

# Construção e validação de instrumento de coleta de dados na pesquisa em Ensino de Ciências

Construction and validation of data collection instrument in science teaching research

Driane Anne Silva de Santana<sup>1</sup>  
Edson José Wartha<sup>2</sup>

## Resumo

Esta pesquisa trata do desenvolvimento metodológico de um instrumento de avaliação de capacidades de Pensamento Crítico com estudantes do Ensino Médio. O objetivo é apresentar e discutir as etapas utilizadas no processo de construção e validação do instrumento por meio de análises psicométricas e semânticas. O instrumento foi validado em quatro domínios (Percentual de Concordância, Índice de Validade de Conteúdo, Coerência e Concordância entre os participantes e Confiabilidade do Instrumento) relacionados aos 20 itens que buscam identificar a manifestação de Capacidades de Pensamento Crítico.

**Palavras chave:** Validade; Confiabilidade; Pensamento Crítico; Análises Quantitativas.

## Abstract

This research deals with the methodological development of an instrument for evaluating Critical Thinking competence with high school students. The purpose of this study is to present and discuss the steps used in the process of construction and validation of the instrument through psychometric and semantic analyzes. The instrument was validated in four domains (Percentage of Agreement, Content Validity Index, Coherence and Agreement between participants and Reliability of the Instrument) related to the 20 items that seek to identify the manifestation of Critical Thinking Capabilities.

**Keywords:** Validity; Reliability; Critical Thinking; Quantitative Analysis.

---

<sup>1</sup> Universidade Federal de Sergipe | drianasantana@gmail.com

<sup>2</sup> Universidade Federal de Sergipe | ejwartha@gmail.com

## Introdução

Considerando que o processo de construção e validação de um questionário de pesquisa é uma das ferramentas mais utilizadas por pesquisadores em Ensino de Ciências e, que, a elaboração das perguntas é etapa crucial para a obtenção de um instrumento que cumpra seu papel na coleta de dados de maneira a revelar, da melhor forma possível, a realidade dos fatos ocorridos dentro do alvo do estudo proposto, procura-se neste estudo apresentar as etapas utilizadas durante o processo de desenvolvimento e validação de um instrumento de avaliação de Capacidades de Pensamento Crítico (CPC). Tecemos algumas considerações, através de esquemas e exemplos, para a formulação de um questionário de pesquisa que a partir da utilização de conceitos extraídos do campo da Psicometria foram necessárias à validação do instrumento e à análise da fidedignidade das escalas desse instrumento.

Em geral, na área da pesquisa em Ensino de Ciências, excluindo análises de dados de avaliações de rendimento escolar realizadas por alguns pesquisadores que usam os dados do INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira) em seus estudos, poucos estudos empregam metodologias quantitativas. No entanto, há problemas e questões que para sua contextualização e melhor compreensão necessitam serem também qualificados através de dados quantitativos, como por exemplo, compreender como a distribuição de variáveis como o gênero, idade, condição socioeconômica, região geográfica, cidade-meio rural, escolaridade dos pais, etc., podem ser usados para compreender melhor um problema.

Ressaltamos que no emprego dos métodos quantitativos, precisamos considerar ao menos dois aspectos como ponto de partida: primeiro, que os números, frequências, medidas, têm algumas propriedades que delimitam as operações que se podem fazer com eles, e que deixam claro seu alcance; segundo, que as boas análises dependem de boas perguntas que o pesquisador venha a fazer, ou seja, da qualidade teórica e da perspectiva epistêmica na abordagem do problema, as quais guiam as análises e as interpretações. O significado dos resultados é dado pelo pesquisador em função de seu estudo teórico.

Neste estudo defendemos a ideia de que os métodos de análise de dados que se traduzem por números podem ser muito úteis na compreensão de diversos problemas na pesquisa em Ensino de Ciências. Mais ainda, a combinação deste tipo de dados com dados oriundos de metodologias qualitativas, podem vir a enriquecer a compreensão de eventos, fatos, processos. As duas abordagens demandam, no entanto, o esforço de reflexão do pesquisador para dar sentido ao material levantado e analisado. Para o processo de construção e validação de questionários, que são utilizados com frequência no levantamento de dados na área de pesquisa em Ensino de Ciências, é importante o uso de análises quantitativas de modo a dar uma maior consistência, validade e fidedignidade ao questionário. É muito comum o processo de validação externa de um questionário, como a análise de conteúdo e análise semântica, que geralmente é realizada por juízes. No entanto, é necessário também verificar a validade interna do questionário, ou seja, se o instrumento mede o que está proposto a medir, já que a falta desse processo pode comprometer os resultados encontrados pelos pesquisadores.

A validação interna atribui qualidade ao instrumento de medição. Uma boa mensuração, segundo Hubley e Zumbo (1996), deve ter duas características essenciais: confiabilidade e validade. A confiabilidade é sinônimo de consistência, estabilidade e previsibilidade. Já a validade tem um sentido de acuracidade,

autenticidade e veracidade. Apesar de ambas apresentarem conceitos diferentes, elas estão relacionadas da seguinte forma: uma observação pode ser confiável sem ser válida, mas não válida sem antes ser confiável. Em outras palavras, confiabilidade é necessária, mas não condição suficiente para a validade.

