



FORMAÇÃO CONTINUADA DE ELABORADORES DE ITENS E QUALIDADE DA AVALIAÇÃO EDUCACIONAL: ANÁLISE DE UM CICLO DE OFICINAS

CONTINUING EDUCATION FOR ITEM WRITERS AND THE QUALITY OF EDUCATIONAL ASSESSMENT: ANALYSIS OF A TRAINING CYCLE

FORMACIÓN CONTINUA DE ELABORADORES DE ÍTEMS Y CALIDAD DE LA EVALUACIÓN EDUCATIVA: ANÁLISIS DE UN CICLO DE TALLERES

PRISCILA DE ARAÚJO DIAS¹

RESUMO

Este estudo analisa um ciclo de oficinas voltado à formação de elaboradores de itens, com o objetivo de compreender suas contribuições para o desenvolvimento profissional e para a qualidade dos instrumentos de avaliação educacional. A formação foi estruturada progressivamente, abordando matrizes de referência, Taxonomia de Bloom Revisada e análise de resultados. Trata-se de um estudo descritivo, de abordagem quantitativa com apoio qualitativo, baseado em questionários com escala Likert aplicados ao final das oficinas. Os dados foram organizados em dimensões relacionadas à condução formativa, compreensão conceitual, aplicação prática e autoeficácia, sendo analisados por medidas descritivas e testes não paramétricos. Os resultados indicam alta avaliação inicial, seguida de redução da concordância em oficinas mais complexas, e posterior recuperação. Esse padrão, estatisticamente significativo, sugere que maior exigência cognitiva favorece postura mais crítica dos elaboradores. Observa-se, também, aumento da autonomia e da capacidade de revisão. Conclui-se que a formação continuada progressiva fortalece a qualidade da avaliação.

Palavras-chave: Avaliação Educacional; Elaboração de Itens; Formação Continuada; Validade; Avaliação em Larga Escala.

ABSTRACT

This study analyzes a workshop cycle designed to train item writers, aiming to understand its contributions to professional development and to the quality of educational assessment instruments. The training was progressively structured, covering reference matrices, the Revised Bloom's Taxonomy, and results analysis. This is a descriptive study with a quantitative approach supported by qualitative data, based on Likert-scale questionnaires administered after each workshop. Data were organized into analytical dimensions related to training quality, conceptual understanding, practical application, and self-efficacy, and analyzed

¹ Mestre em Ecologia e Evolução pela Universidade Federal de Goiás (EcoEvol-UFG) e Analista de Instrumentos de Avaliação da Supervisão de Desenvolvimento de Instrumentos de Avaliação e Certificação (SDIAC) no Centro Brasileiro de Pesquisa em Avaliação e Seleção e de Promoção de Eventos (Cebbraspe). E-mail: priscila.dias@cebraspe.org.br

using descriptive measures and nonparametric tests. Results show a high initial evaluation, followed by a decrease in agreement levels in more complex workshops and subsequent recovery. This statistically significant pattern suggests that increased cognitive demand promotes a more critical stance. Improvements were also observed in autonomy and item review capacity. It is concluded that progressive continuous training strengthens assessment quality.

Keywords: Educational Assessment; Item Writing; Professional Development; Validity; Large-Scale Assessment.

RESUMEN

Este estudio analiza un ciclo de talleres orientado a la formación de elaboradores de ítems, con el objetivo de comprender sus contribuciones al desarrollo profesional y a la calidad de los instrumentos de evaluación educativa. La formación se estructuró progresivamente, abordando matrices de referencia, la Taxonomía de Bloom Revisada y el análisis de resultados. Se trata de un estudio descriptivo, con enfoque cuantitativo y apoyo cualitativo, basado en cuestionarios con escala Likert aplicados al final de cada taller. Los datos se organizaron en dimensiones analíticas relacionadas con la conducción formativa, la comprensión conceptual, la aplicación práctica y la autoeficacia, y se analizaron mediante medidas descriptivas y pruebas no paramétricas. Los resultados muestran una alta evaluación inicial, seguida de una disminución en los niveles de concordancia en talleres más complejos y una posterior recuperación. Este patrón sugiere que una mayor exigencia cognitiva favorece una postura más crítica. También se observan avances en autonomía y revisión de ítems. Se concluye que la formación continua progresiva fortalece la calidad de la evaluación.

