

REDUÇÃO FATORIAL HÍBRIDA COMO TÉCNICA DE MINERAÇÃO DE DADOS DA AVALIAÇÃO EDUCACIONAL EM LARGA ESCALA

<https://dx.doi.org/10.5902/2318133870944>

Simone Mágná Menezes Carneiro Lopes¹
José Carlos da Silveira Freire²

Resumo

A avaliação em larga escala tem despertado o interesse de pesquisadores, de governos e da sociedade. Tal interesse reside no fato de que seus resultados têm orientado as políticas públicas de gestão e financiamento da educação básica. Entretanto, os dados produzidos pelo Sistema de Avaliação da Educação Básica do Brasil são subutilizados pela gestão escolar como ferramenta de diagnóstico e de promoção da aprendizagem. Nessa perspectiva, buscou-se identificar fatores que influenciaram no índice do Ideb na avaliação do ano de 2017, para o 9º ano do ensino fundamental, entre escolas da rede estadual do Tocantins. Para isso, foi realizada uma análise fatorial multivariada exploratória para identificar os fatores os quais estão relacionados a um maior ou menor desempenho na avaliação do Ideb. Diante da variação das escalas dos questionários da Saeb e das notas do Ideb que variam de 0 a 10, houve a necessidade da dicotomização das escalas, por isso a análise fatorial multivariada pautou-se, tanto na extração fatorial por meio do método dos componentes principais a partir da correlação de Pearson, quanto na obtida por meio da correlação tetracórica. Concluiu-se que aplicação do método irá auxiliar o gestor em compreender os índices levantados e traçar uma perspectiva com pontos específicos em que se deve melhorar, além de possibilitar a extração de informações importantes para que a gestão possa intervir, de forma focalizada, na aplicação de recursos e nortear políticas públicas.

Palavras-chave: avaliação educacional; Ideb; Saeb; análise fatorial multivariada.

HYBRID FACTOR REDUCTION AS A DATA MINING TECHNIQUE FOR LARGE SCALE EDUCATIONAL ASSESSMENT

Abstract

Large-scale assessment has aroused the interest of researchers, governments and civil society. Such interest lies in the fact that its results have guided public policies on management and financing of basic education. However, the data produced by the Brazilian Basic Education Evaluation System are underused by school management as a diagnostic and learning promotion tool. In this perspective, we sought to identify factors that influenced the Ideb index in the 2017 assessment for the 9th grade of elementary education among schools in the state network of Tocantins. To this end, an exploratory multivariate factor analysis was performed to identify the factors that are related to a higher or lower performance in the Ideb assessment. Given the variation in the scales of the Saeb questionnaires and the Ideb scores that vary from 0 to 10, there was the need to dichotomize the scales. Therefore, the multivariate factor analysis was based both on the factor extraction by means of the principal components method from Pearson's correlation and on that obtained by means of the tetrachoric correlation. It was concluded that the application

¹ Universidade Federal do Tocantins, Brasil. E-mail: simone.lobes@uft.edu.br.

² Universidade Federal do Tocantins, Brasil. E-mail: cfreire@uft.edu.br.

of the method will help the manager to understand the indexes raised and draw a perspective with specific points where improvement is needed, besides making it possible to extract important information so that the management can intervene in a focused way in the application of resources and guide public policies.

Key-words: educational assessment; Ideb; Saeb; multivariate factor analysis.

Introdução

Ao manusear os resultados da avaliação do Índice de Desenvolvimento da Educação Básica - Ideb - para as escolas públicas de ensino fundamental do Estado do Tocantins, foi possível notar que, até mesmo em escolas de uma mesma esfera administrativa, há relevante defasagem em suas notas. Assim, buscou-se compreender as particularidades deste fenômeno.

Pressupõe-se que o desenvolvimento educacional, apresenta-se como um meio para o direcionamento eficiente de políticas públicas, com foco no desenvolvimento econômico. Foi com essa justificativa que a avaliação educacional começou a ser implantada no Brasil nos anos 1960, havendo uma preocupação específica com os processos avaliativos escolares baseados em critérios mais claramente enunciados e em instrumentos que poderiam garantir, até certo ponto, seu nível de realização pelo uso das técnicas 'objetivas' (Gatti, 2002).

Em 1991, foi instituído o Sistema Nacional de Avaliação da Educação Básica - Saeb. Desde então, desenvolvem-se pesquisas com relação à avaliação escolar, que foram desde formatos de avaliação focalizados apenas em desempenho dos estudantes, até a adoção da Teoria de Resposta ao Item - TRI -, para a elaboração de testes e análise de resultados para compreender o avanço do desempenho dos alunos, buscando novas formas de avaliá-los. Em 2007 foi lançado o Ideb, que em tese é um instrumento de análise e condução de políticas públicas em prol da qualidade da educação (Mello e Bertagna, 2020).

Segundo Nunes (2018), os resultados do Ideb podem ser utilizados para a análise da qualidade do ensino, pois o Ideb permite traçar metas de qualidade educacional para os sistemas. Essa avaliação conta com ferramentas estatísticas que permitem correlacionar o fluxo de alunos na escola, considerando a evasão e a aprovação do aluno, de maneira proforma sem obter o desempenho mínimo de aprendizado. Nesse sentido, Travitzki (2020) relembra que apesar do Ideb ter uma nota que vai de 0 a 10, não necessariamente aquelas escolas com nota acima de 7 são as únicas consideradas boas ou que notas abaixo de 5 caracterizam as escolas ruins. Por isso, a interpretação razoável dos resultados do Ideb deve ser feita de maneira relativa, para que interpretações precipitadas não sobressaiam, como por exemplo, não considerar a influência socioeconômica ao classificar a qualidade de uma escola em boa ou má.

Assim, o problema desta pesquisa é evidenciado pela utilização de métodos de análises insuficientes ou ainda superficiais para interpretação dos resultados do Ideb, muitas vezes negligenciados, que traz como consequências a subestimação da avaliação em larga escala. Isso nos levou a questionar se é possível empregar metodologia eficaz e prática, que seja palpável ao gestor educacional e que lhe permita ter uma visão sistematizada dos resultados das avaliações em larga escala.

A partir disso, buscou-se identificar fatores que influenciaram na nota do Ideb na avaliação aplicada no ano de 2017 para o 9º ano/8ª série do ensino fundamental entre escolas da rede estadual do Tocantins, a partir de uma análise fatorial multivariada, utilizando como ferramenta de redução fatorial uma associação entre a correlação de Pearson e a correlação tetracórica. Entende-se que, uma vez identificados estes fatores, a gestão pode intervir de forma focalizada na aplicação de recursos, bem como nortear políticas públicas.