Neste estudo usamos como parâmetros para a validação interna do instrumento, a análise fatorial e a análise da consistência interna do instrumento. Lembramos que todas as medidas são arbitradas, criadas, inventadas e que não podem ser tomadas como sendo a própria natureza das coisas. Isto também se aplica às categorias dos estudos de análise de conteúdo e de outras análises qualitativas. O que se procura ao criar uma tradução numérica ou categorial de fatos, eventos, fenômenos é que esta tradução tenha algum grau de validade racional, teórica, no confronto com a dinâmica observável dos fenômenos (GATTI, 2004).

Pesquisas de análises quantitativas e qualitativas não são polos opostos e antagônicos, são complementares e oferecem diferentes perspectivas. Uma possibilidade para compreender melhor é pensarmos em um modelo escalonado. Para Gatti (2004) investir em pesquisas de caráter misto (ou quali-quantitativo) contribui para a compreensão de relações mais amplas. Um bom exemplo de como pesquisas de caráter misto pode ser melhor elaborado é um instrumento recorrente na pesquisa quantitativa é o *survey*, que consiste, basicamente, em um conjunto de questões padronizadas apresentadas a uma amostra populacional. Para Babbie (1999), questionários de pesquisa precisam estar apoiados em referências teóricas, decorrente de um acurado exame da literatura específica e, que no seu processo de construção e validação sejam dados devida atenção aos aspectos operacionais, tais como: formatação, clareza da linguagem, ordenamento das questões, bem como para o favorecimento da entrada das respostas. Ortigão e Pereira (2016) salientam que outro fator a ser considerado é a confiabilidade dos itens. Isto tem a ver com o comportamento semelhante de diferentes respondentes diante de um mesmo item.

Assim, queremos esclarecer que há muito a ser considerado na produção de um instrumento de pesquisa, e que tal instrumento não é o princípio nem o fim de uma pesquisa. A utilização de um instrumento validado garante que ele tenha sido corrigido o suficiente para assegurar sua pertinência e legitimidade.

## Etapas realizadas no processo de construção e validação do instrumento

O processo de desenvolvimento de um instrumento para mensurar as Capacidades de Pensamento Crítico (CPC) envolveu as seguintes fases:

**i) Construção do instrumento de avaliação de CPC:** elaboração dos itens e evidências de validade baseadas no conteúdo. Para esta fase foram realizadas três etapas: 1) identificação dos referenciais teórico-metodológicos; 2) construção dos itens; 3) organização do instrumento. Estas etapas tiveram como objetivo a especificação das categorias comportamentais que representam o objeto psicológico a ser medido e a operacionalização dos construtos em itens. Na etapa 1, no caso do CPC, a escolha dos atributos foi determinada pelas indicações da literatura (*California Critical Thinking Skills Test*, o *Watson-Glaser Critical Thinking Appraisal* e o *Cornell Critical Thinking Test* (RIEGEL, CROSSETTI, 2018) e pela análise documental dos documentos oficiais, como os Parâmetros

Curriculares Nacionais e a Base Nacional Comum Curricular. Neste caso, foram observados os modelos de questões, as capacidades avaliadas e as habilidades e competências que tem relação com as CPC. Na etapa 2, a elaboração do instrumento, teve início com a definição das dimensões do teste, em termos constitutivos e operacionais e, por fim, a elaboração dos itens. No caso das CPC, os principais referenciais teóricos que balizaram as definições dos construtos foram os estudos relacionados sobre Pensamento Crítico (ENNIS, 1985, TENREIRO-VIEIRA; VIEIRA, 2000, 2004, 2005) e os estudos sobre psicometria (NACIMENTO; GOMES, 2017; PASQUALI, 2009; ANDRIOLA, 2001; HAIR *et al.*, 1995, KIM, 2009). Na etapa 3, para criação dos itens e de sua organização no instrumento, procedeu-se da seguinte maneira: foram levantados os instrumentos disponíveis que avaliavam o mesmo construto ou construtos correlatos (*Cornell Critical Thinking Test – Level X*) e com o uso de um tema controverso (Fosfoetanolamina: a pílula do câncer) foi construído uma narrativa no qual os estudantes, para responderem às questões, deveriam apelar para algumas CPC. O instrumento foi construído com 20 itens de modo a avaliar as seguintes dimensões das CPC: indução (fazer e avaliar induções), observação (fazer e avaliar observações-considerações importantes), credibilidade (avaliar a credibilidade de uma fonte), dedução (fazer e avaliar deduções), assunções (identificar assunções). O instrumento foi construído com quatro itens para avaliar cada CPC.

ii) **Avaliação pelo Comitê de Juízes:** para os procedimentos da análise de juízes da versão piloto do instrumento, contou-se com professores de química da Educação Básica, pesquisadores em Ensino de Ciências e especialistas em Pensamento Crítico. Os procedimentos incluíram a apresentação dos itens, da definição do construto e das dimensões das CPC cobertas pelo teste. Com base nisso, os especialistas indicavam se o item era pertinente ao traço latente a que teoricamente deveria se referir. Indicavam se havia alguma faceta do construto não coberta pelo instrumento e se a redação do item era inteligível para o público alvo. A análise de conteúdo tem como base a necessidade do julgamento realizado por um grupo de juízes experientes na área inserida do instrumento, a qual caberá aos especialistas investigar se o conteúdo está correto e adequado ao que se propõe (PASQUALI, 2009). Um estudo de validade de conteúdo é essencial para promover informações sobre a representatividade e clareza de cada questão com a colaboração de especialistas, no entanto, vale destacar que existem limitações nesses tipos de validade visto que análise dos juízes é subjetiva (PASQUALI, 2009). Assim, um possível procedimento para diminuir tal subjetividade é o cálculo do Índice de Validação de Concordância (IVC), que pode ser compreendido como a medida da consistência entre o valor absoluto das classificações dos avaliadores. Neste estudo empregamos para calcular o IVC a estatística Kappa de Cohen (coeficiente de Kappa), que indica o grau de consenso entre os avaliadores expressos num índice variando de 0 a 1. Para Stemler (2004), o valor de 0,75 é considerado o mínimo de concordância aceitável na validação de um instrumento para uso em testes de desempenho. Quanto a avaliação, solicitamos aos juízes que analisassem cada item os domínios de CPC e quanto a três critérios: clareza, pertinência e aparência atribuindo a eles, valores de 1 a 5, em que o valor 1 representa baixa concordância e valor 5 alta concordância. Após a avaliação pelo Comitê de Juízes com a proposta de alteração de itens, foi elaborada a segunda versão do instrumento, que novamente foi encaminhada a outra avaliação pelo Comitê de Juízes. Após esta fase o instrumento foi reelaborado e encaminhado para a avaliação semântica.

