



V ENJIE

Encontro Nacional de Jovens
Investigadores em Educação

Livro de Atas

Ficha Técnica

Título: Livro de Atas do V ENJIE: Investigação em Educação e Responsabilidade Social - vozes dos jovens investigadores

Organizadoras: Maria Helena Araújo e Sá & Lina Morgado

Design e paginação: Joana Pereira

Editora:

UA Editora

Universidade de Aveiro

Serviços de Documentação, Informação Documental e Museologia

1ª edição - fevereiro 2022

ISBN: 978-972-789-731-5

DOI: <https://doi.org/10.48528/tr7a-j538>

Os textos desta publicação são decorrentes dos resumos apresentados pelos autores durante V Encontro de Jovens Investigadores em Educação, publicados no Livro de Resumos disponível em https://enjie.pt/2021/wp-content/uploads/2021/04/V-ENJIE_Layout-livro-resumos_VF_compressed.pdf, e foram revistos e aprovados por pares. As versões finais são da inteira responsabilidade dos seus autores.

Financiado por Fundos Nacionais através da FCT – Fundação para a Ciência e a Tecnologia, I.P., no âmbito dos projetos UIDB/00194/2020 e UIDP/00194/2020, UIDB/04372/2020 e UIDP/04372/2020

RECOLHA DE EVIDÊNCIAS DE VALIDADE DE CONTEÚDO PARA A CONSTRUÇÃO DE UM INSTRUMENTO PILOTO DE AVALIAÇÃO DA LITERACIA CIENTÍFICA

Marcelo Alves Coppi

Centro de Investigação em Educação e Psicologia da Universidade de Évora
mcoppi@uevora.pt

Isabel Fialho

Centro de Investigação em Educação e Psicologia da Universidade de Évora
ifialho@uevora.pt

Marília Cid

Centro de Investigação em Educação e Psicologia da Universidade de Évora
mcid@uevora.pt

Resumo

A literacia científica é o principal objetivo do ensino de ciências. Entretanto, raros são os instrumentos que apresentam evidências de validade e sejam capazes de avaliar, em conjunto, as dimensões da natureza da ciência, do conhecimento do conteúdo da ciência e do impacto da ciência e da tecnologia na sociedade, principalmente no 3.º ciclo do ensino básico. A validade é principal indicador de qualidade técnica de um instrumento de avaliação e reflete o grau em que cada evidência e a teoria suportam a precisão das interpretações dos resultados dos testes para os usos propostos. Este estudo tem como objetivo a recolha de evidências de validade baseadas no conteúdo, utilizada para a elaboração de um instrumento piloto de avaliação da literacia científica de alunos no final do 3.º ciclo do Ensino Básico. A recolha das evidências de validade foi realizada em seis etapas: 1) definição dos domínios cognitivos, os quais foram estabelecidos a partir da taxonomia atualizada de

Bloom, selecionando três deles: compreender, analisar e avaliar problemas e situações cotidianas que envolvem o conhecimento nas ciências físicas e naturais; 2) definição do universo do conteúdo, que foi determinado com base na análise dos documentos curriculares da área das ciências físicas e naturais vigentes em Portugal, em consonância com os *benchmarks for science literacy*; 3) definição da representatividade do conteúdo, estabelecida por meio da análise de similaridade dos conteúdos entre os documentos supracitados; 4) elaboração da tabela de especificação, etapa em que se atribuiu a correspondência entre os conteúdos selecionados e os domínios cognitivos estipulados; 5) construção do instrumento, a qual compreendeu as decisões sobre o formato, a configuração e as diretrizes técnicas de construção de itens; e 6) análise teórica dos itens, realizada por um painel de especialistas, responsável por analisar as diversas características dos itens e, conseqüentemente, a determinação do grau de concordância entre as análises, recorrendo ao índice de validade de conteúdo. Como resultado, foram selecionados para compor o instrumento piloto 35 itens, no formato “verdadeiro-falso-não sei”, que avaliam os domínios cognitivos de compreender, analisar e avaliar competências de literacia científica presentes nos principais documentos curriculares portugueses vigentes das Ciências Físicas e Naturais.

