

USO DA ANÁLISE MULTIVARIADA (PCA) E DA ÁLGEBRA MULTILINEAR (PARAFAC) PARA AVERIGUAR A ASSOCIAÇÃO DO CLIMA ESCOLAR COM O DESEMPENHO ACADÊMICO DE ALUNOS

Luzia Queiroz Hippolyto¹
Wagner Bandeira Andriola²
Albano Oliveira Nunes³

RESUMO

O estudo apresenta resultados decorrentes do uso de procedimentos de análises multivariada e multilinear para a modelagem matemática, com o fito de identificar fatores do clima escolar associados ao desempenho acadêmico de alunos por meio de dois modelos de decomposição de dados: o modelo bilinear PCA (*Principal Component Analysis*) e o modelo trilinear PARAFAC (*Parallel Factor Analysis*). Empregou-se microdados de 2400 alunos do Ensino Médio de escolas do município de Fortaleza (Ceará, Brasil) com respostas ao questionário de Clima Institucional Escolar (CIE). Os resultados obtidos foram promissores, indicando a viabilidade da aplicação de ambos os modelos nas Ciências Humanas, havendo destaque para o ganho obtido através da decomposição tensorial na interpretação dos fatores extraídos na investigação das dinâmicas educacionais.

Palavras-chave: Análise Multivariada, Análise Multilinear, Decomposição Tensorial, Avaliação Educacional.

USE OF MULTIVARIATE ANALYSIS (PCA) AND MULTILINEAR ALGEBRA (PARAFAC) TO CHECK THE ASSOCIATION OF SCHOOL CLIMATE WITH STUDENTS' ACADEMIC PERFORMANCE

ABSTRACT

The study presents results arising from the use of multivariate and multilinear analysis procedures for mathematical modeling, with the aim of identifying school climate factors associated with students' academic performance through two data decomposition models: the PCA bilinear model (*Principal Component Analysis*) and the trilinear model PARAFAC (*Parallel Factor Analysis*). Microdata from 2400 high school students from schools in the city of Fortaleza (Ceará, Brazil) were used with responses to the Institutional School Climate (CIE) questionnaire. The results obtained were promising, indicating the viability of applying both models in Human Sciences, highlighting the gain obtained through tensor decomposition in the interpretation of factors extracted in the investigation of educational dynamics.

Keywords: Multivariate Analysis, Multilinear Analysis, Tensor Decomposition, Educational Evaluation.

¹ Mestre em Matemática e Doutora em Educação Brasileira. Professora do Sistema Estadual de Educação do Ceará e do Instituto Federal de Brasília (IFB). E-mail: luziaufc@hotmail.com.

² Doutor em Filosofia e Ciências da Educação (*Universidad Complutense de Madrid*). Bolsista de Produtividade em Pesquisa 1B. E-mail: wagner.andriola@pq.cnpq.br / ORCID: 0000-0002-5527-0097.

³ Doutor em Engenharia de Teleinformática (UFC); Mestre em Ensino de Ciências Naturais e Matemática (UFRN); Pós-doutor em Ensino (UERN). Professor do Sistema Estadual de Educação do Ceará. Professor do Centro Universitário do Vale do Jaguaribe (UniJaguaribe). Professor Colaborador do POLEDUC (UFC). Membro da Rede-Ter. Membro da Sociedade Brasileira de Física. E-mail: albano@fvj.br. ORCID: 0000-0002-6543-2512.

1 INTRODUÇÃO

No uso de ferramentas matemáticas para a análise da mente humana, o destaque inicial pode ser creditado ao psicólogo inglês, *Sir Charles Edward Spearman*, um dos gigantes a incitar diversas pesquisas nessa área, muito conhecido pelo seu trabalho na área da estatística (ANDRIOLA, 2002b). Foi pioneiro no emprego da análise fatorial e o criador do coeficiente de correlação de postos de Spearman. Em 1904, publicou notório artigo sobre a análise fatorial da inteligência, introduzindo relevantes considerações sobre o conceito de variável latente (PASQUALI, 2003).

A análise fatorial sofreu avanços significativos na década de 1930, com as formulações teóricas acerca da inteligência humana, propostas pelo Engenheiro e Doutor em Psicologia, Louis Leon Thurstone, que recusava a existência de uma inteligência geral, opinando que existiriam múltiplas aptidões (ANDRIOLA; PASQUALI, 1995). Ao aplicar testes de aptidão, os dados foram submetidos à análise fatorial e revelaram sete aptidões primárias: aptidões espaciais e visuais, rapidez perceptual, aptidão numérica, compreensão verbal, memória, fluidez verbal e raciocínio (ANDRIOLA, 1997). Após esse estudo, L. L. Thurstone concluiu que os indivíduos podem ter aptidões distintas, em áreas diversificadas do intelecto humano (ANDRIOLA; CAVALCANTE, 1999).

Todavia, com o advento da computação e dos computadores, nos anos 1960, que permitiam a execução de operações matemáticas de grande porte, com velocidade e precisão até então inexistentes, os cálculos e as transformações de matrizes de autovalores e auto-vetores tornaram-se menos complicadas e mais ágeis (ANDRIOLA, 1995). Como a maior parte das descobertas no ramo das ciências, o desenvolvimento dos métodos para a redução de dados foi conduzido a partir da necessidade de se analisar conjuntos de informações com muitas variáveis correlacionadas (ANDRIOLA, 2009a). Nas últimas três décadas, as técnicas analíticas multivariadas têm sido fortemente utilizadas em estudos de diversificados campos do saber, com o intuito de estudar o comportamento de duas ou mais variáveis simultaneamente (VASCONCELOS et al., 2013; ANDRIOLA, 2002a; ANDRIOLA, 2009b).

Dessa forma, o presente trabalho destaca o emprego da Análise de Componentes Principais ou PCA (*Principal Componente Analysis*), como uma potencial ferramenta na extração das características latentes de um conjunto de dados analisados, estendendo-se, ainda, para uma análise multilinear, com destaque para o método de decomposição tensorial em Fatores Paralelos (PARAFAC). Assim sendo, a ideia central do presente estudo consiste em

verificar quais fatores de uma escala empregada para avaliar o Clima Institucional Escolar (CIE) estão associados ao desempenho escolar de alunos do ensino médio. Nesta perspectiva indaga-se: *o PARAFAC e o PCA são ferramentas úteis para identificação desses fatores ou componentes?* Diante dessa indagação, a proposta do estudo é metodológica, uma vez que se propõe a comparar o achado resultante da mineração dos dados a partir do emprego do PCA e do PARAFAC, referentes às respostas de 2400 alunos do ensino médio, oriundos de escolas públicas, convencionais e profissionais, sediadas no município de Fortaleza (CE). O próximo tópico descreverá a base matemática dessas duas ferramentas estatísticas.