iii) **Análise semântica:** o instrumento construído após a análise de conteúdo pelo Comitê de Juízes foi apresentado para uma amostra de 40 estudantes semelhantes ao público-alvo, de ambos os sexos. Além de responderem aos itens, os estudantes deveriam tecer comentários sobre as questões, como por exemplo, o grau de dificuldade, termos específicos, bem como deveriam dizer o que cada item expressava e, para aumentar a validade aparente, sugerir eventuais modificações na linguagem do item. Esta análise busca identificar se há compreensão dos itens pelo público-alvo, ou seja, a população a que se destina o teste, tendo relevância quanto a inteligibilidade dos itens e a validade aparente.

iv) **Administração da versão piloto do instrumento:** 297 estudantes do Ensino Médio responderam aos itens do instrumento, sendo 162 estudantes do sexo feminino e 135 do sexo masculino. Os dados obtidos por meio do instrumento construído e validado semanticamente foram agrupados em um banco de dados, utilizando-se do aplicativo *Microsoft Office Excel*® 2010 e analisados pelo software estatístico SPSS®, versão 23. No processo de validação interna dos itens foram realizadas análises estatísticas como o Alfa de Cronbach e a Análise Fatorial Exploratória. Para a análise da consistência interna do instrumento, utilizou-se o teste de Alfa de Cronbach ( $\alpha$ ), por se apresentar como uma das ferramentas estatísticas mais importantes e difundidas em pesquisas envolvendo a construção de testes e de sua validação. É um índice utilizado para medir a confiabilidade do tipo consistência interna de uma escala, ou seja, para avaliar a correlação entre as respostas do instrumento aplicado por meio do perfil das respostas dadas pelos sujeitos da pesquisa (ALMEIDA *et al.*, 2010).

Pasquali (2009) aponta algumas dificuldades que podem surgir no processo de validação e elaboração do instrumento, visto que, se perpassa em três níveis que são em relação à teoria, da coleta empírica da informação e da própria análise estatística da informação. Para o autor, um teste é válido se de fato ele mede o que deveria medir. A confiabilidade de um teste depende da função do instrumento, o público alvo que é administrado e as circunstâncias, ou seja, o mesmo teste pode não ser confiável em diferentes situações. Sendo que a confiabilidade pode ser afetada pelos avaliadores, pelas características da amostra, tipo de instrumento, método de administração e pelo método estatístico que foi adotado (SOUZA *et al.*, 2017).

No processo de validação interna utilizamos também a análise fatorial exploratória (AFE), visto que é um dos métodos estatísticos mais usados no desenvolvimento, avaliação e aprimoramento de instrumentos psicométricos. Conceitua-se esta análise sendo como um conjunto de técnicas multivariadas que tem como meta encontrar a estrutura subjacente em uma matriz de dados e determina o número e a natureza dos fatores que melhor representam um conjunto de variáveis (DAMÁSIO, 2012). Pode-se dizer que a principal função da AFE é diminuir o número de fatores observados, que de acordo com Hair *et al.*, (2014) é uma forma de encontrar uma maneira de condensar a informação contida em um número de variáveis originais em um conjunto menor de variáveis (fatores) com uma perda mínima de informações.

A primeira etapa para implantar AFE é verificar se os dados podem de ser submetidos ao processo de análise fatorial (PASQUALI, 1999). Com este objetivo, utiliza-se dois métodos estatísticos, o teste de Kaiser-Meyer-Olkin (KMO) e o teste de Esfericidade de Bartlett. O índice de KMO é um teste de adequação da amostra que sugere a proporção de variância dos itens que pode ser explicada por uma variável latente, sendo que este índice indica se é adequado aplicar a AFE para o conjunto de dados (DAMÁSIO, 2012; FERREIRA, HONGYU,

2018; PINTO, 2012). O teste de esfericidade de Bartlett averigua em que medida a matriz de (co)variância é similar a uma matriz-identidade. Para Hair *et al.*, (2014) esse teste avalia a significância de todas as correlações de uma matriz de dados. Os valores desse teste com níveis de significância  $p < 0,05$  indicam que a matriz é favorável. Os testes de KMO e Bartlett tem a tendência de serem uniformes, aceitando ou negando a possibilidade da AFE da matriz de dados. Se o conjunto de dados é passível de análise fatorial, o pesquisador deve prosseguir avaliando a amostra para que possa escolher o método de extração apropriado. Geralmente utiliza-se a máxima Varimax como método de extração e principais componentes fatoriais (DAMÁSIO, 2012).

v) **Aspectos éticos:** o referido estudo foi encaminhado ao Comitê de Ética em Pesquisa da Universidade Federal de Sergipe sob o instrumento número CP 06958018.3.0000.5546. Deste modo, este estudo pauta-se nas determinações da Resolução 466/12 que regulamenta a pesquisa com seres humanos (BRASIL, 2012).

