

ANÁLISE PSICOMÉTRICA DO TESTE DE ALFABETIZAÇÃO CIENTÍFICA BÁSICA SIMPLIFICADO

PSYCHOMETRIC ANALYSIS OF SIMPLIFIED TEST OF BASIC SCIENTIFIC LITERACY

PATRICK ALVES VIZZOTTO*
LUIZ FERNANDO MACKEDANZ**

RESUMO

Medir a Alfabetização Científica permite avaliar como este construto consegue ser percebido nas inúmeras atividades que visam promovê-la. Porém, a extensão dos testes utilizados é um fator que complica sua correta aplicação. Assim, realizou-se uma análise dos aspectos psicométricos de uma versão reduzida do Teste de Alfabetização Científica Básica de Laugksch e Spargo. O questionário foi aplicado em 653 ingressantes de cursos Técnicos e de Graduação e apresentou índices que sugerem adequação para avaliar o que se propõe medir. É indicado o emprego do teste para mensurar o nível de Alfabetização Científica Básica de maneira quantitativa. Sugere-se a sua aplicação em outras amostras dessa população a fim de verificar a estabilidade dos itens, e corroborar a estrutura teórica das componentes.

Palavras-chave: Letramento Científico; Psicometria; Pesquisa Quantitativa.

ABSTRACT

Measuring scientific literacy allows evaluating how this construct can be perceived in the countless activities that aim to promote it. However, the extent of the tests used is a factor that complicates the correct application. Thus, we conducted an analysis of the psychometric aspects of a reduced version of the basic scientific literacy test of Laugksch and Spargo. The questionnaire was applied to 653 freshers of technical and undergraduate courses and presented indexes that suggest adequacy to evaluate what is proposed to measure. It is indicated the use of the test to measure the level of basic scientific literacy in a quantitative manner. It is suggested its application in other samples of this population in order to verify the stability of the items, and corroborating the theoretical structure of the components.

Keywords: Scientific literacy; Psychometrics Quantitative research.

* Doutor em Educação em Ciências. Bolsista de Pós-doutorado da Universidade do Vale do Taquari. E-mail: patrickvizzotto@gmail.com. Orcid: <http://orcid.org/0000-0003-1613-4858>.

** Doutor em Ciências. Professor Associado na Universidade Federal do Rio Grande. E-mail: mackedanz@gmail.com. Orcid: <http://orcid.org/0000-0002-0138-0119>.

INTRODUÇÃO

Alfabetização Científica é um construto muito utilizado em produções acadêmicas e em documentos oficiais da educação. Diferentes concepções compõem o seu significado, tanto no contexto internacional quanto no nacional, e são numerosas as propostas que buscam alcançá-la. No entanto, percebe-se como sutil o intento de avaliá-la ou medi-la em produções da literatura (RIVAS; MOÇO; JUNQUEIRA, 2017).

No decorrer dos anos, junto ao progresso do debate sobre o significado do termo “*Scientific Literacy*”, evoluiu também, as iniciativas para mensurar indicativos desse conceito em diferentes setores da população (MILLER, 1983).

Uma das produções de grande proeminência foi a dos Sul-Africanos Laugksch e Spargo (1996), que relata a criação e validação do questionário denominado de Teste de Alfabetização Científica Básica (TACB). Os autores criaram este instrumento com itens fechados, dicotômicos (verdadeiro ou falso). Esses itens versam sobre situações cotidianas, contextualizadas nas áreas da Física, Biologia, Química, Saúde, Epistemologia e Meio Ambiente. Essas situações são descritas de maneira correta ou incorreta, do ponto de vista científico, assim devendo o respondente usar seus saberes para julgar a coerência científica de cada assertiva.

O TACB foi elaborado tendo como fundamentação teórica a visão de Alfabetização Científica de Miller (1983). Esse autor entende que para alguém ser considerado alfabetizado cientificamente deve manifestar uma compreensão mínima, de maneira concomitante, em três aspectos, denominados pelo autor de eixos: entender os conteúdos da ciência; a natureza da ciência; e o impacto da ciência e tecnologia na sociedade e ambiente.

Desde o seu surgimento, em 1996, até o presente, esse instrumento foi empregado por pesquisadores de diversos países para aferir, de maneira, quantitativa, o nível de Alfabetização Científica de múltiplos grupos da sociedade.

No Brasil, o questionário foi traduzido do inglês por Nascimento-Schulze (2006), e desde o seu surgimento, devido ao processo de adaptação ao contexto brasileiro, muitos trabalhos consideraram desafiador o uso do teste devido a sua grande extensão (110 itens). Por isso, no decorrer dos anos, a literatura nacional apresentou três tentativas de reduzi-lo.

Nascimento-Schulze, Camargo e Wachelke (2006) partiram de uma perspectiva qualitativa de análise para decidir quais itens deveriam ser mantidos no teste. Para isso, pareceristas da área julgaram a relevância de cada um deles para o contexto brasileiro. Com os itens considerados satisfatórios, realizou-se testes estatísticos para verificar os índices de consistência interna da versão reduzida. De 110, o questionário diminuiu para 77 itens, sendo então, chamado de Teste de Alfabetização Científica Básica Reduzido- TACB-R.

Cinco anos depois, Oliveira e Silva-Forsberg (2011) também realizaram um estudo de redução. As autoras submeteram o instrumento à especialistas da área e assim, reduziram-no para 60 questões. No entanto, a estrutura inicial do TACB não foi mantida, pois os itens ligados ao Eixo 2 (entendimento da natureza da ciência) foram eliminados.

Sete anos depois, Vizzotto e Mackedanz (2018) apresentaram outra metodologia de redução instrumental, sob uma perspectiva quantitativa de análise. A partir de um teste piloto e fazendo uso de técnicas estatísticas, investigou-se os itens que menos estavam correlacionados com o instrumento como um todo. Assim, os autores reduziram o questionário para 45 itens. Esse foi chamado

de TACB-S (Teste de Alfabetização Científica Básica Simplificado) e manteve a mesma proporção de itens para os três eixos originais.

Dessa maneira, a fim de contribuir com a literatura nacional dessa área, julga-se relevante investigar as características psicométricas do questionário reduzido, a fim de aferir as características de validade e fidedignidade de uma versão exequível ao contexto nacional. Para isso, optou-se por utilizar aquele dos três que mais obteve êxito em reduzir a quantidade de itens. O último estudo, que chegou a uma quantidade final de 45 questões, foi a versão analisada neste manuscrito.

Assim, realizou-se um teste piloto, fazendo uso do TACB-S, para responder se “os itens do TACB-S são compostos de características psicométricas que atestem a validade e confiabilidade do seu uso no contexto nacional?”. Dessa maneira, o objetivo desse artigo é apresentar os resultados da análise psicométrica realizada.

Espera-se que o manuscrito possa contribuir para a área da pesquisa e ensino das ciências escolares, com a análise de um instrumento que potencialmente pode contribuir para a tarefa de aferir aspectos da educação científica de alunos dentro do cotidiano escolar.

Mensurar a Alfabetização Científica de um grupo é um empreendimento desafiador, que pode variar seus fundamentos, instrumentos e objetivos, a depender da referência teórica e metodológica empregada. Contudo, esse esforço é fundamental para a produção da área, pois julga-se relevante dispor de ferramentas que obtenham êxito, de maneira objetiva, válida e confiável, na tarefa verificar o proveito das metodologias utilizadas em produtos de ensino que se destinam a promoção desse construto.

