Revista Eletrônica
Acervo Saúde
Electronic Journal Collection Health ISSN 2178-2091

Desempenho em testes com consulta compostos por questões abertas no curso de medicina

Performance in open-book open-ended question tests in medical education

Desempeño en pruebas con consulta de preguntas abiertas en el curso de medicina

Paula Ferdinanda Conceição de Mascena Diniz Maia[1,2], José Miguel Gomes Moreira Pêgo[3].

RESUMO

Objetivo: Comparar o desempenho de estudantes de medicina em testes com consulta compostos por questões abertas (TCCQA), descrevendo variáveis relacionadas ao planejamento, execução, tempo de prova, tipo de resposta e domínio cognitivo das questões. **Métodos:** Estudo observacional retrospectivo, baseado na comparação entre o desempenho dos estudantes nos TCCQA aplicados sem (2023.1) e com supervisão e restrição de acesso às fontes de informação (2024.1), em um curso de medicina de instituição privada brasileira. Foram consideradas como variáveis exploratórias: tempo disponível para realização dos testes, tipo de resposta e domínio cognitivo das questões segundo a Taxonomia de Bloom. A análise estatística foi realizada no SPSS 26.0 e Excel 365, adotando-se nível de significância de 5% ($p \leq 0,05$). Utilizou-se o teste de Mann-Whitney para comparação entre os grupos. **Resultados:** Foram analisados 1.160 TCCQA-2023 e 1.146 TCCQA-2024. A maioria dos testes supervisionados apresentou médias de notas inferiores ($p < 0,05$), apesar do maior tempo de prova e da predominância de domínios cognitivos mais baixos. Predominaram respostas longas nos TCCQA supervisionados. **Conclusão:** Houve diferenças de desempenho entre os contextos avaliativos, indicando que as condições de aplicação influenciam a interpretação dos resultados e a confiabilidade do desempenho individual dos estudantes.

Palavras-chave: Avaliação educacional, Desempenho acadêmico, Faculdades de medicina.

ABSTRACT

Objective: To compare the performance of medical students in open-book tests composed of open-ended questions (OBTOEQ), describing variables related to planning, execution, test duration, response type, and the cognitive domain of the questions. **Methods:** This retrospective observational study compared student performance in OBTOEQ administered without supervision (2023.1) and with supervision and restricted access to information sources (2024.1) in a medical course at a private Brazilian institution. The following exploratory variables were considered: time available to complete the tests, type of response, and cognitive domain of the questions according to Bloom's Taxonomy. Statistical analysis was performed using SPSS 26.0 and Excel 365, adopting a significance level of 5% ($p \leq 0.05$). The Mann-Whitney test was used to compare the groups. **Results:** A total of 1,160 OBTOEQ-2023 and 1,146 OBTOEQ-2024 were analyzed.

[1] Faculdade Pernambucana de Saúde (FPS), Recife – PE.
[2] Universidade Federal de Pernambuco (UFPE), Recife – PE.
[3] Escola de Ciências da Saúde – Universidade do Minho, Região do Minho – Portugal

SUBMETIDO EM: 6/2026 | **ACEITO EM: 7/2026** | **PUBLICADO EM: 9/2026**

Most supervised tests showed lower mean scores ($p < 0.05$), despite the longer test duration and the predominance of lower cognitive domains. Long-answer responses predominated in the supervised OBTOEQ. **Conclusion:** Differences in performance were observed between the assessment contexts, indicating that the application conditions influence the interpretation of results and the reliability of students' individual performance.

Keywords: Educational assessment, Academic performance, Medical school.

RESUMEN

Objetivo: Comparar el desempeño de estudiantes de medicina en pruebas con consulta compuestas por preguntas abiertas (PCPA), describiendo variables relacionadas con la planificación, la ejecución, el tiempo de examen, el tipo de respuesta y el dominio cognitivo de las preguntas. **Métodos:** Estudio observacional retrospectivo, basado en la comparación del desempeño de los estudiantes en las PCPA aplicadas sin supervisión (2023.1) y con supervisión y restricción de acceso a las fuentes información (2024.1), en un curso de medicina de una institución privada brasileña. Se consideraron como variables exploratorias: el tiempo disponible para realizar las pruebas, el tipo de respuesta y el dominio cognitivo de las preguntas según la Taxonomía de Bloom. El análisis estadístico se realizó utilizando SPSS 26.0 y Excel 365, adoptándose un nivel de significancia del 5% ($p \leq 0,05$). Se utilizó la prueba de Mann-Whitney para comparar los grupos. **Resultados:** Se analizaron 1.160 PCPA-2023 y 1.146 PCPA-2024. La mayoría de las pruebas supervisadas presentó medias de calificaciones inferiores ($p < 0,05$), a pesar del mayor tiempo disponible para la prueba y del predominio de dominios cognitivos más bajos. Predominaron las respuestas largas en las PCPA supervisadas. **Conclusión:** Se observaron diferencias de desempeño entre los contextos evaluativos, lo que indica que las condiciones de aplicación influyen en la interpretación de los resultados y en la confiabilidad del desempeño individual de los estudiantes.

Palabras clave: Evaluación educacional, Rendimiento académico, Facultades de medicina.

INTRODUÇÃO

Avaliação, no contexto educacional, consiste em um conjunto de atividades planejadas e sistemáticas que evidenciam e subsidiam a análise do desempenho do estudante, integradas ao processo de aprendizagem (Elshama SS, 2020; Panúncio-Pinto MP e Troncon LEA, 2014). Para ser capaz de promover ajustes no processo de formação do estudante, a avaliação deve ser orientada por princípios como: pertinência na escolha dos métodos avaliativos, análise do desempenho baseada em referenciais previamente definidos, prática de feedback formativo e acompanhamento dos resultados de aprendizagem (Elshama SS, 2020).

Considerando as finalidades da avaliação, a avaliação formativa tem como propósito acompanhar e orientar o processo de aprendizagem, contribuindo com o estudante para desenvolver a motivação e a capacidade de autoavaliação. Já a avaliação somativa apresenta como intenção o julgamento do desempenho do estudante, produzindo resultados mensurados por conceitos ou notas que fundamentam decisões quanto à progressão acadêmica, como aprovação ou reprovação (Panúncio-Pinto MP e Troncon LEA, 2014).

Para definir o conteúdo e o formato de uma avaliação, é essencial identificar os objetivos educacionais em cada etapa da formação. Nesse contexto, a Taxonomia de Bloom revisada orienta a elaboração de instrumentos avaliativos ao organizar as habilidades cognitivas em níveis crescentes de complexidade, desde recordar e compreender até analisar, avaliar e criar (Panúncio-Pinto MP e Troncon LEA, 2014; Ferraz AP e BELHOT RV, 2010).

