

IV ENJIE

Encontro Nacional de Jovens Investigadores em Educação

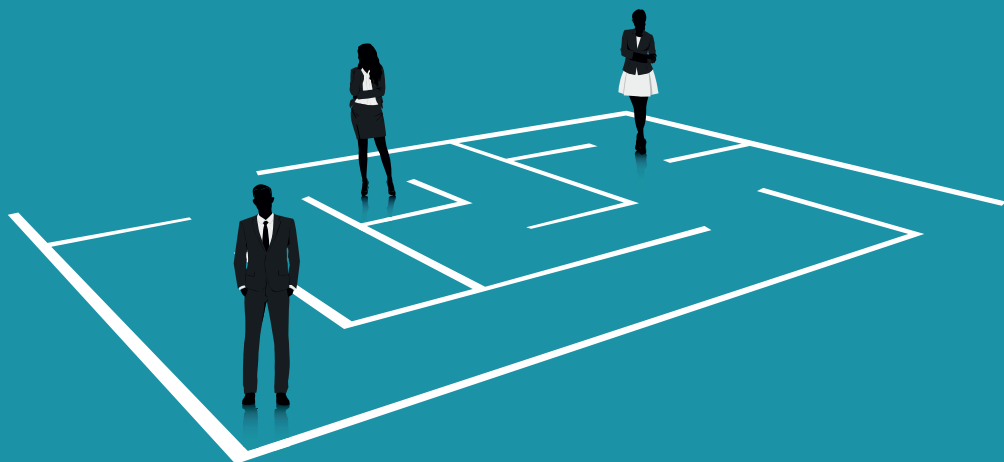
7 e 8 de fevereiro de 2020

Universidade de Évora

Auditório do Colégio do Espírito Santo

LIVRO DE ATAS

Coordenação:
Luís Sebastião | Hugo Rebelo



www.enjie.pt

**IV Encontro Nacional de Jovens
Investigadores em Educação: livro
de atas.**

(Coordenação)

Luís Sebastião

Hugo Rebelo

FICHA TÉCNICA

TÍTULO:

IV Encontro Nacional de Jovens Investigadores em Educação: livro de atas.

COORDENADORES:

Luís Sebastião & Hugo Rebelo

Edição:

© Centro de Investigação em Educação e Psicologia da Universidade de Évora (CIEP | UE), 1.ª Edição, Évora, 2020 www.ciep.uevora.pt

Morada:

Colégio Pedro da Fonseca, Rua da Barba Rala, n.º 1, Parque Industrial e Tecnológico de Évora, 7005-345 Évora, Portugal

Revisão e paginação:

Mafalda Pequeno

Capa e Design gráfico:

©mr-creative.net

Impressão e acabamento

VASP Digital Printing Services – www.vasp.pt

ISBN (versão digital)

978-972-778-167-6

Depósito legal

493045/21

É expressamente proibido reproduzir esta obra, na totalidade ou em parte, sob qualquer forma ou meio, exceto para fins de ensino e investigação. Autorizações especiais podem ser requeridas para ciep@uevora.pt

«Este trabalho é financiado por fundos nacionais através da FCT – Fundação para a Ciência e a Tecnologia, I.P., no âmbito do projeto UIDB/04312/2020»

LITERACIA CIENTÍFICA DE ALUNOS PORTUGUESES NO FINAL DO TERCEIRO CICLO DO ENSINO BÁSICO: DESENVOLVIMENTO E VALIDAÇÃO DE UM INSTRUMENTO DE AVALIAÇÃO