O artigo está organizado da seguinte forma: na sequência, apresentamos a fundamentação teórica na qual o TACB-S está sustentado. Em um segundo momento, será delineado a metodologia empregada na análise psicométrica dos itens. Adiante, apresenta-se os resultados e suas discussões, e por fim, as considerações sobre a pesquisa relatada.

OS FUNDAMENTOS DO TESTE DE ALFABETIZAÇÃO CIENTÍFICA BÁSICA

O teste foi construído tendo como estrutura teórica as ideias de Alfabetização Científica de Miller (1983) e as orientações para a educação científica postuladas pela Associação Americana para o Avanço das Ciências (AAAS), descritas por meio do documento intitulado “*Science For All Americans (SFAA)*” (AAAS, 1989). Esse documento sugere as competências científicas nas quais seria importante egressos da educação básica demonstrarem aptidão para atuarem na sociedade de maneira crítica e responsável.

Para Miller (1983), é indispensável a alguém alfabetizado cientificamente compreender a dimensão conceitual dos conhecimentos científicos. Esse enfoque é englobado pelo eixo 1, chamado de “entendimento dos conteúdos da ciência”. Além disso, a dimensão epistemológica do saber também deve ser envolvida, a fim de ocupar-se da compreensão do processo de construção e evolução do conhecimento. A esse eixo 2 foi dado o nome “entendimento da natureza da ciência”. Por fim, o último eixo inclui a aptidão de fazer uso dos saberes científicos aliados à dimensão social e ambiental. Ele expressa que a ciência não está dissociada de um contexto. O eixo 3 é denominado de “entendimento do impacto da ciência e tecnologia na sociedade e ambiente”.

Os eixos 2 e 3 abrangem perspectivas sócio-científicas. Esses pontos, de caráter social, cultural, econômico e ambiental, auxiliam, junto ao entendimento de cunho conceitual, na prática de leitura de mundo que se espera de um cidadão apto a tomar decisões responsáveis em sua realidade.

No Brasil, há estudos que defendem que os eixos 2 e 3 são os menos abordados dentro da escola, sendo o eixo 1, sobre os conteúdos, o de maior ênfase dentro das disciplinas das Ciências da Natureza (RIVAS, 2015; CAMARGO *et al.*, 2011). Não obstante, somente o enfoque dos conteúdos da ciência não contribui para o que Miller (1983) entende por Alfabetização Científica, pois não sustenta uma visão de mundo crítica, atribuindo a aprendizagem de ciências apenas a aspectos de memorização conceitual.

Devido a pertinente estrutura teórica postulada por Miller (1983), os elaboradores do TACB, Laugksch e Spargo (1996) a adotaram como referencial para a escrita dos itens que fizeram parte do teste.

O questionário integral continha 110 itens, compostos por afirmações verdadeiras ou falsas sobre diferentes tópicos do conhecimento científico e seus processos. Esses itens foram alocados dentro dos 3 eixos supracitados. Cada respondente deve fazer uso dos seus saberes para julgar a coerência científica de cada questão. Ao final, o seu score é determinado pelo somatório de acertos de cada eixo. É necessário obter um desempenho mínimo de 60% de acertos em cada eixo para ser considerado alfabetizado cientificamente. Segundo os autores, esse quantitativo foi determinado pois 60% é o piso de aproveitamento no qual a maioria das instituições de ensino básico e superior adotam para determinar a aprovação ou reprovação de seus alunos. Assim, se em algum dos eixos o participante não obtiver o aproveitamento de 60%, este não poderá ser considerado alfabetizado cientificamente.

O processo de redução do TACB-S (VIZZOTTO; MACKEDANZ, 2018) manteve a estrutura teórica tridimensional do instrumento integral. Assim, as 45 questões do TACB-S são subdivididas da seguinte maneira: eixo 1 (1 a 3; 14 a 26; 35 a 45); eixo 2 (4 a 13); e eixo 3 (27 a 34).

Embora o processo de redução instrumental dessa versão tenha apresentado características satisfatórias de validade e confiabilidade, uma análise psicométrica de cada item não foi apresentada no estudo, justificando a abordagem aqui realizada.

METODOLOGIA

CARACTERÍSTICAS DO ESTUDO

Essa é uma pesquisa de base, de caráter quantitativo, descritiva e faz uso de questionário como procedimento de coleta de dados (GIL, 2008). Inserida dentro da temática da Alfabetização Científica (MILLER, 1983), teve como público-alvo 653 egressos do Ensino Médio, ingressantes de cursos Técnicos e de Graduação. Os dados foram obtidos por meio do Teste de Alfabetização Científica Básica Simplificado - TACB-S e analisados através de procedimentos da estatística, via Teoria Clássica dos Testes (FIELD, 2009; HAIR *et al.*, 2009; PASQUALI, 2017).

VALIDADE E CONFIABILIDADE DOS TESTES

Sempre que um instrumento de coleta de dados é elaborado, testa-se, primeiramente, os seus aspectos de validade e confiabilidade. Para Souza, Alexandre e Guirardello (2017), a qualidade da informação fornecida pelos instrumentos depende, em parte, de suas propriedades psicométricas. Essas propriedades, comumente, são aferidas por meio de técnicas estatísticas.

A psicométrica é um ramo da psicologia no qual faz uso de testes para representar, por meio de indicadores, aspectos intrínsecos do ser humano. Muitos autores estabelecem erroneamente essa

área dentro do campo da estatística, pois para fazer uso de testes e reduzir constructos em números é necessário procedimentos matemáticos sofisticados (PASQUALI, 2017). As áreas da Saúde e da Educação, por exemplo, beneficiam-se da psicometria para fundamentar a criação de testes válidos e fidedignos para inferir aspectos da saúde e de conhecimentos, que antes, não seria possível devido à falta de medidas indiretas confiáveis. Assim, as propriedades psicométricas podem ser aferidas por meio de procedimentos básicos de estatística, com fins de testar se os dados mensuram com qualidade o que se propõe a inferir.

Concebe-se que um questionário possui validade se ele consegue medir aquilo a que se propõe. Já a confiabilidade, também denominada de fidedignidade, é certificada se as medidas de um instrumento apresentarem escores semelhantes, em um mesmo respondente, ao aplicá-lo em momentos diferentes.

Na área da Educação é impossível que uma mesma pessoa não seja influenciada por fatores externos entre um intervalo de tempo e outro (MOREIRA; ROSA, 2007). Isso inviabilizaria o processo de testagem antes e depois, com vistas à aferição da confiabilidade. Assim, há diferentes formas de verificar a fidedignidade de instrumentos de coletas de dados. Segundo Souza *et al.* (2017) a maioria dos questionários testam essa característica por meio da análise da sua consistência interna, sendo o coeficiente Alfa de Cronbach o teste mais comum.

A TEORIA CLÁSSICA DE TESTES E A TEORIA DE RESPOSTA AO ITEM

O coeficiente citado na seção anterior faz parte de uma série de procedimentos estatísticos presentes na Teoria Clássica de Testes (TCT). A TCT possui como síntese medir a partir do somatório de acertos dos itens de um questionário. Dentro do âmbito da Educação, essa teoria auxilia na aferição da qualidade de testes que buscam explicar determinado evento baseado no escore final. Isso significa fazer uso de um conjunto de questões (que podem ser observáveis) para medir um processo mental (que não pode ser observado).

Na TCT é necessário considerar o instrumento integral para atribuir qualidade ao teste. Isso acontece por meio da discriminação dos participantes. Assim, de acordo com Vianna (1978), a TCT segue alguns pressupostos, como a necessidade de os mesmos itens serem utilizados por todos os respondentes, nas mesmas condições, e quanto maior a quantidade de itens e participantes, maior também serão os resultados e menor o erro.