De maneira complementar, na educação em saúde, a pirâmide de Miller estrutura a avaliação das competências clínicas em quatro níveis — "saber", "saber como", "mostrar como" e "fazer" —, direcionando

a escolha de métodos avaliativos conforme o nível de competência a ser mensurado, especialmente no domínio cognitivo e na aplicação do conhecimento (DAWSON P, et al., 2024; SEFER CC, et al., 2019; CRUESS RL, et al., 2016; MILLER GE, 1990).

Entre os métodos utilizados para medir a aquisição de conhecimentos no domínio cognitivo, destacam-se os testes de múltipla escolha (TQME) e os testes compostos por questões abertas (TQA) (Panúncio-Pinto MP e Troncon LEA, 2014). Os TQME permitem avaliar a compreensão e o raciocínio através da seleção da melhor resposta, sendo amplamente utilizado devido à facilidade de aplicação e à correção rápida e objetiva (DAWSON P, et al., 2024; ELSHAMA SS, 2020; DURNING SJ, et al., 2016). Os TQA possibilitam respostas em formato de palavra, uma frase curta ou texto longo. Embora demandem mais tempo para correção e envolvam maior subjetividade no julgamento das respostas, os TQA apresentam maior proximidade com a vivência profissional (ELSHAMA SS, 2020).

Além da finalidade da avaliação e do tipo do método avaliativo, as condições de realização do teste também constituem um aspecto relevante no planejamento da avaliação. Os testes podem ser aplicados com ou sem a permissão de consulta a materiais de apoio, caracterizando-se, respectivamente, os testes com consulta (TCC) e os testes sem consulta (TSC) (DENEEN C, 2020). Quando elaborados para mensurar habilidades cognitivas mais complexas, os TCC tendem a desencorajar respostas baseadas na memorização e na aprendizagem factual, também são capazes de incentivar o raciocínio crítico; além de estimular a capacidade de interpretação e de síntese e a utilização de referências relevantes e validadas, competências indispensáveis à futura prática profissional (ER HM, et al., 2023; BLADT F, et al., 2022; SARKAR S, et al., 2019).

Durante a pandemia da COVID-19, a educação médica enfrentou desafios significativos relacionados à continuidade do processo de ensino-aprendizagem e à realização das avaliações em contexto de suspensão das atividades presenciais (ER HM, et al., 2023; BLADT F, et al., 2022; SARKAR S, et al., 2019). Nesse período, a maioria das escolas médicas adaptou suas avaliações para o formato remoto (DAWSON P, et al., 2024; NG CK, 2020; SARKAR S, et al., 2019), sendo os TCC aplicados online considerados uma alternativa viável para mensurar o desempenho discente (ER HM, et al., 2023). Entretanto, a ausência de supervisão remota e de controle das fontes de consulta utilizadas durante a realização dos testes pode ter comprometido a integridade do processo avaliativo, tornando necessária a validação dos resultados obtidos (VAZQUEZ JJ, et al., 2021; SARKAR S, et al., 2019).

Na avaliação do estudante na área da saúde, a qualidade dos resultados produzidos depende de múltiplos atributos, incluindo a definição clara do que se pretende avaliar, o nível cognitivo das questões, o tipo de resposta exigida, o tempo concedido para realização do teste, os critérios de correção previamente estabelecidos, a disponibilidade de recursos e o reconhecimento da avaliação como adequada e justa pelos participantes envolvidos no processo. Esses elementos estão diretamente relacionados aos critérios de validade, confiabilidade, viabilidade e aceitabilidade do processo avaliativo (MUKURUNGE E, et al., 2024; ABUZIED AIH e NABAG WO, 2023).

No contexto das Instituições de Ensino Superior do Brasil, os TCC compostos por questões abertas (TCCQA), aplicados online e sem nenhum tipo de supervisão, constituíram uma estratégia avaliativa utilizada com maior frequência na pandemia. Após o retorno das atividades presenciais, algumas instituições mantiveram essa estratégia avaliativa, porém com modificações nas condições de aplicação. Os TCCQA passaram a ser realizados de forma presencial, sob supervisão docente e com controle total das fontes de consulta, mantendo a finalidade somativa da avaliação.

Apesar do crescente uso dos TCCQA na educação médica, as pesquisas têm se concentrado em avaliar a percepção dos docentes e dos estudantes e em comparar TCCQA com testes sem consulta, apresentando menor enfoque na análise do desempenho discente em TCCQA aplicados sob diferentes condições de realização. Variáveis relacionadas ao planejamento e à execução da avaliação, como supervisão, tempo de prova, tipo de resposta e critérios de correção são descritos de forma isolada, sem análise da possível influência desses critérios com as pontuações obtidas.

Considerando que o desempenho discente pode ser influenciado tanto pelas condições de realização da avaliação quanto por atributos do instrumento avaliativo; torna-se relevante descrever, de forma sistemática, as variáveis relacionadas ao planejamento e à execução dos TCCQA aplicados em diferentes contextos de realização, em cenários acadêmicos equivalentes.

A proposta desse estudo foi comparar o desempenho dos estudantes de medicina nos TCCQA aplicados em diferentes contextos avaliativos, descrevendo variáveis relacionadas ao planejamento, execução, tempo de prova, tipo de resposta e domínio cognitivo das questões.

MÉTODOS

Desenho do estudo e objetivos

Trata-se de um estudo observacional retrospectivo, de natureza comparativa cujo objetivo foi analisar o desempenho discente nos TCCQA aplicados em diferentes contextos avaliativos. A análise realizada se baseou na comparação entre o desempenho dos estudantes nos TCCQA, aplicados sem supervisão (2023.1) e com supervisão e restrição de acesso às fontes de informação (2024.1), no curso de medicina de uma instituição privada do Brasil.

Como variáveis exploratórias, foram consideradas as seguintes características dos TCCQA: o tempo disponível para realização dos testes, o tipo de resposta e o domínio cognitivo da questão de acordo com a classificação da Taxonomia de Bloom.

População de estudo

A população do estudo foi composta por dados acadêmicos secundários referentes ao desempenho dos estudantes, regularmente matriculados, do curso de Medicina do 1º ao 8º período da Faculdade Pernambucana de Saúde (FPS). Foram incluídos no estudo os TCCQA do 1º ao 4º ano (do 1º ao 8º período) de Medicina dos módulos 1 e 3 realizados no primeiro semestre de 2023 e 2024.