## Resultados do processo de construção e validação do instrumento

Os resultados deste estudo estão apresentados seguindo as etapas do processo de construção do instrumento de avaliação de CPC com estudantes do Ensino Médio em sua fase teórica: avaliação pelo Comitê de Juízes, análise semântica e a fase empírica, piloto e validação interna do instrumento.

Na avaliação pelo Comitê de Juízes foi encaminhado a primeira versão do instrumento com 20 itens e uma tabela contendo as dimensões de cada grupo de itens relacionados com as CPC, no qual deviam indicar valores de 1 a 5. Sendo denominados como: juiz 1 (J1), juiz 2 (J2), juiz 3 (J3), juiz 4 (J4) e juiz 5 (J5). De maneira geral, os principais ajustes solicitados pelos juízes foram a substituição de palavras tanto no texto introdutório do teste como em alguns itens. Além disso, a modificação de palavras e frases feitas pelos juízes foi com o intuito de utilizar uma linguagem que fosse mais próxima da faixa etária do público-alvo dessa pesquisa. Visto que o estudo de Araújo (2007) aponta que o pesquisador deve atentar para a linguagem ou se não causará problemas na compreensão das perguntas pelos respondentes.

Tabela 1. Índice de validação de concordância entre os juízes em relação as CPC avaliadas em cada conjunto de itens.

Capacidade de Pensamento Crítico	Itens	Índice de validação de concordância (IVC)
Credibilidade	1 a 5	1,00
Assunções	6 a 10	0,75
Observação	1-4,11-14	1,00
Dedução	11-15,18-20	0,50
Indução	16-20	0,75

Fonte: Resultado do IVC das capacidades de Pensamento Crítico aplicado ao Comitê de Juízes.

Pelo IVC foi possível identificar a necessidade de repensar a relação entre alguns itens e a dimensão da CPC que deveriam mensurar. Apenas nos itens da dimensão Credibilidade e

Observação há unanimidade na distribuição dos valores, ou seja, alta concordância entre os juízes. Nos demais itens verifica-se uma dispersão entre as notas atribuídas a cada dimensão. Consideramos que IVC acima de 0,70 adequados ao instrumento e que aqueles abaixo de 0,70 seriam considerados na administração da versão piloto do instrumento.

Em relação a análise semântica, realizada por uma amostra de estudantes que responderam ao teste e fizeram um comentário sobre os itens do instrumento, foi observado que os principais comentários se referem ao tamanho do texto utilizado em cada item. Os comentários eram na sua maioria relacionados ao tamanho e a quantidade de questões. Um aspecto positivo foi eles indicarem dificuldades em compreender palavras como: “assolavam” que se encontrava no texto introdutório e o termo científico “neoplasia maligna” na qual solicitaram uma melhoria ou substituição dessas palavras. Assim, foi possível substituir os termos indicados pelos estudantes e reduzido a redação naqueles itens em que era possível redução sem perda de informação.

O instrumento piloto foi administrado a uma amostra de 297 estudantes que não participaram da análise semântica. A tabela 2 mostra a caracterização dos participantes.

Tabela 2. Caracterização dos estudantes que responderam ao instrumento de CPC.

Características		N	%
Sexo	Masculino	135	45,5
	Feminino	162	54,5
Trabalha	Sim	54	18
	Não	243	82
Série	1	156	52,5
	3	141	47,5

Fonte: – Resultado relativo ao questionário socioeconômico aplicado no instrumento piloto.

Após a organização dos dados na planilha do *Microsoft Office Excel*®, estes foram submetidos as análises estatísticas descritivas e inferências usando o programa estatístico SPSS versão 23. Na análise da consistência interna do instrumento, utilizou-se o teste de Alfa de Cronbach ( $\alpha$ ), que para a versão piloto o valor encontrado foi de 0,56 considerando os 20 itens padronizados. Valores de Alfa de Cronbach menores que 0,6 são considerados com baixo valor de confiabilidade. Há pesquisadores como Freitas e Rodrigues (2005) que consideram valores próximos a 0,6 como aceitáveis em instrumentos da área de Ciências Sociais e Humanas. Por outro lado, é necessário considerar que “a elevada consistência na presença de multidimensionalidade indica que os itens que compõem as diferentes dimensões de uma medida estão fortemente correlacionados” (MAROCO; GARCIA-MARQUES, 2006, p. 67–70) e que uma baixa consistência é indicador que os itens não estão fortemente relacionados. Assim, valores de Alfa de Cronbach de 0,50 a 0,60 podem ser considerados aceitáveis desde que os resultados obtidos com esse instrumento sejam interpretados com precaução e tenham em conta o contexto de computação do índice e, que acima de tudo seja uma investigação preliminar (DEVELLIS, 1991).

Na análise fatorial exploratória (AFE) buscamos identificar se a estrutura de fatores do instrumento faz sentido, consistindo em verificar se uma série de itens pode ser reduzida idealmente a uma única dimensão ou variável. A carga fatorial gerada na presente análise indica a covariância entre o item e o fator, e quando o item se aproxima de 100% de covariância ele pode ser considerado ainda mais eficiente (PASQUALI, 2010). O primeiro

passo durante a implementação de AFE é verificar se a aplicação da AFE tem validade para as variáveis escolhidas, sendo justificado pela pouca quantidade de respondentes da pesquisa. Para isso, dois métodos de avaliação são mais comumente utilizados, o critério de Kaiser-Meyer-Olkin (KMO) e o Teste de Esfericidade de Bartlett (DZIUBAN; SHIRKEY, 1974; DAMÁSIO, 2012).