Palavras-chave: literacia científica, avaliação, instrumentos, validade de conteúdo, evidências de validade de conteúdo

Introdução

A literacia científica é atualmente o principal objetivo do ensino de ciências. A expressão literacia científica emergiu na década de 1950, após a Segunda Guerra Mundial e início da corrida espacial. Contudo, foi apenas após a década de 1980, a seguir a publicação do artigo de Miller (1983), na qual o autor propôs uma definição multidimensional para a literacia científica e sugeriu estratégias para avaliá-la, que maiores esforços foram empreendidos a fim de definir o conceito e também de avaliá-lo. Desde então, diversos instrumentos de avaliação da literacia científica vêm sendo desenvolvidos e utilizados em pesquisas nacionais e internacionais (Laugksch & Spargo, 1996b).

De acordo com os autores, a maioria dos instrumentos disponíveis na literatura carecem de mecanismos de validade, ou apresentam validades inespecíficas, e analisam individualmente as dimensões propostas por Miller (1983), não sendo capazes, portanto, de avaliar a literacia científica em sua totalidade. Gormally, Brickman e Lutz (2012)

corroboram o exposto, alegando que muitos destes instrumentos avaliam dimensões ou disciplinas específicas e apenas alguns são capazes de avaliar a literacia científica de uma forma mais ampla. Dentre estes, os mais utilizados são o *Test of Basic Scientific Literacy* (TBSL) (Laugksch & Spargo, 1996a), o *Test of Scientific Literacy Skills* (TOSLS) (Gormally et al., 2012) e o o *Scientific Literacy Assessment* (SLA) (Fives et al., 2014).

O TBSL foi desenvolvido para avaliar ideias e atitudes importantes dos alunos no final do Ensino Secundário sobre a ciência. É composto por 110 itens no formato “verdadeiro-falso-não sei”, baseados em cinco seções do programa *Science for All Americans* desenvolvido pelo Projeto 2061 da *American Association for Advancement of Science*. O TOSLS avalia a qualidade da informação e dos argumentos científicos expostos na mídia por cientistas, através de evidências e dados. Este instrumento, composto por 28 itens de escolha múltipla, foi desenvolvido para os alunos dos cursos de licenciatura de biologia e foi baseado em documentos de políticas educacionais, como a *American Association for Advancement of Science*, a *National Academy of Science*, e a *Organisation for Economic Co-operation and Development*. Por fim, o SLA contém 51 itens de escolha múltipla e escala likert e avalia componentes cognitivos e afetivos da literacia científica e foi desenvolvido para os alunos 3.º ciclo do Ensino Básico, com base na *American Association for Advancement of Science*, na *National Science Teachers Association*, na *Organisation for Economic Co-operation and Development* e na *National Research Council*.

Fives, Huebner, Birnbaum e Nicolich (2014) afirmam que além de avaliarem dimensões específicas da literacia científica, poucos instrumentos contemplam as necessidades dos estudantes do 3.º ciclo do Ensino Básico. Em Portugal não há registos de instrumentos de avaliação da literacia científica para este ciclo de ensino, o qual é marcado pela transição de uma única disciplina científica, as Ciências Naturais, ministrada, normalmente, por um único professor, para duas disciplinas científicas, Ciências Naturais e Ciências Físico-Químicas, cada uma com as suas particularidades e professores específicos. Além disso, o 3.º ciclo do Ensino Básico representa, muitas vezes, o último contato dos alunos com disciplinas científicas. Embora o SLA seja destinado para este ciclo de ensino, este é composto por duas seções distintas, as quais avaliam dimensões distintas daquelas propostas por Miller (1983).

Considerando a escassez de instrumentos de avaliação que abranjem as dimensões de literacia científica propostas por Miller (1983) e que sejam direcionadas aos alunos do 3.º ciclo do Ensino Básico,

desenvolveu-se um projeto de pesquisa no âmbito do doutoramento em curso, cujo objetivo é desenvolver e validar um instrumento de avaliação do nível de literacia científica dos estudantes no final do 3.º ciclo do Ensino Básico português. Essa pesquisa visa oferecer uma ferramenta de avaliação para os professores, os quais poderão utilizar os dados provenientes do instrumento para reformular suas metodologias e estratégias de ensino, e, também, fornecer indicadores que auxiliem a monitorização do progresso da educação científica a níveis regional e nacional.