Foram excluídos do estudo: os testes de recuperação, de segunda chamada e os testes finais; os testes dos módulos cujos conteúdos não apresentaram equivalência temática entre os períodos analisados e testes elaborados por coordenações distintas entre os semestres. A unidade de análise foi a pontuação alcançada pelo estudante nos testes.

Dos oito períodos do curso de medicina da Faculdade Pernambucana de Saúde (FPS), foram elegíveis os períodos do 2º ao 8º para o módulo 1 e do 4º ao 8º para o módulo 3. O motivo para exclusão do 1º período nos módulos 1 e 3 foi a elaboração dos testes por coordenadores de tutoria distintos e a temática dos módulos diferentes entre 2023 e 2024. Para o 3º período, houve modificação da matriz curricular com consequente diferença entre os temas abordados no primeiro semestre de 2023 e de 2024 no módulo 3, mantendo o tema apenas para o módulo 1.

A temática abordada em cada período para o módulo 1 foi: 2o período: Farmacocinética e farmacodinâmica; 3º: Estudo da função dos órgãos e sistemas III; 4º: Estudos dos desequilíbrios na infância III; 5º: Estudo do desenvolvimento psíquico e das relações humanas; 6º: Transtornos psiquiátricos e comportamentais; 7º: Segurança do paciente; 8º: Estudo dos Desequilíbrios no adulto: anemia e sangramento. Para o módulo 3: 4º período: Estudo dos desequilíbrios na gestação; 5º: Ética nas práticas de saúde; 6º: Diversidade, desigualdades e interculturalidade; 7º: Desequilíbrios no trauma cirúrgico e cirurgia minimamente invasiva; 8º: Saúde Global e Ética Planetária.

Local de realização do estudo e contexto educacional

O estudo foi conduzido no curso de Medicina da FPS, localizada em Recife - Pernambuco - Brasil, que adota um currículo orientado pelo desenvolvimento de competências e pela formação integral do estudante. O processo de aprendizagem ocorre por meio da Aprendizagem Baseada em Problemas (ABP), metodologia ativa centrada no estudante, baseada nos princípios do construtivismo, da colaboração, da contextualização e da autonomia.

As atividades teóricas são organizadas em módulos integrados e sequenciais, planejados conforme a progressão das competências previamente desenvolvidas pelo estudante ao longo do curso.

Elaboração dos TCCQA

Os TCCQA foram elaborados pela equipe de professores e coordenador de cada período. Cada equipe apresentava, em média, 6 professores e 1 coordenador. As questões foram construídas com base nos objetivos de aprendizagem previamente definidos para os casos (problemas) de cada módulo e discutidas coletivamente quanto à pertinência do conteúdo e ao grau de complexidade cognitiva, seguindo blueprint avaliativo institucional.

Os TCCQA apresentavam uma a duas questões abertas estruturadas em subitens de resposta classificadas como resposta curta (1 linha) ou resposta longa (2 ou mais linhas). As questões foram corrigidas pelos docentes que elaboraram os testes com base nos espelhos de respostas definidos previamente. A pontuação variou de 0 a 10 e apresentaram caráter somativo.

Análise das questões com consulta e coleta de dados

As questões dos módulos 1 e 3 de cada período selecionado para o estudo foram enviadas à pesquisadora, pela secretaria da faculdade, de forma randomizada.

A classificação das questões quanto ao domínio cognitivo foi realizada com base na Taxonomia de Bloom Revisada[4], considerando o processo cognitivo predominante exigido para a resolução de cada item dissertativo. Foram adotadas definições operacionais prévias para os níveis cognitivos (lembrar, compreender, aplicar, analisar, avaliar e criar), priorizando a demanda cognitiva principal da questão.

A categorização foi conduzida pela pesquisadora principal, docente com experiência em Aprendizagem Baseada em Problemas (ABP) e elaboração de instrumentos avaliativos, reconhecendo-se a possibilidade de subjetividade inerente ao processo de classificação.

Os espelhos dos TCCQA, contendo as respostas corretas, foram solicitados também à secretaria e, nos espelhos das respostas, estava definido o tempo disponibilizado para realização de cada teste. O tempo disponibilizado em cada teste era de conhecimento do estudante, em momento anterior à realização dos testes.

As notas foram solicitadas ao setor de tecnologia e de registro de notas da faculdade.

Condições de aplicação dos testes com consulta compostos por questões abertas

Em 2023.1 e 2024.1, os TCCQA foram aplicados em plataforma online. Entretanto, diferiram quanto às condições de aplicação. Em 2023.1, a realização foi remota, sem nenhum tipo de supervisão e sem controle de acesso às fontes de informação, possibilitando o uso livre das anotações pessoais e de recursos online não estabelecidos como fontes oficiais de consulta. Em 2024.1, apesar de ter sido mantido o formato digital, a aplicação ocorreu presencialmente nas dependências da instituição, em equipamentos próprios da faculdade, com acesso apenas à biblioteca digital da instituição e sob supervisão docente.

Análise dos dados

Foram utilizados os softwares SPSS versão 26.0 (Statistical Package for the Social Sciences) para Windows e Microsoft Excel 365. Foram considerados apenas os dados válidos, excluindo-se os registros com informações ausentes. As variáveis categóricas foram apresentadas por frequências absoluta e relativa, e as variáveis numéricas, por medidas de tendência central e dispersão. A normalidade das distribuições foi avaliada pelo teste de Shapiro-Wilk. Como as variáveis quantitativas não apresentaram distribuição normal, utilizou-se o teste de Mann-Whitney para comparar as notas individuais entre grupos independentes de estudantes de 2023 e 2024. Em cada comparação por período e módulo, cada estudante contribuiu com uma única nota, sem sobreposição de participantes entre os grupos. Adotou-se nível de significância de 5% ($p \leq 0,05$). Considerando o caráter exploratório das comparações realizadas entre períodos e módulos, não foi aplicada correção para múltiplos testes, e os resultados foram interpretados como exploratórios.

Aspectos éticos

O estudo foi aprovado pelo Comitê de Ética e Pesquisa da FPS com CAAE de número: 82988524.2.0000.5569. Houve dispensa do Termo de Consentimento Livre e Esclarecido, por se tratar de pesquisa retrospectiva baseada exclusivamente em dados acadêmicos secundários, previamente existentes e anonimizados.

RESULTADOS

A FPS, no primeiro semestre de 2023, apresentava um total de 818 estudantes e, no primeiro semestre de 2024, 801 estudantes cursando medicina do 1º ao 8º período, todos submetidos às avaliações propostas pela faculdade. Foram avaliados em 2023, 691 (84,5%) TCCQA no módulo 1 e 469 (57,3%) no módulo 3. Em 2024, foram 667 (83,3%) TCCQA no módulo 1 e 479 (59,8%) no módulo 3.