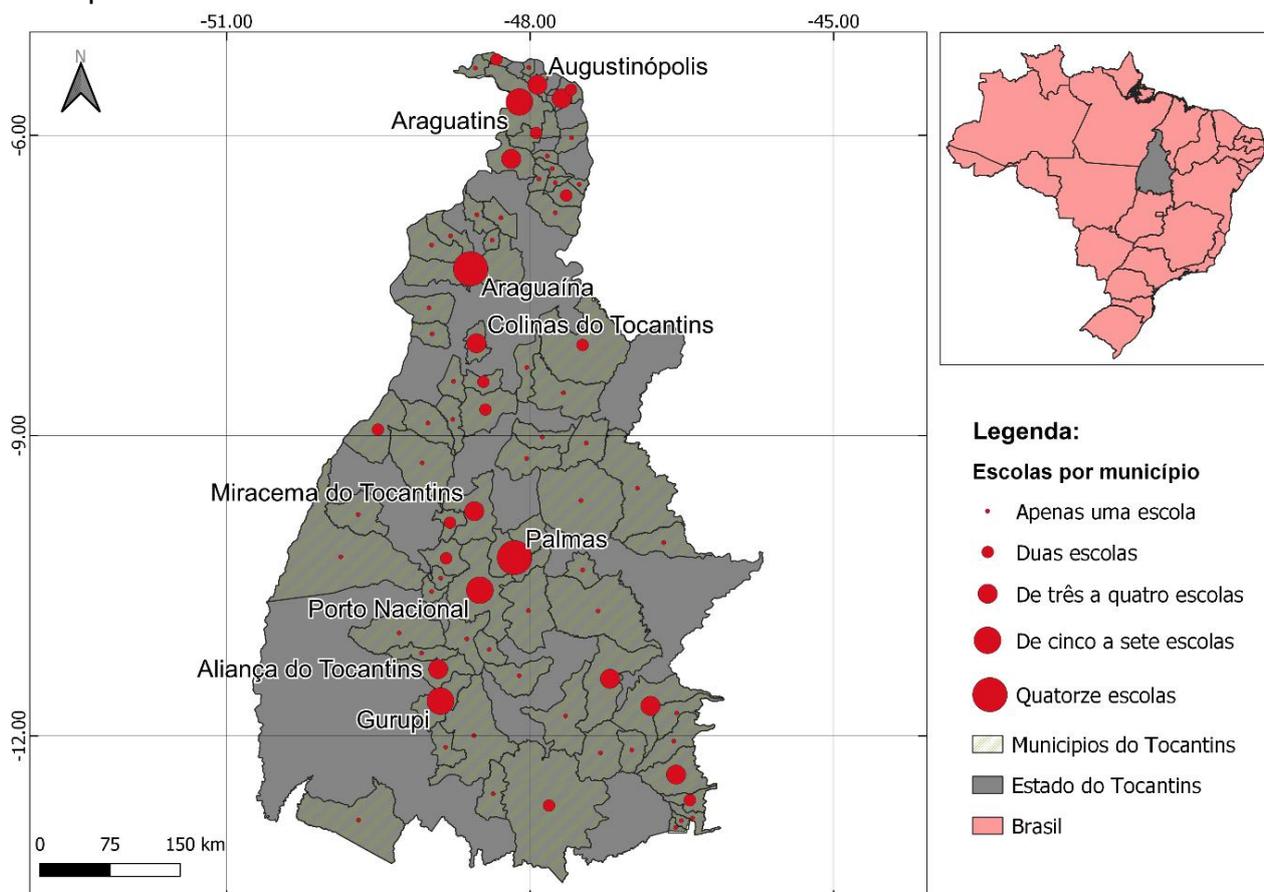
A hipótese desta pesquisa sustenta ser possível que a partir da sistematização de dados quantitativos de forma objetiva e a interpretação relativa das notas do Ideb, os gestores possam gerir melhor as escolas, concentrando esforços em áreas que requeiram maior atenção. O objetivo da pesquisa foi demonstrar uma sistemática de análise e gestão dos resultados de avaliações em larga escala baseada em métodos quantitativos de correlação com os índices de desempenho na avaliação do Ideb, entre escolas estaduais de nível fundamental, na perspectiva do ensino de qualidade.

Campo de pesquisa

Esta pesquisa teve como foco as escolas de nível fundamental da rede estadual de ensino do Estado do Tocantins. Dos dados coletados no sistema Saeb, detectou-se que 400 escolas de ensino fundamental da rede estadual de ensino participaram do Saeb 2017, sendo que destas, apenas 157 tiveram resultados divulgados da avaliação do Inep 2017 para o 9º ano do ensino fundamental. Isso aconteceu porque algumas escolas não participaram da avaliação do Ideb e outras não tiveram uma participação mínima de alunos que fosse suficiente para a divulgação de notas.

O mapa da figura 1 demonstra como as 157 escolas objeto deste estudo estão distribuídas no território do Tocantins, sendo que 64 (41%) destas concentraram-se em apenas nove municípios, sendo que sete estão entre os municípios mais populosos do Estado. Palmas e Araguaína, os dois maiores municípios do Estado, foram os que tiveram o maior número de escolas analisadas - 14 escolas para cada um -, o que corresponde a cerca de 18% dos dados. Destaca-se a presença do município de Aliança do Tocantins que contribuiu com quatro escolas, apesar de ser apenas o 64º município mais populoso, com 5.434 habitantes.

Figura 1 -
Proporção de escolas por município tocantinense que participou tanto do Saeb 2017 quanto do Ideb 2017.



Fonte: autores, a partir de dados vetoriais de IBGE (2020).

Metodologia

Nos dados dos questionários do Saeb, fornecidos pelos diretores das unidades escolares, constam informações sobre o perfil e a experiência dos gestores, as atividades desenvolvidas, os recursos disponíveis e a infraestrutura do estabelecimento. Esse conjunto de perguntas permitiu conhecer mais sobre a realidade da educação das escolas pesquisadas (Inep, 2020).

A partir da análise dos questionários do diretor, do professor, da escola e dos resultados do Ideb, foram selecionadas questões que se enquadrem em grupos específicos, que estatisticamente são chamados de fatores. Esta divisão por fatores objetivou identificar como estes fatores têm relação com a nota dos 25% e 50% das escolas de ensino fundamental que tiveram maiores e menores índices de desempenho na avaliação do Ideb para o Estado do Tocantins. Esperando-se, com isso, que estes fatores a serem obtidos tenham uma correlação contrária, nas escolas com maior índice na avaliação do Ideb e nas escolas com menores índices.

No quadro 1 pode-se identificar os itens que fizeram parte desta pesquisa e as 29 questões selecionadas dos questionários da avaliação do Saeb, sobre esferas inerentes à gestão escolar, classificadas da seguinte forma: relação com a comunidade, gestão democrática, recursos financeiros, dinâmica e formação do corpo docente, infraestrutura e o fator socioeconômico.

