ISSN: 2358-8829

ALFABETIZAÇÃO CIENTÍFICA EM ESTUDANTES DOS ANOS FINAIS DO ENSINO FUNDAMENTAL: DESENVOLVIMENTO E VALIDAÇÃO DE UM INSTRUMENTO MULTIDIMENSIONAL[1]

Geovane Barbosa dos Santos [2]
Antônio Azambuja Miragem [3]

INTRODUÇÃO

A alfabetização científica (AC) consolidou-se como uma das finalidades centrais do ensino de Ciências na Educação Básica, por sua relação com a formação cidadã e com a compreensão crítica de fenômenos naturais e sociais mediados pela ciência e pela tecnologia (DeBoer, 2000; Sasseron; Carvalho, 2011). Para além de memorizar conteúdos, é esperado que os estudantes aprendam a ler, interpretar e problematizar o mundo em que vivem (Chassot, 2000; Lorenzetti; Delizoicov, 2001). Nos anos finais do Ensino Fundamental, isso significa se apropriar de conceitos de Ciências e, ao mesmo tempo, desenvolver o pensamento crítico necessário para reconhecer evidências, avaliar fontes e compreender as relações entre ciência e sociedade.

Embora a literatura convirja quanto ao caráter multidimensional do construto (Miller, 1983; Brossard; Shanahan, 2006; Duschl, 2008), sua avaliação permanece frágil. Instrumentos centrados no domínio conceitual ou em habilidades procedimentais isoladas produzem retratos parciais da AC, o que relega a segundo plano seus aspectos epistêmicos e sociais (Brossard; Shanahan, 2006). Revisões sistemáticas apontam, ainda, lacunas de validação e concentração de testes voltados ao ensino superior (Coppi; Fialho; Cid, 2023). No Brasil, o *Test of Basic Scientific Literacy* (TBSL) tem sido utilizado, mas com limites de adaptação a determinados públicos escolares (Vizzotto; Del Pino, 2020). Diante de resultados como os do PISA 2022, em que apenas cerca de 45% dos estudantes brasileiros

[1] Trabalho resultante de pesquisa de doutorado desenvolvida no Programa de Pós-Graduação em Educação em Ciências da Universidade Federal do Rio Grande do Sul (UFRGS), com apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

[2] Doutorando do Curso de Educação em Ciências da Universidade Federal do Rio Grande do Sul - UFRGS, RS, bgeovane.2011@gmail.com;

[3] Professor orientador: Doutor em Fisiologia, Universidade Federal do Rio Grande do Sul - UFRGS, RS, antonio.miragem@iffarroupilha.edu.br.

de 15 anos atingiram o nível 2 de proficiência em ciências (OECD, 2023), justifica-se a construção de instrumentos diagnósticos concebidos, desde a origem, para os anos finais do Ensino Fundamental.

Este trabalho, recorte de uma pesquisa de doutorado, tem como objetivo apresentar o desenvolvimento do Teste de Alfabetização Científica para Estudantes do Ensino Fundamental (TAC-EF), instrumento multidimensional destinado a estudantes do 6º ao 9º ano, e as evidências iniciais de sua validade de conteúdo. Tem como objetivos específicos: objetiva-se: (a) descrever a matriz de especificação do instrumento, organizada em eixos e níveis de complexidade; (b) explicitar o modelo de pontuação por eixo e global; e (c) examinar a validade de conteúdo dos itens por meio do Coeficiente de Validade de Conteúdo (CVC).