Contudo, o escore bruto não propicia a comparação de um participante que obteve um bom desempenho em um teste fácil de outro que apresentou baixo escore em um teste difícil. À vista disso, é fundamental que se construa parâmetros para descrever, além da qualidade do item, a sua dificuldade, ou ainda, procedimentos de maior robustez, que permitam realizar tais comparações. Com o objetivo de suprir este objetivo, a Teoria de Resposta ao Item (TRI) apresenta mecanismos que possibilitam obter êxito neste particular.

A TRI, dentro da pesquisa em Educação, tem como meta aferir a proficiência do respondente e não apenas um somatório bruto. Mede-se traços latentes que são considerados aspectos da proficiência do que se está querendo aferir. De acordo com Pasquali (2017), a TRI baseia-se na probabilidade de o respondente acertar um item de acordo com o seu conhecimento.

Ela é contemporânea à TCT, porém, é complementar, e não, substitutiva. Geralmente, os resultados obtidos em uma teoria apresentarão o mesmo padrão na outra. A sua vantagem consiste nas possibilidades de comparação entre o nível de proficiência de diferentes grupos, mesmo que tenham respondido a diferentes instrumentos.

Para esta pesquisa, como o foco centra-se em avaliar os aspectos psicométricos de itens e não os usar em testes para comparar proficiências de diferentes grupos amostrais, utilizaremos a TCT como base para nossas análises.

OS TESTES UTILIZADOS

As medidas de tendência central e dispersão auxiliam na visão geral dos dados em um primeiro momento de análise. Caso a **Média** de acertos mostre-se alta, pode significar que há pouca discriminação dos itens. Além disso, para itens dicotômicos, a média de acertos pode fornecer uma leitura do índice de dificuldade das questões.

O **Desvio-padrão**, por sua vez, apresenta a dispersão em torno da média de acertos. Se um item apresentar baixo valor de Desvio-padrão, significa que há baixa variabilidade. Isso também pode significar pouca capacidade de discriminação (SOARES, 2018).

O **Índice de dificuldade do item** auxilia a observar a proporção de participantes que acertaram cada questão. Divide-se o número daqueles que responderam corretamente determinado item pela quantidade total de participantes. Quanto menor for a porcentagem de acertos, maior será o índice de dificuldade. Para Vilarinho (2015), itens com dificuldade igual ou superior a 0,7 são fáceis; entre 0,7 e 0,3 são moderados; e abaixo de 0,3 são considerados difíceis.

O **Índice de discriminação do item** afere a capacidade de a questão diferenciar os respondentes de maior desempenho dos de menor desempenho. Pessoas com diferentes escores em um instrumento devem diferir também no construto analisado. Ou seja, um item discrimina de maneira satisfatória quando pessoas com determinada habilidade apresentam desempenho superior quando comparadas com pessoas sem tal habilidade.

Analisa-se, devido a distribuição normal, os quantitativos dos 27% desses dois grupos. Esse cálculo consiste na diferença entre a proporção de acertos daqueles de maiores escores e os de menores escores, dividido pelo número de pessoas correspondentes a esses 27%. Essa análise gera um índice que varia de -1 até +1. Itens com discriminação igual ou menor que 0,2 são considerados ineficientes; de 0,2 a 0,3 sugere-se a revisão do item; de 0,3 a 0,4 são aceitáveis; Acima de 0,4 são considerados satisfatórios.

Como o índice de discriminação é influenciável pela dificuldade dos itens, sugere-se aplicar um **Test t para amostras independentes** (para amostras distribuídas de maneira normal) ou um **Teste U de Mann-Whitney** (para amostras que fogem da tendência normal) para verificar a existência de diferença entre as médias dos dois grupos (HAIR *et al.*, 2009). Para interpretar ambos os testes, verifica-se o p-valor, que se for menor que 0,05, sugere-se a rejeição da hipótese nula que afirma igualdade entre os grupos analisados.

De acordo com Pasquali (2017), quando as questões são consideradas muito fáceis ou muito difíceis, a análise da discriminação dos itens pode se tornar pouco confiável, pois esse índice tende a se aproximar de zero. Isso pode provocar a decisão errônea de excluí-lo do teste. Em virtude disso, é sugerido, como complemento, o cálculo do coeficiente de correlação ponto-bisserial, pois esse sofre menos influência do nível de dificuldade das questões.

Assim, o **coeficiente de correlação ponto-bisserial** afere se aqueles que tiveram um bom desempenho no teste tiveram maior tendência de assinalar as alternativas corretas. Para isso, verifica-se a correlação entre o desempenho geral e as alternativas de cada item.

É desejável que esta correlação seja positiva para a alternativa correta e negativa para a incorreta. Esse coeficiente é obtido por meio do índice de correlação de Pearson entre a variável dicotômica

(acerto ou erro) e desempenho individual no questionário (FIELD, 2009). Para Soares (2018), esse índice é capaz de identificar questões com formulação inadequada e com erros de gabarito. Ainda segundo este autor, se o coeficiente for negativo para a alternativa correta, significa que essa questão está sendo mais acertada por respondentes de menor desempenho. Geralmente considera-se o item como pobre se as correlações forem menores que 0,15 para a alternativa correta; razoável se for entre 0,15 e 0,3; e se for maior que 0,3, é considerado com bom potencial de discriminação.

O **coeficiente Alfa de Cronbach** é comumente utilizado como teste para verificar a consistência interna do questionário. O cálculo relaciona a correlação entre cada item e o instrumento como um todo. Ele analisa o grau de variância entre os itens do teste. Assim, quanto menor a soma da variância das questões, mais consistente é o instrumento. Tem por premissa que os itens do teste são maneiras paralelas de aferir o construto que se está testando. O coeficiente gera um valor entre 0 e 1 e quanto mais próximo de 1, mais consistente é o teste. A literatura geralmente considera valores a partir de 0,6 aceitáveis para considerar um instrumento confiável (HAIR *et al.*, 2009; FIELD, 2009; PASQUALI, 2017).

A **correlação item-total**, demonstra o quanto a questão está correlacionada com o questionário como um todo. Dessa maneira, itens que contribuem para um alto índice de confiabilidade, apresentarão também, altos valores de correlação item-total. Ao mesmo tempo, se itens apresentarem valores negativos dessa correlação, podem estar contribuindo para um baixo nível de consistência interna (FIELD, 2009).

O **Teste T² de Hotteling** afere a existência do efeito halo. Esse efeito consiste na possibilidade de que a avaliação de um item, seja capaz, por efeito de um viés, de influenciar a concepção do respondente na análise das demais questões, corrompendo o resultado integral. De acordo com Rosa e Kamakura (2001), este efeito retrata uma tendência nas respostas de um questionário, desencadeada devido ao fato de elas serem aferidas em uma mesma coleta de dados:

Por exemplo, se numa pesquisa de satisfação primeiro mede-se a satisfação geral com o serviço, e depois, mede-se a percepção do serviço em quesitos específicos, é provável que as medidas de percepção sejam influenciadas pela avaliação geral, criando uma correlação espúria entre todas as variáveis. Essa correlação espúria faz com que as respostas dadas pelo consumidor sejam correlacionadas simplesmente pelo fato de terem sido obtidas dentro da mesma entrevista (ROSA; KAMAKURA, 2001, p. 3).

