

Realidade ou simulação? Análise do desempenho de estudantes de Medicina em avaliações práticas distintas

Reality or simulation? Performance analysis of medical students through different practical assessments

Matheus Santarosa Cassiano¹ matheuscassiano76@gmail.com
Sílvia Maria Riceto Ronchim Passeri¹ spasseri@unicamp.br
Nelson Afonso Lutaif¹ nlutaif@unicamp.br

RESUMO

Introdução: Avaliar habilidades clínicas é um desafio no curso médico. A heterogeneidade na escolha dos pacientes somada à falta de critérios objetivos resultou em mudança metodológica para utilização do Objective Structured Clinical Examination (OSCE).

Objetivo: O objetivo deste estudo foi identificar se o OSCE resulta em distribuição de frequência de notas de desempenho de padrão gaussiano em comparação ao modelo tradicional.

Método: Analisaram-se as notas de 239 estudantes da disciplina de Semiologia e Propedêutica de um curso de Medicina entre 2016 (modelo tradicional) e 2017 (OSCE) pelos testes *Kolmogorov-Smirnov bidimensional* e *t* de Student para verificar a correlação delas com o coeficiente de rendimento (CR).

Resultados: As notas da prova no modelo tradicional ($p < 0,0001$; $KS = 0,1881$) estão mais distantes da normalidade do que as da prova do modelo OSCE ($p = 0,0010$; $KS = 0,1134$) e são mais correlatas com o CR ($p < 0,0001$; $r = 0,45$) do que no modelo OSCE ($p = 0,31$; $r = 0,06$).

Conclusão: O OSCE pode proporcionar informações mais fidedignas sobre o desempenho do estudante em estágios práticos do curso médico.

Palavras-chave: Avaliação Educacional; Educação Médica; Desempenho Acadêmico.

ABSTRACT

Introduction: Assessing clinical skills represents a challenge in medical training. The heterogeneity in the selection of patients associated to a lack of objective criteria has resulted in a methodological shift toward using the objective structured clinical examination (OSCE).

Objective: This study aimed to identify whether the OSCE results in a Gaussian frequency distribution of performance assessment marks compared to the traditional model.

Method: 239 students from the Semiotics module of the medicine course between 2016 (traditional model) and 2017 (OSCE) were analyzed using the two-dimensional Kolmogorov-Smirnov test and Student's *t*-test for correlation with the performance coefficient (PC).

Results: The test scores in the traditional model ($p < 0.0001$; $KS = 0.1881$) are more discrepant in relation to the normal distribution than the OSCE test scores ($p = 0.0010$; $KS = 0.1134$) and show more correlation with the PC ($p < 0.0001$; $r = 0.45$) than the OSCE model ($p = 0.31$; $r = 0.06$).

Conclusion: OSCE might provide a more accurate assessment of a student's performance during medical internship.

Keywords: Educational Measurement; Medical Education; Academic Performance.

¹ Universidade Estadual de Campinas, Campinas, São Paulo, Brasil.

Editora-chefe: Daniela Chiesa

Editor associado: Kristopherson Lustosa Augusto

Recebido em 31/05/20; Aceito em 10/11/20.

Avaliado pelo processo de *double blind review*.

INTRODUÇÃO

Uma das questões mais importantes do ensino médico, além do interesse em examinar a qualidade do ensino de graduação, é a metodologia utilizada para a avaliação do conteúdo oferecido pelas disciplinas^{1,2}. A metodologia permite que um processo avaliativo possa mostrar ao aluno se o caminho que ele percorreu durante a disciplina foi condizente com os conteúdos exigidos, além de servir como guia para os temas mais importantes e determinantes para uma boa formação.

Toda avaliação é uma forma de aprendizagem que deve servir como guia e suporte para mostrar e solucionar as necessidades de aprendizagem do aluno e aumentar a autoconfiança e capacidade de aprender dele³. A partir do caráter sociopedagógico indissociável de qualquer prática educacional, transformar a avaliação em processo que garanta a promoção e autonomia do envolvidos⁴ é um imperativo que deve ser buscado incessantemente.