Marcelo Alves Coppi | Universidade de Évora | marcelo.coppi@hotmail.com

Isabel Fialho | Universidade de Évora | ifialho@uevora.pt

Marília Cid | Universidade de Évora | mcid@uevora.pt

Resumo

O termo literacia científica surgiu na década de 1950, passando a ser estudado de forma mais concreta a partir de 1980. Ainda que seja um termo polissêmico, neste trabalho é definido como a compreensão do empreendimento científico e a utilização consciente dos conhecimentos científicos e tecnológicos para a resolução de problemas e explicação de fenómenos naturais do cotidiano, assim como para a participação ativa em debates de assuntos científicos que envolvem a sociedade, permitindo ao indivíduo atuar como cidadão. Esta definição baseia-se em três dimensões: a natureza da ciência, o conteúdo da ciência e o impacto da ciência e da tecnologia na sociedade. Diversos instrumentos de avaliação foram elaborados a partir deste estudo. No entanto, poucos avaliam estas três dimensões em conjunto. Além disso, raros são aqueles desenvolvidos para o 3.º ciclo do ensino básico, que representa uma etapa fundamental para a literacia científica dos alunos do ensino básico em Portugal. Assim sendo, o objetivo deste trabalho é desenvolver e validar um instrumento de avaliação da literacia científica, nas suas três dimensões, dos alunos que concluem este ciclo de ensino. A fim de que este seja aplicado no contexto português, os itens foram elaborados com base nas Orientações Curriculares e nas Aprendizagens Essenciais da área das Ciências Físicas e Naturais do 3.º ciclo do ensino básico. Levando em consideração a quantidade de competências presentes nestes documentos e, a fim de garantir que todos estes itens possam ser avaliados no tempo disponível de uma aula de 60 minutos, optou-se pelo formato de item verdadeiro-falso, o qual é capaz de avaliar diferentes processos cognitivos e permitir a cobertura de uma maior quantidade de competências. O instrumento será validado por meio da coleta de evidências de validade baseadas no conteúdo, realizada mediante a análise dos itens por um painel de especialistas da área das Ciências Físicas e Naturais, os quais forneceram um parecer através do preenchimento de um formulário de revisão da qualidade dos itens e pela recolha de evidências de validade baseadas no construto, a qual será feita após os testes piloto e final, por meio dos cálculos dos índices de dificuldade dos itens e pelo cálculo da consistência interna do teste. Pretende-se que o instrumento possa fornecer indicadores capazes de auxiliar a monitorização do progresso da educação científica do 3.º ciclo a níveis regional e nacional.

Palavras-chave: literacia científica, instrumento de avaliação, terceiro ciclo do ensino básico.

Introdução

A literacia científica é atualmente o principal objetivo do ensino de ciências e caracteriza-se pela capacidade de utilizar os conhecimentos científicos para atuar como cidadão e para lidar com os problemas do cotidiano. A sua emergência no período pós Segunda Guerra Mundial, quando muitos cientistas e governantes testemunharam a abrangência e a capacidade do acervo científico e militar utilizado (Miller, 1983). Desde então, muitos esforços vêm sendo realizados a fim de esclarecer a sua definição.

Em 1983, Miller propôs uma definição multidimensional para o termo. Segundo o autor, a literacia científica é definida por três dimensões: a compreensão do empreendimento da ciência, o conhecimento dos principais conteúdos da ciência e a consciência do impacto da ciência e da tecnologia na sociedade (Miller, 1983).

Essa definição suscitou avanços importantes para a sua avaliação (Laugksch, 2000). Laugksch e Spargo (1996a) alegam que, após essa publicação, muitos estudos têm sido conduzidos nos Estados Unidos, Comunidade Europeia, China, Japão, Suécia, Canadá, Austrália e Coreia do Sul com o objetivo de avaliar o conhecimento dos estudantes em relação às três dimensões de literacia científica.

Entretanto, os autores ressaltam que a maior parte dos instrumentos apresenta validade inespecífica e que avaliam as dimensões separadamente. Gormally, Brickman e Lutz (2012) reforçam esse fato, referindo que, embora tenham sido desenvolvidos diversos instrumentos, a maioria deles avalia as três dimensões de forma individual. Além disso, Fives, Huebner, Birnbaum e Nicolich (2014) apontam que nenhum dos instrumentos, até então desenvolvidos, contempla as necessidades da literacia científica dos estudantes do 3.º ciclo do ensino básico.

