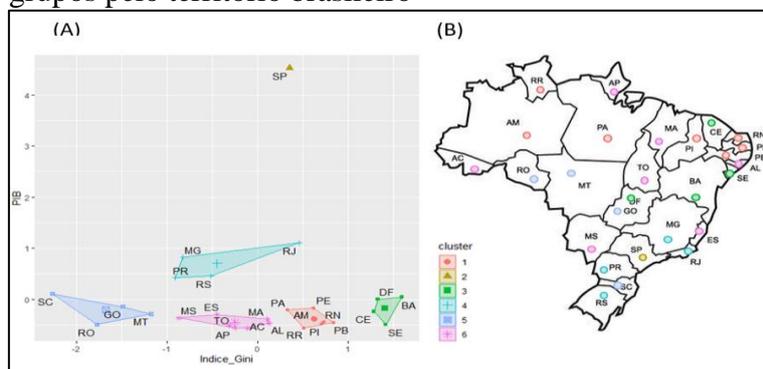


Nota: k corresponde ao número de agrupamentos

Fonte: Resultados originais da pesquisa

Logo, para esse conjunto de dados é possível observar que o mais indicado é a análise considerando 6 agrupamentos. Os resultados detalhados estão disponíveis na Figura 6.

Figura 6 - (A) Estados brasileiros agrupados de acordo com os valores de PIB e Índice de Gini e (B) distribuição dos grupos pelo território brasileiro



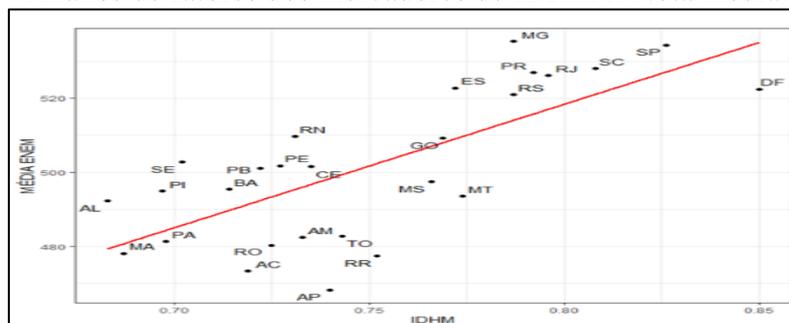
Fonte: Resultados originais da pesquisa

Mesmo neste agrupamento por semelhanças, o Estado de São Paulo permanece isolado, sendo definido como Cluster 2. Em primeira instancia, conclui-se que SP, por destoar dos demais Estados, não é uma boa generalização das condições do país (ao menos por Índice de Gini e PIB) e planos de ação nacionais precisam trabalhar esta informação adequadamente. O Cluster 3 se destaca por ter as maiores concentrações de renda, sendo que a economia dos Estados contidos neste grupo se baseia, em geral, em turismo (IBGE, 2019b). O Cluster 4 é um agrupamento contendo os maiores PIBs do país (exceto SP) e com atividades econômica diversas, contando com indústria, serviços, turismo e agropecuária fortes (IBGE, 2019b). O Cluster 5 se destaca por ter as menores concentrações de renda, sendo que as principais atividades econômicas dos estados são, em geral, agropecuárias e setor de serviços, contando com algumas indústrias (IBGE, 2019b). O Cluster 6 é outro grupo intermediário, porém seus valores de PIB são comparativamente mais baixos que a maioria dos Estados brasileiros. Suas principais atividades são turismo e agropecuária (IBGE, 2019b).

Média ENEM, IDHM (e suas componentes)

Outro ponto interessante obtido através da análise do mapa de calor é que o IDHM tem uma das maiores influências sobre a média do ENEM nos estados, conforme já visto. Assim, quando esses dois índices foram submetidos à análise de fatores, obteve-se a Figura 7:

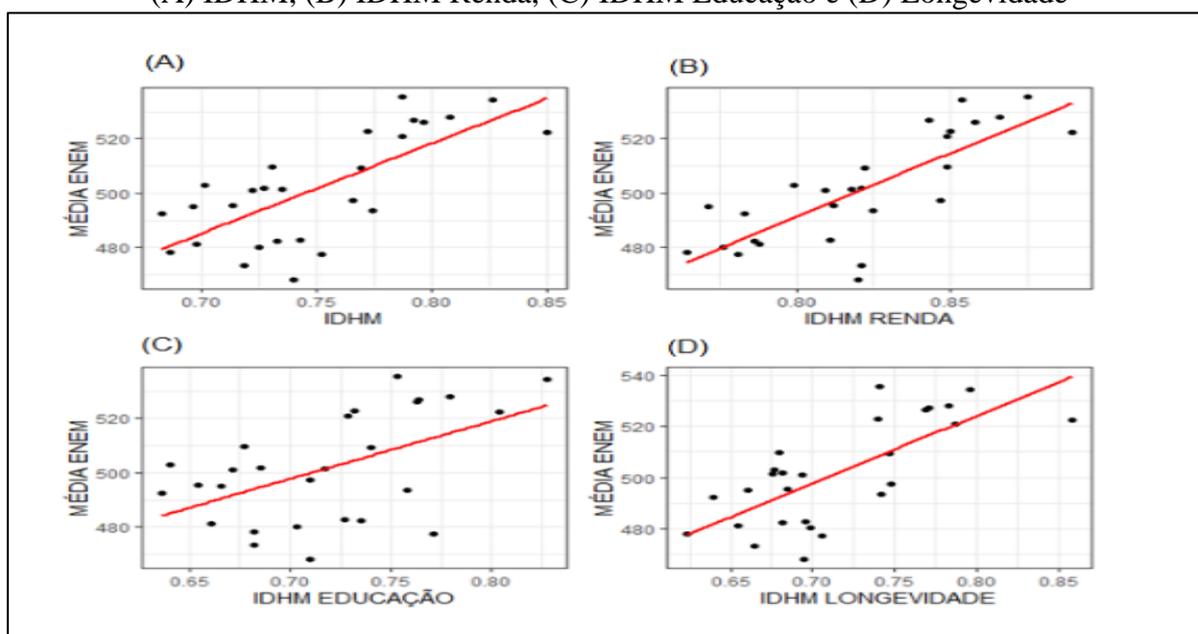
Figura 7 - Análise de fatores dos indicadores de IDHM x Nota média do ENEM



Fonte: Resultados originais da pesquisa

Essa relação já era esperada visto que a qualidade e investimento em educação fazem parte do cálculo do IDHM (Fully et al., 2019). Entretanto, se observada influência das diferentes componentes do IDHM (Renda, Educação e Longevidade) quando relacionadas à média do ENEM, obteve-se o mesmo resultado, conforme a Figura 8:

Figura 8 - Análise de fatores das diferentes componentes do IDHM x Nota média do ENEM, onde: (A) IDHM, (B) IDHM Renda, (C) IDHM Educação e (D) Longevidade