2 BREVE RETROSPECTIVA HISTÓRICA ACERCA DA ANÁLISE MULTIVARIADA

2.1 Conceituação e tipologia

Como supracitado, a tecnologia computacional possibilitou que muitas informações de natureza psicológica e educacional fossem armazenadas e, por conseguinte, tratadas analiticamente, de forma adequada (ANDRIOLA; BARRETO, 1997). Na posse de uma grande quantidade de informações um questionamento que surge ao pesquisador é: *Como interpretá-las e, obedecendo à natureza multivariada, como extrair informações que sejam relevantes à sua correta compreensão?*

A análise multivariada consiste em um conjunto de métodos que permite a compreensão simultânea de medidas múltiplas para cada indivíduo ou objeto (TABACHNICK; FIDELL, 2019). Consoante Hairet al. (2005), não é fácil definir análise multivariada, pois de um modo geral, ela refere-se a todos os métodos estatísticos que simultaneamente analisam múltiplas medidas sobre cada indivíduo ou objeto de investigação. Salientam, ainda:

“muitas técnicas multivariadas são extensões de técnicas da análise univariada e da análise bivariada. Outras técnicas, contudo, são unicamente projetadas para lidar com questões multivariadas” (HAIR et al., 2005, p.26).

Os métodos multivariados são escolhidos de acordo com os objetivos da pesquisa. Nesse sentido, Tabachnick e Fidell (2019) afirmam que a estatística multivariada conta com diversos métodos, em que cada um tem sua própria fundamentação matemática e aplicabilidade. Assim, como um dos intuits da pesquisa é exemplificar o uso da Análise de Componentes Principais (PCA) como modelo de decomposição bilinear eficiente, ele será discutido com maior ênfase na seção seguinte.

2.1.1 Análise de Componentes Principais (PCA)

O PCA foi introduzido no campo da pesquisa psicológica por Karl Pearson, em 1901, com o objetivo de encontrar linhas e planos que melhor se ajustassem a um conjunto de pontos em um espaço p -dimensional (FANG; KUANG; QIAO, 2017). No início de 1930, o PCA foi desenvolvido por Harold Hotelling, tendo sido, posteriormente, rotulada de a Transformada de Hotelling (SILVA, 2018).

O PCA tem por finalidade transformar um conjunto de dados de modo a reduzir sua dimensionalidade possibilitando escolhas mais representativas a partir de combinações lineares das variáveis originais (HOTELLING, 1933). Assim, estando os dados correlacionados, a partir da aplicação do método do PCA, gera-se um conjunto de variáveis descorrelacionadas, o que possibilita uma compreensão mais apurada acerca de relações não tão claras contidas nesse conjunto de informações. É importante salientar que, de modo prático, o algoritmo baseia-se na matriz de variância-covariância, ou ainda na matriz de correlação, da qual serão extraídos os autovalores e autovetores (KOLDA; BADER, 2009).

Matematicamente o PCA é definido como a transformação linear ortogonal em que o conjunto de dados original é modificado para um novo sistema de coordenadas, de modo que a maior variância, por qualquer projeção dos dados, seja alocada ao longo da primeira coordenada, a segunda maior variância se aloque ao longo da segunda coordenada, e assim sucessivamente (SILVA et al. 2013).

Dessa forma, eliminando a redundância de informações com peso desprezível, por meio de uma decomposição em base linear ortogonal, as variáveis latentes serão destacadas. Sob esta ótica, as componentes principais $y_1, y_2, y_3, \dots, y_n$, são definidas por combinações lineares, não correlacionadas, das variáveis observadas originais $x_1, x_2, x_3, \dots, x_p$.

Ressalta-se que, em ordem decrescente, as componentes principais explicam proporções máximas da variação dos dados originais. Ou seja: y_1 explica o máximo da variância de todas as combinações lineares de $x_1, x_2, x_3, \dots, x_p$. Já y_2 explica o máximo da variância residual não explicada pela primeira componente e ao mesmo tempo não correlacionada, isto é, ortogonal a y_1 e, assim por diante.

Portanto, algebricamente, as componentes principais são calculadas com base em combinações lineares das variáveis originais, conforme as equações 1 a 3.

$$y_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p \quad (1)$$

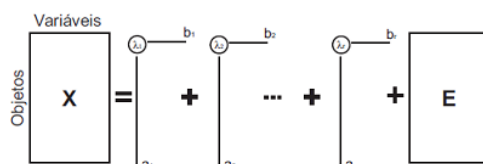
$$y_2 = a_{21}x_1 + a_{22}x_2 + \dots + a_{2p}x_p \quad (2)$$

$$y_n = a_{n1}x_1 + a_{n2}x_2 + \dots + a_{np}x_p \quad (3)$$

A soma dos quadrados dos coeficientes é 1, pois utiliza-se a variância total das variáveis observadas, retendo-se todas as informações das variáveis originais, sendo derivadas as componentes ortogonais. Entretanto, dos p componentes principais calculados pelo método multivariado, pode-se reter os n componentes principais que explicam, aproximadamente, a variabilidade total das p variáveis, com $n < p$.

A Figura 1 traz a representação pictórica do PCA, concebida como uma linguagem mais fácil, de comunicação visual.

Figura 1- Representação Pictórica do PCA



Fonte: Silva *et. al* (2013).

Analogamente, de forma algébrica a PCA pode ser escrita como uma entrada arbitrária de x_{ij} :

$$x_{ij} = \sum_{r=1}^n \lambda a_{ir} b_{jr} + e_{ij}$$

Assim, a matriz original é decomposta em uma adição de produtos externos associados a um autovalor. Salienta-se que os autovalores indicarão a importância dos vetores associados ao PCA de acordo com os dados da base original.

No que tange ao PARAFAC, é extensão baseado na PCA, com a finalidade de possibilitar inovações nos estudos envolvendo a análise multivariada, em diversas áreas do conhecimento humano. As discussões mais profundas sobre essa decomposição trilinear serão tratadas mais adiante.

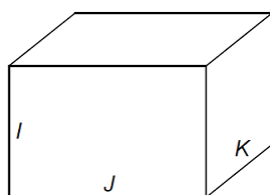
2.2 Álgebra Multilinear

Na álgebra multilinear, o conceito de decomposição tensorial foi introduzido por Frank L. Hitchcock em 1927, conforme Kolda e Bader (2009). Em 1944, as noções dessa decomposição foram aprofundadas por Raymond Bernard Cattell que introduziu o conceito de *perfis paralelos proporcionais* em relação à rotação de soluções da análise fatorial (CATTELL, 1944). Já Ledyard R. Tucker propôs uma forma para calcular os parâmetros de um modelo que se tornou uma forte ferramenta para análise de três vias de matrizes de dados (TUCKER, 1963). A generalidade do modelo Tucker3 e o fato de que ele cobre o modelo PARAFAC, como um caso especial, tornou um procedimento frequentemente utilizado para a decomposição, compressão e interpretação de dados multivariados.

Todavia, os interesses pelos modelos de decomposição tensorial se intensificaram na década de 1970, quando J. Douglas Carrol e Jih-Jie Chang desenvolveram a decomposição Canônica denominada CANDECOMP que consistia em uma decomposição tridimensional em que cada componente do tensor poderia ser representado por uma composição de três vetores (CARROL; CHANG, 1970). Richard A. Harshman, no mesmo período, mas de forma independente ao trabalho de Carrol e Chang (1970), desenvolveu o modelo PARAFAC (caso particular do modelo Tucker3), similar ao CANDECOMP, que ordena os dados e decompõe a soma de produtos triplos, com alternância de procedimentos dos mínimos quadrados (HARSHMAN, 1970).