A avaliação do nível da literacia científica neste ciclo de ensino justifica-se por três principais motivos: o facto de que este é um ciclo de transição, marcado pela mudança do ensino de ciências realizado normalmente por um único professor, para um ensino de duas disciplinas, Ciências Naturais e Ciências Físico-Química, cada uma com as suas particularidades e professores específicos; a tendência da diminuição do interesse dos estudantes para as disciplinas científicas na transição do 1.º ciclo para os 2.º e 3.º ciclos do ensino básico (Fives et al., 2014); e o facto de as disciplinas científicas desse ciclo representarem, muitas vezes, o último contacto formal dos estudantes com disciplinas da área das Ciências Físicas e Naturais.

No contexto de Portugal, Saraiva (2017) aponta que a educação científica recebeu maior atenção após a década de 1990, quando o ensino de ciências passou a ser alvo de debates relacionados com os resultados nas avaliações internacionais PISA e TIMSS, os quais se mostraram crescentes e positivos. No entanto, Galvão, Freire, Faria, Baptista e Reis (2017) salientam que os resultados continuam aquém das expectativas. Os autores alegam que, embora a implementação do currículo para o 3.º ciclo do ensino básico tenha seguido

as recomendações internacionais para melhoria da qualidade do ensino de ciências, “continua a persistir uma discrepância entre aquilo que seria esperado em termos dos desempenhos dos alunos em exames internacionais [...] e aquilo que de facto os alunos alcançam” (Galvão et al., 2017, p. 15).

Nesse sentido, e levando em consideração a carência de instrumentos que avaliam, em conjunto, as três dimensões de literacia científica propostas por Miller (1983), a escassez de avaliações direcionadas para o 3.º ciclo do ensino básico e a necessidade de aprimorar a qualidade do ensino de ciências em Portugal, o objetivo deste trabalho é apresentar os resultados das primeiras etapas do processo de validação do instrumento, mais especificamente da recolha de evidências de validade baseadas no conteúdo. Pretende-se, também, apresentar uma breve explicação das etapas futuras. Este trabalho insere-se em uma proposta de investigação no doutoramento em curso, com o objetivo desenvolver e validar um instrumento de avaliação do nível de literacia científica dos estudantes no final do 3.º ciclo do ensino básico português.

A pesquisa

O projeto de investigação constitui-se com um estudo extensivo. O seu caráter inovador resulta da elaboração de um instrumento de avaliação de sala de aula, ou seja, um instrumento que, diferentemente do testes em larga escala, tem por finalidade “acompanhar os processos de aprendizagem escolar, compreender como eles estão se concretizando, oferecer informações relevantes para o próprio desenvolvimento do ensino na sala de aula em seu dia-a-dia, para o planeamento e replaneamento contínuo da atividade de professores e alunos” (Gatti, 2003, p. 99), referenciado a critério e que será validado mediante a recolha de três tipos de evidências de validade, conforme proposto por Popham (2017), as quais utilizarão a triangulação de técnicas de coleta de dados a fim de que os resultados sejam considerados válidos para os usos propostos.

A elaboração de um instrumento de avaliação exige a utilização de procedimentos que garantam indicadores confiáveis e que evidenciem uma alta qualidade técnica (Popham, 2017). Para o autor, a característica mais importante para a qualidade de um instrumento de avaliação é a validade. Russel e Airasian (2014) respaldam a ideia, argumentando que “a característica mais importante de uma boa avaliação é a sua habilidade de ajudar o professor a tomar as decisões adequadas. Essa característica é chamada de validade” (p. 26).