Para facilitar a compreensão dos principais achados, apresenta-se inicialmente uma visão geral dos resultados obtidos, antes da descrição detalhada das análises por módulo e por período (**Tabela 1**).

Tabela 1 - Síntese dos resultados da comparação entre os TCCQA aplicados em 2023 e 2024 nos módulos 1 e 3 do curso de Medicina.

Módulo	TCCQA 2023	TCCQA 2024	Tendência geral	Principal achado
Módulo 1	691	667	Predomínio de médias inferiores em 2024	Diferenças estatisticamente significativas foram observadas na maior parte das comparações entre os anos.
Módulo 3	469	479	Seis das sete turmas apresentaram médias inferiores em 2024	Predomínio de menores médias em 2024 nas comparações entre os contextos avaliativos.

Fonte: Maia PFCMD e Pêgo JMGM, 2026.

Análise comparativa do desempenho dos estudantes nos TCCQA sem supervisão e com supervisão do módulo 1, associando ao tempo de prova e ao tipo de resposta esperada

Foram analisados os TCCQA com e sem supervisão do total de 10 turmas e 7 períodos no módulo 1. Observaram-se que as médias de desempenho dos estudantes foram, de forma consistente, inferiores nos testes realizados sob supervisão em 2024, em comparação com os realizados sem supervisão em 2023. A comparação global entre os grupos evidenciou diferença estatisticamente significativa ($p<0,05$).

Em análises estratificadas por turma, a maioria apresentou média inferior em 2024. Entretanto, essas análises devem ser interpretadas com cautela devido ao caráter exploratório e ao risco de erro associado a múltiplas comparações.

Em relação às características dos instrumentos avaliativos, observou-se predominância de maior tempo disponibilizado para realização das provas em 2024 em todos os períodos analisados, bem como maior frequência de questões com respostas longas em 80% das turmas (**Tabela 2**).

Tabela 2 - Comparação das médias das notas dos TCCQA (módulo 1) de 2023 em relação a 2024. Os dados são apresentados como média ± desvio padrão (mediana; tempo de prova em minutos; tipo de resposta).

Módulo 1				
Período	Turma	2023 (sem supervisão)	2024 (com supervisão)	p-valor
2º	Única	9,68 ± 1,38 (10,0; 10 min; C[a])	8,64 ± 0,99 (8,75; 60 min; L)	**<0,001**
3º	1	9,94 ± 0,30 (10,0; 20 min; C)	8,40 ± 2,49 (9,50; 60 min; L)	**<0,001**
	2	9,95 ± 0,20 (10,0; 20 min; C)	9,42 ± 1,03 (10,0; 60 min; C)	**0,001**
4º	Única	8,50 ± 1,87 (10,0; 25 min; L[b])	8,24 ± 1,60 (8,75; 100 min; C/L)	0,089
5º	1	10,00 ± 0,00 (10,0; 9 min; C)	9,56 ± 1,16 (10,0; 18 min; C)	**0,003**
	2	9,87 ± 0,74 (10,0; 9 min; C)	8,43 ± 1,84 (9,00; 60 min; C/L)	**<0,001**
6º	Única	9,61 ± 1,31 (10,0; 15 min; C)	8,89 ± 1,18 (8,40; 80 min; C/L)	**<0,001**
7º	1	8,41 ± 1,06 (8,50; 10 min; C)	6,66 ± 1,76 (7,00; 60 min; L)	**<0,001**
	2	7,92 ± 1,82 (8,50; 10 min; C)	7,83 ± 2,50 (10,0; 60 min; C/L)	0,823
8º	Única	8,96 ± 1,11 (9,00; 25 min; C/L[c])	7,72 ± 2,34 (8,50; 75 min; C/L)	**0,002**

*Legendas: a) C = respostas curtas; b) L = respostas longas; c) C/L = Curtas/Longas
Fonte: Maia PFCMD e Pêgo JMGM, 2026.

Análise comparativa do desempenho dos estudantes nos TCCQA sem supervisão e com supervisão do módulo 3, associando ao tempo de prova e tipo de resposta esperada

Em relação ao módulo 3, das 7 turmas avaliadas, 6 (85,7%) apresentaram média das notas dos TCCQA supervisionados menor em relação aos testes com consulta sem supervisão com diferença estatisticamente significativa entre os contextos avaliados (p<0,05). O tipo de resposta esperada predominante nas questões de cada turma no ano de 2023 e no ano de 2024 foi a resposta longa com 71,4% (5 turmas) e 85,7% (6 turmas), respectivamente. O tempo de prova foi maior em todas as turmas do ano 2024 (**Tabela 3**).

Tabela 3 - Comparação das médias das notas dos TCCQA (módulo 3) de 2023 em relação a 2024. Os dados são apresentados como média ± desvio padrão (mediana; tempo de prova em minutos; tipo de resposta).

Módulo 3				
Período	Turma	2023 (sem supervisão)	2024 (com supervisão)	p-valor
4º	Única	9,26 ± 1,05 (10,0; 30 min; L[b])	5,99 ± 2,12 (5,67; 90 min; L)	**<0,001**
5º	1	9,65 ± 0,87 (10,0; 9 min; L)	7,16 ± 2,45 (5,00; 60 min; L)	**<0,001**
	2	9,96 ± 0,28 (10,0; 9 min; L)	9,41 ± 1,28 (10,0; 60 min; L)	**0,002**
6º	Única	9,93 ± 0,49 (10,0; 10 min; C[a])	8,66 ± 2,13 (10,0; 45 min; C)	**<0,001**
7º	1	9,32 ± 0,84 (9,50; 10 min; C/L[c])	8,92 ± 1,76 (10,0; 40 min; C/L)	0,326
	2	9,74 ± 0,88 (10,0; 10 min; C/L)	9,56 ± 0,83 (10,0; 40 min; C/L)	**0,018**
8º	Única	8,81 ± 1,09 (9,00; 15 min; C)	9,29 ± 1,29 (10,0; 30 min; L)	**0,001**