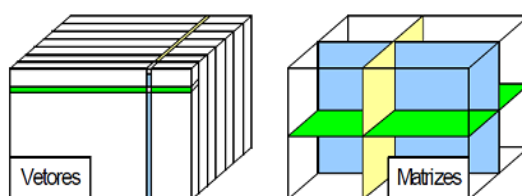
Kolda e Bader (2009) definem um tensor como um ordenamento de dados multidimensionais que varia de acordo com a ordem. Assim, um tensor de primeira ordem é um vetor; de segunda ordem é uma matriz; tensores de ordem $n \geq 3$ são chamados *tensores de ordem superior*. A estrutura tensorial é uma organização natural da informação para o caso em que a informação possui características multidimensionais. Nesse sentido, o tensor é uma forma de representar um conjunto de dados com características multivariadas.

Contudo, assim como em outras formas de decomposição, por vezes uma reorganização do conjunto de dados é necessária. É comum a utilização de uma organização matricial das variáveis (atributos) presentes em um tensor. Essa reorganização é em forma de um desdobramento matricial, uma vez que os arranjos matriciais são facilmente tratados nos programas computacionais. A Figura 2 apresenta uma estrutura tensorial de ordem 3, em que $I = 1, 2, \dots, I$; $J = 1, 2, \dots, J$ e $K = 1, 2, \dots, K$.

Figura 2 – Estrutura tensorial de ordem três.

Fonte: Adaptado de Kolda e Bader (2009).

Nas representações matriciais, também conhecidas como *slices* ou fatias, têm-se três conjuntos, a saber: conjunto das matrizes na horizontal, matrizes laterais e matrizes frontais, consoante a Figura 3, a seguir, cuja determinação de tais variáveis dará origem aos dados de três vias.

Figura 3– Tensores de ordem três.

Fonte: Adaptado de Lacerda (2005).

Conforme Lacerda (2005), a análise tensorial oferece soluções que possuem manipulações matemáticas simples e de fácil interpretação, cuja estrutura oferece vantagens em relação às decomposições matriciais. Prossegue, asseverando que:

“Algumas estruturas tensoriais são únicas, implicando na não necessidade de imposições de constantes e/ou restrições ao modelo. Considerando o fato de que o modelo é apropriado para a organização dos dados, o uso da estrutura tensorial implica em uma organização parcimoniosa. Essa organização apresenta uma riqueza de informação com relação aos efeitos multivariáveis, facilitando o tratamento e a interpretação dos dados, além de fornecer maiores possibilidades de explorá-los” (LACERDA, 2005, p. 12).

A seguir apresenta-se breve descrição do PARAFAC, por ser um método eficaz para o tratamento de dados com o intuito de realizar análises exploratórias, identificação de componentes e calibrações, sendo de interesse para a presente pesquisa.

2.2.1. Decomposição em Fatores Paralelos (PARAFAC)

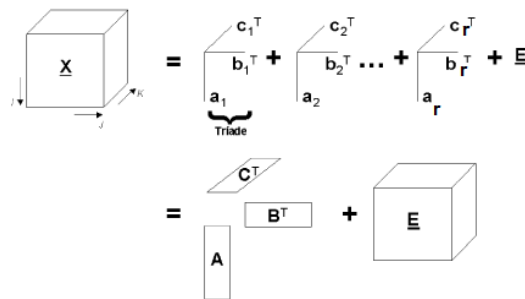
Conforme Bro (1997), o PARAFAC consiste na decomposição de dados multilinear (multidimensional ou *multi-way*). Assim, por exemplo, na decomposição de ordem três, a

informação contida no tensor é decomposta em componentes trilineares da forma x_{ijk} e, dessa decomposição, surgirão três matrizes A, B e C, com os respectivos elementos $a_{i,r}$, $b_{j,r}$ e $c_{k,r} \forall X$ (tensor) que possuem dimensões I x J x K (Vide Equação 4).

$$x_{i,j,k} = \sum_{r=1}^R a_{i,r} b_{j,r} c_{k,r} \quad (4)$$

Vasconcelos *et al.* (2013) afirmam que o modelo trilinear deve ser adotado para minimizar a soma dos quadrados dos resíduos $e_{i,j,k}$, utilizando o algoritmo de mínimo quadrados alternados (ALS). A Figura 4 representa o modelo PARAFAC.

Figura 4 – Decomposição de arranjo tridimensional em r tríades de vetores peso.



Fonte: Adaptado de Sena, Trevisan e Poppi (2005).

Como se pode observar, a decomposição fatoriza um tensor em uma soma de componentes *rank-one* tensores, com muita semelhança entre o PCA e o PARAFAC, posto que o PARAFAC é uma extensão do PCA em n dimensões (ZARI, OZDEMIR, IWEN, AVIYENT, 2018). A seguir, será descrito o método empregado no presente estudo.

3 METODOLOGIA

Tratou-se de estudo *ex-post facto* que empregou microdados de 2400 alunos do Ensino Médio de escolas do município de Fortaleza (Ceará) com as respostas ao questionário de Clima Institucional Escolar (CIE). O referido instrumento possui 40 itens com escala de resposta do tipo Likert de quatro níveis: 1 - Concordo Totalmente; 2 - Concordo Pouco; 3 - Discordo Pouco; 4 - Discordo Totalmente. O número de discentes partícipes do estudo respeita a razão 10 : 1 (nº de sujeitos por cada item do questionário), valor este que é o recomendado para a realização de uma análise fatorial do tipo exploratória, consoante Hill e Hill (2002), Kahn (2006), Worthingtton e Whittaker (2006).

3.1 Fases do Estudo

A pesquisa seguiu as seguintes etapas: i) Definição da amostra, que totalizou 2400 estudantes do Ensino Médio de escolas públicas; ii) Aplicação do PCA ao conjunto de respostas dos alunos ao CIE; iii) Emprego da análise tensorial via PARAFAC ao conjunto de respostas dos alunos ao CIE; iv) Comparação dos resultados tanto no que concerne aos tipos de escolas (escolas convencionais e escolas profissionais) como nas formas de decomposição (bilinear e trilinear), utilizando-se das técnicas de análise multivariadas e multilinear.

4 RESULTADOS ORIUNDOS DAS ANÁLISES DE COMPONENTES PRINCIPAIS (PCA)

Na validação do questionário de Clima Institucional Escolar (CIE) utilizou-se o PCA para determinar o número de componentes a reter na solução inicial, o número de itens que lhes estão associados e a consistência interna dos mesmos, através do SPSS (versão 21.0). Na primeira extração via PCA, observou-se que os itens 43, 45, 49 e 51 possuíam peso fatorial inferiores a 0,4 ensejando a retirada destes das análises posteriores.

Os primeiros resultados encontrados são expostos na matriz de correlação, cujo objetivo é mostrar a magnitude da associação de uma determinada variável em detrimento de outra variável observada. Altas correlações estão associadas ao fato de que as variáveis envolvidas estão sob influência de um mesmo fator. O determinante dessas matrizes necessita ser maior que zero, uma vez que se não, não haverá solução analítica, pois ele define se dada matriz quadrada terá inversa. Ressalta-se que a matriz não invertível é denominada de matriz singular⁴ e a sua ocorrência indica um sistema de matrizes instável.

Se o determinante da matriz for nulo então existe pelo menos uma dependência linear na matriz. Quando isso ocorre a matriz deixa de ser quadrada. O resultado encontrado, no presente estudo, foi: determinante da matriz cuja base era a escola convencional = 1,21E007 e determinante da matriz cuja base era a escola profissional = 5,58E007. Ambos os valores validam o uso do PCA para este caso.