Demo e Klein⁵ compõem um grupo de especialistas que defendem a necessidade de testes e medidas para que se possa conhecer suficientemente bem a situação do aluno e auxiliá-lo em sua progressão, entretanto ponderam que os critérios insuficientemente claros e mal construídos de muitas avaliações as tornam afastadas das premissas da função diagnóstica e formativa³.

No que se refere à avaliação da aprendizagem em disciplinas que desenvolvem a prática médica, há diversos relatos de sucesso de experiências de avaliações práticas com pacientes simulados e *checklist* padronizado que utilizam como referência o *Objective Structured Clinical Examination* (OSCE). O OSCE permite também fornecer *feedback* ao estudante, o que é fundamental para a compreensão das deficiências dele⁶.

É importante ressaltar que evidências científicas comprovam o impacto do *feedback* na formação do estudante, e, portanto, quando pensamos em novos instrumentos de avaliação na educação médica, é necessário pensar também em formas de oferecer o *feedback* com o instrumento escolhido. O *feedback* contribui de forma efetiva para o aumento da retenção do conhecimento⁷.

Muitas disciplinas do curso médico lidam com os dilemas de como oferecer a avaliação ideal e com o melhor papel formativo, como é o caso da disciplina de Semiologia e Propedêutica da nossa faculdade, ministrada no terceiro ano do curso de Medicina e desenvolvida por meio de atividades teórico-práticas e de assistência diária aos doentes do complexo hospitalar da universidade.

Durante os dois meses de vigência do estágio, os objetivos principais são: aprimorar as habilidades da anamnese geral e do exame físico geral e especial; caracterizar os achados do exame físico nos principais quadros sindrômicos e a etiologia

e fisiopatologia de cada um deles; desenvolver habilidade de evolução diária dos doentes internados nas enfermarias clínicas; promover o desenvolvimento do raciocínio clínico, das hipóteses diagnósticas e dos principais diagnósticos diferenciais; introduzir conceitos relacionados aos principais exames complementares da propedêutica armada e suas indicações mais prevalentes; inserir os conceitos fundamentais da terapêutica; e desenvolver e aperfeiçoar os aspectos da relação médico-paciente pautados pelos conceitos éticos e morais. Todas essas habilidades visam ao desenvolvimento da prática clínica exigida no exercício profissional diário.

Essa disciplina é bastante emblemática do ponto de vista do processo avaliativo, uma vez que o propósito do curso é desenvolver uma série de habilidades diversas com pacientes internados em enfermarias clínicas, como observação, raciocínio e técnicas manuais de semiotécnica.

Na literatura médica, há dificuldade tanto na grande diversidade de conteúdo e técnicas empregadas no ensino dessa disciplina quanto no processo avaliativo. Entretanto, os estudos mais recentes concordam, embora sem um consenso claro, que a uniformização do conteúdo e a utilização de *checklists* tanto no ensino quanto na avaliação têm tido os melhores resultados⁸⁻¹⁰.

No caso da nossa disciplina, isso também era regra. Embora na disciplina de Semiologia e Propedêutica houvesse certa homogeneidade no conteúdo, a metodologia do ensino era bastante diversificada por conta de uma grande autonomia oferecida aos preceptores para transmitir esse conteúdo. No método avaliativo, essa heterogeneidade era mais significativa porque decorria de uma média final composta de notas de portfólio, conceitos dos preceptores e uma prova de competência técnica realizada à beira do leito. Todas essas notas apresentavam graus variados de subjetividade que transpareciam em uma perda na variabilidade das médias finais do grupo. Essa perda de variabilidade é confirmada com uma distribuição de frequências das notas demasiado distante de uma curva gaussiana esperada para esse conjunto de dados.