Tabela 3. Valores dos Teste de KMO e Bartlett para a administração do piloto.

Teste de KMO e Bartlett	
Medida Kaiser-Meyer-Olkin de adequação de amostragem.	0,63
Teste de esfericidade de Bartlett	471,25
gl	190
sig	0,00

Fonte: Elaborado pelos autores a partir dos dados do instrumento piloto.

Os valores dos testes aferem que os dados do instrumento são adequados à análise fatorial. O índice de KMO (amostra) encontrado foi de 0,63, ou seja, superior ao patamar crítico de 0,5. Da mesma forma o teste de esfericidade de Bartlett com nível de significância  $p < 0,05$ . O teste de Bartlett foi altamente significativo [ $X^2(190) = 471,25$ ,  $p$  menor que 0,001]; portanto, a realização da análise fatorial é apropriada.

A etapa seguinte, do processo de análise, foi determinar o número de fatores adotando a AFE com a rotação Varimax. Como resultado obtivemos oito fatores com autovalores maiores a 1 (um).

Tabela 4: Distribuição dos autovalores para os fatores da variância total explicada

Componente	Autovalores iniciais			Somadas de extração de carregamentos ao quadrado		
	Total	% de variância	% cumulativa	Total	% de variância	% cumulativa
1	2,585	12,92	12,92	2,585	12,92	12,92
2	1,578	7,88	20,81	1,578	7,88	20,81
3	1,386	6,93	27,74	1,386	6,93	27,74
4	1,297	6,48	34,22	1,297	6,48	34,22
5	1,224	6,11	40,34	1,224	6,11	40,34
6	1,161	5,80	46,15	1,161	5,80	46,15
7	1,095	5,47	51,62	1,095	5,47	51,62
8	1,022	5,10	<b>56,74</b>	1,022	5,108	<b>56,74</b>

Fonte: Elaborado pelos autores a partir dos dados do instrumento piloto.

Os oito fatores obtidos explicam 56,74% da variância dos itens estudados. Com a ajuda do gráfico de escarpa (Figura 1), estes fatores principais resumem efetivamente a variância total e podem ser utilizados para o estudo do conjunto de dados. Para fazer a redução a dimensões, precisamos saber o quanto de variância dos nossos dados são variâncias comuns (cumulativa). Todavia, a única maneira de sabermos a extensão da variância comum é reduzir as variáveis em dimensões. Desse modo, na análise dos componentes principais utiliza-se a variância total e assume-se que a comunalidade de cada variável é 1, transpondo os dados originais em componentes lineares constituintes. Na análise de fatores principais apenas a variância comum é usada e vários métodos de estimação das comunalidades podem ser usados, comumente se utiliza o quadrado da correlação múltipla de cada variável com todas as outras.



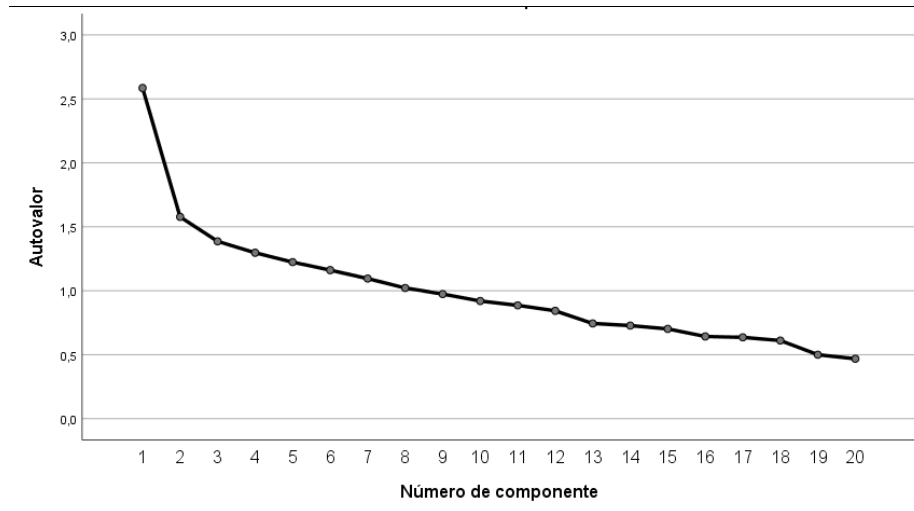


Figura 1: Gráfico de escarpa. Fonte: Resultado da variância das questões no instrumento piloto.

A figura 1 ilustra a dispersão dos componentes. A linha pontilhada ilustra o critério de Kaiser considerando autovalores maiores que 1. Pelo critério do *scree plot*, o número de fatores a ser extraído é o número de fatores à esquerda do ponto de inflexão - neste caso, 3 fatores.

Com o resultado da AFE foi possível reagrupar alguns itens e renomear aspectos dimensionados na fase inicial da pesquisa. Na tabela 5 são apresentados os fatores, os itens e a carga fatorial a eles associados. Ressaltando que cada fator os itens foram ordenados de forma decrescente de acordo com sua carga fatorial.

Tabela 5: Análise Fatorial Exploratória para os 20 itens do instrumento piloto.

