RBEM
Revista Brasileira de Educação Médica

ARTIGO ORIGINAL

DOI: https://doi.org/10.1590/1981-5271v46.supl.1-20220303

Análise da adequação dos itens do Teste de Progresso em medicina

Item analysis of Progress Test in medicine

Edlaine Faria de Moura Villela[1] edlaine@alumni.usp.br
Miguel Angelo Hyppolito[1] mahyppo@fmrp.usp.br
Julio Cesar Moriguti[1] moriguti@fmrp.usp.br
Valdes Roberto Bollela[1] vbollela@fmrp.usp.br

RESUMO

Introdução: A avaliação do estudante deve induzir aprendizagem e ser baseada em competência, ou seja, avaliar (habilidades cognitivas, psicomotoras e afetivas). Para avaliar conhecimento e a habilidade para sua utilização no contexto profissional, o Teste de Progresso (TP) tem sido usado em larga escala, com finalidade somativa e principalmente formativa.

Objetivo: Este estudo teve como objetivo verificar a adequação e qualidade de itens que compõem os TP realizados pelos estudantes.

Método: Trata-se de estudo exploratório descritivo e retrospectivo que analisou todos os itens de seis provas do TP aplicado a estudantes de Medicina do primeiro ao sexto ano da Faculdade de Medicina de Ribeirão Preto/USP, no período de 2013 a 2018. Os sete indicadores de boas práticas foram: 1. abordar tema relevante na formação médica; 2. ter enunciado maior que as alternativas; 3. avaliar aplicação do conhecimento; 4. definir pergunta clara para o item no enunciado; 5. avaliar apenas um domínio do conhecimento em cada item; 6. ter resposta correta e distratores homogêneos e plausíveis; 7. ausência de erros no item que acrescentam dificuldade desnecessária ou dão pistas da resposta correta. Dois avaliadores independentes analisaram as questões e, quando necessário, revisavam em conjunto os itens discordantes.

Resultado: A análise das provas permitiu identificar boa qualidade técnica na maioria dos itens das seis provas, além de indicar que a não adesão foi mais frequente nos indicadores 4 e 5, que podem comprometer tanto a validade quanto a interpretação dos resultados da prova em termos de lacunas do conhecimento por parte dos estudantes.

Conclusão: A qualidade das questões das provas analisadas é muito boa, mas foi possível identificar oportunidades de melhoria no processo de elaboração de itens, que servem de base para o desenvolvimento docente dos elaboradores da instituição.

Palavras-chave: Educação Médica; Avaliação Educacional; Questões de Prova; Docentes.

ABSTRACT

Introduction: *Assessment drives learning and should follow a competence-based approach. The Progress Test (PT) has been used on a large scale for summative and mainly formative purposes to assess knowledge and the ability to use it in the professional context.*

Objective: *To check the adequacy and quality of the items and that make up the progress tests sat by students.*

Method: *Descriptive and retrospective exploratory study that analyzed all the items of six PT exams applied to medical students from the first to the sixth year of the Faculty of Medicine of Ribeirão Preto/USP, from 2013 to 2018. The seven indicators of good practices were: 1. Addresses a relevant topic in medical training; 2. Statement longer than key answer and distractors; 3. Application of knowledge evaluated; 4.Clear lead-in defined for the item in the statement; 5. Only one domain of knowledge assessed in each item; 6.Plausible and homogeneous key answer and distractors; 7.Absence of flaws that add unnecessary difficulty or give clues to the correct answer. Two independent evaluators analyzed the items and, if necessary, they jointly reviewed any disagreement.*

Result: *The analysis showed a good technical quality of most items in the six PT exams. In addition, they indicated that non-adherence was a bit more frequent for indicators 4 and 5, which can compromise both the validity and the interpretation of the test results in terms of knowledge gaps on the part of students.*

Conclusion: *In general, the quality of the items was very good but there are some opportunities for improvement in the process of item writing based on faculty development within the institution.*

Keywords: *Medical Education; Educational Measurement; Examination Questions; Faculty.*

[1]Faculdade de Medicina de Ribeirão Preto, Ribeirão Preto, São Paulo, Brasil.