Ou seja, quando a avaliação do respondente para determinado item ou teste é constituída unicamente desse viés, esse julgamento é chamado de efeito halo. No intuito de que cada item seja respondido de maneira independente, deseja-se que este efeito não seja identificado. O T² de Hotteling testa a hipótese de existência desse efeito. Assim, se o p-valor for menor que 0,05, sugere-se a rejeição da hipótese nula de que determinado item possui influência sobre outros (SOARES, 2018).

A **Análise fatorial exploratória (AFE)** é utilizada para averiguar a unidimensionalidade do instrumento. Isto é, se os itens do questionário estão mensurando um mesmo fator. Essa condição é fundamental para legitimar os resultados ponderados nos índices de dificuldade e discriminação. Um instrumento é unidimensional se apenas um fator for responsável cognitivamente pela realização de todos os itens do teste. Por certo, há diferentes fatores incumbidos pela ação de responder determinado questionário. No entanto, esta premissa propõe que há uma habilidade dominante responsável pelo conjunto de aptidões que serão utilizados para responder os itens do teste.

Uma AFE permite ainda averiguar se o questionário realmente mensura o que se propõe medir. Ou seja, este procedimento pode gerar um indicativo da validade do instrumento. Segundo Soares (2018), se, por exemplo, os itens de um instrumento sobre avaliação de dois componentes curriculares tiverem sido bem elaborados e testados previamente, provavelmente na análise desses itens eles se dividirão, devido à diferença de valor da carga fatorial, e representarão as duas áreas avaliadas. Isso proporciona afirmar a existência da validade de construto daquele conjunto de questões.

A APLICAÇÃO DA PESQUISA

O TACB-S foi aplicado em egressos do Ensino Médio, ingressantes de cursos Técnicos e de Graduação de Instituições de Ensino Superior do Rio Grande do Sul durante o ano de 2019. A maioria dos respondentes ainda não havia tido contato com disciplinas de cunho científico após o término de seus estudos na Educação Básica.

A aplicação ocorreu na seguinte ordem: presencialmente, em sala de aula, os objetivos da pesquisa foram explanados, salientando o anonimato dos respondentes. Em seguida, entregou-se o termo de consentimento livre e esclarecido, contendo as informações éticas sobre a pesquisa. Na sequência, foi entregue o teste. O tempo de um período escolar foi dedicado para a atividade. Após esse processo, os questionários foram corrigidos e tabulados em uma planilha eletrônica e por fim, os desempenhos dos itens foram analisados no Software *Statistical Package for the Social Sciences - SPSS®* versão 23, para Windows.

Na próxima seção apresentamos os resultados das análises realizadas. Salientamos que para assegurar a lisura na apresentação dos resultados obtidos, utilizou-se as figuras geradas pelo próprio SPSS. Isso justifica a eventual formatação própria e linguagem em inglês de algumas imagens.

RESULTADOS

TESTES PSICOMÉTRICOS

Em um primeiro momento, apresenta-se a análise das medidas de tendência central e dispersão, com o objetivo de expor uma visão geral do comportamento dos itens.

Conforme pode-se observar na Figura 1, a maioria das médias foram consideradas nem altas nem baixas. Isso é considerado, inicialmente, um indicativo de que não há itens distorcendo negativamente a respostas do teste (SOARES, 2018).

Figura 1 - Média e Desvio-Padrão de cada item.

Estatísticas dos Itens											
Item	Média	Desvio Padrão									
Q1	0,76	0,42	Q13	0,66	0,48	Q25	0,83	0,38	Q37	0,52	0,50
Q2	0,51	0,50	Q14	0,73	0,44	Q26	0,82	0,39	Q38	0,47	0,50
Q3	0,68	0,47	Q15	0,72	0,45	Q27	0,81	0,39	Q39	0,48	0,50
Q4	0,77	0,42	Q16	0,81	0,39	Q28	0,74	0,44	Q40	0,91	0,29
Q5	0,38	0,49	Q17	0,77	0,42	Q29	0,61	0,49	Q41	0,86	0,34

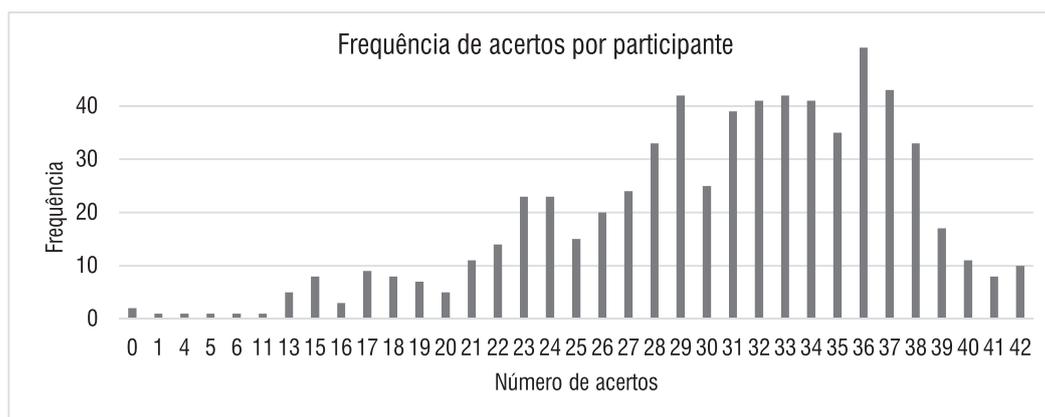
Q6	0,74	0,44	Q18	0,57	0,50	Q30	0,47	0,50	Q42	0,77	0,42
Q7	0,63	0,48	Q19	0,63	0,48	Q31	0,74	0,44	Q43	0,80	0,40
Q8	0,56	0,50	Q20	0,53	0,50	Q32	0,60	0,49	Q44	0,89	0,31
Q9	0,69	0,46	Q21	0,83	0,37	Q33	0,71	0,45	Q45	0,82	0,39
Q10	0,31	0,46	Q22	0,50	0,50	Q34	0,63	0,48			
Q11	0,88	0,32	Q23	0,65	0,48	Q35	0,72	0,45			
Q12	0,68	0,47	Q24	0,49	0,50	Q36	0,78	0,41			

Fonte: autores.

Ainda, ao analisarmos o Desvio-padrão, podemos notar que nenhum valor pode ser considerado próximo de zero, o que poderia sinalizar indícios de pouca variabilidade dos itens, e consequente inutilidade dos mesmos para diferenciar bons e maus desempenhos.

O Gráfico 1 apresenta a frequência geral de acertos. Essa visualização contribui na percepção do desempenho dos participantes, bem como auxilia a notar se a distribuição dos acertos segue uma curva normal.

Gráfico 1 - Frequência de acertos por participante.



Fonte: autores.

Ao analisar a imagem, percebe-se que provavelmente não há uma distribuição normal nos índices de acertos. Para confirmar esta hipótese, procedeu-se com o Teste Normalidade de Kolmogorov-Smirnov. A Figura 2 apresenta o resultado do SPSS para esse teste.

Figura 2 - Teste de Normalidade.

	Tests of Normality					
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
SCORE	,099	653	,000	,941	653	,000

a. Lilliefors Significance Correction

Fonte: autores.

O teste de normalidade, ao apresentar o p-valor de significância menor que 0,0001, confirma que a distribuição de acertos não possui uma distribuição normal. Isso significa que eventuais testes de inferência realizados nas análises deverão seguir as orientações das estatísticas não-paramétricas (PASQUALI, 2017).

Na sequência, a Figura 3 apresenta um compilado de algumas características psicométricas das 45 questões do TACB-S.

Figura 3 - Índices de Dificuldade, Discriminação, Correlação item-total e Correlação ponto-biserial.