Nº	F1	F2	F3	F4	F5	F6	F7	F8
Q5	0,757							
Q4	0,634							
Q13	0,583							
Q 3	0,437							
Q18		0,686						
Q19		0,593						
Q12		0,565						
Q20		0,409						
Q10			0,622					
Q8			0,616					
Q9			0,508					
Q11			0,416					
Q17				0,696				
Q14				0,673				
Q7					0,742			
Q16						0,805		
Q15						0,511		
Q2						0,494		
Q6							0,799	
Q1								0,794

Fonte: Resultado da AFE a partir dos dados do instrumento piloto.

Para considerar um item representativo de um fator, seguimos a orientação de Hair *et al.*, (2014) que indica carga fatorial mínima de 0,35 em amostras de pelo menos 250 respondentes. Sendo que os itens do instrumento piloto apresentaram cargas fatoriais com os respectivos domínios que variaram de magnitude moderada a forte, ou seja, de 0,416 a 0,805. A partir da AFE destacamos alguns pontos importantes que foram observados ao longo da análise: os valores das cargas fatoriais foram maiores que 0,40 para todos os itens e que alguns itens são consistentes aos domínios das CPC propostos inicialmente, enquanto outros há necessidade de alocar em outros domínios.

Assim, os itens de determinado domínio que inicialmente era proposto para determinados itens, com a AFE, itens que não se enquadram neste agrupamento necessitam se alocarem em outro domínio. Levando em consideração o ICV indicado pelo Comitê de Juízes e o resultado na análise da AFE, vimos que domínio das CPC para aquele conjunto de itens proposto inicialmente deverá sofrer alterações para um novo agrupamento de itens para avaliar a dimensão de CPC, conforme a tabela 6.

Tabela 6: Relação entre os domínios de CPC e os itens que se propõem a avaliar.

Domínios (CPC)	Itens do piloto	Itens do novo instrumento
Credibilidade	1-5	3-5,13
Assunções	6-10	12,18-20
Observação	1-4,11-14	8-11
Dedução	11-15,18-20	17,14
Indução	16-20	2, 15-16

Fonte: Resultado do reagrupamento dos itens nos domínios a partir dos dados da AFE.

Após o processo de construção e validação do instrumento, podemos afirmar que no que concerne à validação do instrumento para mensurar Capacidades de Pensamento Crítico, é imprescindível que instrumentos que validem conteúdos na área de Ensino de Ciências apresentem confiabilidade e validade adequadas. As análises quantitativas podem contribuir para que as pesquisas do tipo *survey* possam ser construídas por pesquisadores em seu contexto de uso e, desse modo evitar a adaptação de instrumentos construídos e validados em outros contextos. Também, é pertinente que instrumentos construídos e validados em outros contextos, quando adaptados ao novo contexto, estes sejam novamente validados (SEPINI; ALONSO; MACIEL, 2014). A escolha dessa abordagem foi incitada pelo desejo de buscar o diálogo entre a análise qualitativa e quantitativa, através de seu entrelaçamento, explorando, assim, as diferentes perspectivas de cada uma, a fim de compreender e aprofundar os vários aspectos do fenômeno a explicar. O teste de Capacidades de Pensamento Crítico, bem como todos os dados e informações relativos ao processo de construção e validação, pode ser acessado em Santana (2019).

## Considerações finais

Este estudo teve como objetivo apresentar e discutir um processo de construção e validação de um instrumento para mensurar Capacidades de Pensamento Crítico com estudantes do Ensino Médio. O processo de validação do instrumento foi apresentado e discutido na validação de conteúdo (Comitê de Juízes), na validação

semântica (análise comentada por amostra de estudantes), na verificação de confiabilidade (Alpha de Cronbach) e na análise detalhada dos itens (análise fatorial exploratória).

O processo de validação, etapa fundamental no processo de construção de medidas e instrumentos confiáveis na área de pesquisa em Ensino de Ciências, apresenta-se como um procedimento metodológico que pode auxiliar o pesquisador a decidir se deve ou não aplicar os resultados em sua área de pesquisa. Ressalta-se que a confiabilidade e a validade não são propriedades fixas e, portanto, variam de acordo com as circunstâncias, população, tipo e finalidade do estudo.

Ressaltamos que o processo de construção e validação do instrumento de CPC apresentou algumas limitações como, por exemplo, o objetivo exclusivo de avaliar capacidades de pensamento crítico em diferentes domínios. Neste caso, a representação comportamental de aspectos latentes, em que os dados, geralmente, são compostos por variáveis não mensuráveis diretamente, ou seja, são percebidos de forma indireta através de indicadores observáveis, como itens de uma escala de avaliação psicológica ou de comportamento e atitudes.

Esta pesquisa procura apresentar os caminhos necessários para o desenvolvimento de um instrumento de avaliação em que nos foi permitido uma série de reflexões sobre os passos que devem ser traçados durante a construção e validação de um instrumento. Vale ressaltar que os valores abaixo de 0,6 do Alpha de Cronbach no piloto e na AFE nos indicam que o instrumento necessita de alguns ajustes, principalmente em relação à organização e arranjo dos itens, bem como no aumento de itens para cada domínio.

Como resultado do processo de validação ficou evidente a necessidade de reformular o instrumento de avaliação desenvolvido neste estudo, seguindo novamente os passos descritos no processo de construção e validação. Ressaltamos que este percurso metodológico deveria ser requisito obrigatório para qualquer instrumento construído com objetivos educacionais para a área de Ensino de Ciências no sentido de apresentar confiabilidade e validade para poder afirmar que de fato o instrumento mede o que se propõe medir.