Quadro 1 -

Questões selecionadas no questionário do Diretor, separadas por grupos específicos inerentes à gestão escolar.

Grupo	Nº	Questão selecionada
Relação com a comunidade	1	Neste ano, a escola promoveu eventos para a comunidade frequentemente ou quase sempre?
	2	Considere as condições existentes para o exercício do cargo de diretor nesta escola: Há apoio da comunidade à sua gestão?
	3	Os espaços desta escola são utilizados para eventos promovidos pela comunidade frequentemente ou quase sempre?
	4	Neste ano, a comunidade colaborou com trabalho voluntário para esta escola frequentemente ou quase sempre?
Gestão democrática	5	Neste ano, o Conselho Escolar se reuniu três vezes ou mais?
	6	Há participação de pais e alunos no Conselho Escolar?
	7	O conselho de classe se reuniu três vezes ou mais?
	8	O projeto Político pedagógico teve participação da equipe escolar?
Recursos financeiros	9	O funcionamento da escola foi dificultado por insuficiência de recursos financeiros?
	10	Esta escola recebeu neste ano apoio financeiro do governo federal?
	11	Esta escola recebeu neste ano apoio financeiro do governo estadual?
	12	Esta escola recebeu neste ano apoio financeiro de empresas ou doadores individuais?
Corpo docente	13	O funcionamento da escola foi dificultado por problemas do alto índice de faltas dos professores?
	14	O funcionamento da escola foi dificultado pela alta rotatividade do corpo docente?
	15	75 % ou mais do número de professores têm vínculo estável nesta escola?
	16	O funcionamento da escola foi dificultado por inexistência de professores para algumas disciplinas ou séries?
Formação do corpo docente	17	Mais de 50% do corpo docente tem título de especialista, mestre ou doutor?
	18	Nos últimos dois anos mais de 50 % do corpo docente participou atividades de desenvolvimento profissional na área de atuação específica?
	19	Nos últimos dois anos mais de 50 % do corpo docente participou atividades de desenvolvimento profissional na área de Educação?
	20	A temática do seu curso de pós-graduação foi na área da Educação?
Infraestrutura	21	As salas de aula estão em bom estado de conservação?
	22	A escola conta com acesso à internet de boa qualidade para uso dos alunos?
	23	A escola conta com projetor de slides/Datashow de bom estado de

		conservação?
	24	A escola conta com laboratório de informática em bom estado de conservação?
	25	A escola conta com laboratório de ciências?
	26	A escola conta com biblioteca em que os livros podem ser manuseados e emprestados?
Fator socioeconômico	27	A escola possui índice de nível socioeconômico (Inse) médio baixo?
	28	A escola possui índice de nível socioeconômico (Inse), variando de médio a muito alto?
	29	A escola está situada em zona urbana?

Fonte: autores.

É fato que existem algumas escolas que se destacam de maneira acentuada em relação à grande maioria, como aquelas escolas que praticam a seleção de alunos por meio de prova de conhecimentos, ou ainda, quando sua nota é muito baixa por algum motivo extraordinário, como a baixa adesão de alunos na realização do exame. Nesse caso, suas notas interferem na confiança da análise estatística, pois existirão dados fora da curva normal, sendo chamados estatisticamente de *outliers*. Para contornar esta situação, foram utilizadas ferramentas estatísticas e gráficos *boxplots* para visualização da existência de *outliers*. Uma vez constatada esta existência, estes dados discrepantes foram excluídos quando $z > 2$.

As informações referentes ao nível socioeconômico de cada escola foram obtidas por meio dos resultados dos questionários da escola, classificados conforme informações obtidas nos questionários dos alunos. Conforme Inep (2015), as fontes utilizadas pelo Saeb para a definição do indicador socioeconômico de cada aluno, foram os dados dos questionários contextuais dos estudantes, fornecidos pelos microdados disponibilizados pelo Inep, e a partir da média aritmética simples dos índices por aluno, obteve-se o nível socioeconômico de cada escola. As questões utilizadas relacionam-se com a renda familiar, à posse de bens e contratação de serviços de empregados domésticos pela família dos estudantes e ao nível de escolaridade de seus pais ou responsáveis (Inep, 2015).

Como as médias das notas do Ideb, para as provas de Língua Portuguesa e Matemática, variam de 0 a 10 e depois essa nota média é multiplicada pela taxa de aprovação da escola, que vai de 0 a 100%, a nota final da escola também irá variar de 0 a 10. As questões do Saeb, por sua vez, têm respostas oriundas de questões de múltipla escolha, constando nos microdados do Saeb apenas respostas que vão da letra 'a' até a letra 'e' em algumas questões e em outras até a letra 'd', ou ainda verdadeiro ou falso. Diante desta diferença de escala nas questões selecionadas, e para viabilizar a análise fatorial multivariada, foi necessário adaptar as questões do Saeb e notas do Ideb, conforme o quadro 1, para que as respostas fossem reduzidas a dados binários, sendo 1 para verdadeiro e 0 para falso.

Seguindo os procedimentos apresentados na metodologia de Engelmann (2010), Clement et al. (2014) e Riscal e Luiz (2016), o tratamento dos dados foi realizado por meio de ferramentas computacionais gratuitas e estruturado em oito etapas: (i) sistematização dos dados brutos oriundos dos metadados do Saeb e notas do Ideb e redução binária utilizando o software *Libre Office Calc*; (ii) importação da planilha binária para o software R; (iii) carregamento dos pacotes necessários para a análise fatorial, *devtools*, *corplot* e *pysich*; (iv) extração da matriz de correlação a partir da ferramenta “*Cor*” em R (R Core Team, 2020); (v) extração da matriz gráfica de correlação utilizando o pacote *corplot* em R; (vi) realização do teste de esfericidade de Bartlett e KMO por meio do pacote *pysich* em R; (vii) extração e análise dos componentes principais para redução fatorial em R e (viii) extração do gráfico Scree por meio da ferramenta *screeplot* em R.

Para a interpretação dos fatores obtidos pela análise fatorial, primeiro observou-se as cargas fatoriais de cada variável da matriz de carga fatorial, em seguida verificou-se as variáveis que apresentavam elevadas cargas fatoriais em cada componente obtido. Como a amostra foi de 155 escolas, foram utilizadas para formar um fator variáveis com cargas fatoriais acima de 0,45, conforme condições estabelecidas por Moraes (2016).