Assim como referido por Alexandre e Coluci (2011), a elaboração de um instrumento de avaliação deve valer-se de indicadores confiáveis que evidenciem uma alta qualidade técnica. Segundo Popham (2017) a validade é o principal atributo da qualidade de um instrumento de avaliação. Russel e Airasian (2014) corroboram a ideia, afirmando 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).

Atualmente, a validade é assumida como um conceito que reflete “o grau em que cada evidência e a teoria suportam a precisão das interpretações dos resultados dos testes para os usos propostos” (AERA et al., 2014, p. 11). Nesta atual concepção, a validade envolve o acúmulo de evidências relevantes, capazes de assegurar a interpretação e o uso dos resultados dos instrumentos de avaliação (AERA et al., 2014; Messick, 1989; Miller et al., 2009; Popham, 2017).

Nesta perspetiva, existem cinco tipos de evidências de validade: a) evidências baseadas no conteúdo, relacionada com o conteúdo, com a elaboração, a redação e o formato dos itens; b) evidências baseadas nos processos de respostas, as quais referem-se aos domínios cognitivos de cada item; c) evidências baseadas na estrutura interna, que indicam a relação entre os itens e os construtos que estão sendo avaliados; d) evidências baseadas na relação com outras variáveis, que envolvem a relação entre os resultados de um instrumento com resultados provenientes de outros instrumentos de avaliação; e e) evidências baseadas nas consequências dos testes, as quais relacionam-se com a interpretação e o uso dos resultados (AERA et al., 2014).

Considerando que as evidências de validade baseadas no conteúdo são as fontes primárias de evidências de um instrumento de avaliação (Kane, 2013) validation as formulated by Kane is fundamentally a simply-stated two-step enterprise: (1 e que, “no caso de aplicação no âmbito da sala de aula, a validade mais importante é o conteúdo”

(Depresbiteris & Tavares, 2009, p. 72), este estudo tem como objetivo a recolha de evidências de validade baseadas no conteúdo, que será utilizada para a elaboração de um instrumento piloto de avaliação da literacia científica dos alunos no final do 3.º ciclo do Ensino Básico, no âmbito de um doutoramento em curso.

Trata-se de um levantamento de informações relevantes sobre os processos de construção, administração e pontuação dos itens. De acordo com a versão atual dos *Standards for Educational and Psychological Testing*, este tipo de evidência de validade inclui análises “lógicas ou empíricas da adequação com a qual o conteúdo do teste representa o domínio do conteúdo e a relevância do domínio do conteúdo para a interpretação da pontuação do teste proposto” (AERA et al., 2014, p. 14).

A pesquisa

O procedimento de recolha das evidências de validade baseadas no conteúdo se deu a partir de seis etapas, propostas por Pasquali (2009): 1) a definição dos domínios cognitivos; 2) a definição do universo do conteúdo; 3) a definição da representatividade do conteúdo; 4) a elaboração da tabela de especificação; 5) a construção do instrumento; e 6) a análise teórica dos itens.

Definição dos domínios cognitivos

Esta etapa corresponde à determinação dos processos cognitivos ou psicológicos que se pretende avaliar (Pasquali, 2009). Processos cognitivos são estratégias pelas quais o conhecimento é adquirido ou utilizado para a resolução de problemas (Anderson et al., 2001).

Para a definição dos domínios cognitivos, foi adotada a taxonomia atualizada de Bloom (Anderson et al., 2001), a saber: lembrar, compreender, aplicar, analisar, avaliar, e criar. A seleção levou em consideração a definição de literacia científica estipulada para o instrumento em desenvolvimento, a qual condiz com a compreensão do empreendimento científico e a utilização consciente dos conhecimentos científicos e tecnológicos para a resolução de problemas e explicação de fenômenos naturais do cotidiano, assim como para a participação ativa em debates de assuntos científicos que envolvem

a sociedade, permitindo ao indivíduo atuar como cidadão. Neste sentido, os domínios cognitivos adotados para a elaboração dos itens foram: compreender, analisar e avaliar situações problema a cerca dos conhecimentos das Ciências Físicas e Naturais.