O processo de construção e validação de um instrumento psicométrico é uma tarefa que envolve amplo estudo e várias etapas de desenvolvimento e testagem. A explicitação de tais etapas e dificuldades encontradas permite compreender os processos inerentes, resultados alcançados e o tipo de instrumento que vai ser colocado à disposição da comunidade, assim como pode orientar novos projetos com objetivo semelhante. Um instrumento de avaliação de CPC, embasado na literatura sobre o tema e nos testes já existentes, para que seja confiável, deve conter itens que possibilitem identificar funções cognitivas complexas. Além disso, a situação problema deve envolver conflitos vivenciados pelos estudantes em seu contexto, que ofereça informações necessárias para se fazer as deduções e inferências, assim como selecionar dados relevantes e irrelevantes que possibilitem uma tomada de decisão que seja resultado do pensamento crítico. Ao mesmo tempo, a situação deve manter um nível de complexidade que exija reflexão e raciocínio para sua solução. Obter este equilíbrio é uma tarefa complexa e dinâmica que exige dos formuladores, criatividade e conhecimento das dinâmicas envolvidas na operação do Pensamento Crítico (PC). Embora a tarefa esteja longe de se considerar concluída, o processo aqui descrito permitiu a equipe um melhor delineamento das etapas de produção necessárias para o desenvolvimento e validação do instrumento proposto. Três ideias são,

então, centrais na reelaboração do instrumento: (i) escolha dos processos cognitivos essenciais à avaliação de CPC sem a necessidade de abarcar todos os processos constantes de instrumentos similares; (ii) um formato de resposta que possibilite aos sujeitos escolher e justificar respostas, evitando situações de “itens de múltipla escolha” com base em alternativas de resposta facultadas ou redação de respostas abertas que dificultam a sua análise e cotação posterior; e (iii) uma prova cuja extensão não seja demasiada, tornando possível uma avaliação em torno dos 30-40 minutos.

## Agradecimentos

O desenvolvimento deste trabalho foi possível graças ao auxílio do EDITAL CAPES/FAPITEC/SE N° 10/2016- Programa de Estímulo a Mobilidade e ao Aumento da Cooperação Acadêmica da Pós-Graduação em Sergipe no projeto “Pluralismo epistemológico e as tecnologias de avaliação do desempenho escolar em ciências: um estudo multidisciplinar”. Agradecimento especial ao Professor Dr. Ulisses Vieira Guimarães (UFS) e a Professora Dra. Suzi Sama Pinto (FURG) pela colaboração e orientação na escolha das análises estatísticas bem como sua interpretação e análises.

## Referências

- ALMEIDA, D.; SANTOS, M.A.R.; COSTA, A.F.B. Aplicação do coeficiente alfa de Cronbach nos resultados de um questionário para avaliação de desempenho da saúde pública. **XXX Encontro Nacional de Engenharia de Produção**, v. 15, p. 1-12, 2010. Disponível em: [http://www.abepro.org.br/biblioteca/enegep2010\\_TN\\_STO\\_131\\_840\\_16412.pdf](http://www.abepro.org.br/biblioteca/enegep2010_TN_STO_131_840_16412.pdf). Acesso em: 04 mai. 2019.
- ANDRIOLA, W. B. Descrição dos principais métodos para detectar o funcionamento diferencial dos itens DIF. **Psicologia: reflexão e crítica**, v. 14, n. 3, p. 643-652, 2001. Disponível em: <http://www.scielo.br/pdf/prc/v14n3/7850.pdf>. Acesso em: 03 mar. 2019.
- ARAÚJO, D. **Enunciados de Provas Escritas: uma investigação dos problemas de compreensão subjacentes a sua leitura**. Dissertação de Mestrado em Educação, Pós-Graduação em Educação, Universidade Federal de Pernambuco, Pernambuco, 2007.
- BABBIE, E. **Métodos de pesquisas de survey**. Belo Horizonte: Ed. UFMG, 1999.
- BRASIL. **Base Nacional Comum Curricular**. Ministério Público. Portaria MEC nº 790, de 27 de julho de 2016.
- BRASIL. **Resolução 466 de 12 de dezembro de 2012. Dispõe sobre pesquisa envolvendo seres humanos**. Conselho Nacional de Saúde.
- DAMASIO, B.F. Uso da análise fatorial exploratória em psicologia. **Aval. psicol.** v.11, n.2, p. 213-228, 2012. Disponível em: <http://pepsic.bvsalud.org/pdf/avp/v11n2/v11n2a07.pdf>. Acesso em: 04 nov. 2018.
- DEVELLIS, R.F. **Scale development: theory and applications**. California: Sage Publicatio, 1991.