REFERENCIAL TEÓRICO

A Alfabetização Científica é um conceito polissêmico, cujos sentidos variam do domínio de conteúdos ao engajamento crítico com a ciência na sociedade (DeBoer, 2000; Laugksch, 2000). Miller (1983) identificou três dimensões estruturantes: a compreensão de conteúdos científicos, da natureza da ciência e dos impactos da ciência e da tecnologia na sociedade. No Brasil, Sasseron e Carvalho (2008) sintetizaram eixos estruturantes convergentes (conceitos científicos básicos; natureza da ciência e fatores éticos e políticos de sua prática; relações entre ciência, tecnologia, sociedade e ambiente), e a apropriação do conceito, influenciada pela pedagogia Freireana e pelo enfoque Ciência-Tecnologia-Sociedade (CTS), enfatiza a formação de sujeitos capazes de compreender de maneira crítica a realidade mediada pela ciência e de intervir sobre ela (Auler; Delizoicov, 2001; Lorenzetti; Delizoicov, 2001).

Perspectivas recentes deslocam a ênfase para práticas epistêmicas, como justificar, avaliar e comparar explicações à luz de evidências (Duschl, 2008; Sampson; Clark, 2008). Norris e Phillips (2003) mostram que o sentido fundamental da AC, ligado à leitura e à escrita de textos científicos, é constitutivo de seu sentido derivado, o conhecimento sobre ciência. Kolstø (2001) destaca que a participação cidadã em questões sociocientíficas exige examinar a credibilidade das informações e os limites das evidências disponíveis. Já Silva e Sasseron (2021) defendem que a AC mobilize os domínios conceitual, epistêmico, material e social do conhecimento científico.

Neste trabalho, a AC é definida como a capacidade de compreender conceitos e explicações científicas fundamentais, entender como o conhecimento científico é produzido e validado e mobilizar esses saberes na análise de questões sociais, tecnológicas e ambientais do cotidiano. Para fins de avaliação, o construto foi desdobrado em cinco eixos correlatos, porém analiticamente distinguíveis: (1) conceitos científicos; (2) natureza da ciência; (3) ciência e sociedade; (4) pensamento crítico e avaliação de fontes; e (5) argumentação com base em evidências. Os três primeiros correspondem aos eixos estruturantes de Sasseron e Carvalho (2008); os dois últimos tornam explícitas, como dimensões próprias, a leitura crítica de informações (Norris; Phillips, 2003; Kolstø, 2001) e a argumentação como prática epistêmica (Sampson; Clark, 2008). Como a AC é um contínuo de desempenhos, e não uma condição dicotômica, cada eixo foi organizado em níveis de complexidade, entendidos como ordenação teórico-operacional, e não como escala empiricamente calibrada.

Quanto à mensuração, a validade não é uma propriedade fixa do teste, mas um juízo sobre a adequação das interpretações e dos usos dos escores, sustentado por evidências de conteúdo, processos de resposta, estrutura interna, relações com outras variáveis e consequências do uso (Messick, 1995; AERA; APA; NCME, 2014). A evidência baseada no conteúdo costuma ser a primeira a ser produzida e depende de uma matriz de especificação e do julgamento sistemático de especialistas (Pasquali, 1998; Alexandre; Coluci, 2011). Nesse contexto, o CVC de Hernández-Nieto (2002) permite avaliar critérios separadamente e corrigir o coeficiente conforme o número de juízes, embora ainda seja pouco difundido na pesquisa em Ensino de Ciências (Vizzotto, 2025).

METODOLOGIA

O estudo caracteriza-se como pesquisa metodológica de desenvolvimento de instrumento, de abordagem mista, que articula decisões teóricas (definição do construto, organização da matriz e elaboração dos itens) e um procedimento quantitativo de julgamento por especialistas (Pasquali, 1998; Clark; Watson, 2019). Foram realizadas três etapas, sendo: (1) construção do instrumento a partir de uma matriz de especificação; (2) avaliação dos itens por juízes; e (3) formalização do modelo de mensuração dos escores, apresentado nos resultados por constituir produto do estudo. Esta fase não incluiu aplicação com estudante. Logo, as inferências restringem-se, portanto, à proposição do instrumento e à adequação inicial de seu conteúdo.