Definição do universo do conteúdo

A segunda etapa consistiu no estabelecimento do universo do conteúdo programático, a qual consistiu em delimitá-lo em unidades e subunidades (Pasquali, 2009). Considerando o contexto português e o ciclo de ensino para o qual o instrumento está sendo elaborado, os conteúdos foram selecionados das Orientações Curriculares de Ciências Físicas e Naturais (OC), das Aprendizagens Essenciais de Ciências Naturais e Físico-Química (AE) do 3.º Ciclo do Ensino Básico e do Perfil dos Alunos à Saída da Escolaridade Obrigatória (PA).

As OC definem as competências que devem ser desenvolvidas pela área das Ciências Físicas e Naturais ao longo do 3.º ciclo do Ensino Básico (Galvão et al., 2001). Este documento estabelece a literacia científica como o objetivo do ensino das ciências neste ciclo de ensino e a qualifica como fundamental e indispensável para o exercício da cidadania (Galvão et al., 2001). De acordo com os autores, este documento deve ser assumido como referência para o trabalho em sala de aula.

As AE são um documento de orientação curricular que corresponde a uma base de referência para a aprendizagem dos alunos (DGE, 2018). Trata-se de um conjunto de competências indispensáveis a serem adquiridas e desenvolvidas nas disciplinas da área das Ciências Físicas e Naturais ao longo dos três anos deste ciclo de ensino (CETIC, 2014).

Já o PA representa um conjunto de competências que favorece o desenvolvimento das diversas literacias (Martins et al., 2017). De acordo com os autores, o PA é um documento “de referência para a organização de todo o sistema educativo, contribuindo para a convergência e a articulação das decisões inerentes às várias dimensões do desenvolvimento curricular” (Martins et al., 2017, p. 8).

Tendo em vista que as OC, as AE e o PA correspondem a uma vasta quantidade de competências, optou-se por utilizar como referência as diretrizes de literacia científica estipuladas pelos *Benchmarks for Science Literacy* (BFSL) (AAAS, 1993). Os BFSL estabelecem recomendações do que os alunos americanos deveriam saber sobre ciências, matemática

e tecnologia no fim de cada ciclo de ensino (AAAS, 1993). A escolha dos BFSL justifica-se pelo facto de que este documento se baseou no programa *Science For All Americans*, o qual se fundamentou na pesquisa de Miller (1983), na qual o autor estabeleceu as três dimensões de literacia científica utilizadas para a construção deste instrumento em desenvolvimento, a saber: natureza da ciência (NC), conteúdo da ciência (CC) e impacto da ciência e da tecnologia na sociedade (ICTS).

Com o intuito de identificar as competências que deveriam ser seleccionadas para compor o instrumento, realizou-se uma análise documental, que se deu em três etapas. A primeira consistiu em elencar todas as competências relacionadas com a área das Ciências Físicas e Naturais de cada documento. A segunda etapa correspondeu ao processo de comparação entre os documentos, identificando a correspondência entre as competências dos documentos curriculares e os conhecimentos estabelecidos pelos BFSL, seleccionando aqueles que apresentavam alguma correlação. Já a terceira etapa teve por objetivo eliminar competências similares, visto que competências semelhantes poderiam gerar uma equivalência entre os itens, fazendo com que um item pudesse servir como pista para a resposta de outro.

Foram seleccionadas 60 competências, 10 das OC e 50 das AE, sendo 17 do 7.º ano, 17 do 8.º ano e 16 do 9.º ano.

Definição da representatividade do conteúdo

Esta etapa também foi estabelecida mediante à comparação dos documentos curriculares da área de Ciências Físicas e Naturais com os BFSL. Desta forma, as 60 competências originaram 64 itens, uma vez que 3 competências permitiram a correspondência com mais de um conhecimento do BSFL.