- DZIUBAN, C.D.; SHIRKEY, E. When is a correlation matrix appropriate for factor analysis? Some decision rules. **Psychological Bulletin**, 81(6), 358-361, 1974. Disponível em: <https://psycnet.apa.org/record/1974-28961-001>. Acesso em: 12 fev. 2019.
- ENNIS, R. A logical basis for measuring critical thinking skills. **Educational Leadership**, n. 43, v. 2, p. 46, 1985. Disponível em: <https://eric.ed.gov/?id=EJ327936>. Acesso em 10 set. 2018.
- FERREIRA, A.A.; HONGYU, K. Análise fatorial exploratória aplicada no questionário do cadastro único. **Revista Biodiversidade**, v.17, n. 1, 2018. Disponível em: <http://periodicoscientificos.ufmt.br/ojs/index.php/biodiversidade/article/view/6533>. Acesso em: 14 mai. 2018.
- GATTI, B. A. Estudos quantitativos em educação. **Educação e Pesquisa**, v. 30, n. 1, p. 11-30, 2004. Disponível em: <http://www.scielo.br/pdf/ep/v30n1/a02v30n1.pdf>. Acesso em: 20 dez. 2018.
- HAIR, J.F.; ANDERSON, R.; TATHAM, R.; BLACK, W. **Multivariate data: Analysis with readings**. New Jersey: Prentice Hall, 1995.
- KIM, Y. Validation of psychometric research instruments: the case of information science. **Journal of the American Society for Information Science and Technology**, v.60, n. 6, 2009. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.21066>. Acesso em: 18 jan. 2019.
- MAROCO, J.; GARCIA-MARQUES, T. Qual a fiabilidade do Alfa de Cronbach? Questões antigas e soluções modernas? **Laboratório de Psicologia**, v. 4, n. 1, p. 65-90, 2006. Disponível em: <http://repositorio.ispa.pt/handle/10400.12/133>. Acesso em: 10 de fev. 2019.
- NASCIMENTO, R.D.; GOMES, A.D.T. Análise da validade e da fidedignidade de um questionário para identificação do conhecimento conceitual sobre plano inclinado e aceleração. **Amazônia: Revista de Educação em Ciências e Matemática**, v. 13, n. 26, 2017. Disponível em: <https://periodicos.ufpa.br/index.php/revistaamazonia/article/view/4284>. Acesso em: 03 jun. 2019.
- ORTIGÃO, M. I. R; PEREIRA, T. V. Homogeneização curricular e o sistema de avaliação nacional brasileiro: o caso do estado do Rio de Janeiro. **Educação, Sociedade & Culturas**, v. 47, n. 1, p. 157-174, 2016. Disponível em: <https://www.fpce.up.pt/ciie/sites/default/files/ESC47Isabel.pdf>. Acesso em 08 de ago. 2018.
- PASQUALI, L. **Instrumentação psicológica: Fundamentos e práticas**. Porto Alegre, Brasil: Artmed, 2010.
- PASQUALI, L. **Instrumentos psicológicos: manual prático de elaboração**. Laboratório de Pesquisa em Avaliação e Medida (LabPAM) - Instituto de Psicologia. Brasília: Universidade de Brasília, 1999.
- PASQUALI, L. Psychometrics. **Revista da Escola de Enfermagem da USP**, v. 43, n. SPE, p. 992-999, 2009. Disponível em: <http://www.revistas.usp.br/reeusp/article/view/40416>. Acesso em 08 de ago. 2018.
- PEREIRA, S.; ALICH, V. A avaliação do pensamento crítico numa perspectiva psicológica. **Revista Pensamento Crítico na Educação: desafios atuais**, p-87-98, 2015.

PINTO, I.R.F. **Atividades promotoras de pensamento crítico: sua eficácia em alunos de ciências da natureza do 5º ano de escolaridade**. 2012. Dissertação (Mestrado em Educação) - Escola Superior de Educação de Lisboa.

RIEGEL, F.; CROSSETTI, M.G.O. Referencias teóricas e instrumentos para la evaluación del pensamiento crítico en la enfermería y en la educación. **Rev. Gaúcha Enferm.** v. 39, 2018.

SANTANA, D. A. **Pensamento crítico: construção e validação de um instrumento de avaliação**. Dissertação. Universidade Federal de Sergipe, 2019. Disponível em: [https://ri.ufs.br/bitstream/riufs/11419/2/DRIANE\\_ANNE\\_SILVA\\_SANTANA.pdf](https://ri.ufs.br/bitstream/riufs/11419/2/DRIANE_ANNE_SILVA_SANTANA.pdf). Acesso em: 03 out. 2019

SEPINI, R. P.; VÁZQUEZ, A. A.; MACIEL, M. D. Mudanças de concepções atitudinais sobre a natureza da ciência e tecnologia em estudantes da escola básica após intervenção didática. **Revista de Educação em Ciências e Matemática**, v. 10, n.20, p. 101-111, 2014. Disponível em: <http://periodicos.ufpa.br/index.php/revistaamazonia/article/view/2317>. Acesso em: 03 out. 2019.

SOUZA, S.S.; CUNHA, C.; WARTHA, E. J.; PAGAN, A. A. **Uma nova abordagem para o desempenho escolar em ciências: vida e ambiente; ser humano e saúde**. 1. ed. Curitiba-PR: CRV, 2017. 264p

STEMLER, S. E. A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. **Practical Assessment, Research & Evaluation**, v. 9, n. 4, 2004. Disponível em: < <https://scholarworks.umass.edu/pare/vol9/iss1/4/> > Acesso em 18 de nov. 2018.

TENREIRO-VIEIRA, C. Produção e avaliação de actividades de aprendizagem de ciências para promover o pensamento crítico dos alunos. **Revista Iberoamericana de Educación**, v. 33, n. 6, p. 1-18, 2004. Disponível em: <<https://rieoei.org/historico/deloslectores/708.pdf>> Acesso em 20 de nov. 2018.

TENREIRO-VIEIRA, C; VIEIRA, R.M. **Promover o pensamento crítico dos alunos: Propostas concretas para a sala de aula**. Lisboa: Instituto Piaget, 2000.

VIEIRA, R.M.; TENREIRO-VIEIRA, C. **Estratégias de ensino/aprendizagem**. Lisboa: Instituto Piaget, 2005.