Em seguida, calculou-se o KMO e o teste de esfericidade de Bartlett. O teste Kaiser- Meyer- Oklin (KMO) avalia a existência de um número suficiente de correlações significativas entre os itens para que seja realizado o uso da análise fatorial, constituindo uma

⁴ Uma [matriz](#) é singular [se e somente se](#) seu [determinante](#) é nulo. Uma matriz é singular se somente se existir um [vetor](#) x não nulo tal que: $Ax = 0Ax = 0$

medida global que indica a força da relação entre itens, por meio de correlações parciais, que representam as correlações entre cada par de itens, após o efeito linear de todos os outros itens. A medida de adequação da amostra do KMO deve ser superior a 0,70, sendo considerada inadequada KMO inferior a 0,60.

No que tange ao teste de esfericidade de Bartlett, este testa a hipótese nula de que a matriz de correlação seja uma matriz identidade. Quanto maiores os valores do teste de Bartlett, maior a probabilidade de que a matriz de correlação não seja matriz identidade conduzindo a rejeição da hipótese nula. Para tal, como mencionado, é recomendado que o valor do teste de KMO seja superior a 0,6. A análise resultou em $KMO = 0.940$, enquanto o valor do Teste *Bartlett* = 0,000. Os resultados do KMO e do Teste de Bartlett apontam para a adequação dos dados ao uso da PCA.

Vale a pena realçar que a PCA é utilizada como técnica de redução de dimensionalidade de dados multivariados em componentes independentes (VIEIRA; RIBAS, 2011). Assim, para Vianna (1968), a primeira tarefa é verificar se as escalas que definem as variáveis possuem fidedignidade ou reprodutibilidade, de modo que as variáveis possam ser reproduzidas em futuros estudos. Nesta etapa utilizou-se o Alfa de Cronbach para calcular a fidedignidade e verificou-se alto grau de confiabilidade das escalas de tal ordem que ele não pode ser melhorado, sendo encontrado $\alpha = 0,863$ para uma estrutura com 36 itens.

Segundo as orientações de Blunch (2008), Brown (2006), Hair *et al.* (2006), Henson e Roberts (2006), Kahn (2006), Preacher e MacCallum (2003;), os critérios de extração dos fatores utilizados foram: a) Fatores com autovalor igual ou superior a 1 (*eigenvalue* $\geq 1,0$); b) Peso fatorial dos itens igual ou superior a 0,4 (*factorloadings* ≥ 0.40); c) A variância explicada pelos fatores retidos deve ser no mínimo de 40%; d) A consistência interna do fator deve ser igual ou superior a 0,70 (alfa de Cronbach $\geq 0,70$); e) A consistência interna do fator não deve aumentar se o item for eliminando; f) Cada fator deve ter pelo menos três itens.

As informações seguintes se referem à comunalidade, ou seja, a proporção da variância de uma variável observada pelos fatores extraídos. As suas estimativas variam de 0 a 1. Um valor elevado, nesse intervalo, indica que os fatores extraídos explicam proporções elevada da variância de determinada variável observada. Comunalidade zero implica que nenhuma porção da variância é explicada pelos fatores extraídos.

Tabela 1: Comunalidades dos itens da ...

Itens	Escolas Convencionais	Escolas Profissionais	Itens	Escolas Convencionais	Escolas Profissionais
RQ_000_024	,273	,383	RQ_000_042	,439	,444
RQ_000_025	,445	,476	RQ_000_044	,305	,246
RQ_000_026	,486	,457	RQ_000_046	,321	,349
RQ_000_027	,378	,403	RQ_000_047	,361	,248
RQ_000_028	,376	,381	RQ_000_048	,180	,266
RQ_000_029	,370	,343	RQ_000_050	,306	,189
RQ_000_030	,310	,282	RQ_000_052	,370	,378
RQ_000_031	,180	,162	RQ_000_053	,450	,372
RQ_000_032	,381	,410	RQ_000_054	,364	,386
RQ_000_033	,402	,356	RQ_000_055	,430	,413
RQ_000_034	,651	,654	RQ_000_056	,393	,405
RQ_000_035	,603	,623	RQ_000_057	,551	,549
RQ_000_036	,661	,661	RQ_000_058	,417	,375
RQ_000_037	,592	,618	RQ_000_059	,271	,384
RQ_000_038	,588	,627	RQ_000_060	,510	,518
RQ_000_039	,418	,527	RQ_000_061	,538	,522
RQ_000_040	,410	,429	RQ_000_062	,410	,339
RQ_000_041	,393	,462	RQ_000_063	,415	,459

Fonte: Pesquisa direta. Extração pelo Método da Análise de Componentes Principais.

As Tabelas 2 e 3 apresentam a variância explicada para as escolas convencionais e profissionais, respectivamente, a qual é composta pelos autovalores (*eigenvalues*), refere-se à variância explicada por cada fator são exibidos os percentuais de variância explicados por cada componente, antes e após o procedimento da rotação. Depois da rotação só são exibidos os autovalores superiores a 1, uma vez que o componente cujo autovalor é inferior a 1 explica menos do que seria explicado por uma variável observada.

Tabela 2: Variância explicada com 3 componentes Escolas Convencionais

Componente	TESTES								
	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% de Variância	% Acumulado	Total	% de Variância	% Acumulado	Total	% de Variância	% Acumulado
1	10,16	28,23	28,23	10,16	28,23	28,23	7,40	20,57	20,57
2	2,53	7,03	35,26	2,53	7,03	35,26	4,76	13,23	33,81
3	2,25	6,26	41,53	2,25	6,26	41,53	2,77	7,71	41,53

Fonte: Extração pelo Método da Análise de Componentes Principais, (2017).

Tabela 3: Variância explicada com 3 componentes Escolas Profissionais

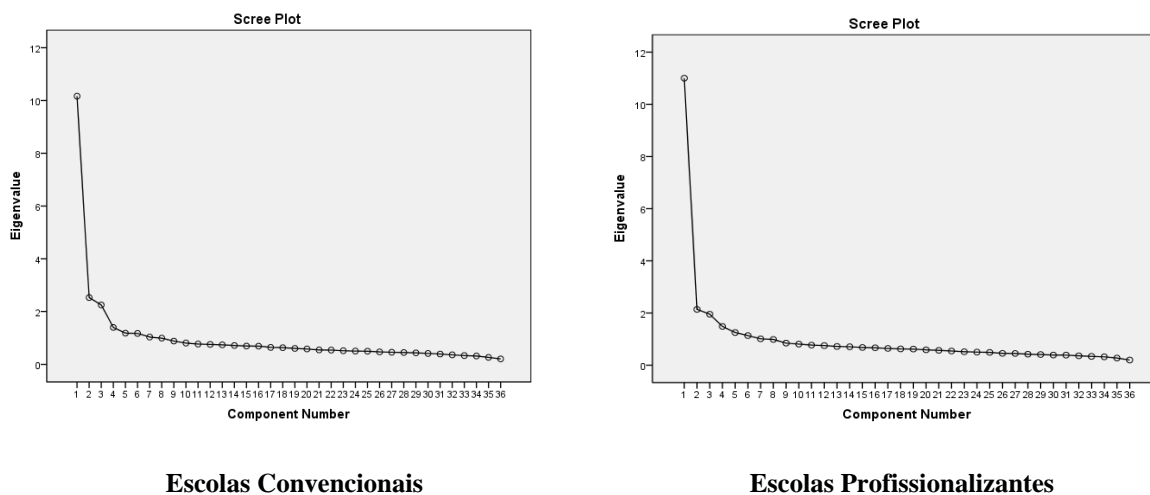
Componente	TESTES								
	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% de Variância	% Acumulado	Total	% de Variância	% Acumulado	Total	% de Variância	% Acumulado
1	11,00	30,56	30,56	11,00	30,56	30,562	7,41	20,60	20,60
2	2,13	5,94	36,50	2,13	5,94	36,502	4,60	12,77	33,37
3	1,95	5,43	41,93	1,95	5,43	41,93	3,07	8,55	41,93

Fonte: Extração pelo Método da Análise de Componentes Principais, (2017).