Para esta análise de redução fatorial, optou-se por utilizar um método híbrido de redução a partir da correlação de Pearson e da correlação tetracórica, pois, a princípio, o método mais usual quando os valores observados estão normalmente distribuídos é a correlação de Pearson (Matos e Rodrigues, 2019; Lira e Chaves Neto, 2004), enquanto a correlação tetracórica é recomendada para variáveis dicotomizadas - ex: sim ou não -, que foi a estratégia utilizada nessa pesquisa para uniformizar as variadas escalas das questões dos questionários do Saeb e notas do Ideb, porém, encontra-se limitações quando a amostra é pequena (Lira e Chaves Neto 2004).

Para auxiliar na identificação dos fatores obtidos foram feitos os gráficos de autovalores - *scree test* -, tanto para redução a partir da correlação de Pearson, quanto para a redução a partir da correlação tetracórica. Para confirmar este fato, também foi realizada a análise paralela por meio das ferramentas *fa.parallel* para correlação de Pearson e *fa.parallel.poly* para a correlação tetracórica em R. A análise paralela é uma técnica alternativa que compara a sequência de fatores dos dados observados com a de uma matriz de dados aleatórios do mesmo tamanho do original, sendo que podem ser considerados componentes acima da curva paralela - *resampled data*.

Resultados e discussão

De posse das notas do Ideb para o 9º ano do ensino fundamental para cada escola em estudo, foi elaborado o gráfico *boxplot*, a fim de estratificar as escolas por faixa de nota e determinar os possíveis *outliers*, que são aquelas escolas que fogem da curva normal. Para o correto procedimento da análise fatorial multivariada pretendida, foi necessário a eliminação preliminar destes *outliers*.

Conforme o gráfico 1, eliminando-se os *outliers*, a nota máxima foi de 5,9 e a mínima de 2,7. Nota-se que a nota média no Ideb das escolas foi de 4,34 e a mediana 4,4. O valor da mediana e os valores máximo e mínimo serviram como parâmetro para delimitação das faixas de estudo que foram divididas em 4:

1) Escolas com nota no Ideb (2017) maiores ou igual a 2,7 e menores que 3,9, correspondendo aos 25% das escolas com menor índice;

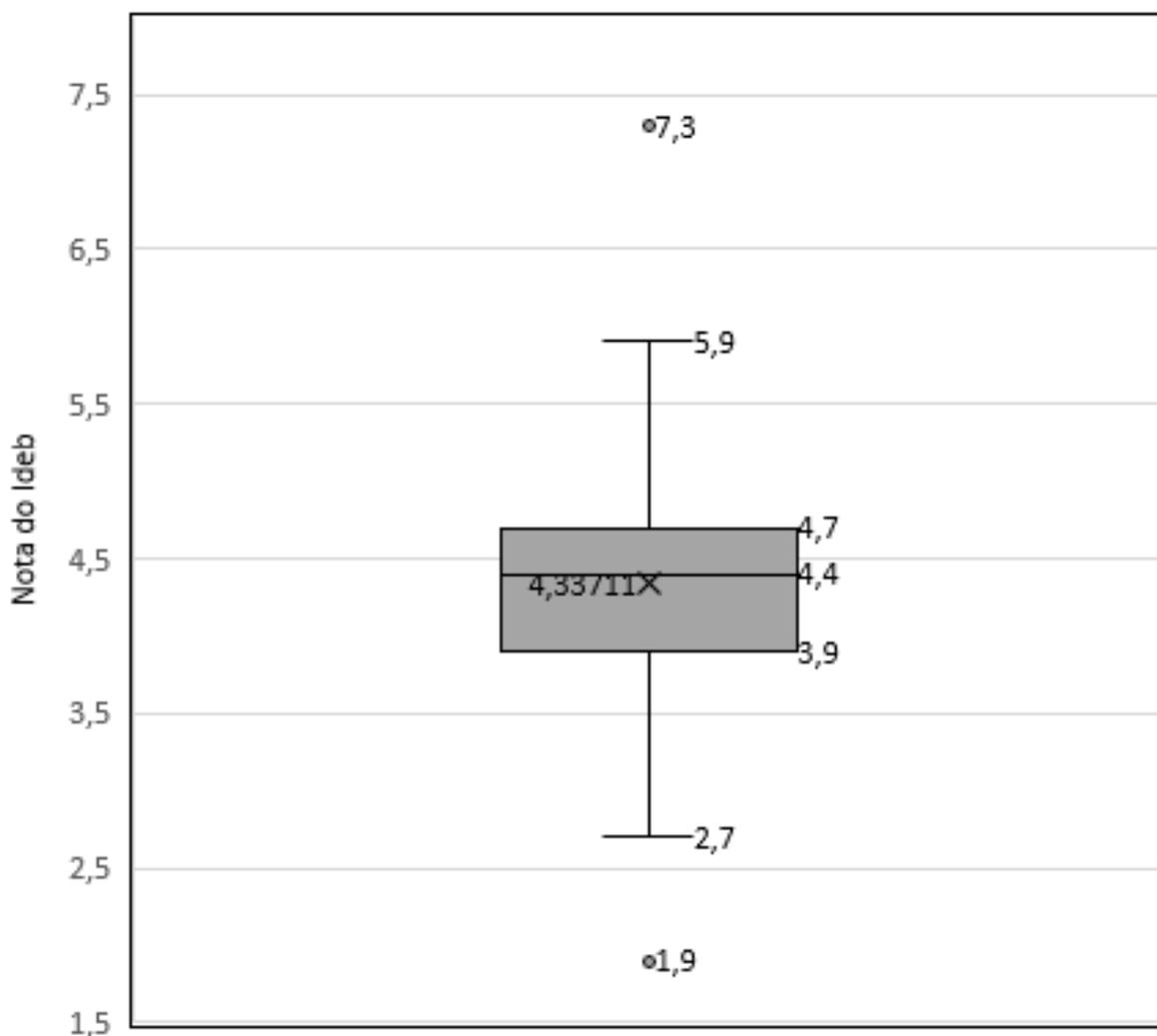
2) escolas com nota no Ideb (2017) maiores ou igual a 2,7 e menores que 4,4, correspondendo aos 50% das escolas com menor índice;

3) escolas com nota no Ideb (2017) maiores ou igual a 4,4 e menores que 5,9, correspondendo aos 50% das escolas com maior índice;

4) escolas com nota no Ideb (2017) maiores ou igual a 4,7 e menores que 5,9, correspondendo aos 25% das escolas com maior índice.

Gráfico 1 -

Boxplot da série de notas do Ideb (2017) das escolas pesquisadas neste estudo.



Fonte: autores.

Os resultados obtidos pela análise do teste de validade da análise fatorial demonstraram a adequabilidade do procedimento utilizado para interpretação dos resultados obtidos. O valor do índice KMO obtido foi de 0,52, o que indica que a análise fatorial é adequada, tendo em vista que autores como Hair et al. (2005) definem que índices KMO acima de 0,5 são valores aceitáveis.