De acordo com a literatura mais recente, a validade corresponde ao grau em que as evidências e a teoria apoiam as interpretações dos resultados dos instrumentos para os usos propostos (AERA, APA, & NCME, 2014; Popham, 2017). Nessa perspectiva, as evidências de validade estão divididas em cinco categorias: evidências baseadas no conteúdo, evidências baseadas nos processos de respostas, evidências baseadas na estrutura interna, evidências baseadas em outras variáveis e evidências baseadas nas consequências das avaliações (AERA et al., 2014).

Por se tratar do processo de construção de um instrumento e levando em consideração que este não se destina a ser aplicado em dois momentos distintos à mesma população, as técnicas de recolha de evidências de validade baseadas em outras variáveis e nas consequências da avaliação não serão abordadas neste estudo. A seguir, são descritas as técnicas que foram e aquelas que serão utilizadas na recolha de evidências de validade para o desenvolvimento do instrumento de avaliação, as quais se fundamentam em Pasquali (2017).

Evidências de validade baseadas no processo de resposta

Esta técnica diz respeito à definição dos processos psicológicos que se pretende avaliar (Pasquali, 2017). Nesse sentido, a recolha de evidências de validade baseadas no processo de resposta foi realizada por meio da definição dos domínios cognitivos dos itens do instrumento. Pasquali (2017) afirma ser importante se respaldar em alguma taxonomia clássica de objetivos educacionais e, com base nela, estabelecer os domínios cognitivos que se pretende medir.

A taxonomia utilizada neste estudo foi a de Anderson et al. (2001), os quais realizaram uma revisão e atualização da Taxonomia de Bloom. Considerando que este instrumento almeja avaliar níveis cognitivos superiores, os domínios cognitivos determinados foram: entender, analisar e avaliar. De acordo com Russel e Airasian (2014), esses são considerados domínios cognitivos de nível superior, já que envolvem mais do que a memorização de informações.

Atendendo à definição de literacia científica adotada neste trabalho – a compreensão do empreendimento científico e a utilização consciente dos conhecimentos científicos e tecnológicos para a resolução de problemas e explicação de fenómenos naturais do cotidiano, assim como para a participação ativa em debates de assuntos científicos que envolvem a sociedade, permitindo ao indivíduo atuar como cidadão – os domínios cognitivos supracitados têm como objetivo avaliar as capacidades de compreensão, análise e avaliação de problemas e situações cotidianas que compreendem o conhecimento das Ciências Naturais e Físico-Químicas.

Evidências de validade baseadas no conteúdo

Esse tipo de evidências de validade envolve a determinação e seriação do conteúdo, a elaboração, a redação e o formato dos itens e o processo de administração e de pontuação dos resultados (Pasquali, 2017). Para a sua recolha foram dados os seguintes passos: 1) a definição do universo do conteúdo; 2) a definição da representatividade do conteúdo; 3) a elaboração da tabela de especificação; 4) a construção do instrumento; 5) a análise teórica dos itens; 6) a aplicação do teste piloto; e 7) a análise empírica dos itens.

Definição do universo do conteúdo.

Estabelecer o universo do conteúdo requer a delimitação do conteúdo em unidades e subunidades de ensino (Pasquali, 2017). Considerando a realidade portuguesa, os

conteúdos foram selecionados a partir dos principais documentos curriculares da área das Ciências Físicas e Naturais: Orientações Curriculares das Ciências Físicas e Naturais (OC), e Aprendizagens Essenciais de Ciências Naturais e Físico-Química (AE).

Considerando o facto de que as OC e as AE correspondem a uma vasta quantidade de conteúdos e competências, optou-se por utilizar as diretrizes de literacia científica estipuladas pelos *Benchmarks for Science Literacy* (BFSL) (AAAS, 1993) como referência para a definição do universo do conteúdo. O BFSL é um documento baseado nas três dimensões de literacia científica de Miller (1983) (Laugksch & Spargo, 1996) e estabelece recomendações sobre os conteúdos e as competências mínimas de ciências, matemática e tecnologia que devem ser desenvolvidos pelos alunos americanos ao fim de cada ciclo de ensino (AAAS, 1993).