Os itens foram redigidos em linguagem acessível, contextualizados em situações cotidianas (saúde, ambiente, consumo, mídia e tecnologia) e alinhados à Base Nacional Comum Curricular (Brasil, 2018). O formato verdadeiro/falso foi escolhido por favorecer a aplicação e a correção objetivas. Para evitar o mero reconhecimento de informações, as afirmações foram formuladas de modo a exigir inferência, identificação de evidências e julgamento conceitual. Cada item foi classificado em um eixo e em um nível de complexidade, segundo a operação cognitiva mobilizada, a distância inferencial exigida e o grau de articulação entre informações. O tempo estimado de aplicação é de 60 minutos, em formato impresso ou digital.

O painel intencional foi composto por três especialistas com doutorado em Educação em Ciências e experiência com os anos finais do Ensino Fundamental. Os juízes receberam material padronizado (objetivo e público-alvo, definição operacional dos eixos e dos níveis, versão completa dos itens e formulário de avaliação) e julgaram cada item quanto à clareza (redação adequada à faixa etária e sem ambiguidades), à pertinência (adequação ao eixo e à competência pretendida) e à relevância (representatividade em relação ao construto), em escala de 1 a 5, com espaço para comentários e sugestões.

O CVC foi calculado segundo Hernández-Nieto (2002), conforme sistematizado por Vizzotto (2025). Para cada item e critério, obteve-se a média das notas dos juízes (M_x) e o coeficiente inicial (CVC_i), dividindo-se a média pelo valor máximo da escala ($V_{máx}$ = 5); em seguida, descontou-se o fator de correção do viés dos juízes (Pe_i), associado à possibilidade de concordância casual, obtendo-se o coeficiente corrigido (CVC_c):

$$M_x = \frac{\sum_{i=1}^{J} x_i}{J} \qquad \mathrm{CVC}_i = \frac{M_x}{V_{\mathrm{máx}}} \quad (1)$$

$$\mathrm{Pe}_i = \left(\frac{1}{J}\right)^J \qquad \mathrm{CVC}_c = \mathrm{CVC}_i - \mathrm{Pe}_i \quad (2)$$

em que x_i é a nota atribuída pelo juiz i e J é o número de juízes. Com J = 3, Pe_i = 0,037, e o maior valor alcançável é CVC_c = 0,963, quando os três juízes atribuem a nota máxima. O CVC de cada eixo e o CVC total do instrumento corresponderam às médias dos coeficientes corrigidos.

Para o trabalho, foram adotados o ponto de corte de 0,80. Itens com $CVC_c \geq 0,80$ nos três critérios seriam mantidos, itens abaixo do corte em um ou dois critérios seriam revisados e reavaliados e, por fim, itens abaixo do corte nos três critérios seriam excluídos (Hernández-Nieto, 2002; Vizzotto, 2025). Os coeficientes foram triangulados com os comentários dos juízes, categorizados em redação e compreensão do enunciado; ambiguidade, dupla afirmação ou generalização; desajuste entre item e eixo; desajuste entre item e nível de complexidade; e recomendações de simplificação linguística. Disso resultou uma matriz de decisão que registra, para cada item, o CVC por critério, a decisão tomada e sua justificativa. A participação dos juízes foi voluntária e formalizada por meio do Termo de Consentimento Livre e Esclarecido, com identificação por códigos (A1, A2 e A3) e tratamento confidencial das pontuações e dos comentários.

Quanto ao uso da Inteligência Artificial Generativa, para os devidos fins de integridade científica e em conformidade com a Política de Integridade na Atividade Científica do CNPq (Portaria CNPq nº 2.664/2026), neste trabalho utilizou-se como suporte à adaptação do texto, ao aprimoramento da escrita acadêmica e à conferência dos cálculos. O conteúdo foi integralmente revisado pelos autores, que assumem total responsabilidade pelos dados, análises, interpretações e conclusões apresentadas.