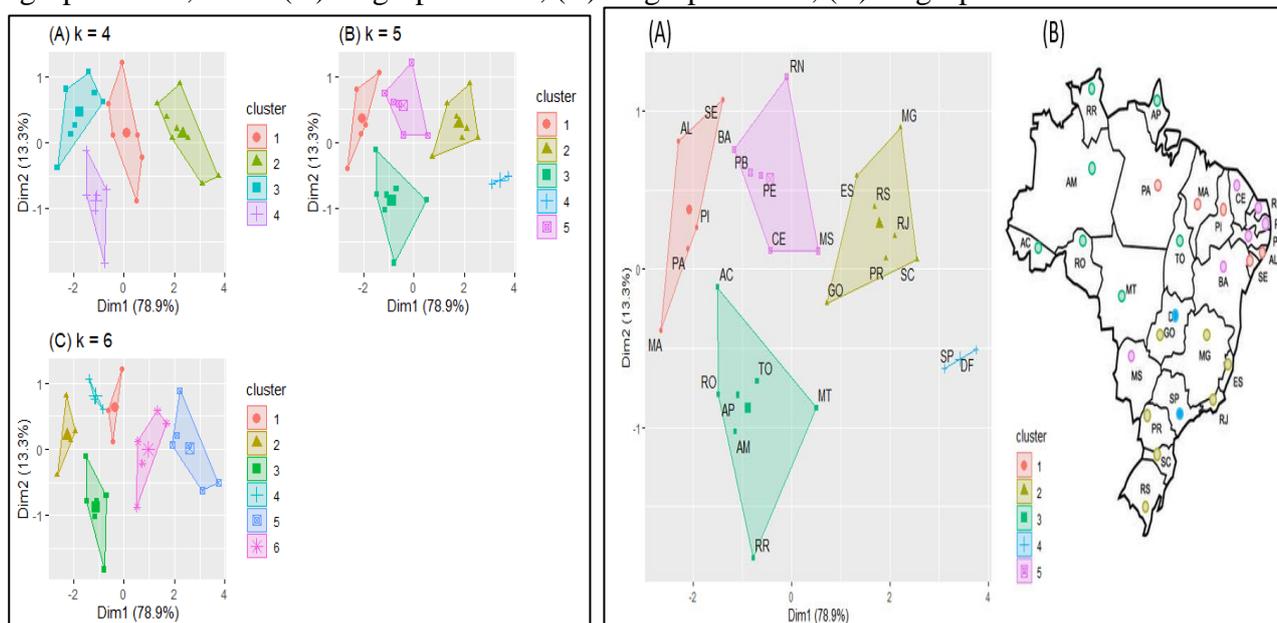
Fonte: Resultados originais da pesquisa

Percebe-se que as componentes do IDHM que mais influenciam os resultados do Enem são, na ordem, renda, longevidade e educação. IDHM Renda sendo o fator mais relevante sugere que as discrepâncias financeiras e econômicas têm maior influência no resultado dos alunos que prestam o Enem. Se observarmos em detalhes os parâmetros utilizados para o cálculo do IDH (PNUD, 2021): número médio de anos de educação recebidos durante a vida por pessoas a partir de 25 anos, e a expectativa em anos de escolaridade para crianças na idade de iniciar a vida escolar. É compreensível que famílias com maiores rendas podem ter acesso a instituições com melhores condições de ensino, além de permitir que o aluno tenha dedicação exclusiva aos estudos e permaneçam por mais tempo na escola (Pieri, 2018).

Uma possível explicação para o IDHM Longevidade ser o segundo fator de maior relevância pode se dar ao fato de que saúde física e mental, e condições insalubres de vida (PNUD, 2021) afetam diretamente o desempenho de alunos. Assim, estudantes nesse tipo de situação, apesar de apresentarem frequência escolar não apresentam bons resultados na aprendizagem (Souza et al., 2019). Portanto, apenas aumentar o investimento por aluno na rede pública, sem o investimento simultâneo nas outras áreas que também formam o IDH, pode não ter o efeito esperado, como já observado por Finger (2019).

Adicionalmente, para melhor compreender as semelhanças entre os estados, novamente aplicou-se o agrupamento sob o método k-means, novamente padronizando as variáveis e iniciando-se pelo método Elbow para definição do número de grupos. O resultado sugeriu visualizar os agrupamentos de 4 a 6 centros (Figura 9).

Figura 9 - Agrupamento dos indicadores IDHM e Média do ENEM em diferentes números de agrupamentos, onde: (A) 4 agrupamentos, (B) 5 agrupamentos, (C) 6 agrupamentos



Nota: k corresponde ao número de agrupamentos

Fonte: Resultados originais da pesquisa

Logo, para esse conjunto de dados é possível observar que o mais indicado é a análise considerando 5 agrupamentos. Os resultados detalhados estão disponíveis na Figura 9 detalhado por estado.