Palabras clave: Evaluación Educativa; Elaboración De Ítems; Formación Continua; Validez; Evaluación A Gran Escala.

INTRODUÇÃO

A qualidade dos instrumentos de avaliação educacional está diretamente relacionada à consistência dos itens que os compõem. Em avaliações em larga escala, os itens constituem a principal unidade de medida do desempenho dos estudantes, sendo responsáveis por operacionalizar as habilidades previstas nas matrizes de referência e gerar evidências que subsidiam a interpretação dos resultados (HALADYNA; DOWNING; RODRIGUEZ, 2002). Nesse contexto, problemas na elaboração de itens — como inconsistências no contexto, no comando, no gabarito ou nos distratores — podem comprometer a qualidade das inferências realizadas a partir dos dados obtidos.

A validade ocupa posição central na teoria da medida educacional, sendo compreendida como o grau em que evidências empíricas e argumentos teóricos sustentam as interpretações e os usos propostos para os resultados de um teste (MESSICK, 1989; AERA; APA; NCME, 2014). Assim, a validade não se restringe a uma propriedade estatística do instrumento, mas envolve a coerência entre o construto

avaliado, o modo como os itens são construídos e as inferências realizadas a partir dos resultados.

Tradicionalmente, a discussão sobre qualidade de itens tem sido fortemente associada às análises estatísticas realizadas após a aplicação das provas, como aquelas baseadas na Teoria Clássica dos Testes (TCT) e na Teoria de Resposta ao Item (TRI) (PASQUALI, 2007). No entanto, embora essas abordagens sejam fundamentais para a verificação do comportamento dos itens, elas não esgotam a discussão sobre validade, uma vez que a qualidade do instrumento também depende das etapas anteriores ao processo de aplicação, especialmente da elaboração dos itens (HALADYNA; DOWNING; RODRIGUEZ, 2002).

Estudos contemporâneos apontam que a validade deve ser compreendida a partir de diferentes fontes de evidência, incluindo aspectos relacionados ao conteúdo, ao processo de resposta, à estrutura interna e às consequências do uso dos resultados (AERA; APA; NCME, 2014; JESUS, 2021). Nessa perspectiva, a elaboração de itens constitui etapa fundamental, pois é nesse momento que se definem as condições sob as quais o desempenho dos estudantes será observado e interpretado.

Nesse sentido, a formação de elaboradores emerge como um componente estratégico para a garantia da qualidade das avaliações. Elaboradores que compreendem os princípios que orientam a construção de itens — incluindo o alinhamento com as matrizes de referência, a adequação dos níveis cognitivos e a clareza na formulação de contextos e comandos — tendem a produzir instrumentos mais consistentes e menos suscetíveis a problemas que comprometam sua validade (HALADYNA; DOWNING; RODRIGUEZ, 2002; AERA; APA; NCME, 2014).

Apesar de sua relevância, a formação de elaboradores ainda é pouco explorada como objeto de investigação em estudos aplicados, sobretudo no que se refere à análise de seus efeitos sobre a prática profissional e sobre a qualidade dos instrumentos produzidos (DESIMONE, 2009; GUSKEY, 2002). Diante desse cenário, este estudo analisa um ciclo de oficinas voltado à formação de elaboradores de itens, com foco na compreensão de suas contribuições para o desenvolvimento profissional dos participantes e para o fortalecimento da qualidade dos instrumentos de avaliação.

METODOLOGIA

Trata-se de um estudo descritivo, com abordagem quantitativa e apoio qualitativo, desenvolvido a partir da análise de dados coletados em um ciclo de oficinas voltado à formação de elaboradores de itens. As oficinas foram realizadas ao longo de quatro meses consecutivos (outubro, novembro, dezembro de 2025 e janeiro de 2026), com periodicidade mensal e duração de três horas cada.

O público participante foi composto por bolsistas do Projeto de Pesquisa e Extensão: Pesquisa e Desenvolvimento de Instrumentos de Avaliação da Educação Básica em parceria com o Cebbraspe, envolvidos no processo de elaboração de itens, totalizando 18 participantes no início do ciclo formativo e 11 participantes na última oficina. A variação no número de respondentes ao longo das oficinas foi

considerada na análise dos dados.