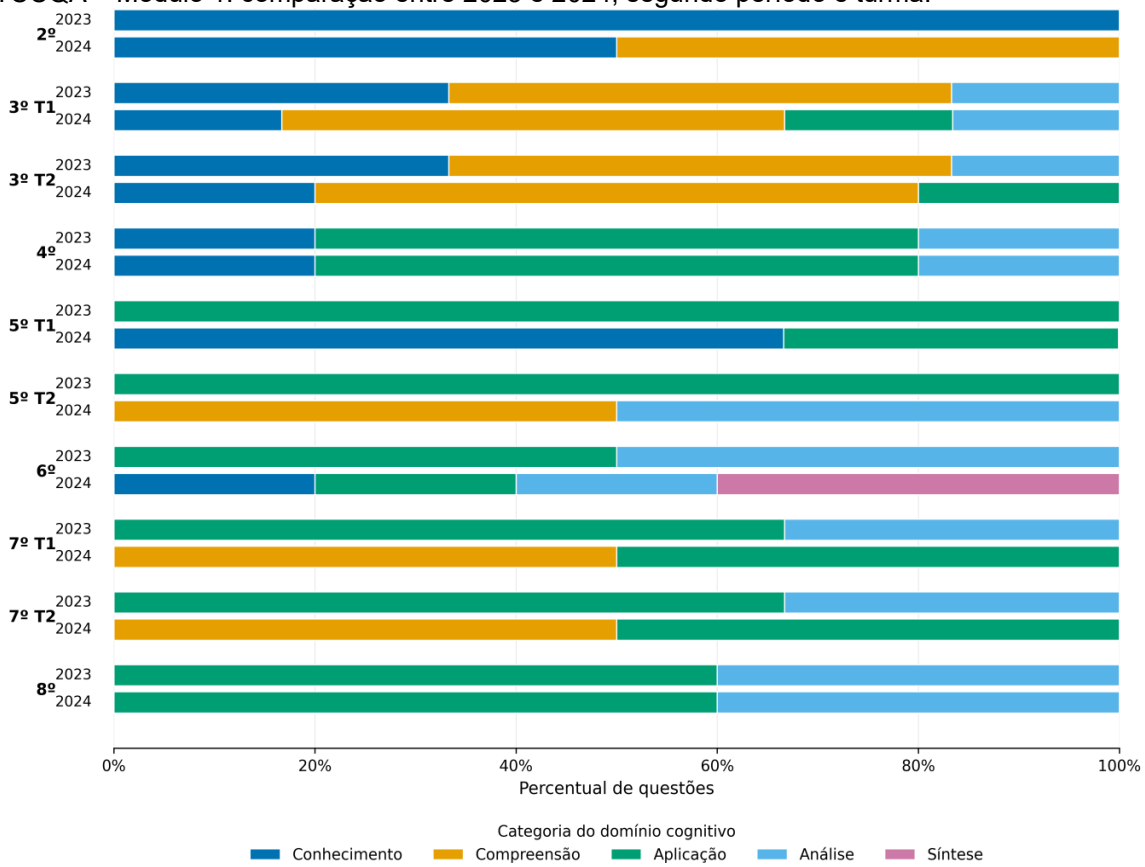
*Legendas: a) C = respostas curtas; b) L = respostas longas; c) C/L = Curtas/Longas

Fonte: Maia PFCMD e Pêgo JMGM, 2026.

Análise comparativa do desempenho dos estudantes nos TCCQA sem supervisão e com supervisão do módulo 1, associando ao domínio cognitivo da Taxonomia de Bloom

Foram analisadas todas as questões de provas dos TCCQA do módulo 1 de todos os períodos elegíveis de 2023 e 2024, totalizando 27 questões. Na análise das questões do módulo 1, foi descrito o percentual da categoria do domínio cognitivo das questões de cada período. Para os períodos com diferença significativa no desempenho dos estudantes, observou-se um predomínio do domínio cognitivo de aplicação em 2023 e dos domínios de compreensão em 2024, ano em que as médias das notas foram menores em relação a 2023. De forma geral, considerando todos os períodos analisados, houve maior percentual de questões formuladas de acordo com os domínios cognitivos de conhecimento, compreensão e aplicação no primeiro semestre de 2023 e 2024 (**Gráfico 1**) e (**Tabela 4**).

Gráfico 1 - Distribuição das categorias cognitivas da Taxonomia de Bloom nas questões do TCCQA – Módulo 1: comparação entre 2023 e 2024, segundo período e turma.



Nota. Os valores exatos estão apresentados na Tabela 4. A categoria avaliação não foi incluída porque não houve questões classificadas nesse domínio.
Fonte: Maia PFCMD e Pêgo JMGM, 2026.

Tabela 4 - Distribuição das questões do TCCQA do Módulo 1 segundo as categorias do domínio cognitivo da Taxonomia de Bloom, comparando 2023 e 2024.

Módulo 1				
Período	Turma	2023 (%)	2024 (%)	p-valor
2º	Única	**C1[a]: 100**	**C1: 50; C2: 50**	**<0,001**
3º	1	C1: 33,3; **C2[b]: 50**; C4[d]: 16,7	C1: 16,7; **C2: 50**; C3: 16,7; C4: 16,7	**<0,001**
	2	C1: 33,3; **C2: 50**; C4: 16,7	C1: 20; **C2: 60**; C3: 20	**0,001**
4º	Única	C1: 20; C3[c]: 60; C4: 20	C3: 42,9; C4: 42,9; C5[e]: 14,3	0,089
5º	1	**C3: 100**	**C1: 66,6**; C3: 33,3	**0,003**
	2	**C3: 100**	**C2: 50; C4: 50**	**<0,001**
6º	Única	**C3: 50; C4: 50**	C1: 20; C3: 20; C4: 20; **C5: 40**	**<0,001**
7º	1	**C3: 66,7**; C4: 33,3	**C2: 50; C3: 50**	**<0,001**
	2	C3: 66,7; C4: 33,3	C2: 50; C3: 50	0,823
8º	Única	**C3: 60**; C4: 40	**C3: 60**; C4: 40	**0,002**