Editor: Aristides Augusto Palhares Neto.

Recebido em 11/09/22; Aceito em 14/10/22.

Avaliado pelo processo de *double blind review.*

DOI: https://doi.org/10.1590/1981-5271v46.supl.1-20220303

INTRODUÇÃO

A avaliação é um fator importante que impulsiona a aprendizagem dos alunos, uma vez que eles tendem principalmente a se concentrar no conteúdo que é avaliado. No contexto do ensino superior (incluindo a educação médica), o método de avaliação adotado pode influenciar na aprendizagem do aluno[1].

A avaliação durante a graduação em Medicina, bem como em outros cursos, não possui uma teoria abrangente ou unificadora. Ela toma como base várias teorias de campos científicos adjacentes, como educação geral, psicologia cognitiva e teorias psicométricas[2].

Um obstáculo relatado na literatura para uma avaliação de qualidade é a realidade encontrada nas salas de aula: turmas com elevado número de alunos e professores pouco familiarizados com princípios e boas práticas na avaliação do estudante, e que simplesmente reproduzem o modelo tradicional de avaliação no ensino superior (avaliação somativa concentrada essencialmente no conhecimento). Essa realidade pode comprometer não somente a avaliação formativa e contínua do desempenho do aluno, mas também o perfil desejado do egresso das escolas médicas.

No campo da educação médica, tem-se priorizado um modelo de avaliação que se concentra na formação profissional do aluno, com enfoque na educação baseada em competências, ou seja, as avaliações estão buscando cada vez mais checar a aquisição de habilidades e a demonstração de atitudes adequadas por parte dos estudantes de graduação em Medicina[3].

A avaliação dos estudantes de Medicina deve ter um caráter somativo e formativo para que consiga reconhecer a capacidade do aluno para a prática profissional e identificar lacunas e corrigi-las durante a graduação, garantindo assim a segurança do paciente no futuro. Para tanto, é necessário adotar mais de um instrumento avaliativo. Esses instrumentos devem ser coerentes com os objetivos de aprendizagem a serem alcançados e garantir que o aluno receba um *feedback* efetivo e regular[4,5].

Historicamente, a aprendizagem cognitiva sempre foi priorizada nos processos de ensino e avaliação, inclusive em habilidades práticas na educação médica[6]. O movimento atual, em termos de avaliação, consiste em incluir todos os domínios da competência por meio da utilização de um conjunto de instrumentos avaliativos que componham um sistema ou programa de avaliação do curso como um todo[7].

A escolha dos métodos deve ser feita de acordo com a finalidade da avaliação e com as dimensões que constituem o foco da avaliação proposta. Nesse sentido, é essencial analisar os atributos gerais dos métodos de avaliação: a validade, a confiabilidade, a viabilidade, a aceitabilidade, a equivalência, o impacto educacional e o efeito dos resultados obtidos na instituição[8].

No contexto da avaliação cognitiva, o Teste de Progresso (TP) tem sido bastante utilizado na educação médica como uma ferramenta que permite avaliar a aprendizagem e viabilizar a realização de intervenções para melhorar a aprendizagem e o ensino, além de discutir padrões educacionais com autores de vários países em busca de aprimorar programas existentes[9]. O TP é uma avaliação cognitiva longitudinal com conteúdo final do curso, que tem por finalidade avaliar a instituição e o desempenho cognitivo dos estudantes. Atualmente, tem sido aplicado em diversas escolas médicas no mundo e no Brasil[10].