Outro fator que garante a confiabilidade do procedimento em questão, foi o valor de significância obtido igual a 0,000. Hair et al. (2005) indicam que valores de significância abaixo de 0,100 são adequados para o método proposto. Ou seja, essa análise

demonstrou que algumas das questões selecionadas nos questionários do Saeb e notas do Ideb estão correlacionadas entre si e que a matriz de correlação das variáveis definidas pelas questões não é uma matriz identidade.

Os resultados obtidos para as duas medidas - Teste de esfericidade de Bartlett e Medida de adequacidade da amostra - indicam que a análise fatorial é adequada, observando-se, contudo, que o índice KMO encontrou-se muito próximo do limite do inaceitável, porém, associando ao teste de esfericidade de Bartlett e ao tamanho pequeno da amostra, que foi de 155 escolas - após eliminação dos outliers -, confirmou-se a validade do método.

Os gráficos *scree* 2 e 3 aproximam-se de uma reta com inclinação relativamente acentuada até o quinto componente principal, após isso ocorre uma quebra abrupta no gráfico, o que demonstra que as 33 variáveis iniciais - 29 oriundas dos questionários do Saeb e quatro relacionadas à nota do Ideb -, podem ser explicadas por apenas cinco fatores, considerando as redundâncias nas respostas. Conforme Matos e Rodrigues (2019), isso é um sinal de que muita variância foi perdida e extração de fatores deve ser interrompida.

As tabelas 1 e 2 mostram que os cinco fatores obtidos a partir das reduções fatoriais têm valor próprio maior que 1, o que garante explicação dos dados a partir dos fatores selecionados. Observa-se, nessas tabelas, que como a matriz não foi rotacionada, os autovalores são a própria variância e, conseqüentemente, o percentual dessa variância é a relação entre o autovalor de cada componente e o somatório dos cinco componentes detectados.

Porém, percebe-se, nas tabelas 1 e 2, que pela percentagem da variância acumulada seguindo imediatamente superior a 60%, conforme proposto por Hair et al. (2005), indica que as 33 variáveis podem ser explicadas pelos três primeiros componentes. Além do mais, o número de componentes deve ser definido a partir da análise em conjunto do teste *scree*, pela percentagem da variância e pelo critério de Kaiser que se refere aos componentes com autovalor maior ou igual a 1 (Matos e Rodrigues, 2019).

Tabela 1 -

Extração dos componentes principais a partir da correlação de Pearson.

Componente	Autovalores iniciais		
	Autovalor	% de variância	% de variância acumulada
1	3.34	28	28
2	2.86	24	51
3	2.20	18	70
4	1.99	16	86
5	1.69	14	100

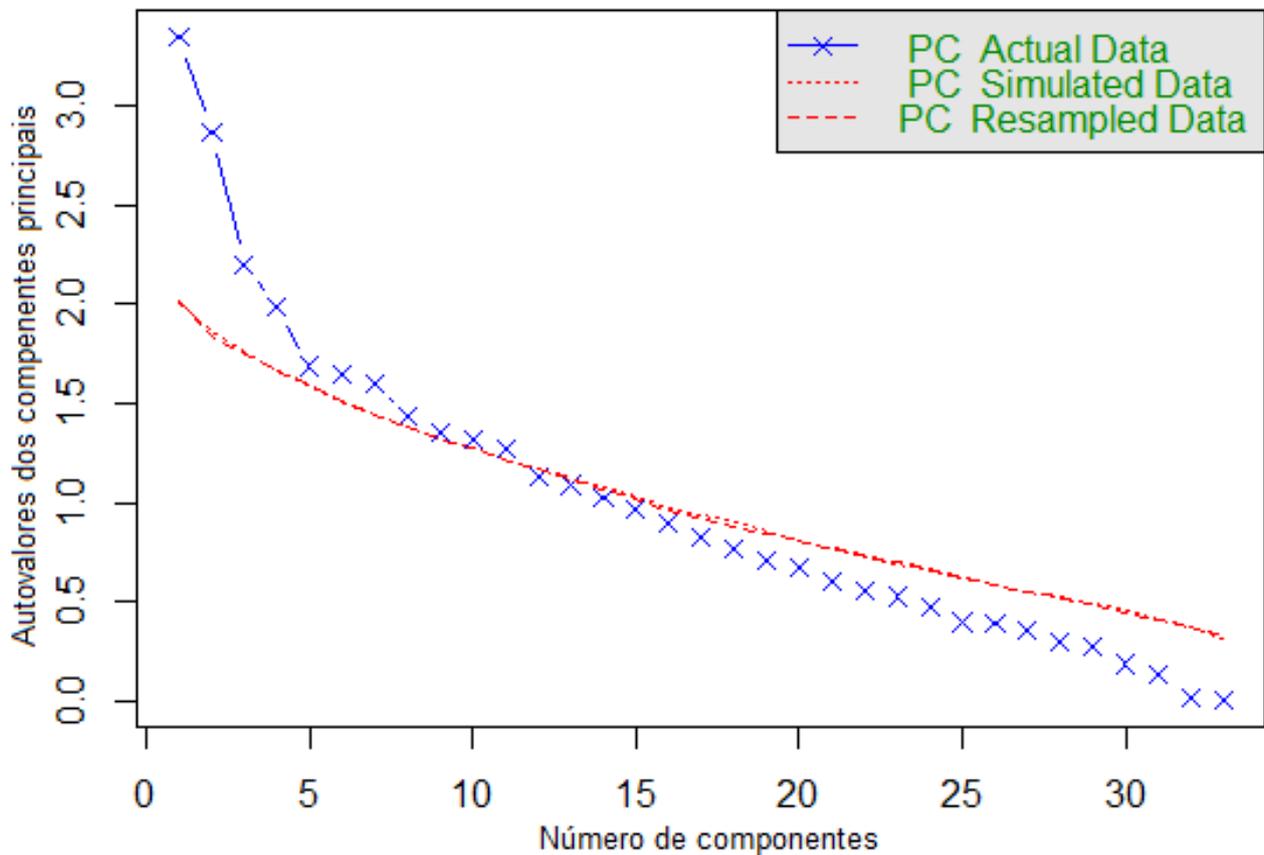
Fonte: autores.

Tabela 2 -
Extração dos componentes principais a partir da correlação tetracórica.