Um gráfico *Scree* exibe os autovalores associados a um componente ou fator em ordem decrescente versus o número do componente ou fator, que pode ser usado para avaliar

visualmente quais componentes ou fatores explicam a maior parte da variabilidade dos dados. A seguir os gráficos mostrados na Figura 6 apresentam os componentes extraídos para os dois tipos de escola.

Figura 6 - Componentes extraídos nos dois tipos de Escolas.



Do exposto, constata-se que, nos dois casos de escolas, sete componentes teriam autovalores maiores do que 1. Contudo, a partir do quarto componente extraído a variância explicada era pouco ampliada. Nesse sentido, optou-se pela extração de três componentes principais. As Tabelas 4 e 5 apresentam a matriz rotacionada que exibe as variáveis agrupadas por cargas elevadas entre as componentes.

Tabela 4: Componentes da matriz rotacionada - Escolas Convencionais

Itens	Componente 1	Componente 2	Componente 3
RQ_000_024	0.47509070618518456		
RQ_000_025	0.622903333436559		
RQ_000_026	0.6139694679592016	0.3098029064408474	
RQ_000_027	0.5850337343345172		
RQ_000_028	0.5845146330157432		
RQ_000_029	0.5968372105585038		
RQ_000_030	0.548602644949667		
RQ_000_031	0.3922069514001406		
RQ_000_032	0.5921380360275578		
RQ_000_033	0.5970358643937335		
RQ_000_034	0.7691546087215148		
RQ_000_035	0.7319733778677517		
RQ_000_036	0.785408798700705		
RQ_000_037	0.7367001593464536		
RQ_000_038	0.7202955622597006		
RQ_000_039	0.5487896163785826	0.34218404137182007	
RQ_000_040	0.5759800989959784		
RQ_000_041			0.6152137082631259
RQ_000_042			0.6445271788675109
RQ_000_044			0.5486019465073287

RQ_000_046			0.5505640146564961
RQ_000_047			0.5970485074890218
RQ_000_048			0.4078259859934241
RQ_000_050			0.5378703203910663
RQ_000_052		0.5728695503602864	
RQ_000_053		0.6360028519630674	
RQ_000_054			0.5430027527986383
RQ_000_055		0.5889412417377463	
RQ_000_056		0.6006246401585945	
RQ_000_057		0.7171752632698738	
RQ_000_058		0.611311518042453	
RQ_000_059		0.45258303878230993	
RQ_000_060		0.666966163898796	
RQ_000_061	0.4438101295107118	0.5843525718460837	
RQ_000_062		0.5638844044325686	
RQ_000_063	0.3515925623953414	0.5300553478803731	

Fonte:

Extração pelo Método da Análise de Componentes Principais, (2017).

Tabela 5: Componentes da matriz rotacionada - Escolas Profissionais

Itens	Componente 1	Componente 2	Componente 3
RQ_000_024	0.5620015070947355		
RQ_000_025	0.677628735779997		
RQ_000_026	0.6018124485317136		
RQ_000_027	0.6137654100082232		
RQ_000_028	0.5344991307020964		
RQ_000_029	0.5207279470900232		
RQ_000_030	0.495969806758741		
RQ_000_031	0.3857812792267626		
RQ_000_032	0.5847986052530764		
RQ_000_033	0.4946454523362015		
RQ_000_034	0.7219968035618122		
RQ_000_035	0.702431468052825		
RQ_000_036	0.7688839941466579		
RQ_000_037	0.7417785179687809		
RQ_000_038	0.739599382795158		
RQ_000_039	0.6593303374061165		
RQ_000_040	0.5768609891588838		
RQ_000_041			0.6512623193674993
RQ_000_042			0.6213622133651667
RQ_000_044			0.4769553836626121
RQ_000_046			0.542553317024451
RQ_000_047			0.4768673168569991
RQ_000_048			0.4686717135222495
RQ_000_050			0.41042463702506804
RQ_000_052		0.4910283701861037	-0.3046306899328078
RQ_000_053		0.5479985141192947	
RQ_000_054			0.5666378263764642
RQ_000_055		0.6034759788967805	
RQ_000_056		0.5806827036481298	
RQ_000_057		0.6773848761666805	
RQ_000_058		0.5627264191262281	
RQ_000_059		0.5323473227734756	
RQ_000_060		0.6602007492750368	
RQ_000_061	0.4299781879227903	0.5676539082955867	
RQ_000_062		0.518611120255098	
RQ_000_063	0.3380753179682927	0.5838235337102634	

Fonte: Extração pelo Método da Análise de Componentes Principais, (2017).

As informações possibilitam a interpretação da análise dos componentes. As cargas fatoriais observadas são distribuídas pelos componentes, muito embora exista um componente, no qual se observa carga mais elevada. As cargas significam o peso de cada variável na escala, isto é, o que mais fortemente influencia determinada variável. Na realidade, cada fator influencia um conjunto de variáveis, permitindo que seja escolhido um título que melhor reflita seu significado. O procedimento de rotação escolhido foi o *Varimax* por possibilitar se encontrar a maior variabilidade possível. O valor 0,30 constitui na prática boa fronteira entre cargas reduzidas e elevadas e, por isso, foi utilizado para eliminar as cargas fatoriais abaixo desse valor. As categorizações, bem como as variáveis pertencentes a cada um dos fatores extraídos estão apresentados na Tabela 6. As estruturas de extração dos itens se mantêm, independentemente do tipo de escola que está sendo estudado. Assim, apresentou-se uma única tabela, a qual valerá para as escolas convencionais e para as escolas profissionais.

Tabela 6: Identificação das Componentes extraídas. Escola Convencional e Escola Profissional

Componente	Identificação	Itens
1	Ambiência Escolar	24, 25, 26, 27, 28,29,30,31, 31,32,33,34,35,36,37,38,39 e 40
2	Atuação Docente	52, 53,55,56,57,58,59,60,61,62 e 63.
3	Gestão das Atividades Escolares	41, 42,44,46,47,48,50 e 54.

Fonte: Extração pelo Método da Análise de Componentes Principais (2017).

De acordo com os resultados encontrados os três componentes, para os dois tipos de escolas, justificam no seu conjunto, em torno de, 42% da variância total dos resultados. De acordo com Hair et al. (2006), este valor é considerado satisfatório para os estudos na área das ciências sociais. Em relação aos pesos fatoriais todos apresentam valores acima de 0.40. Ressalta-se que vários autores (Hair et al., 2006; Kahn, 2006.) indicam 0.30 como valor mínimo para que um item seja retido, de modo que o pesquisador decidirá em manter ou não a variável. O peso de cada variável, portanto, significa a importância dela no fator. Note ainda que os itens 43, 45, 49 e 51 foram retirados da amostra. A justificativa da retirada foi em virtude do peso fatorial destes itens ser menor que 0.4 (*factor loadings* \geq 0.40).