Com efeito, em estudos populacionais, uma curva de distribuição de frequências não gaussiana pode ser interpretada estatisticamente como sendo uma amostragem com vícios de coleta ou de geração. Consequentemente, a média e o desvio padrão não serão os melhores parâmetros para representar o comportamento do grupo como um todo, e muitas das conclusões derivadas da análise dessas notas deverão ser interpretadas com diversas ressalvas¹¹. Inúmeros estudos corroboram essa constatação da falta de homogeneidade nos critérios subjetivos usados pelos docentes para atribuir notas e falhas na detecção de dificuldades de aprendizagem de seus estudantes^{12,13}.

Diante disso, visando à efetividade no sistema

de avaliação da aprendizagem, o objetivo deste estudo comparativo é identificar se a prova prática com paciente simulado e *checklist* padronizado resulta em distribuição de frequência de notas de desempenho de padrão gaussiano, uma vez que um dos principais motivos da padronização é minimizar os possíveis vieses da subjetividade na visão do avaliador, comparado ao desempenho dos estudantes na prova prática com paciente real.

MÉTODO

Trata-se de um estudo descritivo, com análise quantitativa dos dados, que selecionou as notas obtidas nas avaliações práticas e o coeficiente de rendimento (CR) acadêmico de 239 estudantes de Medicina (124 em 2016 e 115 em 2017) que cursaram a disciplina de Semiologia e Propedêutica e que participaram da avaliação prática aplicada ao final da matéria.

A avaliação do desenvolvimento de habilidades na disciplina, até 2016, era realizada com prova prática à beira do leito com pacientes, a partir da observação do docente em relação ao desempenho do estudante. A partir de 2017, essa prova prática foi reformulada e passou a ser aplicada a pacientes simulados e *checklist* padronizado para os avaliadores, utilizando como referência o OSCE, que é indicado para avaliação de competências médicas¹⁴⁻¹⁶.

Além da análise dessa nota na avaliação prática, que é o objetivo deste estudo, analisamos também o CR desses estudantes. O desempenho geral do estudante no curso de Medicina gera um CR que é calculado com base nas notas obtidas em cada uma das disciplinas e na carga horária delas no curso. O CR é um indicativo que subsidia muitos estudos relacionados ao desempenho do estudante e das turmas.

O modelo estatístico utilizado para a verificação do desempenho foi a média aritmética simples; para a dispersão das notas o desvio padrão; e para a verificação da normalidade de distribuição de notas o teste Qui-quadrado. Avaliamos a correlação entre os CRs dos estudantes nos anos de 2016 e 2017 por meio do teste de Pearson. Quanto mais próximo de 1 estiver o "p de Person", isso será indicativo de uma correlação perfeita positiva entre duas variáveis; e quanto mais próximo estiver de -1, confirmar-se-á uma perfeita correlação negativa entre duas variáveis. Um $p = 0$ significa que as duas variáveis não dependem linearmente uma da outra.

Utilizamos o teste Kolmogorov-Smirnov bidimensional para comparar as notas nas provas práticas entre as duas turmas e, em seguida, avaliamos a correlação entre as notas das diferentes turmas nos diferentes métodos de avaliação utilizando o teste não paramétrico de Spearman.

Com o objetivo de identificar se há uma tendência de um grupo de notas apresentar-se com valores maiores do que

o outro, utilizamos o teste de Mann-Whitney e, finalmente, correlacionamos as notas obtidas pelos estudantes nas provas práticas dos diferentes métodos de avaliação com o CR da turma, utilizando o teste t de Student. Essa é uma forma para analisar se há correlação entre o desempenho médio ao longo da graduação com o desempenho avaliado na prova prática.