A distribuição dos itens se deu da seguinte forma: 6 itens com competências da dimensão da NC, 7 itens com competências da dimensão do ICTS e 51 itens com competências da dimensão do CC. Esta última dimensão compreendeu competências específicas das disciplinas científicas de biologia, geologia, física e química, a saber: alterações ambientais, universo e sistema solar, geodinâmica interna e externa, temperatura e mudanças de estado físico da matéria, átomos e elementos químicos, substâncias e misturas, reações químicas, energia, ondas, força, gravidade e movimento, ecologia, evolução, células e fisiologia.

Elaboração da tabela de especificação

A tabela de especificação foi elaborada por meio da atribuição da correspondência entre as competências das OC, AE e do PA, indicadas na etapa da definição do universo do conteúdo, com os domínios cognitivos da Taxonomia revisada de Bloom (Anderson et al., 2001). Dentre os 64 itens, 26 pertencem ao domínio cognitivo de avaliar, 21 de compreender e 17 de analisar, conforme consta na Tabela 1.

Tabela 1 - Tabela de especificação.

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

Construção do instrumento

A construção do instrumento envolveu a tomada de decisões sobre o formato dos itens, as diretrizes técnicas utilizadas e a configuração das frases dos itens. O formato de item definido foi o verdadeiro-falso. Estes são “afirmações declarativas simples absolutamente verdadeiras ou falsas” (Haladyna, 2018, p. 4) exigem que o aluno crie mentalmente um contraexemplo da afirmação e opte pela opção verdadeira ou falsa (Haladyna, 2004).

De acordo com Frisbie (1973) e Haladyna (2004), muitos autores da área de avaliação aprovam o uso de itens de verdadeiro-falso em avaliações de sala de aula. Dentre as suas principais características, são itens relativamente fáceis e menos demorados para construir e apresentam uma alta eficiência, já que um maior número de itens pode ser respondido em um tempo limitado (Ebel, 1979; Haladyna, 2004; Rush et al., 2016).

No entanto, este formato de item recebe críticas sobre a alta probabilidade de um indivíduo acertar a resposta por adivinhação. Devido ao seu formato, “os alunos têm 50% de chance de responder corretamente um item sem conhecimento do conteúdo” (Rush et al., 2016). Em contrapartida, Frisbie e Becker (1991) afirmam que “autores que consideram os efeitos da adivinhação em testes de

verdadeiro-falso como uma das fraquezas primárias falharam completamente em considerar evidências racionais e empíricas” (p. 74).

Com o intuito de reduzir a adivinhação, optou-se por utilizar uma versão adaptada do formato de itens de VF, o formato de “verdadeiro-falso-não sei”. De acordo com Ebel e Frisbie (1991), ao adicionar a opção “não sei”, que deve ser assinalada caso o aluno não tenha conhecimento do conteúdo ou da competência solicitada pelo item, diminui-se significativamente o número de respostas corretas obtidas apenas por meio de adivinhação.

Uma vez estabelecido o formato, deu-se início ao processo de elaboração dos itens. Foram utilizadas duas principais diretrizes, a de Haladyna (2004) e a de Ebel e Frisbie (1991). Haladyna (2004) apresenta instruções detalhadas sobre as diretrizes de conteúdo, as orientações de estilo e de formato e a elaboração do enunciado para a construção de itens objetivos. Ebel e Frisbie (1991) apresentam diretrizes técnicas específicas para o desenvolvimento de itens no formato verdadeiro-falso. Nestas, os autores indicam que os itens devem: testar apenas uma ideia central; avaliar a compreensão e a explicação dos eventos e não apenas a simples memorização de factos triviais; apresentar uma resposta correta que seja defensável cientificamente; apresentar uma resposta correta que não seja óbvia para qualquer aluno, mas para aqueles que realmente detenham determinado conhecimento; serem escritos de forma clara e concisa; e não apresentar a dupla negação. A fim de que este instrumento não apresentasse, ou minimizasse ao máximo, falhas na elaboração, estas diretrizes foram seguidas de forma rigorosa.