Item	Dificuldade	Discriminação	Correlação item-total	Correlação ponto-biserial	
				Alt. Correta	Alt. Incorreta
Q1	0,76	0,31	,288	0,35	-0,35
Q2	0,51	0,26	,148	0,22	-0,22
Q3	0,68	0,20	,146	0,21	-0,21
Q4	0,77	0,18	,145	0,21	-0,21
Q5	0,38	0,25	,126	0,20	-0,20
Q6	0,74	0,28	,251	0,31	-0,31
Q7	0,63	0,43	,299	0,36	-0,36
Q8	0,55	0,32	,178	0,25	-0,25
Q9	0,69	0,40	,298	0,36	-0,36
Q10	0,31	0,15	,055	0,12	-0,12
Q11	0,88	0,20	,302	0,35	-0,35
Q12	0,68	0,24	,196	0,26	-0,26
Q13	0,66	0,35	,241	0,31	-0,31
Q14	0,73	0,52	,418	0,47	-0,47
Q15	0,72	0,45	,339	0,40	-0,40
Q16	0,81	0,26	,240	0,30	-0,30
Q17	0,77	0,39	,333	0,39	-0,39
Q18	0,57	0,60	,422	0,48	-0,48
Q19	0,63	0,41	,287	0,35	-0,35
Q20	0,53	0,54	,334	0,40	-0,40
Q21	0,83	0,29	,312	0,36	-0,36
Q22	0,50	0,40	,275	0,34	-0,34
Q23	0,65	0,49	,366	0,43	-0,43
Q24	0,49	0,44	,292	0,36	-0,36
Q25	0,83	0,25	,252	0,30	-0,30
Q26	0,82	0,38	,383	0,43	-0,43
Q27	0,81	0,32	,331	0,38	-0,38
Q28	0,74	0,24	,234	0,29	-0,29
Q29	0,61	0,42	,266	0,33	-0,33
Q30	0,47	0,44	,243	0,31	-0,31
Q31	0,74	0,20	,097	0,16	-0,16
Q32	0,60	0,39	,249	0,32	-0,32
Q33	0,71	0,28	,214	0,28	-0,28

Q34	0,63	0,49	,318	0,38	-0,38
Q35	0,72	0,41	,333	0,39	-0,39
Q36	0,78	0,50	,468	0,51	-0,51
Q37	0,52	0,60	,412	0,47	-0,47
Q38	0,47	0,40	,212	0,28	-0,28
Q39	0,48	0,46	,309	0,37	-0,37
Q40	0,91	0,22	,375	0,41	-0,41
Q41	0,86	0,32	,431	0,47	-0,47
Q42	0,77	0,35	,307	0,36	-0,36
Q43	0,80	0,52	,497	0,54	-0,54
Q44	0,89	0,24	,357	0,40	-0,40
Q45	0,82	0,36	,375	0,42	-0,42

Fonte: autores.

Ao analisar a Figura 3, podemos notar que quanto ao índice de dificuldade dos itens, 51% deles puderam ser considerados fáceis, com valores acima de 0,7. Os restantes foram classificados como moderados (índices entre 0,3 e 0,7) e nenhum como difícil (abaixo de 0,3). As questões 40, 44, 11, 41, e 21 foram consideradas, em ordem decrescente, as de maior facilidade. Por outro lado, os itens 10, 5, 38, 30, e 39 obtiveram os menores índices de acertos.

No que se refere ao índice de discriminação, os itens 3, 4, 10, 11, e 31 foram apontados, à priori, como ineficientes, pois seus valores foram menores do que 0,2. Por outro lado, 65% das questões restantes foram classificadas como satisfatórias e aceitáveis. No teste, os itens 18, 37, 20, 43 e 14 foram os que mais conseguem diferenciar bons de maus respondentes.

A fim de confirmar se há, de fato, uma diferença nas médias dos dois grupos testados no índice de discriminação (27% daqueles de melhor e pior desempenhos) realizamos um teste para aferir a hipótese nula de igualdade entre a distribuição de acertos das duas categorias de respondentes.

Conforme referido anteriormente, devido ao fato de a distribuição de acertos não seguir uma tendência à normalidade, deve-se utilizar testes não-paramétricos. Assim, procedeu-se a análise com o Teste U de Mann-Whitney (FIELD, 2009). O p-valor obtido no teste foi menor do que 0,05. Isso sugere a rejeição da hipótese anteriormente descrita. Ou seja, há diferença estatisticamente significativa entre os dois grupos.

A média de acertos do grupo de desempenho superior foi de 37,71, com desvio-padrão de 1,776. Por outro lado, o grupo de desempenho inferior, teve média de acertos de 21,54, com um desvio-padrão de 5,201. Essa constatação legitima as considerações advindas da análise de discriminação dos itens.

A correlação Item-total, por sua vez, apontou que 17% dos itens (2, 3, 4, 5, 8, 10, 12, 31) caracterizaram-se abaixo do ponto de corte arbitrariamente estabelecido. Ou seja, essas questões foram as que menos contribuíram para a consistência interna do instrumento como um todo. No entanto, todos os itens apresentaram valores de correlação positivos, ou seja, sua permanência contribui e não diminui o valor de fidedignidade do questionário.

Como um todo, o questionário apresentou um índice de consistência interna considerado dentro dos parâmetros desejáveis. A figura abaixo apresenta a saída do SPSS para o Teste KR-20 de Kuder-Richardson que, no software utilizado, consiste em um caso particular (quando há respostas dicotômicas aos itens) do Coeficiente Alfa de Cronbach (PASQUALI, 2017).

Figura 4 - Saída do Teste de Kuder-Richardson do SPSS.

Reliability Statistics		
Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,827	,834	45

Fonte: autores.

Foi possível analisar também no Teste KR-20 que nenhuma remoção de itens acarretaria no aumento da consistência interna do questionário. Ou seja, o conjunto dos 45 itens contribuiu inteiramente para o valor obtido pelo teste e é o maior índice possível em qualquer circunstância.

Na sequência, testou-se ainda a hipótese de existência do Efeito Halo, por meio do Teste T² de Hotteling, conforme exibido na Figura 5.

Figura 5 - Saída do SPSS para o Teste T² de Hotteling.

Hotelling's T-Squared Test				
Hotelling's T-Squared	F	df1	df2	Sig
3185,155	67,616	44	609	,000

Fonte: autores.

Como pode-se notar, o p-valor do Teste T² de Hotteling aponta para a não existência do efeito halo entre os respondentes (p-valor sig. <0,05), o que pode corroborar em favor da validade dos itens e do rigor metodológico no qual o TACB-S foi aplicado.

É importante ressaltar que alguns dos índices apresentados até então, como Dificuldade, Discriminação e Correlação Item-total, são influenciáveis de acordo com o desempenho geral do grupo estudado. Por isso, para complementar o julgamento realizado nos itens, na sequência, será apresentado o Teste de Correlação Ponto-bisserial e a Análise Fatorial Exploratória, pois, de acordo com a literatura da área (FIELD, 2009; PASQUALI, 2017; SOARES, 2018), esses testes são menos influenciados pela variabilidade do escore dos respondentes e, assim, podem apresentar uma visão de maior rigor metodológico para as conclusões de fidedignidade dos itens e do teste, de modo geral.