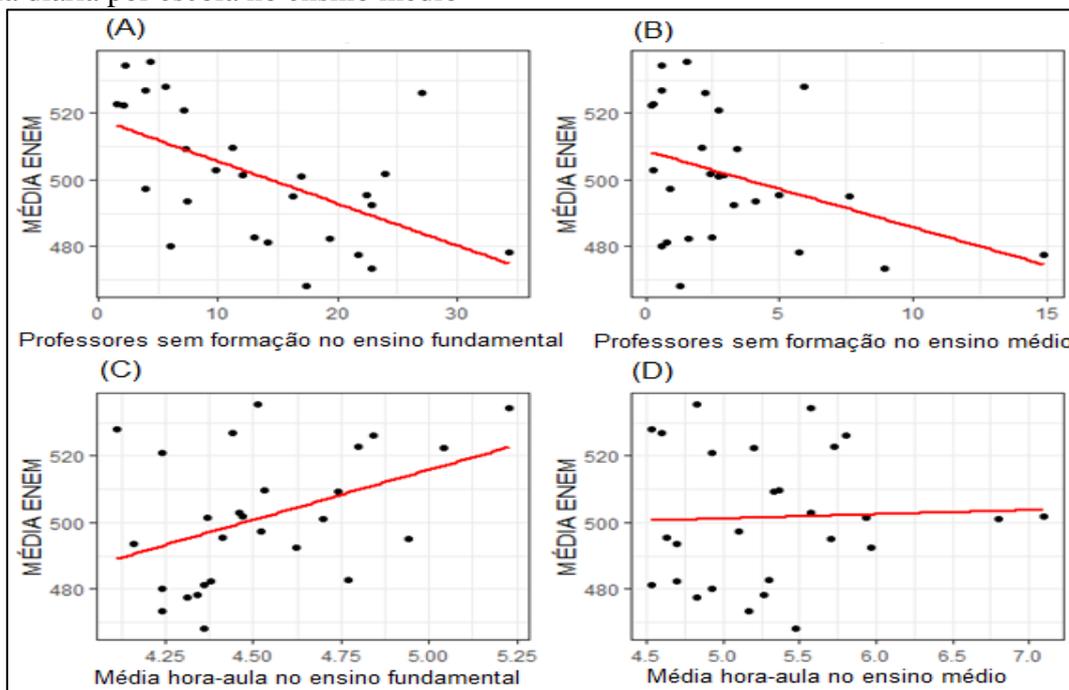
Como se trata de um agrupamento considerando múltiplos fatores, a interpretação se torna mais complexa: O Grupo 1 contém os Estados com menores IDHM e médias no ENEM abaixo da média do país, apesar de não serem as mais baixas entre os Estados. São Estados geograficamente próximos, por vezes compartilhando fronteiras. São regiões com maior densidade demográfica (IBGE, 2010). Já o grupo 3 contém a pior média do país (Amapá) e, em geral, as piores médias do país. Todos os Estados possuem IDHM baixos comparativamente aos demais Estados. Em sua grande maioria são de estados do norte do país, com menor densidade demográfica e com diversas áreas de preservação ambiental (IBGE, 2010). O Grupo 4 é composto apenas por São Paulo e Distrito Federal, com médias acima de 520 no ENEM e IDHM maiores que a média do país. Embora não dividam fronteiras, são próximos geograficamente. SP se destaca por ter a maior densidade demográfica do país (IBGE, 2010). Os grupos 2 e 5 são grupos intermediários, com grande variedade de Médias no ENEM e, diferem entre si pelo IDHM. Em geral, o grupo 2 apresenta melhores médias e IDHM e são compostos pelos Estados das regiões sudeste e sul, onde os setores industrial e de serviços são mais fortes (IBGE, 2010).

Notas do ENEM x Estrutura do Sistema Educacional

Dentre os indicadores do sistema educacional selecionados, o que mais afeta a média do ENEM é a qualificação do docente do ensino fundamental, sendo que neste estudo foi avaliada a porcentagem de docentes sem formação superior lecionando no fundamental e médio (Figura 10).

Figura 10 - Análise de fatores das diferentes componentes da Nota média do ENEM com os indicadores: (A) % de professores sem formação no ensino fundamental, (B) % de professores sem