Por fim, seguem os resultados da tabela 9 em que verificamos que o alfa de Cronbach apresenta bons valores de consistência interna nos 3 fatores. Destaca-se que apesar de na generalidade dos casos se estabelecer o valor .70 como critério para uma razoável consistência interna, podemos aceitar valores até .60, em especial se estamos a realizar uma análise exploratória (Hair et al., 2006).

Tabela 7: Consistência interna de cada componente

Componente	Consistência Interna do Componente
1	$\alpha = 0,92$
2	$\alpha = 0,86$
3	$\alpha = 0,70$

Fonte: Extração pelo Método da Análise de Componentes Principais (2017).

Em suma, se analisarmos os resultados apresentados, quer da análise exploratória, quer da consistência interna, à luz dos critérios de determinação dos fatores estabelecidos, podemos afirmar que a escala de CIE possui boas qualidades psicométricas iniciais. E, assim sendo a sua estrutura com 36 itens deve ser mantida. Na sequência, são apresentadas discussões sobre a identificação dos fatores extraídos e ainda a decomposição PARAFAC.

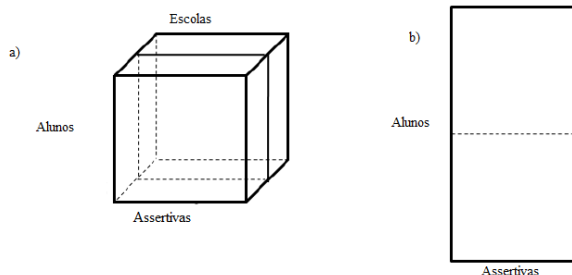
5 ANÁLISE DA DECOMPOSIÇÃO DE FATORES PARALELOS (PARAFAC) DO CLIMA INSTITUCIONAL ESCOLAR

No tratamento dos dados houve a preocupação da retirada de sujeitos que não responderam o questionário por completo e, ainda, de verificar a existência de possíveis combinações lineares nas respostas da escala. Esta inquietação deve-se ao fato de que nos modelos bilineares, a exemplo do PCA, existe o conhecido problema de liberdades rotacionais. Assim, as cargas em uma decomposição bilinear refletem os valores analíticos puros mensurados, e não é possível, sem informações externas, avaliar esses valores sem a rotação. Todavia, na maioria dos casos, o modelo PARAFAC é exclusivamente identificado por meio das permutações das componentes. Na prática só não ocorre a unicidade de solução no modelo PARAFAC caso se tenham vetores de cargas exatamente iguais. Isto é: a decomposição PARAFAC não é única se dois vetores de carregamento em qualquer um dos três modos forem proporcionais ou ainda se o produto de todas as matrizes de carregamento não tiver como posto a coluna completa. Conforme já descrito foram analisados 2400 sujeitos, pertencentes a dois tipos de escola e que responderam a 40 itens de uma escala, a qual denominou-se de Escala de Clima Institucional Escolar.

Os dados de todos os 2400 sujeitos podem ser decompostos em uma matriz de dimensões $I \times J \times K$ de três vias de tamanho específico $2400 \times 40 \times 2$. O primeiro modo refere-se aos sujeitos respondentes da pesquisa. Os 2400 alunos são oriundos de escolas públicas, sendo 1200 pertencentes as escolas convencionais e 1200 pertencentes as escolas profissionais.

O segundo modo trata dos 40 itens da escala empregada. Por fim, o terceiro modo trata os tipos de escola analisados. A Figura 7 ilustra tal processo analítico.

Figura 7 - Representação pictórica da estrutura tensorial



Fonte: Adaptado de Kolda e Bader (2009).

A adaptação de um modelo PARAFAC visa encontrar a melhor aproximação de baixa patente de uma matriz de dados. Salienta-se a possibilidade de não se encontrar uma solução, caso algumas restrições não sejam impostas, a exemplo da não negatividade. (TOMASI; BRO, 2004). Ademais, conforme dito na seção que trata o PARAFAC uma das principais vantagens em sua aplicabilidade é a unicidade. O modelo é parcimonioso. Neste sentido, ele apresenta a melhor solução, sem necessidade imediata de retirada de informações.

Em estatística dado um conjunto de dados, a variância é uma medida de dispersão que mostra o quão distante cada valor desse conjunto está do valor central (médio). A figura 8a aborda a análise dos sujeitos e como ocorre este comportamento na matriz de dados A. Por sua vez, a figura 8b, traz a Matriz B que trata do comportamento dos itens da escala e que será alvo de estudo mais aprofundado nos tópicos seguintes. A Figura 9 apresenta a decomposição por tipo de escola.

Figura 8a: Alunos respondentes e variância

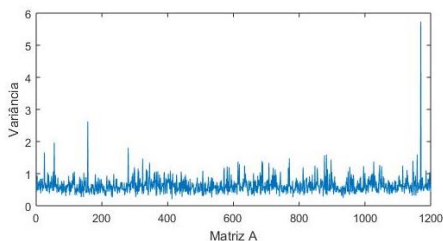
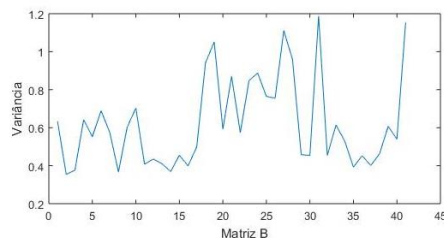
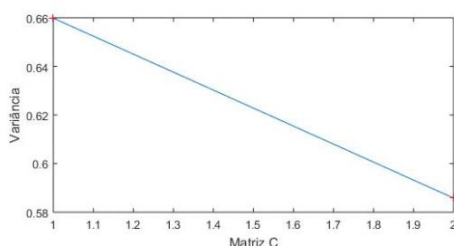


Figura 8b: Itens da escala e a variância



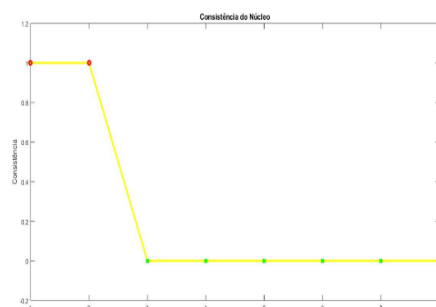
Fonte: Extração (MATLAB®, 2017).

Figura 9: Tipos de escola e a variância



Fonte: Extração (MATLAB®, 2017).

O critério de convergência é definido pela determinação de um erro mínimo de estimação de (10^{-6}) a partir da análise da diferença das matrizes. A decomposição PARAFAC se mostrou eficaz para a aplicação na base de dados coletada. Desta forma os achados desta tese na análise utilizando-se da decomposição PARAFAC serão mostrados a seguir. A Figura 10 traz a consistência interna do núcleo.

Figura 10 - Consistência interna do núcleo e componentes extraídos

Fonte: Extração (MATLAB®, 2017).