RESULTADOS

Iniciamos a análise por meio do CR dos estudantes, que, para o grupo de 2016 ($n = 124$), resultou em uma média de 8,17, com mínimo de 6,7, máximo de 9,04 e desvio padrão de 0,44; enquanto para os discentes da turma de 2017 ($n = 115$) resultou em uma média de 8,06, com mínimo de 6,7, máximo de 9,3 e desvio padrão de 0,43. Na amostra deste estudo, podemos observar, por meio do Gráfico 1, que as turmas 2016 e 2017, apesar de possuírem as médias de CRs muito próximas, o teste de Pearson nos apresenta que não há correlação significativa entre os CRs das duas turmas analisadas ($p = 0,66$; $r = 0,04$) e o teste t de Student nos mostra que não há uma tendência de uma turma ter CRs maiores do que a outra ($p = 0,06$; $r = 0,01$).

A análise das notas da prova prática de 2016 resultou em uma média aritmética simples de 8,85 e um desvio padrão de 0,844 para uma amostra de 124 alunos. Na prova prática de 2017, a análise resultou em uma média aritmética simples de 8,18 e um desvio padrão de 1,2 para uma amostra de 115 alunos.

Gráfico 1. Distribuição dos coeficientes de rendimento das turmas 2016 e 2017

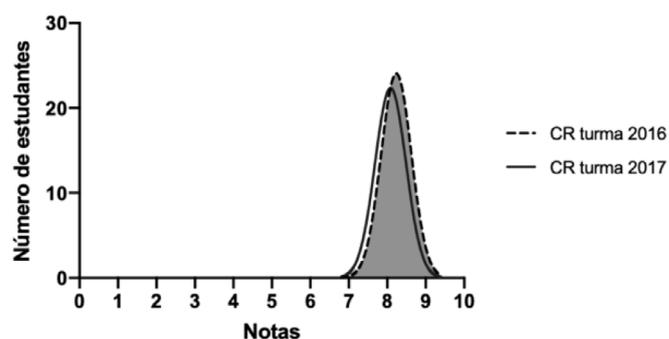
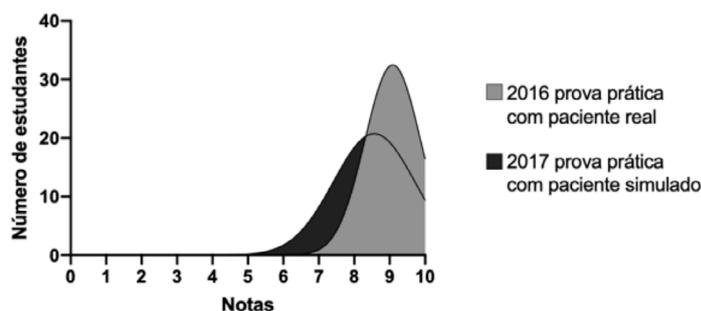


Gráfico 2. Comparativo da distribuição das notas dos estudantes nas provas práticas de 2016 e 2017



A comparação entre a distribuição de notas nas provas práticas 2016 e 2017 demonstra que a prova prática com pacientes simulados apresenta-se com uma distribuição de notas mais próxima de uma curva de Gauss do que as notas da prova com pacientes reais, e isso se comprova na análise que fizemos por meio do teste Kolmogorov-Smirnov bidimensional. Por meio dessa análise, identificou-se que os dados de 2016 e 2017, referentes a essas notas, apresentaram-se com distribuição desviada, mas uma diferença entre os valores p sinaliza que as notas de 2016 ($p < 0,0001$ e distância $KS=0,1881$) estão mais distantes da normalidade do que as notas de 2017 ($p = 0,0010$ e distância $KS = 0,1134$).

Para a análise da correlação entre as notas dos estudantes nas provas práticas em 2016 e 2017, utilizamos o teste não paramétrico de Spearman com intervalo de confiança de 95%. A análise apresentou uma correlação significativa ($p < 0,0001$) entre as notas nas provas práticas com paciente real e as notas nas provas práticas com paciente simulado. Entretanto, trata-se de uma correlação baixa, com escore de correlação de 0,43. Considerando o tamanho da amostra, fizemos a mesma correlação por meio do teste paramétrico de Pearson que

resultou em $p < 0,0001$ e $r = 0,52$.