Componente	Autovalores iniciais		
	Autovalor	% de variância	% de variância acumulada
1	4.44	28.00	28.00
2	3.72	23.00	51.00
3	3.11	19.00	70.00
4	2.60	16.00	87.00
5	2.15	13.00	100.00

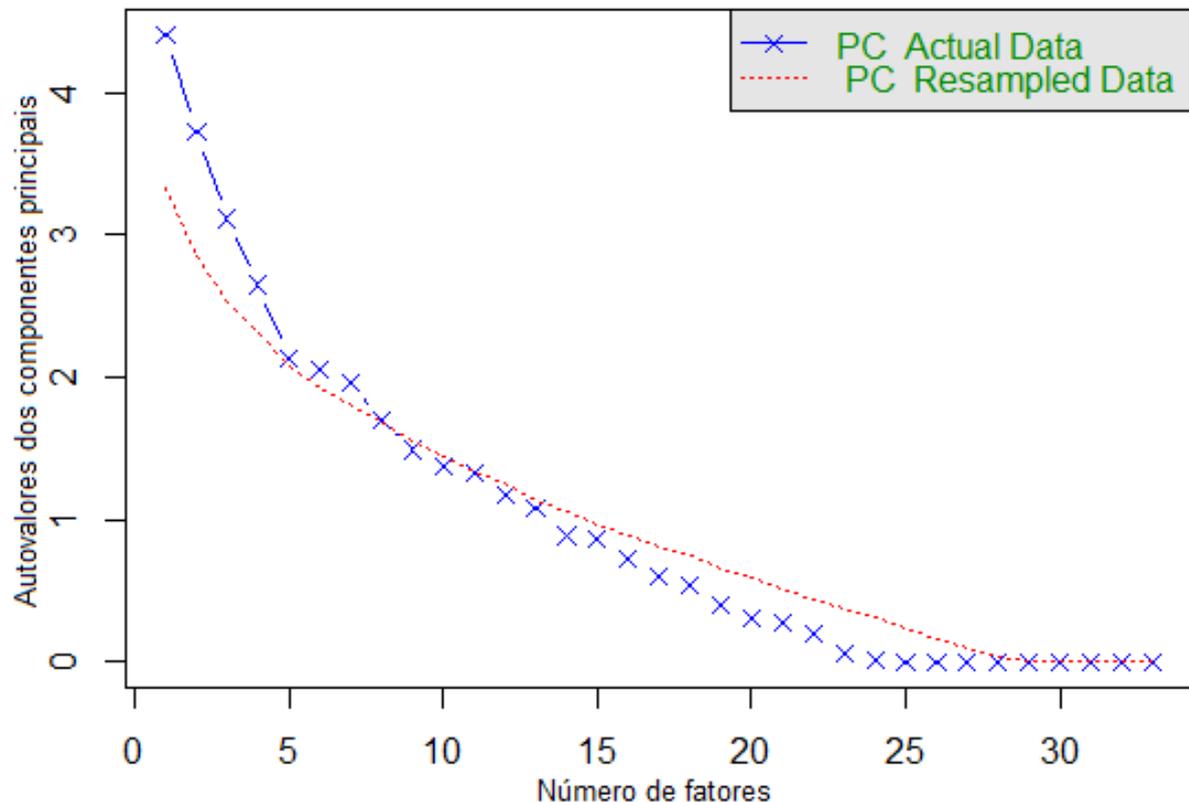
Fonte: autores.

Gráfico 2 -
Gráfico de autovalores a partir da correlação de Pearson.



Fonte: autores.

Gráfico 3 -
Gráfico de autovalores a partir da correlação tetracórica.



Fonte: autores.

Utilizando a correlação de Pearson, o fator 1 diz respeito às variáveis que se correlacionaram com as escolas com notas do Ideb acima da mediana, ou seja, os 50% das escolas com maior desempenho. Podendo-se destacar variáveis relacionadas à infraestrutura, como: salas de aulas em bom estado de conservação, internet de boa qualidade para alunos e professores e laboratórios de informática e de ciências. Além disso verificou-se que o índice socioeconômico de médio a alto correlacionou-se com as escolas que obtiveram maior desempenho no Ideb.

Os fatores 2 e 3, dizem respeito às variáveis que se correlacionaram com a variável correspondente às escolas abaixo da mediana e a variável correspondente às escolas abaixo do quartil inferior. Já ambos os fatores, indicam variáveis que devem ser avaliados pelas escolas visando à melhora das notas do Ideb. Dentre as mais relevantes, destaca-se as correlações com a insuficiência de recursos financeiros, alto índice de faltas dos professores, alta rotatividade do corpo docente, inexistência de professores para séries específicas e gestão pouco participativa, a partir da correlação contrária com o número de reuniões maiores que três vezes no ano do conselho escolar e abertura da escola para eventos promovidos pela comunidade.

Com relação a falta de efetividade do conselho escolar, entende-se que existe coerência na correlação com as escolas com menor desempenho, tendo em vista que no trabalho de Riscal e Riscal (2014) foram obtidos resultados satisfatórios com uma análise correlata ao caso. Os referidos autores constataram que o conselho escolar só se torna

significativo a partir de três ou mais reuniões ao ano, o que significa que um conselho realmente presente na gestão da escola e atuante é importante fator para elevação da nota do Ideb (Riscal; Riscal, 2014).

Outra variável que se correlacionou com as escolas com menores desempenhos, foi a que questiona se o conselho de classe se reuniu três vezes ou mais durante o ano. Desse dado, verifica-se coerência, pois a nota do Ideb é diretamente proporcional ao percentual de aprovações e é sabido que um número maior de reuniões do conselho de classe implica em maior número de reprovações.

Observou-se, também, que se correlacionou com as escolas mais mal ranqueadas no Ideb a existência de bibliotecas em que os livros possam ser emprestados e manuseados pelos alunos. Constatou-se que isso ocorreu em detrimento à laboratórios de informática e internet de qualidade que, por sua vez, correlacionou-se com as escolas que alcançaram maiores índices de desempenho. Este fato reforça a hipótese de que as bibliotecas estão sendo utilizadas como uma alternativa para suprir outros déficits para propiciar aos alunos condições de pesquisa e estudo. De todo modo, é inegável que o vasto acervo gratuito disponível na internet não poderia ser substituído por bibliotecas físicas, além do mais, apenas o fato da existência da biblioteca não garantiria ao aluno o acesso a material de pesquisa atualizado. O que se denota como uma deficiência do questionário do Saeb, sendo que não basta questionar a existência de biblioteca, mas se essa efetivamente conta com acervo em quantidade e qualidade suficientes e atualizado.

A extração fatorial por meio da correlação de Pearson foi complementada pela extração fatorial utilizando a correlação tetracórica. Pôde-se observar que para esta extração, apesar de ter sido considerado o método da maior percentagem de variância, não se pôde verificar agrupamento fatorial para a variável referente às escolas situadas abaixo do quartil inferior, de forma que ocorresse de maneira análoga à correlação de Pearson. Portanto, foram selecionados dois fatores, correspondendo às escolas com desempenho acima da mediana - fator 1 - e às escolas com desempenho abaixo da mediana - fator 2.