RESULTADOS E DISCUSSÃO

Matriz de especificação do TAC-EF

O TAC-EF resultou em 70 afirmações do tipo verdadeiro/falso, distribuídas nos cinco eixos e em quatro níveis gerais de complexidade: (N1) reconhecimento ou identificação de informações explícitas; (N2) aplicação contextualizada de conceitos em situações simples; (N3) interpretação e análise de dados, relações e evidências; e (N4) integração crítica e argumentação. Esses níveis foram especificados para cada eixo, conforme o Quadro 1.

Quadro 1 – Eixos, níveis de complexidade e exemplos de itens do TAC-EF

Eixo (itens)	Níveis de complexidade (N1 → N4)	Exemplo de item (nível; gabarito)
1. Conceitos científicos (1–15)	Reconhecimento; Aplicação; Interpretação; Análise integrada.	Item 3 (N1; V): “Quando cientistas testam um novo medicamento, eles comparam os resultados com um grupo que não recebeu o tratamento para garantir resultados confiáveis.”

2. Natureza da ciência (16–29)	Visão ingênua da ciência; Reconhecimento da dinamicidade; Compreensão das práticas científicas; Visão crítica e contextualizada.	Item 23 (N2; F): "Se um estudante encontra um resultado diferente em um experimento, isso significa que ele errou e deve ser desconsiderado."
3. Ciência e sociedade (30–44)	Reconhecimento superficial; Aplicação em situações sociais simples; Compreensão da articulação ciência-sociedade; Integração crítica e sustentável.	Item 42 (N3; F): "É aceitável divulgar resultados de pesquisas sem revisar os dados, desde que a conclusão pareça coerente."
4. Pensamento crítico e avaliação de fontes (45–60)	Aceitação acrítica da informação; Reconhecimento da necessidade de verificação; Aplicação de critérios de confiabilidade; Autonomia crítica informacional.	Item 45 (N1; F): "Quando um influenciador digital diz que um suco natural cura gripe sem apresentar estudos científicos, essa é uma forma segura de propagar conhecimento."
5. Argumentação com base em evidências (61–70)	Ausência de exigência por evidências; Reconhecimento da importância de evidências; Uso de dados na construção de argumentos; Valorização da argumentação científica.	Item 68 (N4; V): "Compreender as escalas usadas em gráficos e mapas é importante para interpretar informações corretamente."

Legenda: N1 a N4 = níveis de complexidade; V = verdadeiro; F = falso.
Fonte: Elaborado pelos autores (2026).

A gradação respeita a natureza de cada dimensão. No eixo 4, por exemplo, vai da aceitação acrítica da informação à autonomia crítica informacional, em consonância com o sentido fundamental de AC proposto por Norris e Phillips (2003). A Tabela 1 apresenta a distribuição dos itens por eixo e nível.

Tabela 1 – Quadro de especificação do TAC-EF (eixo × nível × número de itens × pesos)

Eixo	Itens	N1	N2	N3	N4	Total	Den_e	N3 + N4 no escore máximo (%)	Itens com gabarito F
1. Conceitos científicos	1–15	3	4	4	4	15	39	71,8	3
2. Natureza da ciência	16–29	3	5	4	2	14	33	60,6	5
3. Ciência e sociedade	30–44	2	6	5	2	15	37	62,2	4
4. Pensamento crítico e avaliação de fontes	45–60	4	5	5	2	16	37	62,2	4
5. Argumentação com base em evidências	61–70	2	2	3	3	10	27	77,8	3
Total	**1–70**	**14**	**22**	**21**	**13**	**70**	**173**	**66,5**	**19**

Nota: pesos por nível $w_1 = 1$, $w_2 = 2$, $w_3 = 3$ e $w_4 = 4$; Den_e = soma dos pesos multiplicados pelo número de itens de cada nível do eixo.
Fonte: Elaborado pelos autores (2026).