A fim de identificar os conteúdos e as competências das OC e das AE a serem selecionados para integrar o instrumento, realizou-se análise documental. Esta foi composta por três etapas: elencar todos os conteúdos e as competências relacionadas com a área das Ciências Físicas e Naturais de cada documento; comparar os documentos entre si e verificar os conteúdos e as competências correspondentes; e eliminar as competências similares.

Foram selecionadas 60 competências, 10 das OC e 50 das AE. Dessas, 17 são referentes aos componentes curriculares do 7.º ano, 17 do 8.º ano e 16 do 9.º ano. É necessário ressaltar que duas competências das OC, e uma competência da AE do 8.º foram utilizadas em mais de um item. Isso ocorreu pelo facto de que estas são mais amplas do que as demais e permitiram encontrar mais de um correspondente nos BFSL.

Definição da representatividade do conteúdo.

A definição da representatividade do conteúdo foi estabelecida mediante a comparação dos documentos curriculares da área de Ciências Físicas e Naturais com os BFSL. O número de itens por conteúdo foi determinado pela quantidade de competências correspondentes entre as OC e as AE e os conteúdos dos BFSL.

Como o instrumento será elaborado com base nas três dimensões de literacia científica estabelecidas por Miller (1983) – a natureza da ciência (NC), o conteúdo da ciência (CC) e o impacto da ciência e da tecnologia na sociedade (ICTS) –, as 60 competências foram convertidas em 64 itens da seguinte forma: 6 itens com competências dos conteúdos da dimensão da NC, 7 itens com competências dos conteúdos do ICTS e 51 itens com competências dos conteúdos do CC. Como a dimensão do CC envolve mais conteúdos específicos das ciências do que as outras duas dimensões, a sua representatividade foi detalhada conforme se ilustra na Tabela 1.

Tabela 1 – Representatividade do conteúdo da dimensão do CC.

Conteúdo	Qtd. de itens	Conteúdo	Qtd. de itens	Conteúdo	Qtd. de itens
Alterações ambientais	3	Átomos e elementos químicos	3	Força, gravidade e movimento	2
Universo e sistema solar	3	Substâncias e Misturas	2	Ecologia	4
Geodinâmica interna	4	Reações químicas	2	Evolução	1
Geodinâmica externa	4	Energia	4	Células	2
Temperatura e mudanças de estado físico da matéria	3	Ondas	3	Fisiologia	11

Fonte: Elaboração própria.

Elaboração da tabela de especificação.

A tabela de especificação (Tabela 2) foi elaborada mediante o paralelismo entre as competências das OC e das AE com os domínios cognitivos da Taxonomia de Bloom revisada.

Tabela 2 – Tabela de especificação após a definição dos domínios cognitivos.

Dimensão do conteúdo	Domínio cognitivo			Total
	Entender	Analisar	Avaliar	
NC	4	1	1	6
ICTS	2	1	4	7
CC	15	15	21	51
Total	21	17	26	64

Fonte: Elaboração própria.

Construção da versão 1 do instrumento.

O processo de construção do instrumento inicial envolveu decisões sobre: o formato dos itens, as diretrizes técnicas de construção e a configuração das frases dos itens (Pasquali, 2017).

O formato dos itens escolhido foi o verdadeiro-falso. Itens de verdadeiro-falso são “afirmações declarativas simples absolutamente verdadeiras ou falsas” (Haladyna, 2018, p. 4). Pelo facto de não apresentar opções alternativas, itens neste formato exigem que o aluno crie mentalmente um contraexemplo da afirmação e opte pelo verdadeiro ou falso (Haladyna, 2004). O autor alega que testes com itens de verdadeiro-falso são bem aceites para avaliações de sala de aula.