Ainda na Figura 3, é exibido o Coeficiente de Correlação Ponto-bisserial. Esse teste demonstrou que houve correlação positiva entre todas as alternativas verdadeiras e o escore de acertos dos respondentes. Da mesma maneira, houve correlação negativa para as alternativas falsas. Isso significa que participantes com melhores desempenhos tenderam a responder, de maneira majoritária, mais os itens corretos do que os incorretos. Uma incoerência nesse quantitativo sinalizaria ao pesquisador uma possível inadequação de algumas das alternativas dos itens, sugerindo sua revisão. Como o TACB-S possui alternativas dicotômicas, observou-se que o valor de correlação foi o mesmo para cada alternativa dos itens, pois os respondentes escolheram uma em detrimento da outra, sendo assim, apenas diferentes nos seus sinais.

ANÁLISE FATORIAL EXPLORATÓRIA

Ao cogitar a realização de uma análise fatorial exploratória, em um primeiro momento, faz-se necessário testar se os dados disponíveis são adequados a este procedimento estatístico. Para isso, realizou-se os testes de adequação de Kaiser-Meyer-Olkin (KMO) e Esfericidade de Bartlett (BTS). A Figura 6 apresenta esses resultados:

Figura 6 - Saída do SPSS para os Testes de adequação à análise fatorial.

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,815
Bartlett's Test of Sphericity	Approx. Chi-Square	5092,951
	df	990
	Sig.	,000

Fonte: autores.

De acordo com os quantitativos apresentados os dados são adequados aos procedimentos de uma análise fatorial ($KMO > 0,8$; $BTS \text{ sig.} < 0,0001$) (HAIR, *et al.*, 2009).

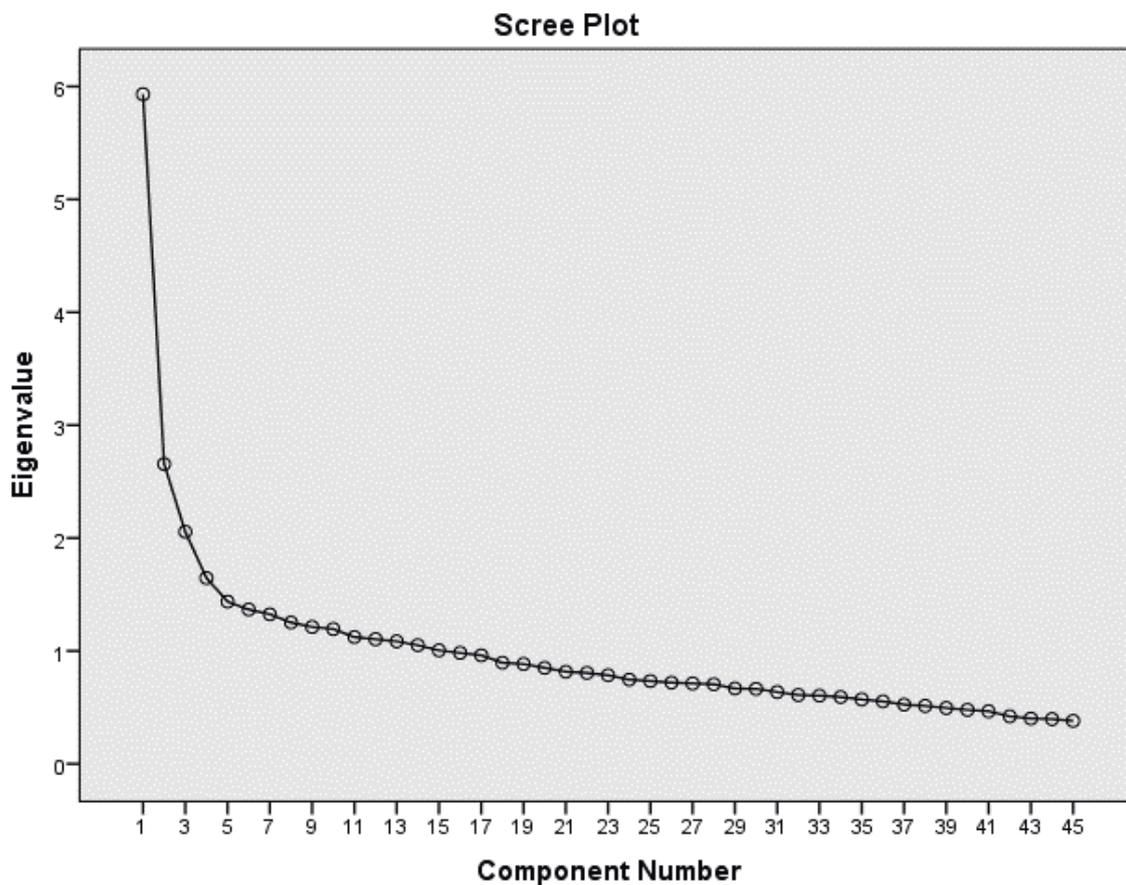
Ao prosseguir com a análise fatorial realizou-se a extração dos fatores por meio do método de Análise das Componentes Principais. A saída revelou que todos os 45 itens apresentaram comunalidades com carga fatorial acima de 0,3, o que confere relevância a cada uma das questões.

Ainda na extração dos fatores, observou-se, por meio do critério de Kaiser (FIELD, 2009), que um total de 15 componentes puderam ser consideradas, pois apresentaram autovalores superiores a 1. Essas componentes explicam 57% da variância total dos dados.

Ao analisar a carga fatorial de cada item em relação às componentes, dentro da matriz fatorial, notou-se que as 3 primeiras continham a maioria das questões do teste. Assim, procedeu-se a uma análise qualitativa com fins de verificar diferenças e semelhanças entre as questões, com vistas a apurar a possibilidade de alocar nas 3 primeiras, os demais itens, classificados em outras componentes. Dessa maneira, pode-se concluir que as 3 primeiras componentes possuíam aptidão para contemplar todos os 45 itens, pois haviam componentes extraídas que continham apenas 1 ou 2 itens.

A fim de verificar se essa escolha arbitrária poderia ser corroborada na análise fatorial, utilizou-se a análise do Gráfico Scree Plot (Gráfico 2), no qual apresenta a distribuição dos autovalores de cada uma das componentes. A literatura costuma delimitar as componentes adequadas para a estrutura do construto tendo como base a visualização do ponto de inflexão desse gráfico (HAIR *et al.*, 2009).

Gráfico 2 - Relação entre autovalores e componentes extraídas.



Fonte: autores. Citar a fonte

De acordo com a análise proposta pelo gráfico Scree Plot, pode-se afirmar que o ponto de inflexão acontece após a quarta componente. No entanto, como o TACB-S possui, em si, uma fundamentação teórica em três eixos, buscou-se analisar se as três principais componentes não estariam, por ventura, classificando os itens de acordo com os eixos de Miller (1983). Essa quarta componente continha itens de ambos os 3 eixos de Miller (1983) e qualitativamente não apresentavam uma coerência lógica entre si. Assim, após outra análise qualitativa dos itens, verificou-se como verdadeira essa hipótese, fazendo com que prevalecesse a estrutura teórica do instrumento. Ou seja, optou-se por manter a distribuição dos itens dentro das 3 componentes de maior autovalor.

Após essa análise confirmatória, classificou-se cada componente com os nomes dos três eixos de Miller (1983): Entendimento do conteúdo da Ciência (eixo 1); Entendimento da Natureza da Ciência (eixo 2); e Entendimento do Impacto da Ciência e Tecnologia na Sociedade e Ambiente (eixo 3). As Figuras a seguir apresentam a distribuição dos itens em cada uma das componentes, bem como, suas cargas fatoriais, resposta correta e, ao final de cada figura, o valor da consistência interna de cada eixo:

Figura 7 - Informações sobre a componente 1 “Entendimento dos conteúdos da Ciência.