Isso se deve ao fato da limitação existente na correlação tetracórica, a qual pré-estabelece que a dicotomização da escala deve ser feita o mais próximo possível da mediana (Lira e Chaves Neto, 2004). E conforme o *boxplot* do gráfico 1, que por um lado percebe-se que a amplitude entre a mediana (4,4) e o quartil inferior (3,9) de 0,5 que se mostrou significativa para extração fatorial a partir da correlação tetracórica; por outro lado, a amplitude entre a mediana e o quartil superior (4,7) de 0,3, mostrou-se insignificante confundindo-se as variáveis referentes aos 50% e aos 25% das escolas com maior desempenho.

Da extração fatorial por meio da correlação tetracórica, percebeu-se que, além das variáveis obtidas a partir da correlação de Pearson, houve a correlação das variáveis referentes aos recursos financeiros. Foi observado que se correlacionou com o grupo das melhores escolas as questões que solicitavam saber do gestor se a escola havia recebido naquele ano, tanto apoio financeiro do governo federal, quanto do governo estadual. Já para o grupo das escolas com menor desempenho em relação à nota do Ideb, correlacionou-se a questão em que a escola informa positivamente que naquele ano houve insuficiência de recursos financeiros e que, apesar de ter correlacionado o apoio financeiro do governo federal, não se verificou a correlação com o apoio de recursos do

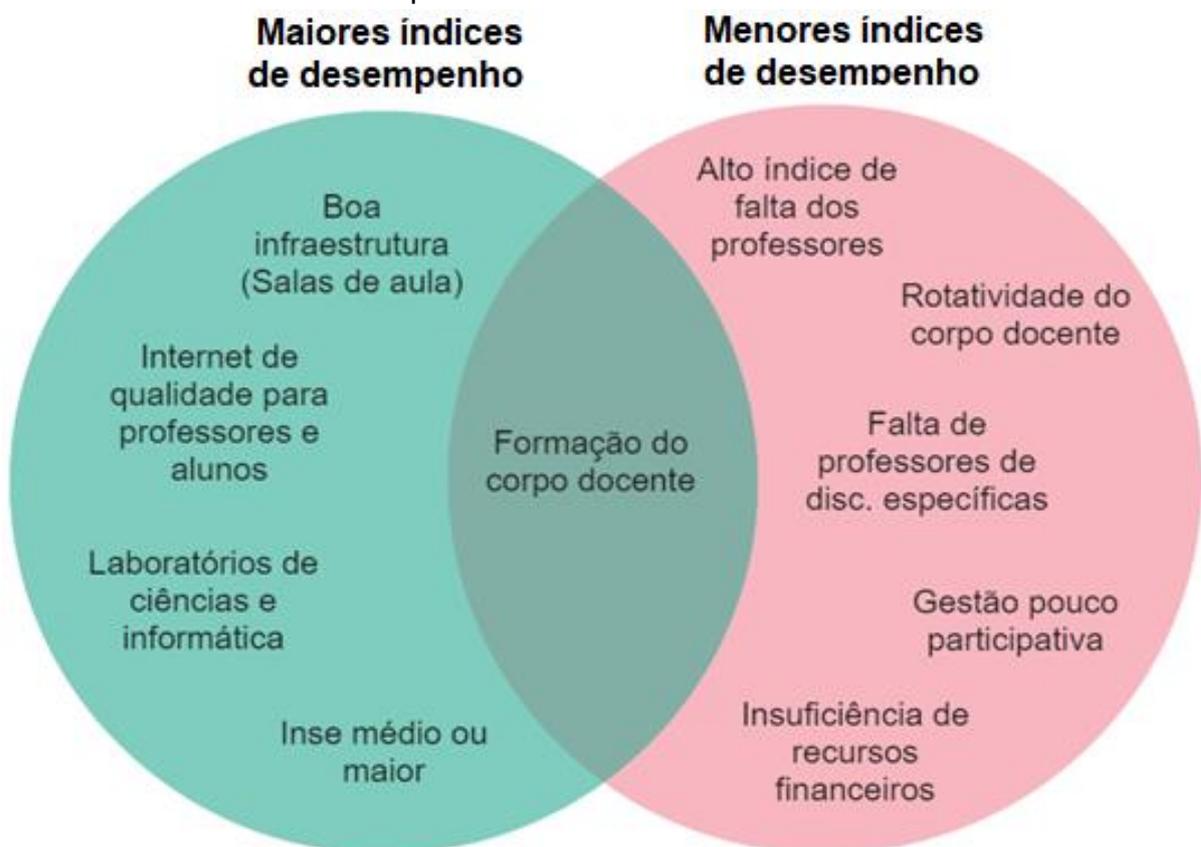
governo estadual. Isso apresenta-se como um indicativo de que parte das escolas que compõem o grupo daquelas que tiveram menores notas no Ideb, podem não ter tido o efetivo repasse de recursos do governo estadual.

Além disso, houve correlação com o grupo das melhores escolas com a variável referente à localização da escola em zona urbana, o que se demonstra coerente, sendo inclusive destacado no trabalho Riscal e Riscal (2014), em que a localização da escola, seja urbana ou rural, influencia nas notas do Ideb, sendo que a localização urbana possibilita uma ascensão no Ideb em relação às escolas rurais.

Na figura 2 consta um resumo sobre as variáveis mais importantes correlacionadas com as escolas com maiores e menores índices de desempenho na avaliação do Ideb. Considerando esta análise fatorial, observa-se que a formação docente se situou na região de interseção, não sendo fator preponderante para melhoria nas notas, e não se correlacionando exclusivamente com um ou outro grupo, e a maior parte das variáveis referentes a formação docente não apresentou carga fatorial significativa. Porém, pela análise das matrizes de correlação, constatou-se que a formação docente, por sua vez, contribuiu para dirimir os problemas inerentes ao alto índice de faltas dos professores, a alta rotatividade do corpo docente e ausência de professores para disciplinas específicas.

Figura 2 -

Resumo das variáveis correlacionadas com as escolas com maiores e menores índices de desempenho na avaliação do Ideb para o ano de 2017 em escolas do ensino fundamental da rede pública do estado do Tocantins.



Fonte: autores.

Considerações finais

A pesquisa demonstrou a aplicabilidade e a importância de se analisar os resultados da avaliação educacional em larga escala de forma multivariada, considerando avaliações aplicadas no ano de 2017 em escolas da rede pública estadual de ensino fundamental do Estado do Tocantins. Nesse contexto, concluiu-se que os microdados do Saeb oferecem um amplo universo de dados que podem ser aplicados considerando as particularidades da área de estudo. Pôde-se perceber que existe uma riqueza de informações nestes questionários e avaliações, materializados em um grande volume de dados. Por isso, requerem que sejam tratados por técnicas de mineração de dados educacionais, visando a encontrar padrões que permitam concluir quais fatores interferem no desempenho das escolas na avaliação do Ideb (Silva et al., 2020).