Modelo de mensuração dos escores

Cada resposta é codificada como acerto ($X_{ij} = 1$) ou erro ($X_{ij} = 0$), e as respostas em branco são tratadas como erro, o que mantém constante o número de itens considerados por eixo e nível. Em seguida, somam-se os acertos do participante *i* em cada eixo *e* e nível *l* (R_{el}) e calcula-se o denominador ponderado do eixo (Den_e), que corresponde ao máximo de pontos possível:

$$R_{el}(i) = \sum_{j \in (e,l)} X_{ij} \qquad \text{Den}_e = \sum_{l=1}^{4} w_l \cdot n_{el} \tag{3}$$

em que n_{el} é o número de itens do eixo *e* no nível *l* e w_l é o peso do nível. O escore principal de cada eixo (S_e) expressa, em escala de 0 a 100, a proporção do total ponderado alcançado, o que permite comparar eixos com números distintos de itens; o escore global (S_{global}), opcional e descritivo, é a média simples dos cinco eixos:

$$S_e(i) = \frac{\sum_{l=1}^{4} w_l \cdot R_{el}(i)}{\text{Den}_e} \times 100 \qquad S_{\text{global}}(i) = \frac{1}{5} \sum_{e=1}^{5} S_e(i) \tag{4}$$

Como complemento, propôs-se um indicador exploratório de discrepância (D_e), que registra apenas as inversões locais da progressão esperada, isto é, situações em que a taxa de acerto em um nível (p_{el}) supera a do nível imediatamente anterior:

$$p_{el}(i) = \frac{R_{el}(i)}{n_{el}} \qquad D_e(i) = \sum_{l=2}^{4} \max\left(0,\ p_{el}(i) - p_{e(l-1)}(i)\right) \tag{5}$$

O indicador assume valor zero quando o padrão de acertos é compatível com uma progressão não crescente ao longo dos níveis. Como exemplo hipotético, no eixo 5 ($\text{Den}_5 = 27$), um estudante que acerta 2/2 itens de N1, 2/2 de N2, 1/3 de N3 e 1/3 de N4 obtém $S_5 = (2 + 4 + 3 + 4)/27 \times 100 = 48{,}1$ e $D_5 = 0$. Outro, que acerta 1/2, 1/2, 2/3 e 3/3, obtém $S_5 = (1 + 2 + 6 + 12)/27 \times 100 = 77{,}8$, mas $D_5 = 0{,}50$, padrão que pode sinalizar acerto casual, lapsos de atenção, dificuldades de leitura ou conhecimento fragmentado. O indicador não corrige nem penaliza o escore, apenas serve como marcador para uma leitura prudente dos resultados.

É importante ressaltar que o modelo tem limites explícitos, visto que os pesos são convenções teórico-operacionais, e não parâmetros empíricos de dificuldade. O escore global não pressupõe unidimensionalidade do construto e, como os anos escolares diferem em oportunidades de aprendizagem, os escores devem ser lidos no contexto da série do estudante, e não como métrica de comparação entre séries. Essa delimitação é coerente com a concepção de validade como juízo sobre interpretações e usos pretendidos (Messick, 1995).

Evidências de validade de conteúdo

Os três juízes avaliaram os 70 itens nos três critérios, totalizando 630 julgamentos. Desses, 603 (95,7%) receberam nota 5, 16 receberam nota 4 e 11 receberam nota 3; nenhuma nota foi inferior a 3. Em 58 itens (82,9%), os três juízes atribuíram a nota máxima nos três critérios. A Tabela 2 sintetiza os coeficientes por eixo e critério.

Tabela 2 – Coeficiente de Validade de Conteúdo (CVC_c) por eixo e critério

Eixo	Itens	Clareza	Pertinência	Relevância	CVC do eixo
1. Conceitos científicos	15	0,936 (0,83)	0,954 (0,90)	0,954 (0,83)	0,948
2. Natureza da ciência	14	0,949 (0,83)	0,949 (0,83)	0,949 (0,83)	0,949
3. Ciência e sociedade	15	0,950 (0,83)	0,954 (0,90)	0,954 (0,90)	0,953
4. Pensamento crítico e avaliação de fontes	16	0,955 (0,83)	0,963 (0,96)	0,963 (0,96)	0,960
5. Argumentação com base em evidências	10	0,950 (0,83)	0,936 (0,83)	0,936 (0,83)	0,941
Instrumento	**70**	**0,948 (0,83)**	**0,952 (0,83)**	**0,952 (0,83)**	**0,951**

Nota: valores expressos como média (mínimo) dos coeficientes corrigidos dos itens; com três juízes, o valor máximo possível é 0,963.
Fonte: Elaborado pelos autores (2026).