Eixo 1: Entendimento dos conteúdos da Ciência			
Item	Descrição	Resposta	Carga fatorial
1	A Terra é tão antiga quanto o universo.	F	0,57
2	A luz da estrela mais próxima ao nosso sol leva apenas alguns minutos para chegar até nós.	F	0,58
3	A maioria do nosso conhecimento sobre o universo advém da observação de fatias muito pequenas do espaço e pequenos intervalos de tempo.	V	0,64
14	<i>Os biólogos classificam os organismos em grupos e subgrupos.</i> Isso é feito de uma forma que não está relacionada com a estrutura e o comportamento dos organismos.	F	0,51
15	Ao obter a energia e a matéria necessárias para a vida, os seres humanos são independentes das teias alimentares.	F	0,51
16	Cada gene é uma sequência específica da molécula de DNA.	V	0,49
17	Muitas das funções básicas de organismos, tais como a extração de energia a partir de nutrientes, são realizadas ao nível da célula.	V	0,54
18	A informação genética codificada em moléculas de DNA não desempenha nenhum papel na montagem de moléculas de proteína.	F	0,54
19	Os processos químicos na célula são controlados de dentro e de fora da célula.	V	0,56
20	A interdependência dos organismos em um ecossistema muitas vezes resulta em um sistema quase estável durante longos períodos de tempo.	V	0,46
21	Os ecossistemas sofrem alterações quando diferentes espécies aparecem.	V	0,56
22	Os organismos vivos não compartilham com outros sistemas naturais os mesmos princípios de conservação de matéria e energia.	F	0,57
23	Apenas uma pequena parte da vida na Terra é mantida por transformações de energia a partir do sol.	F	0,44
24	Os elementos que compõem as moléculas dos seres vivos são continuamente reciclados.	V	0,55
25	O carvão e o petróleo foram formados há milhões de anos.	V	0,54
26	A seleção natural costuma resultar em organismos com características bem adaptadas para sobrevivência em ambientes específicos.	V	0,57
35	A forma como os átomos se conecta é determinada pela disposição dos elétrons no exterior de cada átomo.	V	0,64
36	No universo, a energia só aparece em um formato.	F	0,45
37	Arranjos de átomos em moléculas não estão relacionados com os diferentes níveis de energia das moléculas.	F	0,55
38	As forças eletromagnéticas que atuam entre os átomos são muito mais fortes do que as forças gravitacionais que atuam entre eles.	V	0,59
39	Na maioria dos aspectos biológicos, os seres humanos são diferentes de outros organismos vivos.	F	0,61
40	O sistema imunológico desempenha um papel importante na autoproteção dos animais em relação às doenças.	V	0,57
41	Muito do aprendizado parece ocorrer através da interação de um novo pedaço de informação com um pedaço de informação já existente.	V	0,55
42	A boa saúde depende do esforço coletivo das pessoas de tomar medidas para manter seu ar, solo e água preservados.	F	0,55
43	Os genes anormais jamais afetam o modo de funcionamento das partes do corpo humano, nem dos seus sistemas.	F	0,57
44	Uma boa saúde mental não está relacionada com a interação dos aspectos psicológicos, biológicos, fisiológicos, sociais e culturais.	F	0,59
45	As anomalias biológicas podem causar alguns tipos de perturbações psicológicas graves.	V	0,53
Coefficiente Kuder-Richardson para o Eixo 1		27 itens	0,809

Fonte: autores.

As questões referentes ao eixo “entendimento dos conteúdos da ciência” são as de 1 a 3, 14 a 26 e as de 35 a 45. Observa-se que o eixo 1 apresenta um valor satisfatório de confiabilidade na sua estrutura (0,809).

Figura 8 - Informações sobre a componente 2 “Entendimento da Natureza da Ciência.

Eixo 2: Entendimento da Natureza da Ciência			
Item	Descrição	Resposta	Carga fatorial
4	Cedo ou tarde, a validade das afirmações científicas é comprovada através da observação de fenômenos.	V	0,59
5	Os cientistas discordam sobre os princípios de raciocínio lógico que conectam as evidências com as conclusões.	F	0,54
6	O processo de propor e testar hipóteses não é uma das principais atividades dos cientistas.	F	0,66
7	<i>Os cientistas tentam dar sentido aos fenômenos dando explicações para eles.</i> Essas explicações raramente usam princípios científicos atualmente aceitos.	F	0,62
8	As teorias científicas devem explicar observações adicionais que não foram utilizadas no desenvolvimento das teorias anteriores.	V	0,65
9	Os cientistas tentam identificar possíveis vieses no trabalho de outros cientistas.	V	0,59
10	Ao levar a cabo uma investigação, nenhum cientista deve sentir que ele / ela deve chegar a um determinado resultado.	V	0,73
11	A disseminação da informação científica não é importante para o progresso da ciência.	F	0,35
12	Os campos científicos como a química e a biologia possuem limites ou fronteiras.	F	0,59
13	Ética científica (ou seja, sistema de moral) está preocupada, entre outras coisas, com os possíveis efeitos nocivos da aplicação dos resultados da investigação.	V	0,72
Coeficiente Kuder-Richardson para o Eixo 2		10 itens	0,618

Fonte: autores.

Por sua vez, o eixo “entendimento da natureza da ciência” é composto pelas questões de 4 a 13. A consistência interna desse eixo apresentou valor ligeiramente menor, porém, também dentro dos limites aceitáveis pela literatura (PASQUALI, 2017).

Figura 9 - Informações sobre a componente 3 “Entendimento do Impacto da Ciência e Tecnologia na Sociedade e Ambiente.

Eixo 3: Entendimento do Impacto da Ciência e Tecnologia na Sociedade e Ambiente			
Item	Descrição	Resposta	Carga fatorial
27	Novos instrumentos e técnicas que estão sendo desenvolvidos através da tecnologia pouco contribuem para a pesquisa científica.	F	0,52
28	A tecnologia apenas fornece ferramentas para a ciência, raramente fornece motivação e direção para as pesquisas.	F	0,58
29	<i>Os efeitos de uma grande quantidade de objetos relativamente simples (por exemplo, fogões solares) podem ser individualmente pequenos.</i> No entanto, estes efeitos podem ser significativos, coletivamente.	V	0,61
30	Apesar da grande complexidade dos sistemas tecnológicos modernos, todos os efeitos colaterais de novos projetos tecnológicos são previsíveis.	F	0,52

31	Não importa quais precauções sejam tomadas ou quanto dinheiro é investido. Qualquer sistema tecnológico pode falhar.	V	0,64
32	As forças sociais e econômicas dentro de um país têm pouca influência sobre quais tecnologias serão desenvolvidas dentro desse país.	F	0,61
33	A tecnologia teve pouca influência sobre a natureza da sociedade humana.	F	0,61
34	O efeito gerado pelas decisões de um grande número de indivíduos distintos pode influenciar na utilização de tecnologia em larga escala, tanto quanto a pressão realizada pelos governos.	V	0,48
Coefficiente Kuder-Richardson para o Eixo 3		8 itens	0,700

Fonte: autores.

Por fim, o eixo “entendimento do impacto da ciência e tecnologia na sociedade e ambiente” compreende as questões da 27 até a 34. A confiabilidade desse último eixo também se mostrou adequada (0,700) para os seus 8 itens.

A AFE explorou as dimensões avaliativas do instrumento. Assim, podemos corroborar com a ideia de que a sua estrutura possui validade. O instrumento foi construído e, posteriormente reduzido, de maneira a preservar os seus aspectos de validade e confiabilidade.