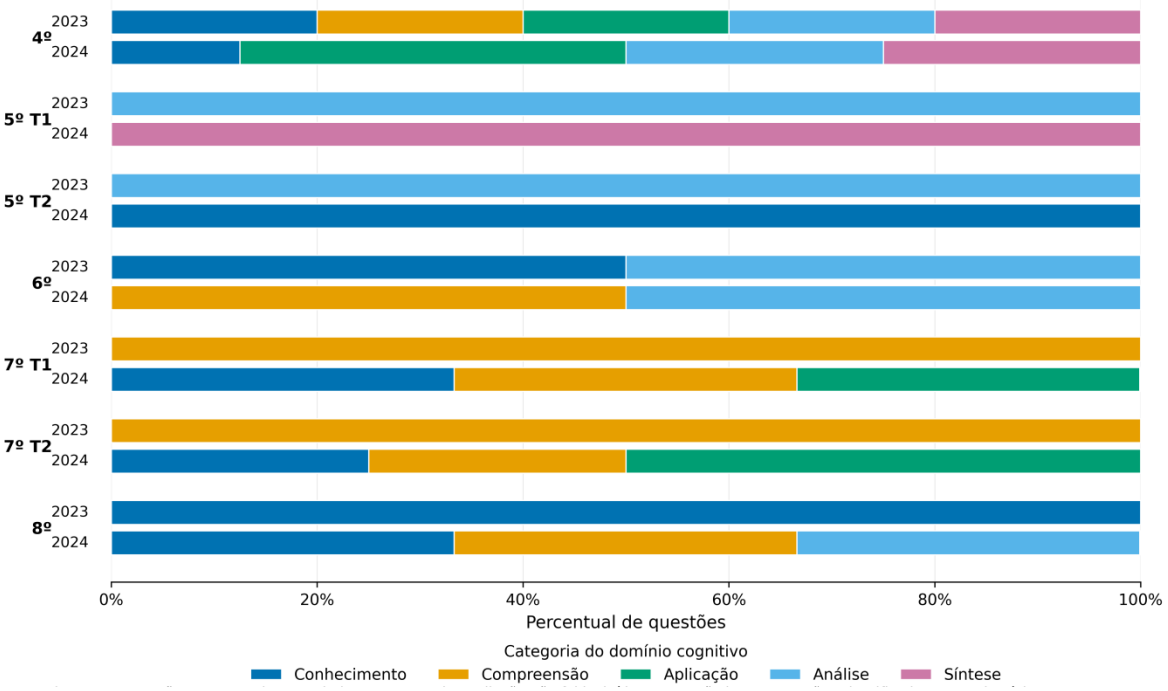
*Legenda: Domínio cognitivo da Taxonomia de Bloom: C1[a]. Conhecimento, C2[b]. Compreensão, C3[c]. Aplicação, C4[d]. Análise, C5[e]. Síntese e C6[f]. Avaliação. O domínio cognitivo de avaliação não está na tabela, pois não houve questão classificada com este domínio cognitivo.
Fonte: Maia PFCMD e Pêgo JMGM, 2026.

Análise comparativa do desempenho dos estudantes nos TCCQA sem supervisão e com supervisão do módulo 3, associando ao domínio cognitivo da Taxonomia de Bloom

Para o módulo 3, foram analisadas também todas as questões dos períodos elegíveis de 2023.1 e 2024.1, totalizando 14 questões. Foi descrito o percentual da categoria do domínio cognitivo das questões

de cada período. Nos períodos com diferença significativa no desempenho dos estudantes, houve maior percentual de questões elaboradas com os domínios cognitivos de análise em 2023 e de conhecimento em 2024, ano em que a média das notas foi menor (**Gráfico 2**) e (**Tabela 5**).

Gráfico 2 - Distribuição das categorias cognitivas da Taxonomia de Bloom nas questões do TCCQA – Módulo 3: comparação entre 2023 e 2024, segundo período e turma.



Nota. Os valores exatos estão apresentados na **Tabela 5**. A categoria avaliação não foi incluída porque não houve questões classificadas nesse domínio.
Fonte: Maia PFCMD e Pêgo JMGM, 2026.

Tabela 5 - Distribuição das questões do TCCQA do Módulo 3 segundo as categorias do domínio cognitivo da Taxonomia de Bloom, comparando 2023 e 2024.

Módulo 3				
Período	Turma	2023 (%)	2024 (%)	p-valor
4º	Única	C1[a]: 20; C2[b]: 20; C3[c]: 20; C4[d]: 20; **C5:20**	C1: 12,5; **C3: 37,5**; C4: 25; C5[e]: 25	**<0,001**
5º	1	**C4: 100**	**C5: 100**	**<0,001**
	2	**C4: 100**	**C1: 100**	**0,002**
6º	Única	**C1: 50; C4: 50**	**C2: 50; C4: 50**	**<0,001**
7º	1	C2: 100	C1: 33,3; C2: 33,3; C3: 33,3	0,326
	2	**C2: 100**	C1: 25; C2: 25; **C3: 50**	**0,018**
8º	Única	**C1: 100**	C1: 33,3; C2: 33,3; **C4: 33,3**	**0,001**

*Legenda: Domínio cognitivo da Taxonomia de Bloom: C1[a]. Conhecimento, C2[b]. Compreensão, C3[c]. Aplicação, C4[d]. Análise, C5[e]. Síntese e C6[f]. Avaliação. O domínio cognitivo de avaliação não está na tabela, pois não houve questão classificada com este domínio cognitivo.
Fonte: Maia PFCMD e Pêgo JMGM, 2026.

DISCUSSÃO

TCCQA em diferentes contextos avaliativos: implicações para validade

Nos módulos 1 e 3, a maioria das turmas dos períodos analisados apresentou um desempenho inferior nos TCCQA supervisionados em relação aos TCCQA não supervisionados, mesmo com a disponibilidade de um tempo maior de prova. Os TCCQA supervisionados apresentaram predominância de questões de respostas longas. Embora essa diferença não permita inferir causalidade direta, devido às variáveis simultâneas presentes no contexto avaliativo, os achados sugerem que ambientes não supervisionados,

especialmente quando associados ao acesso irrestrito a fontes externas de informações, podem facilitar práticas colaborativas indevidas, indicando resultados de baixa confiabilidade em relação ao desempenho individual do estudante (AHMED Y, et al., 2024; NG CK, 2020; DURNING SJ, et al., 2016). Outras pesquisas com avaliações remotas apontaram a ausência de controle ao acesso às fontes de consulta como um fator facilitador de comportamentos inadequados, comprometendo os valores éticos de honestidade e responsabilidade imprescindíveis para a realização individual das avaliações acadêmicas, com impacto direto na validade das medidas de desempenho (ALGUACIL M, et al., 2024; NEWTON PM e ESSEX K, et al. 2024; SUSNJAK T e MCINTOSH TR, 2024). A supervisão foi uma medida apontada por estudantes para garantir resultados mais fidedignos (ELSALEM L, et al., 2021; BEARMAN M, et al., 2020; GAMAGE KA, et al., 2020; GUANGUL FM, et al., 2020; JERVIS CG e BROWN LR, 2020; NG CK, 2020; TUAH NA e NAING L, 2020).

Dessa forma, o melhor desempenho em TCCQA observado em contextos sem supervisão não necessariamente reflete maior domínio cognitivo. Os resultados alcançados podem ser interpretados como decorrentes, por exemplo, de uma maior capacidade de mobilização de recursos externos desfavoráveis ao propósito da avaliação.