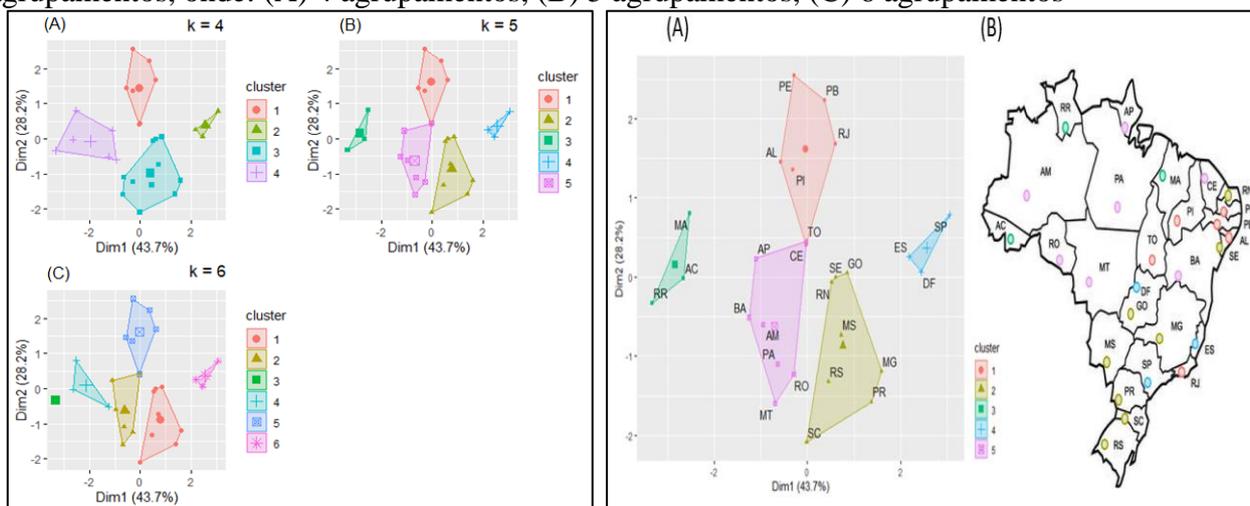
formação no ensino médio, (C) Média hora-aula diária por escola no ensino fundamental e (D) Média hora-aula diária por escola no ensino médio



Fonte: Resultados originais da pesquisa

Já por esta análise, é possível verificar que quanto maior a proporção de professores com formação superior no ensino fundamental [EF] ou do ensino médio [EM], menores as notas do ENEM. Além disso, nota-se que a média de hora-aula diária no ensino fundamental tem maior influência que a média de hora-aula no ensino médio. Desta primeira análise já se pode inferir que a educação de base inicial tem grande peso nas etapas seguintes. Avaliando Método Elbow para definição do número de grupos optou-se por gerar a visualização para variando de 4 a 6 centros (Figura 11).

Figura 11 - Agrupamento dos indicadores IDHM e Média do ENEM em diferentes números de agrupamentos, onde: (A) 4 agrupamentos, (B) 5 agrupamentos, (C) 6 agrupamentos



Nota: k corresponde ao número de agrupamentos

Fonte: Resultados originais da pesquisa

Logo, para esse conjunto de dados é possível observar que o mais indicado novamente é a análise considerando 5 agrupamentos. Os resultados detalhados estão disponíveis na Figura 11.

O grupo 1, em geral, possui médias maiores no ENEM, porém ainda tem alto volume de docentes sem formação superior. O grupo 2 contém estados com boa formação de seu corpo docente, porém com grande variação na média do ENEM. Em linhas gerais apresenta resultados superiores à média do país. Já grupo 3 contém os estados cuja proporção de docentes sem formação superior é maior e, portanto, investimentos na formação desses professores seriam necessários. No grupo 4 contém as menores proporções de docentes sem formações e com médias maiores no ENEM. Analisando isoladamente estes fatores, este grupo pode servir como base para estudos de aplicação de políticas educacionais para melhorar os resultados no ENEM. O Grupo 5, em geral, apresenta resultados intermediários em todos os fatores.

De um modo geral, conclui-se dois pontos. Primeiramente que a adequação do corpo docente do ensino fundamental possuiu maior peso se comparado ao ensino médio (confirmando o exposto pelo mapa de calor) e que aponta na prioridade do ensino fundamental como fator de redução de desigualdade social, conforme explicado por Oliveira (2007).

Considerações Finais

A partir do estudo de correlações de PIB e Índice de Gini pode-se concluir que o volume de riquezas de um estado não interfere com a má distribuição dos recursos financeiros e, portanto, políticas de redução de desigualdade podem ser aplicadas amplamente em território nacional, independente do PIB.