O TP permite que não somente os estudantes sejam avaliados, mas também o próprio curso de graduação, viabilizando a análise do conteúdo e a estrutura curricular durante o processo de desenvolvimento dos alunos, que descrevemos como avaliação diagnóstica ou informativa[7]. Ademais, o TP é uma excelente ferramenta de avaliação formativa podendo identificar lacunas a serem trabalhadas ao longo da formação discente. Cabe destacar que a qualidade dos itens dos TP adotados influencia os resultados e o desempenho dos estudantes. Portanto, é fundamental que sejam estabelecidos critérios para elaboração, aplicação e análise das questões[10].

Diante do contexto apresentado, observa-se a expectativa de um egresso com perfil diferenciado, ou seja, com habilidades cognitivas que garantam uma base sólida para a prática profissional segura. Se tais habilidades cognitivas têm sido valorizadas, nada mais prudente que garantir avaliações que identifiquem o progresso da aprendizagem dos estudantes de Medicina. Assim, surge o interesse em verificar a adequação e qualidade de itens que compõem os TP realizados pelos estudantes.

MÉTODO

Local de realização do estudo

Este estudo analisou as provas do TP realizadas pelos estudantes da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (FMRP-USP).

População e tipo de estudo

Trata-se de um estudo exploratório de abordagem quantitativa. A população do estudo contou com uma amostra de exames completos do TP, referentes ao período de 2013 a 2018.

Coleta e análise dos dados

Realizou-se um estudo descritivo retrospectivo por meio de análise documental dos exames realizados pelos

estudantes. Adotaram-se as seguintes etapas: revisão qualitativa dos itens para toda a prova, revisão qualitativa dos itens por área e análise estatística descritiva dos dados. Para a realização da estatística descritiva, cada item foi classificado como: *completamente adequado*, quando contemplava os sete indicadores; *parcialmente adequado*, quando apresentava uma vinheta clínica ou um problema no enunciado e buscava avaliar a aplicação do conhecimento (indicador 3), mas não era objetivo, pois a pergunta do teste era muito aberta e a resposta correta incluía diferentes dimensões do conhecimento (não adesão aos indicadores 4 e 5); e *inadequado*, quando não havia situação-problema ou vinheta clínica no enunciado, ou trazia um enunciado, mas o teste poderia ser respondido sem a presença do mesmo (indicador 3), e também não contemplava os indicadores 4 e 5.

Também foi feita análise conjunta dos resultados obtidos pelos estudantes nas avaliações realizadas durante a graduação (TP).

Indicadores de qualidade dos testes de múltipla escolha (TME) ou itens

Foram definidos sete indicadores de boas práticas de acordo com duas referências sobre o tema[8,11]:

1) Abordar conceito relevante para a formação e atuação médica (alinhado à matriz de competência).

2) Escrever enunciado mais longo de modo a contextualizar o que se pretende avaliar, seguido por alternativas mais curtas.

3) Avaliar preferencialmente a aplicação do conhecimento ou interpretação de dados. Evitar questões que requerem apenas memorização de conteúdo.

4) Definir uma pergunta clara para o item ao final do enunciado indicando o foco do que se está avaliando. Por exemplo: mecanismo de doenças, diagnóstico, investigação complementar, manejo/tratamento, prevenção/reabilitação ou promoção da saúde.

5) Cada item deve avaliar apenas uma dimensão do conhecimento (ver indicador anterior), evitando questões muito abertas que abordam aspectos epidemiológicos, mecanismo de doença, prognóstico, diagnóstico, tratamento e prevenção em um único item da prova.

6) Ter resposta correta e distratores homogêneos e plausíveis.

7) Evitar erros de elaboração que acrescentam dificuldade desnecessária (confundem o estudante) ou que dão pistas da resposta correta (induzem acerto mesmo sem conhecimento do que está sendo perguntado).

Os indicadores 3, 4 e 5 foram escolhidos para nortear a classificação dos itens por serem estruturantes em um TME com uma única alternativa correta voltado à aplicação de conhecimento. O contexto, que deve estar presente no enunciado, é a base do raciocínio clínico e da tomada de decisão. A ausência de contexto está relacionada a itens que avaliam apenas memorização, que é o mais baixo nível na taxonomia de Bloom[12].