CONSIDERAÇÕES

O artigo descreveu uma pesquisa que teve por objetivo realizar uma análise psicométrica dos itens de uma versão reduzida (VIZZOTTO; MACKEDANZ, 2018) do Teste de Alfabetização Científica Básica de Laugksch e Spargo (1996). Um teste piloto foi realizado com 653 egressos do Ensino Médio, público-alvo para qual o teste foi, inicialmente, elaborado. Por meio de técnicas estatísticas, aferiu-se, via Teoria Clássica de Testes, os índices de dificuldade, discriminação, consistência interna, correlação item-total, efeito halo, e também, a hipótese de unidimensionalidade, através de uma análise fatorial exploratória.

A justificativa para este trabalho centra-se na necessidade de assegurar que testes realizados em pequena, média e grande escala sejam considerados válidos e confiáveis para que seus resultados assumam legitimidade. Assim, devido a abrangência de aplicações que o TACB-S pode alcançar, como por exemplo, contribuir para a tarefa de aferir aspectos da educação científica de alunos dentro do cotidiano escolar, julgou-se importante conhecer suas qualidades psicométricas. Para isso, buscou-se responder se “os itens do TACB-S são compostos de características psicométricas que atestem a validade e confiabilidade do seu uso no contexto nacional?”.

As características psicométricas apresentadas mostram que o TACB-S possui indicativos de que mensura o que se propõe mensurar, bem como, suas medidas apresentariam valores confiáveis se aplicadas a um mesmo participante em diferentes intervalos de tempo, ou seja, possui confiabilidade.

Alguns itens revelaram baixo potencial de discriminação e correlação com o instrumento em geral. No entanto, análises paralelas reforçam a importância da manutenção de tais questões, pois contribuem, em um contexto integral do teste, para a aferição unidimensional de diferentes características da Alfabetização Científica de egressos do Ensino Médio (MILLER, 1983).

A análise fatorial demonstrou que a distribuição de cada eixo, postulada pela sua estrutura original, pode ser observada por meio da análise estatística, reforçando a estrutura tridimensional do teste. Sobretudo, as 3 componentes observadas corroboram para a estrutura de validade unidimensional

do teste. Assim, a análise fatorial exploratória observou que 3 componentes integram a estrutura que visa aferir um fator em comum. A esse fator, denominamos de Alfabetização Científica Básica.

Para estudos subsequentes, sugere-se a aplicação do instrumento em outras amostras dessa população a fim de verificar a estabilidade desses itens, bem como, corroborar a estrutura teórica das componentes.

REFERÊNCIAS

American Association for the Advancement of Science (AAAS). **Science for all Americans: A project 2061 report on literacy goals in science, mathematics and technology**. Washington, AAAS, v. 1, p. 256, 1989.

CAMARGO, Andrea Norema Bianchi de; PILAR, Fabiana Dias; RIBEIRO, Marcus Eduardo Maciel; FANTINEL, Mirian; RAMOS, Maurivan Guntzel. Alfabetização Científica: A evolução ao Longo da formação de Licenciandos Ingressantes, Concluintes e de Professores de Química. **Momento - Diálogos em Educação**, v. 20, n. 2, p. 19-29, 2011.

FIELD, Andy. **Descobrimo a estatística usando o SPSS**. Porto Alegre: Bookman Editora, 2009.

GIL, Antônio Carlos. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Editora Atlas SA, 2008.

HAIR JÚNIOR, Joseph; BLACK, William; BABIN, Barry; ANDERSON, Rolph; TATHAM, Ronald. **Análise Multivariada de Dados**. 6ª ed. Porto Alegre: Bookman Editora 2009.

LAUGKSCH, Rüdiger.; SPARGO, Peter. Construction of a paper-and-pencil Test of Basic Scientific Literacy based on selected literacy goals recommended by the American Association for the Advancement of Science. **Public Understanding of Science**, v. 5, p. 331-359, 1996.

MILLER, Jon. Scientific literacy: A conceptual and empirical review. **Daedalus: Journal of the American Academy of Arts and Sciences**, v. 112, n. 12, p. 29-48, 1983.

MOREIRA, Marco Antônio; ROSA, Paulo Ricardo Silva. **Uma introdução à pesquisa quantitativa em Ensino**. Porto Alegre: ed. dos Autores, 2007.

NASCIMENTO-SCHULZE, Clélia Maria. Um estudo sobre Alfabetização Científica com jovens catarinenses. **Psicologia: teoria e prática**, v. 8, n. 1, p. 95-117, 2006.

NASCIMENTO-SCHULZE, Clélia Maria; CAMARGO, Brigido Vizeu; WACHELKE, João Fernando Rech. Alfabetização científica e representações sociais de estudantes de ensino médio sobre ciência e tecnologia. **Arquivos Brasileiros de Psicologia**, v. 58, n. 2, p. 24-37, 2006.

OLIVEIRA, Will Fadul Alencar de; SILVA-FORSBERG, Maria Clara. Níveis de Alfabetização Científica de estudantes da última série do Ensino Fundamental. **Anais do VIII Encontro nacional de pesquisa em educação em ciências (ENPEC)**, Campinas, Abrapec, p. 1-11, 2011.

PASQUALI, Luiz. **Psicometria: teoria dos testes na psicologia e na educação**. Petrópolis: Editora Vozes, 2017.

RIVAS, Marcela Ines Espinosa. **Avaliação do nível de Alfabetização Científica de estudantes de biologia**. Trabalho de conclusão de curso. Graduação em biologia. Universidade Federal do Rio Grande do Sul - UFRGS, 2015.

RIVAS, Marcela Ines Espinoza; MOÇO, Maria Cecília de Chiara; JUNQUEIRA, Heloisa. Avaliação do nível de alfabetização científica de estudantes de biologia. **Revista Acadêmica Licenciaturas**, v. 5, n. 2, p. 58-65, 2017.

ROSA, Fernando; KAMAKURA, Wagner Antônio. Pesquisas de satisfação de clientes e efeito halo: interpretações equivocadas. **Encontro Nacional da Associação Nacional dos Programas de Pós-Graduação e Pesquisa em Administração**, v. 25, p. 1-15, 2001.

SOARES, Denílson. Júnior Marques. **Teoria clássica dos testes e teoria de resposta ao item aplicadas em uma avaliação de matemática básica**. Dissertação (Mestrado em Estatística Aplicada e Biometria). Universidade Federal de Viçosa, Viçosa, 2018.

SOUZA, Ana Cláudia de; ALEXANDRE, Neusa Maria Costa; GUIRARDELLO, E. B. Propriedades psicométricas na avaliação de instrumentos: avaliação da confiabilidade e da validade. **Epidemiologia e Serviços de Saúde**, v. 26, n. 3, p. 649-659, 2017.

VIANNA, Heraldo. Impactos dos testes sobre os sistemas e objetivos educacionais: a experiência brasileira. **Cadernos de Pesquisa**, n. 27, p. 69-71, dez. 1978.

VILARINHO, Ana Paula Lima. **Uma proposta de análise de desempenho dos estudantes e de valorização da primeira fase da OBMEP**. Dissertação (Programa de Mestrado Profissional em Matemática). Universidade de Brasília, Brasília, 2015

VIZZOTTO, Patrick Alves; MACKADANZ, Luiz Fernando. Teste de Alfabetização Científica Básica: processo de redução e validação do instrumento na língua portuguesa. **Revista Prática Docente**, v. 3 n. 2, p. 575-594, 2018.

RECEBIDO EM: 14 jan. 2020

CONCLUÍDO EM: 19 abr. 2020