Analisamos então, por meio do teste de Mann-Whitney, as notas nas provas práticas de 2016 e 2017 para identificarmos se há a tendência de uma delas ter valores maiores do que a outra. O resultado se mostrou significativo ($p < 0,0001$) com mediana de 9,0 (2016) e 8,5 (2017) apresentando diferença de -0,5, ou seja, a prova prática com pacientes reais tem uma tendência a ter valores maiores do que a prova prática realizada com pacientes simulados.

Correlacionamos as notas obtidas pelos estudantes nas provas práticas dos diferentes métodos de avaliação com o CR da turma. Considerando o tamanho da amostra das turmas de 2016 e 2017, fizemos a análise estatística utilizando o teste t de Student com intervalo de confiança de 95%.

A análise estatística mostrou haver significância, apesar de apresentar índice baixo, entre as notas da prova prática de 2016 com o respectivo CR ($p < 0,0001$; $r = 0,45$). Em 2017, quando a avaliação foi realizada com pacientes simulados, não houve correlação significativa entre as notas obtidas pelos estudantes na prova prática com o respectivo CR ($p = 0,31$; $r = 0,06$).

Gráfico 3. Correlação entre as notas dos estudantes nas provas práticas de 2016 e 2017

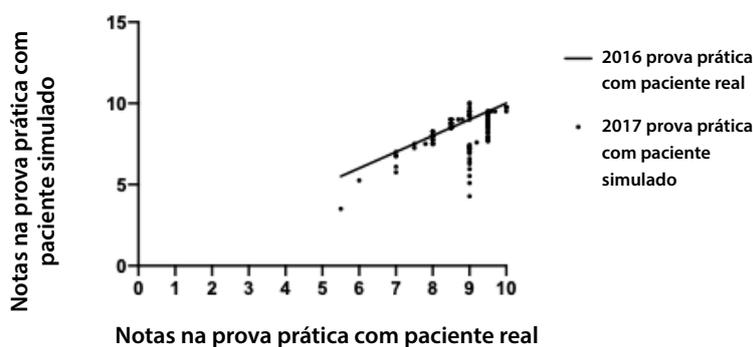
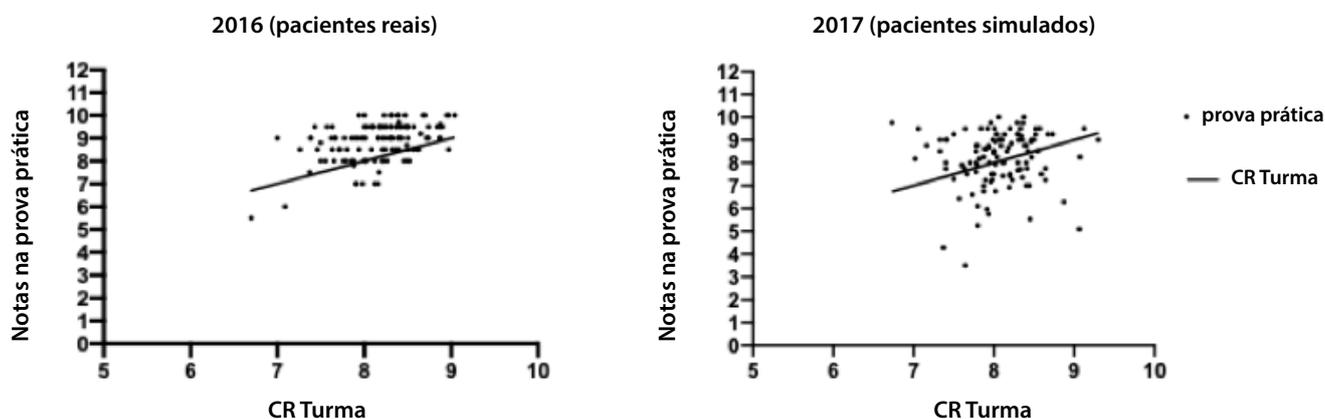