Contudo, esse formato de item recebe diversas críticas de especialistas da área de avaliação e elaboração de itens. De acordo com Ebel e Frisbie (1991), a maior parte dessas

críticas esta associada com as chances de um indivíduo acertar um item por simples adivinhação, já que o formato verdadeiro-falso possui apenas duas alternativas de respostas. Desta forma, optou-se por utilizar uma versão adaptada desse formato, incluindo uma terceira opção de resposta, o “não sei”. Será solicitado aos alunos que selecionem essa opção quando, de facto, não souberem a resposta do item.

Para Ebel e Frisbie (1991) inserir esta opção em questionários de pesquisas diminui significativamente o número de respostas obtidas por meio da adivinhação. Além disso, a análise das respostas dos alunos utilizando a opção “não sei” pode resultar em dados importantes sobre o conhecimento dos alunos acerca do conteúdo e da competência que está sendo avaliada pelo item.

As diretrizes técnicas utilizadas para a elaboração dos itens foram a de Haladyna (2004, 2018) e de Ebel e Frisbie (1991). Haladyna (2004, 2018) apresenta instruções detalhadas para a elaboração de itens objetivos, relacionadas com as diretrizes de conteúdo, as orientações de estilo e o formato e a elaboração do enunciado. Já Ebel e Frisbie (1991) divulgam orientações específicas para itens no formato verdadeiro-falso, direcionadas para a diminuição da trivialidade, ambiguidade, obviedade, clareza e precisão dos itens.

Com relação à configuração dos itens, optou-se pelos itens interpretativos com o intuito de avaliar os domínios cognitivos definidos anteriormente. Estes fornecem informações em forma de textos, gráficos, quadros, imagens ou tabelas, quais servem de base para que os alunos possam respondê-los (Russel & Airasian, 2014).

A configuração dos itens deste instrumento apresentou as informações em forma de texto. Desse modo, todos os itens são compostos por uma ou mais frases destacadas em negrito que descrevem uma situação, um caso ou um evento, seguidas de uma afirmação, sem destaque, que deve ser analisada e respondida como verdadeira ou falsa, conforme o exemplo abaixo.

Numa experiência, o professor amarrou um balão bem esticado numa das extremidades abertas de uma lata e colocou uma pequena bola de esferovite sobre o balão. Em seguida, tocou um tambor perto da lata. É correto afirmar que a vibração do toque do tambor originou uma onda que se propagou pelo ar até à lata, fazendo com que o balão vibrasse, movimentando a bola de esferovite.

Análise teórica dos itens.

A análise teórica tem como objetivo a verificação da “representatividade dos itens em relação às áreas de conteúdo e à relevância dos objetivos a medir” (Raymundo, 2009, p. 87). Considerando a inexistência de um teste específico para este tipo de análise, realizou-se uma abordagem qualitativa, seguida de uma abordagem quantitativa, conforme proposto por Alexandre e Coluci (2011).

Para a abordagem qualitativa, foi constituído um painel de especialistas da área das Ciências da Natureza, responsável por avaliar a correspondência entre o item e os

documentos curriculares, a veracidade das afirmações, a presença de ambiguidades lógicas e científicas, a adequação da linguagem e do vocabulário para o público-alvo e a relevância do item para a literacia científica.

A avaliação foi realizada mediante o preenchimento de um formulário de revisão composto por cinco perguntas: (1) Há correspondência entre o item e as orientações curriculares/aprendizagens essenciais/benchmark? (2) A afirmação é claramente verdadeira ou claramente falsa? (3) O item apresenta ambiguidades lógicas e/ou científicas? (4) A linguagem, o vocabulário ou a estrutura da frase podem ser de difícil interpretação/entendimento para estudantes do 3.º ciclo? (5) Qual a relevância deste item para a literacia científica que foi definida para este instrumento?

Já a abordagem quantitativa compreendeu a quantificação do grau de concordância entre as respostas dos especialistas por meio do Índice da Validade de Conteúdo (IVC). O critério de exclusão dos itens levou em consideração aqueles que obtiveram valores menores do que 0.8 no IVC. Como resultado, foram selecionados 35 itens, de uma amostra de 64, para compor o instrumento inicial.