Por fim, para a configuração dos itens, optou-se pela construção de itens interpretativos. De acordo Russel e Airasian (2014) estes itens fornecem informações que servem de base para os alunos os responderem, já que precisam “interpretar, compreender, analisar, aplicar ou sintetizar as informações apresentadas” (p. 146) e, por esse motivo, são itens capazes de acessar domínios cognitivos superiores.

Desta forma, os itens foram estruturados com uma ou mais frases destacadas, as quais descrevem uma situação problema, seguidas de uma afirmação sem destaque, que devem ser analisadas e respondidas como verdadeira ou falsa, conforme o exemplo a seguir: “*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 dos itens foi realizada por uma abordagem qualitativa, seguida de uma quantitativa, conforme sugerido por Alexandre e Coluci (2011). Na etapa qualitativa, foi constituído um painel especialistas que avaliou a correspondência entre os itens 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 sua relevância para a literacia científica. Os membros do painel foram selecionados de acordo com a área e nível de formação e área e nível de atuação, considerando especialistas as áreas das Ciências da Educação e das Ciências Físicas e Naturais.

Desta forma, foram selecionados 10 especialistas, 4 professores do Ensino Básico e/ou Secundário e 6 professores universitários. Cada especialista respondeu a um formulário contendo as OC ou as AE e os BFSL referentes a cada item, o item propriamente dito e a sua respetiva opção de resposta correta. Para cada item, foram elaboradas cinco perguntas: 1) Há correspondência entre o item e as OC/AE/BFSL? 2) A afirmação é claramente verdadeira ou claramente falsa? 3) O item apresenta ambiguidades lógicas e/ou científicas? 4) A linguagem, o vocabulário e/ou a estrutura da frase podem ser de difícil interpretação/entendimento para alunos do 9.º ano? 5) Qual a relevância deste item para a literacia científica que foi definida para este instrumento?

Já na etapa quantitativa, foi utilizado o método do Índice de Validade de Conteúdo (IVC), o qual “mede a proporção ou percentagem de juízes que estão em concordância sobre determinados aspetos do instrumento e de seus itens” (Alexandre & Coluci, 2011, p. 3065). Levando em consideração que as quatro primeiras perguntas estavam associadas às características técnicas dos itens e que apenas a quinta pergunta se referiu à relevância dos itens para a literacia científica, o IVC foi calculado apenas para esta última. As respostas das outras quatro perguntas foram utilizadas para refinar o enunciado dos itens, assim como para diminuir a ambiguidade e possíveis problemas técnicos que estes apresentavam.

A fim de compor o instrumento apenas com os itens considerados essenciais para a avaliação da literacia científica, foram excluídos aqueles com IVC menor que 0.8. Consequentemente, o instrumento piloto foi constituído por 35 itens, 6 da dimensão da NC, 6 da dimensão

do ICTS e 23 da dimensão do CC, conforme a tabela de especificação do instrumento piloto (Tabela 2).

Tabela 2 - Tabela de especificação do instrumento piloto.

Dimensão do conteúdo	Domínio cognitivo			Total
	Entender	Analisar	Avaliar	
NC	4	1	1	6
ICTS	1	1	4	6
CC	5	9	9	23
Total	10	11	14	35

Conclusões

Com o intuito de recolher de evidências de validade baseadas no conteúdo para a elaboração do instrumento piloto de avaliação da literacia científica dos alunos no final do 3.º ciclo do Ensino Básico, este estudo realizou a definição dos domínios cognitivos, a definição do universo e da representatividade do conteúdo, a elaboração da tabela de especificação, a construção do instrumento e a análise teórica dos itens.

Como resultado, foram elaborados 64 itens, no formato “verdadeiro-falso-não sei”, que avaliam os domínios cognitivos de compreender, analisar e avaliar problemas e situações cotidianas que envolvem as competências presentes nos principais documentos curriculares portugueses vigentes das Ciências Físicas e Naturais. Dentre estes, após a última etapa de recolha de evidências de validade de conteúdo realizada, a análise teórica, foram selecionados 35 itens para compor a primeira versão do instrumento de avaliação da literacia em desenvolvimento, a qual será submetida ao teste piloto a fim de recolher evidências de validade baseadas na estrutura interna, gerando maior robustez para a definição dos itens que integrarão o instrumento final.