Na segunda análise, concluiu-se que quando comparadas as notas do Enem com o IDHM, a componente de renda é a mais significativa com relação ao resultado dos alunos neste exame nacional que direciona os estudantes para o ensino superior. Também se conclui que regiões com menor densidade demográfica tendem a apresentar os piores resultados com relação à média do ENEM.

Isso reforça que uma melhor condição financeira traz melhores resultados educacionais, independente do Estado. Por fim, analisando ENEM x indicadores do sistema educacional, conclui-se que investimentos no ensino fundamental tem maior potencial de melhorar os resultados dos estudantes no ENEM, principalmente no quesito adequação do corpo docente.

Referências

Bueno, E. P. (2009). O índice de desenvolvimento humano (idh): avaliação de seus pressupostos teóricos e metodológicos. *Boletim Goiano de Geografia* 27: 49–69.

Campello, T.; Gentili, P. (2017). As múltiplas faces da desigualdade, p. 10-15. In: Campello, T. *Faces da desigualdade no Brasil*. FLACSO, Brasília, DF, Brasil.

Carvalho, A. R.; Souza, L.R.; Gonçalves, S. L.; Almeida, E.R.F. (2021). Vulnerabilidade social e crise sanitária no Brasil. *Cadernos de Saúde Pública* 37: e00071721

Cestero, E.V.; Caballero, A.M. (2018). *Data science y redes complejas: Métodos y aplicaciones*. Editorial Centro de Estudios Ramon Areces, Madrid, Espanha.

Dorsa, A. C. C.; Constantino, M. A. (2018). Convergências entre indicadores de desenvolvimento local, índice de desenvolvimento humano e índice de Gini, *Revista Contribuciones a las Ciencias Sociales*.

Disponível em: <<https://www.eumed.net/rev/cccss/2018/07/desenvolvimento-local-humano.html>>
Acesso em 15 de mai de 2022.

Finger, E. (2019). Relação entre despesas com educação, desempenho escolar e o IDH: um estudo exploratório nos municípios do Vale do Caí em 2017. Trabalho de conclusão de curso do bacharelado em Ciências Contábeis. Universidade de Caxias do Sul, São Sebastião do Caí, RS, Brasil.

Fully, F.M.P.; Guimarães, A.A.B.; Silva, A.F.O.; Pimentel, L.S. (2019). Os gastos públicos sobre o IDH na microregião de caratinga no período de 2014 a 2016. *Brazilian Journal of Development* 7: p.10585-10605.

Instituto Brasileiro de Geografia e Estatística - IBGE. (2010). Densidade demográfica. Disponível em: <<https://www.ibge.gov.br/geociencias/cartas-e-mapas/sociedade-e-economia/15955-densidade-demografica.html?=&t=acesso-ao-produto>>. Acesso em: 21 de mai. 2022.

Instituto Brasileiro de Geografia e Estatística - IBGE. (2019a). Produto Interno Bruto dos Municípios. Disponível em <<https://www.ibge.gov.br/estatisticas/economicas/contas-nacionais/9088-produto-interno-bruto-dos-municipios.html?=&t=destaques>>. Acesso em: 21 de mai. 2022.

Instituto Brasileiro de Geografia e Estatística - IBGE. (2019b). Sistema de contas regionais: Brasil: 2019. Disponível em: <<https://biblioteca.ibge.gov.br/visualizacao/livros/liv98881.pdf>>. Acesso em: 21 de mai. 2022.

Instituto Brasileiro de Geografia e Estatística – IBGE. (2020). Síntese de Indicadores Sociais – Uma análise das condições de vida da população brasileira 2020. Disponível em: <<https://biblioteca.ibge.gov.br/visualizacao/livros/liv101760.pdf>>. Acesso em: 21 de nov. 2021.

Instituto Brasileiro de Geografia e Estatística - IBGE. (2021a). PIB. Disponível em: <<https://www.ibge.gov.br/explica/pib.php>>. Acesso em: 21 de nov. 2021.