Aplicação do teste piloto.

Com o intuito de construir uma amostra significativa de respondentes ($n \geq 350$) para o teste piloto, este será aplicado aos estudantes do 9.º ano das escolas públicas da região sul de Portugal continental no final dos anos letivos de 2019/2020, 2020/2021 e 2021/2022.

Análise empírica dos itens.

A análise empírica será realizada após a aplicação do teste piloto e consistirá na análise dos índices de dificuldade e de discriminação dos itens. De acordo com Pasquali (2017), essas são características dos instrumentos de avaliação tradicionalmente analisadas pela psicometria. Ambas serão concebidas e tratadas por meio do modelo logístico de dois parâmetros da Teoria de Resposta ao Item (TRI).

A escolha pela utilização desse recurso justifica-se pelo fato de que a TRI tem interesse específico em cada um dos itens e pode ser utilizada para saber quais são os fatores que afetam a probabilidade de cada item individualmente ser respondido correta ou incorretamente (Baker, 2001; Pasquali, 2017).

No modelo logístico de dois parâmetros da TRI, o índice de dificuldade do item representa o nível de habilidade necessário para que a probabilidade de um indivíduo responder corretamente ao item seja de 50% (Baker, 2001). Já o índice de discriminação descreve quão bem um item pode diferenciar os respondentes que dominam dos que não dominam a habilidade requerida pelo item (Sartes & Souza-Formigoni, 2013).

Evidências de validade baseadas na estrutura interna

Este tipo de evidências corresponde ao grau de relação entre os itens do instrumento e os construtos estabelecidos previamente. Vianna (2014, p. 35) argumenta que a recolha desse tipo de evidência “é de grande relevância na área educacional, tendo em vista o fato de que a avaliação se vale, frequentemente, de construtos, que, após sua definição operacional, são medidos por intermédio de testes”.

Pasquali (2017) afirma que esse procedimento permite explicar as diferenças individuais nas pontuações de um instrumento de medida e que não é expresso em termos de apenas um coeficiente quantitativo, mas resulta da convergência de diferentes métodos. Nesse sentido, serão utilizadas três técnicas distintas: a análise fatorial, a consistência interna e a curva de informação da TRI.

A análise fatorial consiste em técnicas de análises multivariadas e de matrizes e é capaz de fornecer informações do quanto um instrumento de avaliação está representando um construto ou uma habilidade (Pasquali, 2017). Já a consistência interna, a qual será calculada pelo coeficiente de Kuder-Richardson (KR20), pode ser caracterizada como a análise da homogeneidade do instrumento e auxilia “na definição do construto, especialmente ao indicar se o teste mede um único traço ou se, ao contrário, mede diversos traços” (Vianna, 2014, p. 40). Por fim, a função de informação da TRI mede a qualidade do item para o construto, fornecendo dados sobre o quão preciso cada nível de habilidade está sendo estimado por um instrumento de medida (Baker, 2001). Após a realização das análises estatísticas, o instrumento final será composto pelos itens que apresentarem valores satisfatórios, de acordo com a literatura, em cada uma das técnicas.

Aplicação do teste final

A fim de obter uma amostra significativa a nível nacional e de patentear o instrumento, o teste final será aplicado aos estudantes do 9.º ano das escolas públicas das cinco regiões de Portugal continental de acordo com a Nomenclatura das Unidades Territoriais para fins Estatísticos (NUTS) III, no final do ano letivo 2022/2023.

Considerações finais

Este estudo compõe um projeto de tese de doutoramento em curso e se propôs a apresentar os principais resultados, encontrados até então, relacionados com a recolha de evidências de validade baseadas nos processos de resposta e nas cinco primeiras etapas da recolha de evidências de validade baseadas no conteúdo. Assim como, identificar as etapas a seguir: aplicação dos testes piloto e final, análise empírica dos itens e a recolha de evidências de validade baseadas na estrutura interna do instrumento.