Todos os itens alcançaram $CVC_c \geq 0,80$ nos três critérios (mínimo de 0,83), e o CVC total do instrumento foi de 0,951. Pela regra de decisão adotada, nenhum item precisou ser reavaliado ou excluído. Os 12 itens com ao menos uma nota inferior a 5 foram os de número 1, 2, 9, 14, 16, 24, 30, 34, 42, 58, 62 e 65. A clareza concentrou ressalvas maiores, onde sete itens obtiveram $CVC_c = 0,83$ nesse critério (2, 9, 14, 16, 42, 58 e 62), contra três em pertinência (16, 62 e 65) e quatro em relevância (9, 16, 62 e 65). Os itens 16 e 62 foram os únicos com esse valor nos três critérios.

Esse padrão é coerente com o desafio de redigir afirmações curtas e inequívocas para estudantes de 11 a 14 anos, já que, em itens dicotômicos, uma generalização pode

tornar o gabarito discutível. Além disso, cinco dos 14 itens de nível 1 receberam ressalvas, proporção superior à dos níveis 2 (1 de 22), 3 (5 de 21) e 4 (1 de 13), o que sugere que as reservas recaíram sobre a formulação dos enunciados do que sobre a complexidade planejada. Os ajustes decorrentes do julgamento, orientados pelos comentários qualitativos, envolveram redação, adequação linguística e refinamento do alinhamento dos itens aos eixos. As ressalvas distribuíram-se de modo equilibrado entre os juízes (9, 10 e 8 notas inferiores a 5 para A1, A2 e A3), o que reduz a hipótese de que os resultados reflitam a severidade ou a complacência de um único avaliador.

Diferentemente do *TBSL* (Laugksch; Spargo, 1996), com o qual compartilha o formato dicotômico, e do *Test of Scientific Literacy Skills* (TOSLS), voltado a universitários (Gormally; Brickman; Lutz, 2012), o TAC-EF foi construído para os anos finais do Ensino Fundamental, com itens organizados em eixos e níveis e regras de pontuação documentadas, o que responde à lacuna apontada por Coppi, Fialho e Cid (2023).

Os resultados devem, contudo, ser lidos com cautela. Com três juízes, o CVC_c máximo é 0,963, e a média obtida (0,951) situa-se próxima desse teto, o que reduz a capacidade do índice de discriminar itens. O painel atende ao número mínimo de três especialistas indicado para o procedimento (Hernández-Nieto, 2002; Vizzotto, 2025), mas um painel maior e mais diversos, que inclua, por exemplo, professores da Educação Básica, tenderia a produzir estimativas mais estáveis. Além disso, a validade de conteúdo atesta a adequação dos itens segundo especialistas, mas não informa como os estudantes interpretam os enunciados nem se a estrutura empírica corresponde aos cinco eixos (AERA; APA; NCME, 2014; Clark; Watson, 2019). Tais evidências dependem de entrevistas cognitivas, aplicação piloto, análises de estrutura interna e estimativas de precisão, as quais serão focos para pesquisas futuras.

CONSIDERAÇÕES FINAIS

Este trabalho apresentou o TAC-EF, que operacionaliza a AC em cinco eixos, com 70 itens dicotômicos distribuídos em quatro níveis de complexidade e um modelo explícito de pontuação por eixo, global e de discrepância. O julgamento por especialistas resultou em coeficientes de validade de conteúdo superiores ao ponto de corte em todos os itens e critérios, o que sustenta a adequação inicial dos itens aos eixos teóricos e ao público-alvo.