A coleta de dados foi realizada ao final de cada oficina, por meio de um formulário eletrônico disponibilizado via *Microsoft Forms*. O instrumento foi composto por afirmativas relacionadas aos conteúdos abordados em cada encontro, às quais os participantes deveriam responder utilizando uma escala do tipo Likert de quatro pontos: “discordo completamente”, “discordo”, “concordo” e “concordo completamente”. Esse tipo de escala é amplamente utilizado em pesquisas educacionais por permitir a mensuração de percepções e atitudes de forma estruturada (LIKERT, 1932) e incluiu afirmativas com diferentes polaridades, sendo consideradas na interpretação dos resultados. O conjunto de afirmativas utilizado no instrumento de coleta de dados é apresentado no Apêndice A, organizado por dimensão analítica.

Os dados foram, inicialmente, exportados em formato Excel e organizados por oficina. Em seguida, procedeu-se à consolidação das bases em uma única planilha e à construção de uma matriz de frequências, na qual foram contabilizadas as respostas para cada afirmativa em cada oficina. Essa organização possibilitou a sistematização dos dados de forma comparável entre os diferentes momentos do ciclo formativo.

A análise estatística foi realizada com o auxílio do *software* RStudio. Inicialmente, os dados foram organizados em uma matriz de frequências e, posteriormente, expandidos para nível individual, permitindo a aplicação de técnicas inferenciais. Foram calculadas medidas descritivas, incluindo média, desvio padrão e intervalos de confiança de 95%, com o objetivo de sintetizar o comportamento das respostas em cada oficina.

Considerando a natureza ordinal dos dados provenientes de escala Likert, optou-se pela utilização de testes não paramétricos (FIELD, 2018). Para verificar as diferenças entre as oficinas, foi aplicado o teste de Kruskal-Wallis, adequado para a comparação de múltiplos grupos independentes sem pressuposto de normalidade.

Quando identificadas diferenças estatisticamente significativas, foram realizadas comparações múltiplas por meio do teste de Wilcoxon com correção de Bonferroni, a fim de identificar quais grupos diferiam entre si.

Adicionalmente, foi calculado o tamanho de efeito (η^2), estimado a partir da estatística H do teste de Kruskal-Wallis, permitindo avaliar a magnitude das diferenças observadas entre as oficinas e sua relevância prática (TOMCZAK; TOMCZAK, 2014). As análises foram conduzidas adotando-se nível de significância de 5% ($\alpha = 0,05$), sendo considerados estatisticamente significativos os resultados com $p < 0,05$.

RESULTADOS

1. AVALIAÇÃO GERAL DAS OFICINAS

A análise dos dados evidencia variação nos níveis de concordância entre as oficinas, conforme a natureza e a complexidade dos conteúdos abordados ao longo do ciclo formativo.

A Oficina 1 apresentou a maior média de concordância ($M = 3,93$; IC 95% $[3,77; 4,08]$), associada a um percentual de concordância próximo da totalidade (99%), indicando avaliação amplamente favorável dos aspectos relacionados à condução formativa, como clareza da linguagem, domínio do conteúdo e suporte oferecido pela equipe.

Nas Oficinas 2 e 3, que abordaram conteúdos mais técnicos relacionados à elaboração de itens e às matrizes de referência, observou-se redução das médias ($M = 2,79$ em ambas), acompanhada de maior dispersão das respostas ($DP = 1,05$ e $DP = 0,89$, respectivamente) e redução dos níveis de concordância (61% e 73%). Esses resultados indicam maior heterogeneidade nas respostas e sugerem aumento da complexidade percebida pelos participantes.

Na Oficina 4, que tratou da Taxonomia de Bloom Revisada, verificou-se recuperação dos indicadores ($M = 3,50$; IC 95% $[3,12; 3,88]$; concordância de 93%), apontando maior consolidação dos conhecimentos trabalhados ao longo do ciclo formativo.