TCCQA e tempo de prova

Embora a pesquisa não apresente uma análise estatística multivariada, foi observada uma tendência de associação entre maior disponibilidade de tempo de prova e menor desempenho nos TCCQA com supervisão.

Um estudo australiano avaliou 2 formatos de TCC. O primeiro teste com duração de 50 minutos e o segundo de 4 horas e 24 minutos. A avaliação com menor duração apresentou notas mais baixas e indicou menor possibilidade de colaboração entre pares durante a realização (NG CK, 2020).

Em contradição, os TCCQA deste estudo com menor tempo para realização estiveram associados a desempenho superior dos estudantes. Considerando a relevância da supervisão para assegurar a realização individual dos testes, é possível sugerir que a ausência de supervisão nos testes com menor tempo disponível pode ter colaborado para ampliar a possibilidade de colaboração entre os estudantes, incluindo o uso de aplicativos de comunicação.

TCCQA e tipo de resposta

O tipo de resposta esperada nas avaliações com consulta é um importante critério de qualidade a ser considerado. As questões com respostas longas inclinam-se para avaliar competências superiores da Taxonomia de Bloom, demandando um processamento profundo do pensamento crítico (ELSAMANOUDY A, et al., 2024; MEE J, et al., 2024; ELDAKHAKHNY B e ELSAMANOUDY AZ, 2023).

No referido estudo, houve maior prevalência de questões com respostas longas nos testes com média de notas menores, apesar da predominância de domínios cognitivos mais baixos. Entretanto, os testes com menor desempenho foram aplicados com supervisão e não foi realizada uma análise isolada para atribuir ao tipo de resposta o desempenho inferior do estudante, viabilizando apenas a hipótese de haver uma possível associação.

TCCQA e Taxonomia de Bloom

A análise dos domínios cognitivos evidenciou predominância de categorias mais complexas nos TCCQA sem supervisão, contrastando com um melhor desempenho. Esse resultado permitiu deduzir que a elaboração de questões utilizando a Taxonomia de Bloom mais avançada, de forma isolada, não garantiu maior qualidade avaliativa.

Embora as categorias cognitivas de aplicação, análise, síntese e avaliação sejam capazes de exigir um conhecimento cognitivo superior, promovendo maior raciocínio crítico (DAVIES DJ, et al., 2022; BULJAN I, et al., 2021), nesta pesquisa, seu potencial discriminativo pode ter sido reduzido devido à ausência de supervisão e ao livre acesso as mais variadas fontes de informações.

Ética na avaliação e implicações na formação profissional

A possibilidade de colaboração indevida em avaliações não supervisionados pode constituir um problema ético e, consequentemente, de formação profissional, além de um problema operacional (SABRINA F, et al., 2022; HOLDEN OL, et al., 2021). A avaliação não tem a função de apenas medir o desempenho do estudante, mas exerce também um papel normativo ao sinalizar comportamentos aceitáveis ou não no processo de formação (GARG M e GAEL A; 2022). Ambientes avaliativos permissivos podem, indiretamente, contribuir para a normalização de práticas que fragilizam a integridade acadêmica, especialmente quando não há mecanismos claros de regulação e responsabilização (ALGUACIL M, et al., 2024; GARG M e GAEL A; 2022).

Na formação médica, a dissociação entre desempenho e responsabilidade individual pode comprometer o desenvolvimento do profissionalismo, uma vez que comportamentos inadequados e tolerados durante a formação tendem a ser reproduzidos na prática clínica (PIENAAR L, et al., 2025; MAK-VAN DER VOSSEN MC, et al., 2019). A supervisão nas avaliações pode atuar como um mecanismo de regulação do ambiente avaliativo, ao reduzir práticas incompatíveis com avaliações individuais e explicitar limites institucionais com relação ao estabelecido como aceitável.

Limitações do estudo

Este estudo apresenta limitações que devem ser consideradas na interpretação dos resultados. O delineamento observacional retrospectivo não permite atribuir as diferenças de desempenho observadas exclusivamente às condições de aplicação dos testes, restringindo as conclusões à comparação entre contextos avaliativos distintos. Além disso, embora tenham sido adotados critérios para selecionar avaliações com conteúdos equivalentes, não foi possível controlar completamente todas as diferenças existentes entre os anos analisados, incluindo aspectos relacionados ao ambiente de aplicação, ao acesso às fontes de informação, às características dos instrumentos avaliativos e a outros fatores potencialmente associados ao desempenho discente. A classificação dos domínios cognitivos das questões segundo a Taxonomia de Bloom Revisada, foi realizada por uma única avaliadora, sem avaliação independente por um segundo avaliador ou análise de concordância, o que pode ter introduzido subjetividade no processo de classificação. Por fim, foram realizadas múltiplas comparações entre turmas e módulos sem aplicação de métodos de correção para controle do erro do tipo I, aumentando a possibilidade de resultados estatisticamente significativos ocorrerem ao acaso. Assim, esses achados devem ser interpretados com cautela e considerados exploratórios.

Apesar dessas limitações, o estudo reúne um número expressivo de avaliações aplicadas em diferentes módulos do curso de Medicina e contribui para ampliar o conhecimento sobre testes com consulta compostos por questões abertas, oferecendo evidências que podem subsidiar o aperfeiçoamento de práticas avaliativas em diferentes contextos da educação médica.

CONCLUSÃO

Conclui-se que houve diferenças de desempenho entre os TCCQA aplicados sem supervisão e aqueles realizados com supervisão e restrição de acesso à informação. Os resultados sugerem que as condições de aplicação podem influenciar a interpretação das notas e a confiabilidade do desempenho individual dos estudantes. Embora a elaboração de questões abertas com diferentes domínios cognitivos permaneça relevante para a avaliação em educação médica, os achados indicam que o contexto de aplicação, a supervisão, o controle das fontes de consulta, o tempo de prova e o tipo de resposta devem ser considerados no planejamento avaliativo. Por se tratar de estudo observacional retrospectivo, não é possível estabelecer causalidade direta, sendo recomendados novos estudos com controle mais rigoroso das variáveis avaliativas.

AGRADECIMENTOS

À Faculdade Pernambucana de Saúde por possibilitar o desenvolvimento da pesquisa.