Aspectos éticos

O presente projeto foi encaminhado ao Comitê de Ética em Pesquisa da FMRP-USP e aprovado: Certificado de Apresentação para Apreciação Ética (CAAE) nº 88929618.8.0000.5440.

RESULTADOS

Analisaram-se seis exames de TP entre os anos de 2013 e 2018, totalizando 720 questões analisadas (120 questões em cada exame). Cada avaliação conta com 20 questões de ciências básicas e 20 questões de cada grande área do conhecimento, com exceção do exame do ano de 2015, o qual apresentou a seguinte composição: 24 questões de clínica médica, 24 questões de cirurgia, 24 questões de ginecologia e obstetrícia, 26 questões de medicina social e 22 questões de pediatria, não apresentando questões de ciências básicas (Tabela 1).

Tabela 1. Distribuição do total de questões dos Testes de Progresso por grande área do conheci-mento no período de 2013 a 2018

Grandes áreas do conhecimento	Número absoluto
Ciências básicas	100
Cirurgia	124
Clínica médica	124
Ginecologia e obstetrícia	124
Medicina social	126
Pediatria	122
Total	**720**

Fonte: Elaborada pelos autores.

 DOI: https://doi.org/10.1590/1981-5271v46.supl.1-20220303

Para cada item, verificou-se a adequação referente aos sete indicadores de qualidade esti-pulados para este estudo. No total, foram 549 itens (76,3%) adequados, 140 (19,4%) parcialmente adequados e 31 (4,3%) inadequados (Tabela 2).

A seguir, pode-se verificar como se deu a distribuição de questões parcialmente adequadas e inadequadas nas seis grandes áreas de conhecimento dos exames de TP (tabelas 3 e 4, respec-tivamente), permitindo identificar, por grande área, aquelas com necessidade de adequação.

Diante da análise da série temporal, pode-se observar a pequena porcentagem de questões inadequadas, pois nenhuma das grandes áreas teve mais que 5% de questões parcial ou totalmen-te inadequadas, o que mostra alta taxa de adesão às boas práticas na elaboração de TME do TP e um bom processo de gestão da prova.

Analisamos também a adequação das questões para cada um dos indicadores de qualidade definidos para este estudo, com o propósito de dar visibilidade aos indicadores que necessitam de maior atenção no momento de elaboração. Na Tabela 5, observamos que os indicadores 4 e 5 apresentaram maior porcentagem de inadequações (19,9% e 20,8%, respectivamente).

Os 720 itens também foram avaliados de acordo com três categorias: 1- aplicação do co-nhecimento, 2 - interpretação de dados, e 3 - memorização de conteúdo. Como o TP é construído com base nos estudantes concluintes, espera-se que avalie mais a aplicação do conhecimento ou a interpretação de dados no contexto da saúde do que simplesmente a memorização do que preci-sa ser aprendido. A grande maioria (97,5%) dos itens foi de aplicação do conhecimento.

Tabela 2. Distribuição da adequação dos itens levando em conta os indicadores de qualidade nos Testes de Progresso realizados entre 2013 e 2018

		Adequadas	Parcialmente adequadas	Inadequadas
Anos	2013	96	19	05
	2014	98	20	02
	2015	88	25	07
	2016	72	25	06
	2017	86	27	07
	2018	93	24	03
	Total	**549 (76,3%)**	**140 (19,4%)**	**31 (4,3%)**

Fonte: Elaborada pelos autores.