A análise inferencial, realizada por meio do teste de Kruskal-Wallis, indicou diferença estatisticamente significativa entre as oficinas ($\chi^2(3) = 18,21$; $p < 0,001$). As comparações múltiplas evidenciaram diferenças significativas entre a Oficina 1 e as Oficinas 2 ($p = 0,0039$) e 3 ($p = 0,0013$), não foram observadas diferenças significativas entre as Oficinas 2 e 3 ($p = 1,000$), nem entre estas e a Oficina 4. O tamanho de efeito foi classificado como grande ($\eta^2 = 0,292$), indicando que as diferenças observadas possuem relevância prática.

A Figura 1 apresenta a média de concordância por oficina, com intervalos de confiança de 95%, evidenciando o padrão observado ao longo do ciclo formativo.

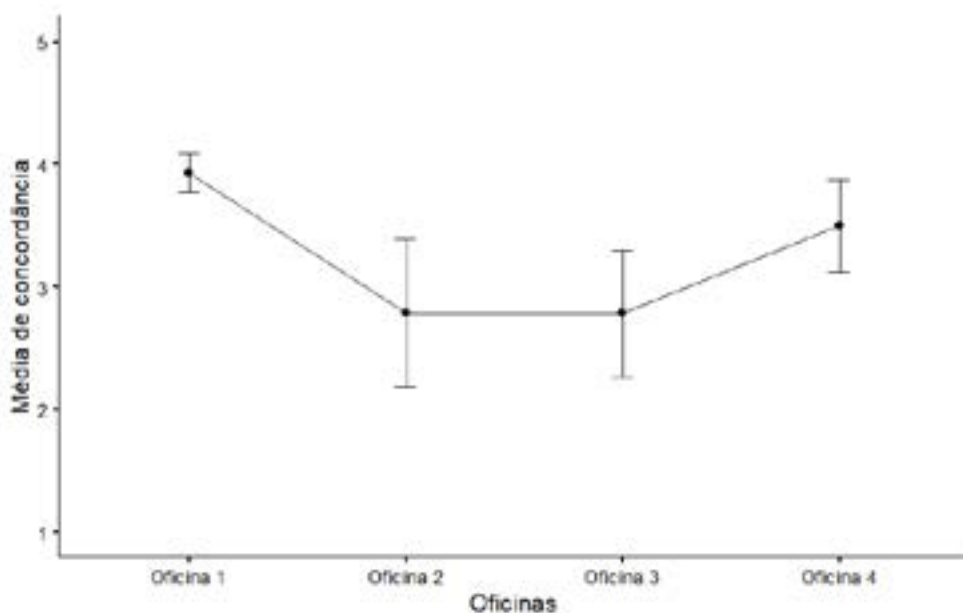


Figura 1 – Média de concordância por oficina, com intervalos de confiança de 95%, em escala Likert de 1 a 4.

2. DIMENSÃO 1 — QUALIDADE DA CONDUÇÃO FORMATIVA

Os dados indicam avaliação altamente positiva da condução da formação, especialmente na Oficina 1, que apresentou os maiores níveis de concordância e menor variabilidade nas respostas. As afirmativas relacionadas à clareza da linguagem, ao domínio do conteúdo, à disponibilidade da equipe e à qualidade dos materiais concentraram respostas nas categorias superiores da escala, evidenciando forte alinhamento entre expectativas dos participantes e execução da formação.

Esse resultado sugere que a estrutura inicial da formação foi bem estabelecida, garantindo condições favoráveis para o engajamento dos participantes e para o desenvolvimento das etapas subsequentes do ciclo formativo.

3. DIMENSÃO 2 — COMPREENSÃO CONCEITUAL

Nas oficinas voltadas à compreensão de conceitos estruturantes da avaliação educacional, observou-se maior variação nas respostas, especialmente nas Oficinas 2 e 3, caracterizadas por médias mais baixas e maior dispersão.

A redução dos níveis de concordância nessas etapas sugere que os conteúdos abordados — particularmente aqueles relacionados às matrizes de referência e às regras de elaboração de itens — demandaram maior esforço cognitivo e domínio técnico por parte dos participantes.

Na Oficina 4, observa-se melhora nos indicadores, evidenciada pelo aumento da média e pela redução da variabilidade, indicando que a abordagem da Taxonomia de Bloom Revisada pode ter contribuído para organizar e consolidar a compreensão conceitual dos participantes.