Como resultado das recolhas de evidências de validade, resultou um instrumento constituído por 35 itens no formato “verdadeiro-falso-não sei”. Os itens possuem uma configuração específica, com frases em destaque, que avaliam os domínios cognitivos de entender, analisar e avaliar fenómenos e situações cotidianas, vinculados aos conteúdos e

competências presentes nos principais documentos curriculares de Ciências Físicas e Naturais portuguesas, no PA e nos BFSL.

Referências

- AAAS. (1993). *Project 2061: benchmarks for science literacy*. Washington, DC: Oxford University Press.
- AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Alexandre, N. M. C., & Coluci, M. Z. O. (2011). Validade de conteúdo nos processos de construção e adaptação de instrumentos de medidas. *Ciência & Saúde Coletiva*, 16(7), 3061–3068.
- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., ... Wittrock, M. C. (2001). *A taxonomy for learning, teaching and assessing: a revision of Bloom's taxonomy of educational objectives*. New York: Addison Wesley Longman.
- Baker, F. B. (2001). *The basics of item response theory*. Washington, DC: ERIC.
- Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Englewood Cliffs: Prentice Hall International, INC.
- Fives, H., Huebner, W., Birnbaum, A. S., & Nicolich, M. (2014). Developing a measure of scientific literacy for middle school students. *Science Education*, 98(4), 549–580.
- Galvão, C., Freire, S., Faria, C., & Baptista, M. (2017). *Avaliação do currículo das ciências físicas e naturais: percursos e interpretações*. Lisboa: Instituto de Educação da Universidade de Lisboa.
- Gatti, B. A. (2003). O professor e a avaliação em sala de aula. *Estudos Em Avaliação Educacional*, 8(27), 97. <https://doi.org/10.18222/eae02720032179>
- Gormally, C., Brickman, P., & Lutz, M. (2012). Developing a test of scientific literacy skills (TOSLS): measuring undergraduates' evaluation of scientific information and arguments. *CBE Life Sciences Education*, 11(4), 364–377.
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3rd ed.). London: Lawrence Erlbaum Associates.
- Haladyna, T. M. (2018). Developing test items for course examinations. *IDEA*, 70, 1–16.
- Laugksch, R. C. (2000). Scientific literacy: a conceptual overview. *Science Education*, 84(1), 71–94.
- Laugksch, R. C., & Spargo, P. E. (1996). 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*, 5(4), 331–359. <https://doi.org/10.1088/0963-6625/5/4/003>
- Miller, J. D. (1983). Scientific literacy: a conceptual and empirical review. *Daedalus*, 112(2), 29–48.
- Pasquali, L. (2017). *Psicometria - Teoria dos testes na psicologia e na educação* (10th ed.). Petrópolis: Vozes.
- Popham, W. J. (2017). Classroom assessment: what teachers need to know. In *Journal of Educational Measurement* (8th ed., Vol. 39). <https://doi.org/10.1111/j.1745-3984.2002.tb01136.x>
- Raymundo, V. P. (2009). Construção e validação de instrumentos um desafio para a psicolinguística. *Letras de Hoje*, 44(3), 86–93.
- Russel, M. K., & Airasian, P. W. (2014). *Avaliação em sala de aula: conceitos e aplicações* (7th ed.). Porto Alegre: AMGH.
- Saraiva, L. (2017). A aprendizagem das ciências em Portugal: uma leitura a partir dos resultados do TIMSS e do PISA. *Mediações*, 5(2), 4–18.
- Sartes, L. M. A., & Souza-Formigoni, M. L. O. (2013). Avanços na psicometria: da teoria clássica dos testes à teoria de resposta ao item. *Psicologia: Reflexão e Crítica*, 26(2), 241–250. <https://doi.org/10.1590/S0102-79722013000200004>
- Vianna, H. M. (2014). Validade de construto em testes educacionais. *Estudos Em Avaliação Educacional*, 25(60), 136–152.