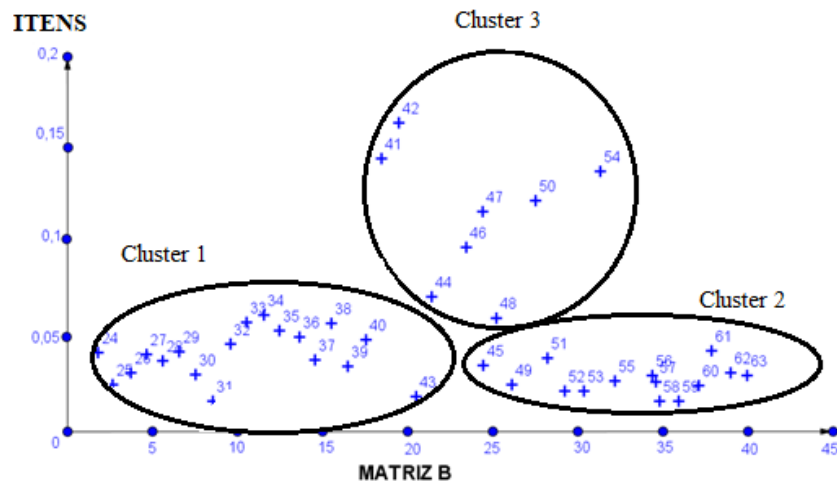
As informações da Figura 10 salientam que os dados podem ser estudados no plano bidimensional, e como o interesse desta pesquisa está relacionado à identificação dos padrões obtidos das assertivas, apenas os resultados do modo 2 (Matriz B) serão analisados.

Vale a pena destacar que, de forma equivalente à análise sobre a importância das contribuições dos dados na composição dos autovalores selecionados no método PCA, o método tensorial PARAFAC também reflete a mesma importância dos dados na composição dos valores singulares (e consequentemente autovalores) selecionados nas componentes principais (sob o critério de maiores valores da variância explicada) para os 3 modos de desdobramento. Mais do que isso, os 2 conjuntos de dados associados as escolas profissional e convencional, estão satisfatoriamente refletidos e associados pelas componentes referentes ao

modo 2 (Matriz B) da decomposição associado às assertivas, cujos 2 valores singulares deste modo refletem as contribuições de todos os dados de cada conjunto em análise.

A Figura 11 retrata a decomposição de interesse para essa pesquisa, o modo 2 que trata sobre a identificação do padrão das assertivas.

Figura 11 - Decomposição Modo 2



Fonte: Extração (MATLAB®, 2017).

A Figura 11 representa a matriz B, a qual exibe os 40 itens aplicados nos dois tipos de escola. Percebem-se três agrupamentos. O primeiro deles é composto pelos itens 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40 e **43**, o segundo contempla os itens **45, 49, 51, 52, 53, 55, 56, 57, 58, 59, 60, 61, 62** e 63, o terceiro grupo, por sua vez, agrega os itens 41, 42, 44, 46, 47, **48, 50** e 54.

Dos agrupamentos, evidencia-se as semelhanças com as componentes 1, 2 e 3 encontrados no PCA. Observemos ainda, que nesta análise, cuja decomposição é o PARAFAC, os itens destacados em negrito, no parágrafo anterior, referem-se aos itens que na análise de componentes principais haviam sido retirados, em virtude de possuírem cargas fatorais menores que 0.30. Contudo, a despeito da implicação do PCA pela retirada, optou-se por manter os itens, nesta análise, uma vez que há grande coerência técnica e pedagógica entre os itens e o agrupamento que o mesmo permaneceu.

De fato, as componentes extraídas na análise de componentes principais mantêm os mesmos itens e acrescenta, para o PARAFAC, os itens que haviam sido retirados, mantendo, porém, uma estrutura sintática semântica, conforme pode-se observar ao comparar os achados.

Neste sentido, as nomenclaturas das componentes permaneceram e, portanto, para este estudo defende-se que o cluster 1 está diretamente associado a ambiência escolar, o cluster 2 a atuação docente e o cluster 3 a gestão das atividades escolares. Por fim, segue a análise de consistência interna para a organização destes três clusters (Tabela 8).

Tabela 8: Consistência interna de cada *cluster*

Cluster	Consistência Interna dos itens
1	$\alpha = 0.91$
2	$\alpha = 0.85$
3	$\alpha = 0.70$

Fonte: Extração (SPSS, 2018).

As análises das consistências internas de cada organização (*cluster*) revelaram-se eficientes. Ademais, quando comparado a consistência interna das componentes extraídas no PCA e a consistência interna dos clusters no PARAFAC, há muita similaridade entre os achados. A decomposição PARAFAC se mostrou competente para este tipo de análise.

6 CONCLUSÕES

A presente pesquisa buscou demonstrar a viabilidade da aplicação de ferramentas matemáticas de análise matricial e tensorial para a área de avaliação educacional, visando verificar a aplicabilidade do PARAFAC para dados na área da avaliação educacional. Foi realizado estudo envolvendo o uso da Análise Multivariada (PCA) e da Álgebra Multilinear (PARAFAC) nos fatores associados ao desempenho acadêmico, com a extração de fatores no campo da avaliação educacional, utilizando-se, uma escala de Clima Institucional Escolar.

Foi realizado estudo envolvendo 2400 estudantes, da 3ª série do ensino médio de escolas públicas convencionais e profissionais; abordando aspectos relacionados CIE. Os resultados obtidos sinalizaram na direção positiva do uso do PARAFAC para a mineração, extração e identificação de componentes na escala CIE. Os resultados da análise por meio do PARAFAC se assemelham aos achados do PCA, ferramenta bastante utilizada no campo da avaliação como técnica de mineração de dados. Salienta-se, contudo, que o PARAFAC possui benefícios em relação ao PCA, a exemplo da unicidade, pois diferente do caso matricial (tensor de ordem dois), no qual o modelo sofre com problemas de ambiguidade de rotação, o tensor de ordem três ou de ordem superior pode possuir uma decomposição única.

Tanto o PCA como o PARAFAC apresentaram três componentes a partir da mineração dos dados do CIE, as quais foram nomeadas neste estudo e referindo-se a: ambiência

escolar, atuação docente, gestão das atividades escolares. Cada uma destas componentes compunha um conjunto de itens da escala, os quais foram bem semelhantes nos resultados das duas decomposições.

Diante do exposto, é evidenciado o fato de que os métodos de análise apresentados (Matriciais: PCA e Tensoriais: PARAFAC), podem contribuir como suporte para avaliar o CIE, permitindo a extração de informações significativas para intervenções voltadas à melhoria da qualidade da educação no estado. Por fim, constata-se que os objetivos iniciais da pesquisa foram alcançados. Trabalhos futuros podem replicar o estudo, empregando o mesmo instrumental e aplicando-o a outras amostras, de modo a comparar os resultados obtidos nos dois momentos. Desse modo, a ciência avança, pois como referiu o emérito filósofo norte-americano John Dewey (1859-1952): *todo grande progresso da ciência resulta da audácia e da imaginação de uma mente preparada.*

7 REFERÊNCIAS BIBLIOGRÁFICAS

ANDRIOLA, W. B. Problemas e perspectivas quanto ao uso dos testes psicológicos no Brasil. **Psique**, Belo Horizonte, v. 6, p. 46-57, 1995.

ANDRIOLA, W. B. Avaliação do raciocínio verbal em estudantes do Segundo Grau. **Estudos de Psicologia**, Natal, v. 2, n. 2, p. 277-285, 1997.

ANDRIOLA, W. B. Principais métodos para detectar o funcionamento diferencial do item (DIF) no âmbito da avaliação educacional. **Revista Educação em Debate**, Fortaleza, Ano 24, v. 2, n. 44, p. 83-97, 2002a.