Tabela 3. Distribuição das questões parcialmente adequadas conforme as seis grandes áreas do conhecimento e no conjunto das provas de Teste de Progresso, 2013-2018

	Grandes áreas do conhecimento						
Anos	**Ciências básicas**	**Clínica médica**	**Cirurgia**	**Ginecologia e obstetrícia**	**Medicina social**	**Pediatria**	**Total**
2013	1	4	0	7	3	4	19
2014	4	3	4	6	2	1	20
2015	0	8	10	1	3	3	25
2016	7	4	2	1	1	10	25
2017	3	5	6	4	0	9	27
2018	4	4	1	7	1	7	24
Total	**19 (13,6%)**	**28 (20%)**	**23 (16,4%)**	**26 (18,6%)**	**10 (7,1%)**	**34 (24,3%)**	**140**
%Prova	**2,6%**	**3,9%**	**3,2%**	**3,6%**	**1,9%**	**4,7%**	**720**

Fonte: Elaborada pelos autores.

 DOI: https://doi.org/10.1590/1981-5271v46.supl.1-20220303

Tabela 4. Distribuição das questões inadequadas conforme as seis grandes áreas do conhecimento nos exames de Teste de Progresso, 2013-2018

	Grandes áreas do conhecimento						
Anos	**Ciências básicas**	**Clínica médica**	**Cirurgia**	**Ginecologia e obstetrícia**	**Medicina social**	**Pediatria**	**Total**
2013	1	0	1	0	3	0	5
2014	0	0	0	0	1	1	2
2015	0	1	0	5	1	0	7
2016	0	0	0	1	6	0	7
2017	0	4	0	0	3	0	7
2018	0	0	1	0	2	0	3
Total	**1 (3,2%)**	**5 (16,1%)**	**2 (6,5%)**	**6 (19,4%)**	**16 (51,6%)**	**1 (3,2%)**	**31**
%Prova	**0,13%**	**0,69%**	**0,27%**	**0,83%**	**2,22%**	**0,13%**	**720**

Fonte: Elaborada pelos autores.

Tabela 5. Adequação dos itens de acordo com os indicadores de qualidade numerados de 1 a 7

Indicador →	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Adequados	720	655	692	577	570	697	654
Inadequados (n)	0	65	28	143	150	23	66
Inadequados (%)	0%	9,0%	3,9%	19,9%	20,8%	3,2%	9,2%
Total	**720**	**720**	**720**	**720**	**720**	**720**	**720**

Fonte: Elaborada pelos autores.

DISCUSSÃO

O TP é uma estratégia de avaliação que analisa o domínio completo do conhecimento considerado pertinente para o egresso de um curso de graduação em Medicina. Por causa da natureza abrangente desse teste, é muito difícil estabelecer uma pontuação de aprovação[13]. É uma avaliação cognitiva sem caráter de seleção ou classificação, constituída de uma prova institucional que avalia individualmente se o ganho de conhecimento por parte do estudante está sendo contínuo e progressivo, e como o conhecimento está sendo consolidado nas áreas básicas e clínicas, importantes para o desfecho do desenvolvimento do profissional[14].

Na educação médica, avaliações de competência que decidem sobre a progressão podem ter consequências de longo prazo, tanto para os estudantes quanto para a sociedade. Se alunos competentes falham em um exame, isso dificulta o progresso de sua carreira, e, se os estudantes não competentes passam no exame, isso pode colocar pessoas em risco[15]. Apesar de mais comumente ser utilizado para propósito formativo, em alguns países e escolas médicas, o TP é adotado como um exame de grande importância para conclusão do curso (*highstake exam*), e isso requer maior precisão nas mensurações para tomada de decisões sobre aprovação/reprovação[15]. Independentemente do propósito da avaliação, a busca pela excelência nas avaliações dos estudantes na formação médica deve ser uma constante, e, por isso, a importância de se elaborarem questões seguindo as recomendações e boas práticas disponíveis tanto na literatura nacional quanto na internacional[8].

A análise das provas permitiu identificar boa qualidade técnica da maioria dos itens nas seis provas aplicadas durante o período do estudo, além de indicar que a não adesão foi mais frequente para os indicadores 4 e 5, o que poderia comprometer tanto a validade do exame quanto a interpretação dos resultados da prova em termos de lacunas do conhecimento por parte dos estudantes[16,17].