A principal contribuição está em oferecer à área de Ensino de Ciências um instrumento construído para os anos finais do Ensino Fundamental, com construto, matriz e regras de cálculo documentados de forma transparente e reproduzível. Para professores e redes de ensino, a leitura por eixos e níveis permite identificar em que dimensões os estudantes apresentam maior domínio ou dificuldade e orientar intervenções didáticas, em vez de reduzir a AC a um único escore.

As inferências, porém, permanecem circunscritas ao conteúdo. Como passos futuro, prevê-se: (a) realizar entrevistas cognitivas e aplicação piloto com estudantes do 6º ao 9º ano; (b) examinar a estrutura interna, a precisão e o funcionamento dos itens; (c) avaliar o efeito do desequilíbrio entre afirmações verdadeiras e falsas e ampliar o número de itens de nível 4; e (d) desenvolver e validar descritores de níveis de desempenho por eixo que apoiem a interpretação pedagógica dos escores. Com esse percurso, espera-se que o TAC-EF contribua para diagnósticos educacionais informativos e para o fortalecimento da avaliação da AC na Educação Básica.

Palavras-chave: Alfabetização científica; Avaliação educacional; Desenvolvimento de instrumento; Ensino de Ciências; Validade de conteúdo.

REFERÊNCIAS

ALEXANDRE, N. M. C.; COLUCI, M. Z. O. Validade de conteúdo nos processos de construção e adaptação de instrumentos de medidas. **Ciência & Saúde Coletiva**, Rio de Janeiro, v. 16, n. 7, p. 3061-3068, 2011. DOI: 10.1590/S1413-81232011000800006.

AMERICAN EDUCATIONAL RESEARCH ASSOCIATION (AERA); AMERICAN PSYCHOLOGICAL ASSOCIATION (APA); NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION (NCME). **Standards for educational and psychological testing**. Washington, DC: AERA, 2014.

AULER, D.; DELIZOICOV, D. Alfabetização científico-tecnológica para quê? **Ensaio: Pesquisa em Educação em Ciências**, Belo Horizonte, v. 3, n. 2, p. 122-134, 2001. DOI: 10.1590/1983-21172001030203.

BRASIL. Ministério da Educação. **Base Nacional Comum Curricular**. Brasília, DF: MEC, 2018.

BROSSARD, D.; SHANAHAN, J. Do they know what they read? Building a scientific literacy measurement instrument based on science media coverage. **Science Communication**, Thousand Oaks, v. 28, n. 1, p. 47-63, 2006. DOI: 10.1177/1075547006291345.

CHASSOT, A. **Alfabetização científica**: questões e desafios para a educação. Ijuí: Ed. Unijuí, 2000.

CLARK, L. A.; WATSON, D. Constructing validity: new developments in creating objective measuring instruments. **Psychological Assessment**, Washington, DC, v. 31, n. 12, p. 1412-1427, 2019. DOI: 10.1037/pas0000626.

COPPI, M.; FIALHO, I.; CID, M. Scientific literacy assessment instruments: a systematic literature review. **Educação em Revista**, Belo Horizonte, v. 39, e37523, 2023. DOI: 10.1590/0102-4698237523-T.

DEBOER, G. E. Scientific literacy: another look at its historical and contemporary meanings and its relationship to science education reform. **Journal of Research in Science Teaching**, Hoboken, v. 37, n. 6, p. 582-601, 2000.

DUSCHL, R. Science education in three-part harmony: balancing conceptual, epistemic, and social learning goals. **Review of Research in Education**, Washington, DC, v. 32, n. 1, p. 268-291, 2008. DOI: 10.3102/0091732X07309371.

GORMALLY, C.; BRICKMAN, P.; LUTZ, M. Developing a Test of Scientific Literacy Skills (TOSLS): measuring undergraduates' evaluation of scientific information and arguments. **CBE—Life Sciences Education**, Bethesda, v. 11, n. 4, p. 364-377, 2012. DOI: 10.1187/cbe.12-03-0026.