Agradecimento

Este trabalho foi financiado por fundos nacionais através da FCT – Fundação para a Ciência e a Tecnologia, I.P., no âmbito da Bolsa de Investigação de Doutoramento em Ciências da Educação com referência UI/BD/151034/2021 e do Projeto UIDB/4312/2020.

Referências

- AAAS. (1993). *Project 2061: benchmarks for science literacy*. Oxford University Press.
- AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. 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., RATHS, J., & Wittrock, M. C. (2001). *A taxonomy for learning, teaching and assessing: a revision of Bloom's taxonomy of educational objectives*. Addison Wesley Longman.
- CETIC. (2014). *Pesquisa sobre o uso das tecnologias da informação e comunicação nas escolas brasileiras – TIC Educação 2013*. Comitê Gestor da Internet no Brasil.
- Depresbiteris, L., & Tavares, M. R. (2009). *Diversificar é preciso...: Instrumentos e técnicas de avaliação de aprendizagem*. Senac São Paulo.
- DGE. (2018). *Aprendizagens Essenciais*. <https://www.dge.mec.pt/aprendizagens-essenciais-0>
- Ebel, R. L. (1979). *Essentials of educational measurement* (3rd ed.). Prentice Hall International, INC.
- Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). 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.
- Frisbie, D. A. (1973). *Multiple choice versus true-false: a comparison of reliabilities and current validities*. 10(4), 297–304.
- Frisbie, D. A., & Becker, D. F. (1991). An analysis of textbook advice about true-false tests. *Applied Measurement in Education*, 4(1), 67–83.
- Galvão, C., Neves, A., Freire, A. M., Lopes, A. M. S., Santos, M. da C., Vilela, M. da C., Oliveira, M. T., & Pereira, M. (2001). *Ciências físicas e naturais - orientações curriculares para o 3º ciclo do ensino básico*. Ministério da Educação.
- 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.). Lawrence Erlbaum Associates.
- Haladyna, T. M. (2018). Developing test items for course examinations. *IDEA*, 70, 1–16.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73. <https://doi.org/10.1111/jedm.12001>

- Laugksch, R. C., & Spargo, P. E. (1996a). 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>
- Laugksch, R. C., & Spargo, P. E. (1996b). Development of a pool of scientific literacy test-items based on selected AAAS literacy goals. *Science Education*, 80(2), 121–143. [https://doi.org/10.1002/\(SICI\)1098-237X\(199604\)80:2<121::AID-SCE1>3.0.CO;2-I](https://doi.org/10.1002/(SICI)1098-237X(199604)80:2<121::AID-SCE1>3.0.CO;2-I)
- Martins, G. O., Gomes, C. A. S., Brocardo, J. M. L., Pedroso, J. V., Carrillo, J. L. A., Silva, L. M. U., Encarnação, M. M. G. A. da, Horta, M. J. do V. C., Calçada, M. T. C. S., Nery, R. F. V., & Rodrigues, S. M. C. V. (2017). *Perfil dos alunos à saída da escolaridade obrigatória*. Ministério da Educação e Ciência - DGE.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 3–209). Macmillan.
- Miller, J. D. (1983). Scientific literacy: a conceptual and empirical review. *Daedalus*, 112(2), 29–48.
- Miller, M. D., Linn, R. L., & Gronlund, N. E. (2009). *Measurement and assessment in teaching* (10th ed.). Pearson Education, Inc. <https://doi.org/10.1007/s13398-014-0173-7.2>
- Pasquali, L. (2009). *Psicometria teoria dos testes na psicologia e na educação* (4th ed.). Vozes.
- Popham, W. J. (2017). *Classroom assessment: what teachers need to know* (8th ed.). Pearson.
- Rush, B. R., Rankin, D. C., & White, B. J. (2016). The impact of item-writing flaws and item complexity on examination item difficulty and discrimination value. *BMC Medical Education*, 16(250), 1–10.
- Russel, M. K., & Airasian, P. W. (2014). *Avaliação em sala de aula: conceitos e aplicações* (7th ed.). AMGH.