REFERÊNCIAS

1. ABUZIED AIH, NABAG WO. Structured viva validity, reliability, and acceptability as an assessment tool in health professions education: a systematic review and meta-analysis. BMC Med Educ., 2023; 23(1): 531.
2. AHMED Y, et al. Evolving assessment in medical education: exploring the role of open-book examinations. J Adv Med Educ Prof., 2024; 12(4): 215-25.
3. ALGUACIL M, et al. Academic dishonesty and monitoring in online exams: a randomized field experiment. J Comput High Educ., 2024; 36: 835–51.
4. BEARMAN M, et al. Ensuring academic integrity and assessment security with redesigned online delivery. Melbourne: Deakin University; 2020. Disponível em: http://dteach.deakin.edu.au/2020/03/23/academic-integrity-online/. Acessado em: 6 de junho de 2026.
5. BLADT F, et al. Medical students' perception of changes in assessments implemented during the COVID-19 pandemic. BMC Med Educ., 2022; 22(1): 844.
6. BULJAN I, et al. Cognitive levels in testing knowledge in evidence-based medicine: a cross-sectional study. BMC Med Educ., 2021; 21(1): 25.
7. CRUESS RL, et al. Amending Miller's Pyramid to include professional identity formation. Acad Med., 2016; 91(2): 180–5.
8. DAVIES DJ, et al. Assessment of factual recall and higher-order cognitive domains in an open-book medical school examination. Adv Health Sci Educ Theory Pract., 2022; 27(1):147-65.
9. DAWSON P, et al. Beyond open book versus closed book: a taxonomy of restrictions in online examinations. Assess Eval High Edu., 2024; 49(2): 262–74.
10. DENEEN C. Assessment considerations in moving from closed-book to open-book exams. Melbourne: Melbourne Centre for the Study of Higher Education, The University of Melbourne; 2020 May. Disponível em: https://melbourne-cshe.unimelb.edu.au/__data/assets/pdf_file/0010/3341944/closed-book-to-open-book-exam_final.pdf. Acessado em: 6 de junho de 2026.
11. DURNING SJ, et al. Comparing open-book and closed-book examinations: a systematic review. Acad Med., 2016; 91(4): 583-99.
12. ELDAKHAKHNY B, ELSAMANOUDY AZ. Discrimination power of short essay questions versus multiple choice questions as an assessment tool in clinical biochemistry. Cureus., 2023; 15(2): e35427.
13. ELSALEM L, et al. Remote E-exams during COVID-19 pandemic: a cross-sectional study of students' preferences and academic dishonesty in faculties of medical sciences. Ann Med Sur., 2021; 62: 326-33.
14. ELSAMANOUDY A, et al. Evaluation of modified essay questions (MEQs) as an assessment tool in third-year medical students' modular summative assessment. BMC Med Educ., 2024; 24:1445.
15. ELSHAMA SS. How to use and apply assessment tools in medical education? Iberoam J Med., 2020; 2(4): 351-9.
16. ER HM, et al. Remote online open book examinations: through the lenses of faculty and students in health professions programmes. BMC Med Educ., 2023; 23(1): 397.
17. FERRAZ AP, BELHOT RV. Taxonomia de Bloom: revisão teórica e apresentação das adequações do instrumento para definição de objetivos instrucionais. Gest Prod., 2010; 17(2): 421-31.
18. GAMAGE KA, et al. Online delivery and assessment during COVID-19: safeguarding academic integrity. Educ Sci., 2020; 10(11): 301.
19. GARG M, GAEL A. A systematic literature review on online assessment security: Current challenges and integrity strategies. Comput Secur., 2022; 113: 102544.
20. GUANGUL FM, et al. Challenges of remote assessment in higher education in the context of COVID-19: a case study of Middle East College. Educ Assess Eval Acc., 2020; 32(4): 519-35.
21. HOLDEN OL, et al. Academic integrity in online assessment: a research review. Front Educ., 202; 6: 639814.
22. JERVIS CG, BROWN LR. The prospects of sitting 'end of year' open book exams in the light of COVID-19: a medical student's perspective. Med Teach., 2020; 42(7): 830–31.
23. Mak-VAN DER VOSSEN MC, et al. A road map for attending to medical students' professionalism lapses. Acad Med., 2019; 94(4): 570-8.

24. MATHIESON G, et al. Could the future of medical school examinations be open-book: a medical student's perspective? Med Educ Online., 2020; 25(1): 1787308.
25. MEE J, et al. An experimental comparison of multiple-choice and short-answer questions on a high-stakes test for medical students. Adv Health Sci Educ Theory Pract., 2024; 29(3):783-801.
26. MILLER GE. The assessment of clinical skills/competence/performance. Acad Med.,1990 Sep; 65(9): S63-7.
27. MUKURUNGE E, et al. Assessment approaches in undergraduate health professions education: towards the development of feasible assessment approaches for low-resource settings. BMC Med Educ., 2024; 24(1): 318.
28. NEWTON, PM, ESSEX, K. How common is cheating in online exams and did it increase during the COVID-19 pandemic? A systematic review. J Acad Ethics., 2024; 22:323–43.
29. NG CK. Evaluation of academic integrity of online open book assessments implemented in an undergraduate medical radiation science course during COVID-19 pandemic. J Med Imaging Radiat Sci., 2020; 51(4): 610-16.
30. PANÚNCIO-PINTO MP, TRONCON LEA. Avaliação do estudante: aspectos gerais. Medicina (Ribeirão Preto)., 2014; 47(3): 314-23.
31. PIENAAR L, et al. Health professions education and unprofessional behaviour in the global south: a scoping review of conceptions, theoretical frameworks, and prevalence. BMC Med Educ., 2025; 25(1): 1733.
32. RADEMAKERS J, et al. Progress testing with short answer questions. Med Teach., 2005; 27(7): 578-82.
33. SABRINA F, et al. Ensuring academic integrity in online assessments: a literature review and recommendations. Int J Inf Educ Technol., 2022; 12(1): 60-70.
34. SARKAR S, et al. Online open-book examination of undergraduate medical students: a pilot study of a novel assessment method used during the coronavirus disease 2019 pandemic. J Laryngol Otol., 2021;135(4): 288-92.
35. SEFER CC, et al. Avaliação do desempenho de estudantes de Medicina em Neurologia utilizando metodologias ativas. REAS [Internet]. 2019[citado 14 mar 2026]; 11(9):e293.
36. SUSNJAK T, MCINTOSH TR. ChatGPT: the end of online exam integrity? *Educ Sci*., 2024; *14*(6): 656.
37. TUAH NA, NAING L. Is online assessment in higher education institutions during COVID-19 pandemic reliable? Siriraj Med J., 2020; 73(1): 61–8.
38. VAZQUEZ JJ, et al. Can we stay one step ahead of cheaters? A field experiment in proctoring online open book exams. J Behav Exp Econ., 2021;90: 101653.