Instituto Brasileiro de Geografia e Estatística - IBGE. (2021b). Indicadores de Desenvolvimento Sustentável -Tabela 5801 - Índice de Gini da distribuição do rendimento mensal das pessoas de 15 anos ou mais de idade, com rendimento - Série Histórica. Disponível em <<https://sidra.ibge.gov.br/tabela/5801>> . Acesso em: 21 de mai. 2022.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - INEP. (2022^a). Indicadores Educacionais. Disponível em: <<https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/indicadores-educacionais>>. Acesso em: 21 de mai. 2022.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - INEP. (2022^b). Avaliações e exames educacionais. Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais>> . Acesso em: 21 de mai. 2022.

Instituto de Pesquisa Econômica Aplicada - IPEA. (2018). Agenda 2030. Disponível em: <http://repositorio.ipea.gov.br/bitstream/11058/8855/1/Agenda_2030_ods_metas_nac_dos_obj_de_desenv_susten_propos_de_adequa.pdf>. Acesso em: 21 de nov. 2021.

Julnes, J. (1999). Principal Component Analysis, Factor Analysis, and Cluster Analysis, p. 549-598. In: Miller, G.J.; Whilker, M.L. Handbook of Research Methods in Public Administration. Marcel Dekker, New York, NY, United States of America.

Malik, A.; Tuckfield B. (2019). Applied unsupervised learning with R. Packt Publishing, Birmingham, United Kingdom

Medeiros, M.C.; Barbosa, R.J.; Carvalhaes, F. (2019). Educação, desigualdade e redução da pobreza no Brasil. Texto para Discussão, No. 2447. 1 ed. Instituto de Pesquisa Econômica Aplicada (IPEA), Brasília, DF, Brasil.

Moita-Neto, J.M.; Moita, G.C. (1998). Uma introdução à análise exploratória de dados multivariados. Química Nova 21: 467-469.

Oliveira, C.L.; Santos, F.A. (2021). Análise de correlações entre indicadores econômicos: PIB, Cesta básica e IDH. Revista Ensino e Pesquisa em Administração e Engenharia 7: 93-109.

Oliveira, F. K. (2012). Migração e desigualdade regional em Sergipe. Revista Paranaense De Desenvolvimento 121: 167-188.

Oliveira, R. P. (2007). Da universalização do ensino fundamental ao desafio da qualidade: uma análise histórica. Educação & Sociedade, v. 28, n. 100

Pieri, F. (2018). Retratos da Educação no Brasil. Insper Instituto de Ensino e Pesquisa, São Paulo, SP, Brasil.

Pinheiro, W.; Vieira, E.T.; Oliveira, E.A.A.Q. (2019). As relações entre o índice de Gini com o índice de desenvolvimento humano e o produto interno bruto dos municípios da RMVPLN sob a ótica do desenvolvimento regional. Revista Humanidades e Inovação 6.

Programa das Nações Unidas Para o Desenvolvimento - PNUD. (2021). O que é o IDH?. Disponível em: <<https://www.br.undp.org/content/brazil/pt/home/idh0/conceitos/o-que-e-o-idh.html>>. Acesso em: 21 de nov. 2021.

Programa das Nações Unidas Para o Desenvolvimento – PNUD; Fundação João pinheiro [FJP]; Instituto de Pesquisa Econômica Aplicada – IPEA. (2017). Ranking. Disponível em: <<http://www.atlasbrasil.org.br/ranking>>. Acesso em: 21 de maio 2022.

Saldanha, R.F.; Barcellos, C.; Pedroso, M.M. (2021). Ciência de dados e big data: o que isso significa para estudos populacionais e da saúde?. Cadernos Saúde Coletiva 29: 51-58.

Silva, J.J.; Bruno, M.A.P.; Silva, D.B.N. (2020). Pobreza multidimensional no Brasil: uma análise do período 2004-2015. Brazilian Journal of Politics and Economy 40: 138-160.

Silva, M.A.S. (2016). Pobreza multidimensional: a educação como fator de superação da pobreza no semiárido brasileiro. 2016. Tese de Doutorado em Desenvolvimento e Meio Ambiente. Universidade Federal do Ceará, Fortaleza, CE, Brasil.

Souza, L.B.; Panuncio-Pinto, M.P.; Fiorati, R.C. (2019). Crianças e adolescentes em vulnerabilidade social: bem-estar, saúde mental e participação em educação. Caderno Brasileiro de Terapia Ocupacional 27: 251-269.