HERNÁNDEZ-NIETO, R. A. **Contributions to statistical analysis**. Mérida: Universidad de Los Andes, 2002.

KOLSTØ, S. D. Scientific literacy for citizenship: tools for dealing with the science dimension of controversial socioscientific issues. **Science Education**, Hoboken, v. 85, n. 3, p. 291-310, 2001. DOI: 10.1002/sce.1011.

LAUGKSCH, R. C. Scientific literacy: a conceptual overview. **Science Education**, Hoboken, v. 84, n. 1, p. 71-94, 2000.

LAUGKSCH, R. C.; SPARGO, P. E. 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**, Bristol, v. 5, n. 4, p. 331-359, 1996. DOI: 10.1088/0963-6625/5/4/003.

LORENZETTI, L.; DELIZOICOV, D. Alfabetização científica no contexto das séries iniciais. **Ensaio: Pesquisa em Educação em Ciências**, Belo Horizonte, v. 3, n. 1, p. 45-61, 2001. DOI: 10.1590/1983-21172001030104.

MESSICK, S. Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. **American Psychologist**, Washington, DC, v. 50, n. 9, p. 741-749, 1995. DOI: 10.1037/0003-066X.50.9.741.

MILLER, J. D. Scientific literacy: a conceptual and empirical review. **Daedalus**, Cambridge, v. 112, n. 2, p. 29-48, 1983.

NORRIS, S. P.; PHILLIPS, L. M. How literacy in its fundamental sense is central to scientific literacy. **Science Education**, Hoboken, v. 87, n. 2, p. 224-240, 2003. DOI: 10.1002/sce.10066.

OECD. **PISA 2022 results (volume I and II)**: country notes: Brazil. Paris: OECD Publishing, 2023.

PASQUALI, L. Princípios de elaboração de escalas psicológicas. **Revista de Psiquiatria Clínica**, São Paulo, v. 25, n. 5, p. 206-213, 1998.

SAMPSON, V.; CLARK, D. B. Assessment of the ways students generate arguments in science education: current perspectives and recommendations for future directions. **Science Education**, Hoboken, v. 92, n. 3, p. 447-472, 2008. DOI: 10.1002/sce.20276.

SASSERON, L. H.; CARVALHO, A. M. P. Almejando a alfabetização científica no ensino fundamental: a proposição e a procura de indicadores do processo. **Investigações em Ensino de Ciências**, Porto Alegre, v. 13, n. 3, p. 333-352, 2008.

SASSERON, L. H.; CARVALHO, A. M. P. Construindo argumentação na sala de aula: a presença do ciclo argumentativo, os indicadores de alfabetização científica e o padrão de Toulmin. **Ciência & Educação**, Bauru, v. 17, n. 1, p. 97-114, 2011. DOI: 10.1590/S1516-73132011000100007.

SILVA, M. B.; SASSERON, L. H. Alfabetização científica e domínios do conhecimento científico: proposições para uma perspectiva formativa comprometida com a transformação social. **Ensaio: Pesquisa em Educação em Ciências**, Belo Horizonte, v. 23, e34674, 2021. DOI: 10.1590/1983-21172021230129.

VIZZOTTO, P. A. O coeficiente de validade de conteúdo (CVC) como ferramenta para validação de questionários: exemplo de uso na pesquisa em Educação e Ensino de Ciências. **Revista Educar Mais**, v. 9, p. 1-23, 2025. DOI: 10.15536/reducarmais.9.2025.4089.

VIZZOTTO, P. A.; DEL PINO, J. C. O uso do Teste de Alfabetização Científica Básica no Brasil: uma revisão da literatura. **Ensaio: Pesquisa em Educação em Ciências**, Belo Horizonte, v. 22, e15846, 2020. DOI: 10.1590/1983-21172020210116.