A elaboração de um TME com única alternativa correta requer uma pergunta focada (*lead-in*). Quando isso não ocorre, cria-se uma dificuldade para a análise dos resultados dos estudantes. A existência de uma pergunta direcionada a um objetivo de aprendizagem permite elaborar distratores plausíveis e semelhantes à alternativa correta. Por exemplo, se

o objetivo de aprendizagem é avaliar o mecanismo de ação de um patógeno que causa uma doença infecciosa (por exemplo, dengue), temos de elaborar o enunciado do item com uma vinheta que traga elementos da manifestação clínica da doença e fazer em seguida a pergunta do teste de forma clara e direta. Assim podemos apresentar uma história clínica compatível com um quadro de dengue grave em paciente que está hipotenso e tem derrame pleural. Informamos no enunciado que o teste diagnóstico (NS1) confirmou a doença e perguntamos: "Que mecanismo fisiopatológico explica esse quadro?". A resposta correta é aumento da permeabilidade capilar e perda de líquidos para o extravascular. Os distratores serão também mecanismos de doença, mas que não são a resposta correta. Por exemplo: dano ao endotélio da microvasculatura; lesão por contiguidade com dano tecidual; redução da concentração de albumina intravascular. Assim, o enunciado do item, a pergunta e as alternativas estão alinhados no intuito de avaliar um objetivo de aprendizagem relevante que consta do programa de ensino.

Um erro muito comum nas provas que utilizam TME com uma única alternativa correta é construir um item em que, a partir do enunciado, perguntamos várias coisas diferentes: mecanismo de doença, diagnóstico, o tratamento indicado, tudo junto na mesma questão da prova. Nesse caso, as alternativas serão inevitavelmente mais longas e compostas por várias possibilidades que costumam se repetir e podem dar pistas de qual é a alternativa correta[8,11].

Diante do caso de dengue grave mencionado, a pergunta poderia ter sido: "Quais são o mecanismo fisiopatológico e a conduta para esse caso?". As alternativas incluiriam: aumento da permeabilidade vascular ou hipoalbuminemia; internação ou tratamento ambulatorial, hidratação ou aminas vasoativas, monitoramento de plaquetas ou avaliar discrasias sanguíneas, entre outras possibilidades. Assim, os distratores serão uma combinação com itens incorretos, e a certa será uma combinação de respostas corretas para as duas perguntas.

Essa opção na elaboração de um item com única alternativa correta não encontra respaldo nas boas práticas e costuma dar pistas aos estudantes, pois as respostas que mais se repetem costumam ser as corretas. Além disso, quando formos analisar o desempenho dos alunos, não saberemos se a dificuldade daqueles que erraram estava mais relacionada a compreensão dos mecanismos de doença, o diagnóstico ou o tratamento. Outra possível consequência não desejada nesta abordagem é misturarmos tantas coisas, que podem confundir os estudantes e resultar em erro por falta de compreensão e não por falta de conhecimento do tema em questão.

Ao analisarmos a adequação das questões de acordo com as grandes áreas do conhecimento, observamos que, na *medicina social*, houve maior número de questões inadequadas (16/31; 51,6%) e menor número de questões parcialmente adequadas (10/140; 7,1%). Já na *pediatria* observamos maior número de questões parcialmente adequadas (34/140; 24,3%) e menor número de questões inadequadas (1/31; 3,2%). Essas informações podem favorecer uma abordagem dos elaboradores dessas áreas e, após a análise dos itens, verificar se existem oportunidades de melhoria que poderiam ser abordadas em uma oficina de desenvolvimento docente específica para qualificar as questões do TP, como preconizado por Pinheiro et al.[18].

Um exame que contém itens bem redigidos apresenta resultados fidedignos no momento de aferir os objetivos de aprendizagem e a qualidade do programa educacional[19]. Dessa forma, fica explícita a importância de a universidade investir no desenvolvimento docente, pensando que irão atuar como elaboradores de itens.