4. DIMENSÃO 3 — APLICAÇÃO NA PRÁTICA

Os resultados indicam que os participantes reconhecem avanços em sua capacidade de aplicar os conhecimentos na elaboração de itens, embora esse processo não tenha ocorrido de forma homogênea ao longo do ciclo formativo.

A variabilidade observada nas respostas, especialmente nas oficinas intermediárias, sugere que a transição entre compreensão conceitual e aplicação prática constitui um dos principais desafios no processo formativo. Esse padrão é consistente com a diferença entre domínio declarativo e domínio procedimental, frequentemente observada em contextos de aprendizagem profissional.

5. DIMENSÃO 4 — AUTOEFICÁCIA DO ELABORADOR

As afirmativas relacionadas à autoeficácia indicam percepção progressivamente mais positiva quanto à capacidade de elaboração, especialmente nas etapas finais do ciclo.

Os participantes demonstram maior segurança na interpretação de habilidades, na identificação de problemas em itens e na revisão de suas próprias produções, o que evidencia desenvolvimento gradual da autonomia profissional.

Esse resultado é consistente com o padrão observado nas análises descritivas e inferenciais, sugerindo que, apesar da redução inicial nos níveis de concordância

em função do aumento da complexidade, houve fortalecimento da confiança dos participantes ao longo do processo formativo.

DISCUSSÃO

Os resultados evidenciam que o ciclo de oficinas contribuiu para o desenvolvimento progressivo das competências dos elaboradores, ainda que de forma não linear. A variação observada entre as oficinas — caracterizada por níveis elevados de concordância na etapa inicial, redução nas etapas intermediárias e recuperação parcial na etapa final — foi estatisticamente significativa ($\chi^2(3) = 18,21$; $p < 0,001$) e apresentou magnitude elevada ($\eta^2 = 0,292$), indicando que as diferenças observadas possuem relevância prática no contexto da formação analisada. Esse padrão pode ser interpretado à luz da complexidade crescente dos conteúdos abordados, aspecto frequentemente associado a processos formativos que demandam maior mobilização cognitiva e reflexão crítica por parte dos participantes (ILLERIS, 2013).

A elevada avaliação da Oficina 1 indica que a estrutura inicial da formação — incluindo organização, clareza da linguagem e suporte da equipe — foi bem estabelecida, criando condições favoráveis para o engajamento dos participantes. Esse resultado é consistente com estudos que apontam a mediação pedagógica e a qualidade da organização didática como fatores determinantes para o sucesso de processos formativos, especialmente em contextos de desenvolvimento profissional (GUSKEY, 2002).

Por outro lado, a redução das médias nas Oficinas 2 e 3, corroborada pelas diferenças estatisticamente significativas em relação à Oficina 1, não deve ser interpretada como fragilidade da formação, mas como indício do aumento da exigência cognitiva e técnica. Os conteúdos abordados nessas oficinas demandam maior domínio conceitual e operacional, o que pode levar os participantes a adotarem uma postura mais crítica em relação à própria prática. Nesse sentido, a diminuição da concordância pode refletir um processo de tomada de consciência sobre a complexidade da elaboração de itens, fenômeno associado ao desenvolvimento profissional e à construção de expertise (ERICSSON; POOL, 2016).

A ausência de diferenças estatisticamente significativas entre as Oficinas 2 e 3 reforça a interpretação de que essas etapas compartilham um mesmo nível de complexidade cognitiva, caracterizando um momento intermediário do processo formativo. Por sua vez, a recuperação dos indicadores na Oficina 4, mesmo sem diferença significativa em relação às oficinas intermediárias, sugere um processo de reorganização cognitiva, no qual os participantes passam a integrar os conhecimentos previamente adquiridos.

Nesse contexto, a introdução da Taxonomia de Bloom Revisada pode ter desempenhado papel estruturante, ao explicitar os níveis cognitivos envolvidos na resolução de tarefas. A taxonomia oferece uma estrutura que favorece a compreensão do raciocínio exigido do participante, funcionando como elemento articulador entre a habilidade avaliada e a forma de construção do item (ANDERSON; KRATHWOHL, 2001).