Percebeu-se que utilização da análise fatorial multivariada como ferramenta de mineração de dados educacionais oriundos de avaliações em larga escala nos moldes aqui apresentados, não deve ser aplicada de maneira indiscriminada. Hair *et al.* (2005) demonstraram que o uso dessa técnica é viável, desde que sejam cumpridos e verificados determinados critérios, como por exemplo, a normalidade dos dados, a exclusão de *outliers*, o teste de esfericidade de Bartlett e o valor KMO. Além disso, a escolha de variáveis dentre os questionários do Saeb, deve ser feita por profissionais da área da educação que saibam discernir sobre qual questão deve ser agrupada inicialmente e a partir da análise fatorial confirmar estes pressupostos. Contudo, a correlação das notas do Ideb com os questionários do Saeb, mostrou-se possível, mesmo havendo registros de manipulação da aplicação destes testes por parte de algumas escolas, pois estes eventos podem ser dirimidos por meio de técnicas estatísticas.

Além da escolha das questões, a dicotomização das escalas de forma adequada foi atividade mandatória na aplicação do método. Tendo em vista a variação das escalas dentre as respostas dos questionários do Saeb e notas do Ideb, que ora são questões com alternativas de 'a' a 'd', ora de 'a' a 'e', ou ainda verdadeiro ou falso, além de ora existirem questões que a resposta positiva é evidenciada seguindo de forma decrescente da 'a' a 'e', ora ocorre o contrário. O fato de os dados serem dicotomizados, obriga o pesquisador a realizar vários métodos de extração fatorial, a partir de diversos métodos de cálculo para a obtenção das matrizes de correlação e ainda considerando sua rotação. Neste caso, mostrou-se pertinente a extração fatorial híbrida, a partir da correlação de Pearson associada à correlação tetracórica, ambas não rotacionadas.

Referências

CLEMENT, Luiz et al. Motivação autônoma de estudantes de física: evidências de validade de uma escala. *Psicologia Escolar e Educacional*, São Paulo, v. 18, n. 1, 2014, p. 45-56.

ENGELMANN, Erico. *A motivação de alunos dos cursos de artes de uma universidade pública do norte do Paraná*. Londrina: UEL, 2010. 124f. Dissertação (Mestrado em Educação) - Universidade Estadual de Londrina.

GATTI, Bernardete Angelina. Avaliação educacional no Brasil: pontuando uma história de ações. *Eccos Revista Científica*, São Paulo, v. 4, 2002, p. 17-41.

HAIR, Joseph et al. *Análise multivariada de dados*. Porto Alegre: Bookman, 2005.

- IBGE. *Malha municipal 2020*. Disponível em: <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/15774-malhas.html?edicao=30138&t=acesso-ao-produto>. Acesso em: 14 dez. 2021.
- INEP. *Nota técnica do indicador de nível socioeconômico (Inse)*. Brasília: Inep, 2015. Disponível em: https://download.inep.gov.br/informacoes_estatisticas/indicadores_educacionais/2015/nota_tecnica/nota_tecnica_inep_inse_2015.pdf. Acesso em: 1 ago. 2022.
- INEP. *Testes e questionários*. Disponível em: <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/saeb/testes-e-questionarios>. Acesso em: 23 set. 2020.
- LIRA, Sachiko Araki; CHAVES NETO, Anselmo. Pontos de dicotomização para a obtenção do coeficiente de correlação tetracórico. *Ciências Exatas e Naturais*, Guarapuava/Irati, v. 7, 2005, p. 195-208.
- MATOS, Daniel Abud Seabra; RODRIGUES, Erica Castilho. *Análise fatorial*. Brasília: Enap, 2019.
- MELLO, Liliane Ribeiro de; BERTAGNA, Regiane Helena. Tensões do Saeb e do Ideb para a educação de qualidade como direito. *Revista Educação em Questão*, Natal, v. 58, 2020, p. 1-24.
- NUNES, Bárbara Cristine Bottino. *Os desafios da gestão democrática através do conselho escolar: um estudo baseado no questionário do Saeb*. Santa Maria: UFSM, 2018. 24f. Monografia (Especialização em Gestão Escolar. Universidade Federal de Santa Maria.
- SILVA, Denise dos Santos Vasconcelos; SOUSA, Francisco Cavalcante de. Direito à educação igualitária e(m) tempos de pandemia: desafios, possibilidades e perspectivas no Brasil. *Revista Jurídica Luso-Brasileira*, Lisboa, v. 4, 2020, p. 961-979.
- R CORE TEAM. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing, 2020.
- RISCAL, José Reinaldo; LUIZ, Maria Cecília. *Gestão democrática e a análise de avaliações em larga escala: o desempenho de escolas públicas no Brasil*. São Carlos: Píxel, 2016.
- RISCAL, Sandra Aparecida; RISCAL, José Reinaldo. A contribuição do Ideb para o estudo do papel da democratização da escola pública na melhoria da qualidade do ensino. *@mbienteeducação*, São Paulo, v. 7, n. 2, 2014, p. 525-536.
- SILVA, Ivonaldo Vicente da et al. Fatores preditivos de desempenho escolar em avaliações do Saeb: influência da gestão escolar. *Research, Society and Development*, Vargem Grande Paulista, v. 9, n. 10, 2020, p. 1-35.
- TRAVITZKI, Rodrigo. Qual é o grau de incerteza do Ideb e por que isso importa? *Ensaio: Avaliação e Políticas Públicas em Educação*, Rio de Janeiro, v. 28, 2020, p. 500-520.

Simone Mágná Menezes Carneiro Lopes é professora na Universidade Federal do Tocantins.

Orcid: <https://orcid.org/0000-0001-7625-6721>.

Endereço: Quadra 404 Norte, Alameda 28, Lote 01-A - 77006-450 - Palmas - TO - Brasil.

E-mail: simone.lopes@uft.edu.br.

José Carlos da Silveira Freire é professor na Universidade Federal do Tocantins.
Orcid: <https://orcid.org/0000-0002-6798-6164>.
Endereço: Quadra 109 Norte, Avenida NS 15, Bloco III, Sala 29 - 77001-090 -
Palmas - TO - Brasil.
E-mail: cfreire@uft.edu.br.

Critérios de autoria: Simone Mágnã realizou a concepção, execução, análise, interpretação, redação e consolidação do artigo. José Carlos realizou a supervisão do planejamento e execução da atividade de pesquisa.

Recebido em 7 de julho de 2022.

Aceito em 25 de setembro de 2022.