Dentre os parâmetros que podemos utilizar para mensurar a qualidade das questões de uma prova, podemos incluir também dados sobre a reprodutibilidade do exame e a qualidade dos itens a partir dos índices de discriminação e de dificuldade[2]. A reprodutibilidade (confiabilidade) da prova indica sua consistência interna, ou se se a amostragem foi adequada e se existe estabilidade na medida que foi feita[19].

Além de aderir às boas práticas na elaboração de itens[8], é preciso que cada item seja capaz de diferenciar estudantes mais proficientes, aqueles que têm maior pontuação na prova e no item, daqueles que são menos proficientes. Assim, se estudantes com melhor desempenho acertam um TME escolha e estudantes com proficiência inferior erram, podemos inferir que esse item tem boa discriminação e contribui de maneira efetiva para avaliar a competência dos alunos na prova[20].

Ainda no que tange às boas práticas, o estudo apresentou a distribuição de questões parcialmente adequadas e inadequadas nas seis grandes áreas de conhecimento dos exames de TP, permitindo identificar, por grande área, aquelas com necessidade de maior adequação dos itens, que acabam por comprometer o desempenho do aluno, independentemente de sua maior ou menor proficiência. Esse diagnóstico é essencial para aprimorar a elaboração dos itens, corroborando outros estudos realizados[17,20,21].

Por fim, o TP, aplicado anualmente, tem potencial tanto de avaliação formativa quanto informativa (informa sobre o programa educacional-currículo), com benefícios não somente para o estudante, mas também para a instituição. Essas informações permitem identificar fortalezas e fragilidades em diferentes áreas do conhecimento e unidades curriculares. Essa informação quando bem trabalhada e compartilhada com gestores dos currículos tem potencial de qualificar a

formação médica. Ao zelar pela boa construção de itens para o TP, a gestão acadêmica garante qualidade na formação e cria oportunidades para a formação de comunidades de práticas que reúnam docentes sensibilizados e interessados em avaliação do estudante por meio da utilização do TP em suas instituições e/ou nacionalmente[18-21].

O estudo realizado apresenta limitações quanto à sua amostra. Os dados coletados são referentes a um período específico de aplicação do TP e a seleção da amostra não foi aleatória, mas sim por conveniência. Ademais não foi utilizada estatística inferencial para identificar associação entre o desempenho durante os TP e o desempenho na prova de residência.

CONCLUSÃO

Diante da necessidade de conhecer a qualidade dos exames nacionais que avaliam os egressos das graduações em Medicina no Brasil, este estudo traz uma proposta de indicadores simples e fáceis de serem utilizados para auxiliar elaboradores de itens e gestores de prova a obter resultados de qualidade na avaliação dos estudantes.

As provas realizadas pelos estudantes de graduação da FMRP-USP que faz parte do mais antigo consórcio de TP incluíram itens relevantes no contexto da formação médica, que estavam de acordo, em sua maioria, com as boas práticas de elaboração de itens com uma única alternativa correta.

Provas de qualidade são instrumentos com potencial para avaliações formativa, somativa e informativa, ou seja, as que auxiliam a rever e melhorar o programa educacional como de fato acontece na instituição.

CONTRIBUIÇÃO DOS AUTORES

Todos os autores contribuíram em maior ou menor grau para o desenvolvimento da experiência descrita e participaram da elaboração do artigo.

CONFLITO DE INTERESSES

Declaramos não haver conflito de interesses.

FINANCIAMENTO

Bolsas de pós-doutorado concedidas no âmbito do acordo da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) e Fundação de Amparo à Pesquisa do Estado de Goiás (Fapeg): Edital Capes/Fapeg n° 09/2018.