Outro aspecto relevante refere-se à relação entre compreensão conceitual e

aplicação prática. Os resultados indicam que, embora os participantes reconheçam avanços em sua capacidade de elaboração, a transição entre compreender os princípios da avaliação e aplicá-los de forma consistente na produção de itens constitui um desafio importante. Esse achado é consistente com a variabilidade observada nas respostas e reforça a necessidade de ações formativas que integrem momentos de estudo conceitual com atividades práticas orientadas e feedback qualificado, condição amplamente reconhecida como essencial para a aprendizagem profissional (DESIMONE, 2009).

No que se refere à autoeficácia, os dados apontam para o desenvolvimento progressivo da confiança dos participantes em sua capacidade de elaborar itens, revisar produções e identificar problemas. Esse resultado é particularmente relevante, uma vez que a percepção de autoeficácia está diretamente relacionada ao desempenho profissional e à capacidade de enfrentar tarefas complexas (BANDURA, 1997). No contexto da avaliação educacional, a autonomia do elaborador constitui fator central para a qualidade dos instrumentos produzidos e para a eficiência dos processos de revisão.

De modo geral, os resultados sugerem que a formação estruturada em etapas progressivas — contemplando aspectos técnicos, cognitivos e evidenciais — favorece a ampliação da compreensão dos elaboradores sobre o processo de avaliação educacional. Ao longo do ciclo, observa-se um deslocamento do foco exclusivo nas regras de elaboração para uma visão mais ampla, que considera o item como instrumento de medida do desempenho dos estudantes, em consonância com abordagens contemporâneas da validade (AERA; APA; NCME, 2014).

Do ponto de vista institucional, esse achado reforça a importância da formação continuada como estratégia de garantia da qualidade das avaliações. Ao promover o desenvolvimento de competências relacionadas à elaboração, à análise e à revisão de itens, a formação contribui para a redução de inconsistências que podem comprometer a validade dos resultados, como itens com problemas de contexto, comando, gabarito ou distratores (HALADYNA; DOWNING; RODRIGUEZ, 2002) incluindo suas implicações sociais e educacionais (SCARAMUCCI, 2011).

Por fim, destaca-se que a organização progressiva do ciclo formativo constitui elemento central para os resultados observados. A articulação entre diferentes dimensões da avaliação — matriz de referência, níveis cognitivos e análise de resultados — possibilita uma compreensão mais integrada do processo avaliativo, favorecendo o desenvolvimento profissional dos elaboradores e fortalecendo os processos institucionais de produção de instrumentos de avaliação.

CONCLUSÃO

Os resultados deste estudo indicam que o ciclo de oficinas contribuiu para o desenvolvimento progressivo das competências dos elaboradores, evidenciado tanto pela ampliação da compreensão conceitual quanto pelo fortalecimento da percepção de autoeficácia na prática de elaboração de itens. A variação observada entre as oficinas — estatisticamente significativa e com magnitude elevada — revela um padrão não linear, caracterizado por redução dos níveis de concordância nas etapas intermediárias e recuperação parcial na etapa final. Esse

comportamento sugere que conteúdos mais técnicos e complexos demandam maior esforço cognitivo e favorecem a adoção de uma postura mais crítica por parte dos participantes, aspecto característico de processos de aprendizagem profissional (ILLERIS, 2013).

A recuperação dos indicadores nas etapas finais do ciclo reforça a importância da organização progressiva da formação, na medida em que a articulação entre diferentes dimensões da avaliação — como estrutura das matrizes, níveis cognitivos e análise de resultados — contribui para a consolidação dos conhecimentos e para o desenvolvimento de uma compreensão mais integrada do processo avaliativo.

Do ponto de vista teórico, os achados dialogam com a compreensão contemporânea de validade como um processo baseado em evidências e interpretações, e não apenas como um atributo estatístico do teste (AERA; APA; NCME, 2014). Nesse sentido, a formação de elaboradores assume papel relevante, ao influenciar diretamente a qualidade das evidências produzidas pelos instrumentos de avaliação.

No âmbito institucional, os resultados reforçam a relevância da formação continuada como estratégia de garantia da qualidade das avaliações. Ao promover o desenvolvimento de competências relacionadas à elaboração, à análise e à revisão de itens, a formação contribui para a redução de inconsistências que podem comprometer a validade dos resultados.