ANDRIOLA, W. B. **Detección del funcionamiento diferencial del item (DIF) en tests de rendimiento**: aportaciones teóricas y metodológicas. 2002. 630 f. Tese (Doutorado em Educação) - Facultad de Educación, Universidad Complutense de Madrid, Madrid, 2002b.

ANDRIOLA, W. B. Psicometria Moderna: características e tendências. **Estudos em Avaliação Educacional**, São Paulo, v. 20, n. 43, p. 319-340, 2009a.

ANDRIOLA, W. B. Fatores institucionais associados aos resultados do Exame Nacional de Desempenho Estudantil (ENADE): estudo dos cursos de graduação da Universidade Federal do Ceará (UFC). REICE. **Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación**, Madrid, n. 1, p. 22-49, 2009b.

ANDRIOLA, W. B.; BARRETO, J. A. E. Análise métrica de um instrumento de medida da aprendizagem através da teoria de resposta ao item (TRI). **Ensaio: Avaliação e Políticas Públicas em Educação**, Rio de Janeiro, v. 5, n. 14, p. 59-74, 1997.

ANDRIOLA, W. B.; CAVALCANTE, L. R. Avaliação do raciocínio abstrato em estudantes do Ensino Médio. **Estudos de Psicologia**, Natal, v. 4, n.1, p. 23-37, 1999.

ANDRIOLA, W. B.; PASQUALI, L. A construção de um Teste de Raciocínio Verbal (RV). **Psicologia: Reflexão e Crítica**, v. 8, n. 1, p. 51-72, 1995.

BENZÉCRI, J. P. **Correspondence analysis handbook**. New York: Marcell Dekker, 1992.

BRO, R. PARAFAC. Tutorial and applications. **Chemometrics and Intelligent Laboratory Systems**, v. 38, p. 149-171, 1997.

BRO, R.; KIERS, H. A. L. A new efficient method for determining the number of components in parafac models, *Journal of Chemometrics*, vol. 17, 2003.

BUSSAB, W. O; MIAZAKI, E. S.; ANDRADE, D. F. **Introdução à análise de agrupamentos**: In: SIMPÓSIO BRASILEIRO DE PROBABILIDADE E ESTATÍSTICA, 1990. São Paulo. **Resumos**. 1990.

CATELL, R. B. PARALLEL PROPORTIONAL PROFILES” AND OTHER PRINCIPLES FOR DETERMINING THE CHOICE OF FACTORS BY ROTATION. *PSYCHOMETRIKA*, V. 9, N. 4, P. 267–283, 1944.

CARROL, J. D.; CHANG, J. **Analysis of individual differences in multi dimensional scaling via an N-Way generalization of Eckart – Young decompositions**, *Psychometrika*, 1970.

CAVALCANTE, S. M., ANDRIOLA, W. B. Avaliação da eficiência dos cursos de Graduação da Universidade Federal do Ceará (UFC) através da Análise Envoltória de dados (DEA). **Revista Iberoamericana de Evaluación Educativa**, v. 5, n. 3, p. 291-314, 2012.

CORREA, S. M. B. B. **Probabilidade e estatística**. 2 ed. Belo Horizonte: PUC Minas Virtual, 2006.

FAN, H.; KUANG, G.; QIAO, L. Fast Tensor Principal Component Analysis via Proximal Alternating Direction Method with Vectorized Technique. **Applied Mathematics**, v. 8, p. 77-86, 2017.

GARSON, G. D. **Statnotes: Topics in Multivariate Analysis**. Disponível em: <http://faculty.chass.ncsu.edu/garson/PA765/statnote.htm>. 2009.

GIORGIO, T.; RASMUS, B. **A comparison of algorithms for fitting the PARAFAC model** *Food Science*, Royal Veterinary and Agricultural University, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark, 2004.

GREENACRE, M. J. **Correspondence Analysis in Practice**. 2 ed. Chapman & Hall/CRC, 2007.

HAIR, J. F.; BLACK, B.; BABIN, B.; ANDERSON, R. E.; TATHAM, R. L. **Análise multivariada de dados**. 5. Ed. Porto Alegre: Bookman, 2005.

HARSHMAN, R. A. Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. **UCLA Working Papers in Phonetics**, 16, 1970 - (p. 1 - 84 - University Microfilms, Ann Arbor, No. 10,085).

HOTELLING, H. Analysis of a complex of statistical variables into principal components. **Journal of Educational Psychology**, v. 24, n. 6, p. 417-441, 1933.

LACERDA NETO, R. L.; ALMEIDA, A. L. F.; FAVIER, G.; MOTA, J. C. M. **PARAFAC Receivers for Multiuser SIMO CDMA-Based**, Wireless Systems, IFAC Symposium on System Identification, Newcastle, Australia, Março, 29-31, 2006.

LACERDA NETO, R. L. **Supervised PARAFAC-based Receivers for Wireless Communication Systems**; UNICE, Master Thesis, Sophia Antipolis, France, Junho 2005.

PASQUALI, L. **Psicometria: teoria dos testes na psicologia e na educação**. Petrópolis: Vozes, 2003.

REESE, D. D.; TABACHNICK, B. G.; KOSKO, R. E. **Using Multivariate Statistics** ; 2007.

SENA, M. M.; TREVISAN, M. M.; POPPI, R. J. PARAFAC: uma ferramenta quimiométrica para tratamento de dados multidimensionais. Aplicações na determinação direta de fármacos em plasma humano por espectrofluorimetria. **Química Nova**, São Paulo, v. 28, n. 5, p. 910-920, 2005.

SILVA, T. E. et al. **Ferramentas de processamento e análise de Informação Aplicadas ao ensino da Engenharia**. Simpósio Brasileiro de Telecomunicações, 2013.

SILVA, R. D. Aprendizagem de máquina aplicada a métodos de classificação de supernovas. **Dissertação de Mestrado apresentada ao curso de Pós-Graduação em Física da Universidade Federal do Espírito Santo (UFES) - (137 páginas)**. Vitória: UFES, 2018.

TABACHNICK, B. G.; FIDELL, L. S. **Using Multivariate Statistics**. 7th Edition. London: Pearson International Edition, 2019.

TUCKER, L. R. Implications of factor analysis of three-way matrices for measurement of change (p. 122–137). In C. W. Harris (ed). **Problems in Measuring Change**, University of Wisconsin Press, 1963.

KOLDA, T. G.; BADER, B. W. Tensor Decompositions and Applications. **Society for Industrial and Applied Mathematics**, v. 51, n. 3, p. 455–500, 2009.

TOMASI, G.; BRO, R. **Modelos Multilineares: Interativos Métodos Giorgio**. 2007.

VASCONCELOS, F. H. L.; SILVA, T. E. V.; SANTOS, W. M.; ALMEIDA, A. L. F.; ANDRIOLA, W. B.; MOTA, J. C. M. Multilinear Decomposition Application into Students' Evaluation of Teaching Effectiveness. **IEEE Global Engineering Education Conference (EDUCON)**, p. 910-916, 2013.

VIEIRA, P. R. C.; RIBAS J. R. **Análise multivariada com uso do SPSS**, Rio de Janeiro, Ed. Ciência Moderna, 2011.

ZARI, A.; OZDEMIR, A.; IWEN, M. A.; AVIYENT, S. Extension of PCA to Higher Order Data Structures: An Introduction to Tensors, Tensor Decompositions, and Tensor PCA. **Proceedings of the IEE**, v. 106, n. 8, p. 1341 – 1358, 2018.