REFERÊNCIAS

1. Htwe, TT, Ismail SB, Low GKK. Comparative assessment of students' performance and perceptions on objective structured practical models in undergraduate pathology teaching. Singapore Med J. 2014;55(9):502-5.
2. Schuwirth LWT, Van Der Vleuten CPM. General overview of the theories used in assessment: AMEE Guide No. 57. Med Teach. 2011;33:783-97.
3. Souza MPG, Rangel M. Avaliação: um impasse na educação médica. Rev Bras Educ Med. 2003;27(3):213-22.
4. Norcini J, Anderson B, Bollela V, Burch V, Costa MJ, Duvivier R, et al. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. Med Teach. 2011;33(3):206-14.
5. Bates J, Konkin J, Suddards C, Dobson S, Pratt D. Student perceptions of assessment and feedback inlongitudinal integrated clerkships. Med Educ. 2013;47:362-74.
6. Swanwick T, organizador. Understanding medical education: evidence, theory, and practice. New York: Wiley-Blackwell; 2016.
7. Norcini J, Anderson MB, Bollela V, Burch V, Costa MJ, Duvivier R, et al. Consensus framework for good assessment. Med Teach. 2018;40(11):1102-9.
8. Bollela VR, Borges MC, Troncon LEA. Avaliação somativa de habilidades cognitivas: boas práticas na elaboração de testes de múltipla escolha e na composição de exames. Rev Bras Educ Med. 2018;42(4):74-85.
9. Wrigley W, Van der Vleuten CPM, Freeman A, Muijtjens A. A systemic framework for the progress test: strengths, constraints and issues: AMEE Guide No. 71. Med Teach. 2012;34(9):683-97.
10. Sakai MH, Ferreira Filho, OF, Almeida MJ, Mashima DA, Marchese MC. Teste de progresso e avaliação do curso: dez anos de experiência da medicina da Universidade Estadual de Londrina. Rev Bras Educ Med. 2008;32(2):254-63.
11. National Board of Medical Examiners. Construindo o teste escrito: questões para ciências básicas e clínicas. Filadélfia: NBME; 2017.
12. Bloom BS. Taxonomy of educational objectives. New York: David Mckay; 1956. 214 p.
13. Verhoeven BH, Van der Steeg AF, Scherpbier AJ, Muijtjens AM, Verwijnen GM, Van der Vleuten CP. Reliability and credibility of an angoff standard setting procedure in progress testing using recent graduates as judges. Med Educ. 1999;33(11):832-37.
14. Verhoeven BH, Snellen-Balendong HAM, Hay IT, Boon JM, Van der Linde MJ, Blitz-Lindeque JJ, et al. The versatility of progress testing assessed in an internatio- nal context: a start for benchmarking global standardization? Med Teach. 2005;27(6):514-20.
15. Lahner FM, Schauber S, Lorwald AC, Kropf R, Guttormsen S, Fischer MR, et al. Measurement precision at the cut score in medical multiple choice exams: theory matters. Perspect Med Educ. 2020;9(4):220-8.
16. Romão GS. Como elaborar questões de múltipla escolha de boa qualidade. Femina. 2019;47(9):561-4.
17. LeClaire EL, Destephano CC, Lerner VT, Chen CCG. Decisions and consequences: Validation of High-Stakes Simulation-Based Assessments in Gynecologic Surgery. J Minim Invasive Gynecol. 2021;28(7):1285-90.
18. Pinheiro OL, Spadella MA, Moreira HM, Ribeiro ZMT, Guimaraes APC, Almeida Filho OM, et al. Teste de Progresso: uma ferramenta avaliativa para a gestão acadêmica Rev Bras Educ Med. 2015;39(1):68-78.
19. Tombi ECNA, Zukowsky-Tavares C, Ferreira-Gerab I. Qualidade dos itens de múltipla escolha utilizados em um teste de progresso. Estud Aval Educ. 2022;33(e07533).
20. Tavakol M, Dennick R. Post-examination analysis of objective tests. Med Teach. 2011;33(6):447-58.
21. Hafferty FW, O'Brien BC, Tilburt, JC. Beyond high-stakes testing: learner trust, educational commodification, and the loss of medical school professionalism. Acad Med. 2020;95(6):833-7.