Adicionalmente, os resultados evidenciam que a formação deve ser concebida como processo progressivo e integrado, articulando aspectos conceituais e práticos. A simples apresentação de regras de elaboração mostra-se insuficiente para o desenvolvimento de competências profissionais mais amplas, sendo necessário promover situações formativas que favoreçam a aplicação dos conhecimentos em contextos reais de produção de itens (DESIMONE, 2009).

Por fim, sugere-se que estudos futuros aprofundem a análise da relação entre formação de elaboradores e qualidade empírica dos itens, incorporando indicadores provenientes de análises estatísticas após a aplicação das avaliações, de modo a ampliar a compreensão sobre os efeitos da formação na validade dos instrumentos e no desempenho dos estudantes.

REFERÊNCIAS

AERA; APA; NCME. **Standards for Educational and Psychological Testing**. Washington, DC: American Educational Research Association, 2014.

ANDERSON, L. W.; KRATHWOHL, D. R. (org.). **A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives**. New York: Longman, 2001.

BANDURA, A. **Self-efficacy: the exercise of control**. New York: Freeman, 1997.

BESSA, N. M. Validade: o conceito, a pesquisa e os problemas de provas. **Estudos em Avaliação Educacional**, São Paulo, v. 18, n. 37, 2007.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). **Guia de elaboração de itens**. Brasília: INEP, 2020.

DESIMONE, L. M. **Improving impact studies of teachers' professional development.** *Educational Researcher*, v. 38, n. 3, p. 181–199, 2009.

ERICSSON, K. A.; POOL, R. **Peak: secrets from the new science of expertise.** Boston: Houghton Mifflin Harcourt, 2016.

FIELD, A. **Discovering statistics using IBM SPSS statistics.** 5. ed. London: Sage, 2018.

GUSKEY, T. R. **Professional development and teacher change.** *Teachers and Teaching*, v. 8, n. 3, p. 381–391, 2002.

HALADYNA, T. M.; DOWNING, S. M.; RODRIGUEZ, M. C. **A review of multiple-choice item-writing guidelines for classroom assessment.** *Applied Measurement in Education*, v. 15, n. 3, p. 309–334, 2002.

ILLERIS, K. **Transformative learning and identity.** London: Routledge, 2013.

JESUS, G. R. de. A questão da validade na avaliação educacional brasileira. **Ensaio: Avaliação e Políticas Públicas em Educação**, Rio de Janeiro, 2021. Disponível em: <https://www.scielo.br>. Acesso em: 17 mar. 2026.

LIKERT, R. **A technique for the measurement of attitudes.** *Archives of Psychology*, 22 140, 55, 1932.

MESSICK, S. Validity. In: LINN, R. L. (ed.). **Educational Measurement.** 3. ed. New York: Macmillan, 1989.

PASQUALI, L. Validade dos testes psicológicos: será possível reencontrar o caminho? **Psicologia: Teoria e Pesquisa**, Brasília, v. 23, n. spe, p. 99-107, 2007.

SCARAMUCCI, M. V. R. **Validade e consequências sociais das avaliações em contextos educacionais.** Linguarum Arena, Porto, 2011.

TOMCZAK, M.; TOMCZAK, E. **The need to report effect size estimates revisited.** *Trends in Sport Sciences*, 2014.

APÊNDICE A

INSTRUMENTO DE COLETA DE DADOS

O instrumento de coleta de dados foi composto por afirmativas organizadas em quatro dimensões analíticas: (1) qualidade da condução formativa, (2) compreensão conceitual, (3) aplicação na prática e (4) autoeficácia do elaborador. Os participantes responderam às afirmativas utilizando escala do tipo Likert de quatro pontos: “discordo completamente”, “discordo”, “concordo” e “concordo completamente”.

DIMENSÃO 1 — QUALIDADE DA CONDUÇÃO FORMATIVA (OFICINA 1)

- **A carga horária indicada para a atividade foi suficiente.**
- **As informações e os materiais de apoio disponibilizados contribuíram para a realização do trabalho.**
- **Os(as) palestrantes e a equipe de apoio demonstraram domínio do conteúdo.**

- A dinâmica estabelecida para o evento foi adequada.
- A linguagem utilizada na oficina foi clara e acessível.
- A equipe esteve disponível para o esclarecimento de dúvidas.
- A equipe foi cordial na relação com os participantes.
- A equipe demonstrou competência em relação ao objeto do evento.
- O sistema utilizado facilitou o trabalho realizado.
- O espaço utilizado contribuiu para a realização do trabalho e para a interação entre os participantes.

DIMENSÃO 2 — COMPREENSÃO CONCEITUAL (OFICINAS 2, 3 E 4)

- Os elementos das matrizes de referência devem ser evidenciados na elaboração de itens de avaliação.
- A correta interpretação das matrizes de referência e das associações entre seus elementos é essencial para a produção de itens de qualidade.
- As taxonomias educacionais se expressam nos itens por meio do tipo de raciocínio mobilizado pelo estudante.
- As matrizes e as regras de elaboração estão alinhadas aos objetivos das avaliações em larga escala.
- O processo de elaboração deve considerar as especificidades do público-alvo da avaliação.
- A qualidade dos itens está diretamente relacionada ao domínio das matrizes de referência.
- Compreendi as regras de elaboração e entendo por que elas devem ser seguidas.
- A prática de elaboração contribuiu para a compreensão do conceito de habilidade.
- As oficinas ampliaram meus conhecimentos sobre o processo de elaboração e interpretação de matrizes de referência.
- A oficina contribuiu para a compreensão da estrutura da Taxonomia de Bloom Revisada.
- Consigo diferenciar as categorias do domínio cognitivo da Taxonomia de Bloom Revisada.

- **Compreendo a relação entre níveis cognitivos e tipos de conhecimento (factual, conceitual, procedural e metacognitivo).**
- **Compreendo como a hierarquia dos níveis cognitivos se organiza de forma cumulativa.**
- **Consigo identificar os níveis cognitivos da Taxonomia de Bloom Revisada a partir do tipo de raciocínio exigido em uma tarefa.**
- **Entendo a importância da Taxonomia de Bloom Revisada para a definição de objetivos de aprendizagem.**
- **A distinção entre lembrar, entender, aplicar, analisar, avaliar e criar ficou clara após a oficina.**
- **A oficina ampliou minha compreensão sobre o uso da Taxonomia de Bloom Revisada na elaboração de itens de avaliação.**

DIMENSÃO 3 — APLICAÇÃO NA PRÁTICA (OFICINAS 2, 3 E 4)

- **As oficinas favoreceram a produção de itens alinhados às matrizes de referência.**
- **Consigo perceber avanços na minha prática de elaboração após participar das oficinas.**
- **Compreendo melhor como avaliar minha área de conhecimento após as oficinas.**
- **Meu processo de elaboração melhorou após receber as considerações da equipe.**
- **Consigo identificar, compreender e corrigir erros no processo de elaboração após receber devolutivas.**
- **Consigo identificar rapidamente quando um item está desalinhado à matriz de referência.**
- **Tenho facilidade para revisar meus próprios itens antes de submetê-los à equipe responsável.**
- **Reconheço quando um item precisa de ajustes para atingir o nível cognitivo adequado.**
- **Consigo relacionar habilidades descritas em matrizes de referência aos níveis cognitivos da taxonomia.**

- **Consigo selecionar verbos adequados ao nível cognitivo desejado.**
- **Compreendo melhor como alinhar contexto, comando e alternativas ao nível cognitivo avaliado.**
- **Consigo perceber avanços na qualidade dos itens que elaboro após a oficina.**

DIMENSÃO 4 — AUTOEFICÁCIA DO ELABORADOR (OFICINAS 3 E 4)

- **Sinto-me seguro(a) para interpretar uma habilidade e transformá-la em um item coerente.**
- **Consigo identificar rapidamente quando um item está desalinhado à matriz de referência.**
- **Tenho facilidade para revisar meus próprios itens antes de submetê-los à equipe responsável.**
- **Reconheço quando um item precisa de ajustes para atingir o nível cognitivo adequado.**
- **Sinto confiança em minha capacidade de elaborar itens de forma autônoma e consistente.**