Gráfico 4. Correlação entre as notas dos estudantes nas provas práticas e os respectivos coeficientes de rendimento acadêmico



DISCUSSÃO

A amostra analisada neste estudo, apesar de não apresentar correlação significativa entre os desempenhos acadêmicos indicados por meio do CR, demonstrou que as duas turmas possuem médias muito próximas (8,17 e 8,06) com praticamente o mesmo desvio padrão, sugerindo que não há grandes discrepâncias quanto ao desempenho acadêmico dos dois grupos estudados.

O principal objetivo deste trabalho foi comparar o padrão de distribuição das notas do modelo de prova prática realizada com o paciente à beira do leito com o modelo de avaliação com paciente simulado e *checklist* padronizado. De acordo com a literatura, a mudança de sistema de avaliação provoca, de maneira inevitável, um impacto educacional¹⁷. Neste estudo, os estudantes que cursaram a disciplina em 2017 se depararam com a necessidade de alterar a maneira como se preparam para a prova prática, que deixou de ser aplicada no modelo com pacientes reais e passou a ser realizada com padronização dos itens avaliados e ênfase nas habilidades clínicas que o estudante deveria ter adquirido até o terceiro ano do curso médico. Tais alterações mostraram-se mais condizentes com os princípios de avaliação da competência do estudante observados na pirâmide de Miller¹⁸.

O primeiro dado observado neste estudo foi a diminuição da média aritmética das notas entre os anos do modelo de avaliação com paciente real (2016) e o modelo com paciente simulado (2017), bem como uma maior variabilidade na distribuição dessas notas, comprovada pelo crescimento significativo do desvio padrão ao longo da análise. Embora haja uma baixa correlação entre as notas das diferentes turmas, as notas na prova prática com pacientes reais tendem a ser mais altas do que na prova com pacientes simulados.

Esse dado nos permite inferir que o *checklist* padronizado da prova prática com pacientes simulados pode ser um indicativo que influencia no desempenho do estudante, uma vez que são vários os professores avaliadores e esse modelo define de forma objetiva quais habilidades devem ser observadas, sendo, portanto, mais criterioso do que uma avaliação geral do atendimento médico realizado.

Outro aspecto que pode influenciar nesses dados é a mudança abrupta do modelo de avaliação que diminuiu a sensação de previsibilidade empregada historicamente na disciplina, culminando em um padrão de média mais condizente com uma avaliação adotada dentro de um grupo heterogêneo de estudantes.

Quanto à distribuição da frequência das notas, a mudança para o modelo com paciente simulado resultou em uma curva mais próxima da normalidade, quando comparada ao modelo de prova anterior, e, nesse sentido, entendemos

que, mais uma vez, o *checklist* padronizado, por minimizar vieses, demonstra maior capacidade de avaliar amostras heterogêneas de alunos e fornecer notas mais condizentes com o desempenho real de cada estudante. Tal análise sugere que o OSCE se mostrou satisfatório para avaliar o estudante do terceiro ano médico, dado que foi possível determinar previamente o nível de complexidade exigido e a avaliação dos todos os alunos transcorreu exatamente nas mesmas condições, sem depender da disponibilidade de casos clínicos de pacientes. Com o *checklist* preestabelecido, houve um aporte considerável de objetividade ao exame¹⁹.

No que concerne à análise da correlação da nota na prova prática com o desempenho acadêmico do estudante ao longo do curso a partir do CR, observou-se que há uma correlação entre CR e prova prática em 2016 quando a prova foi aplicada com pacientes reais, ou seja, quando o cenário reproduziu o próprio ambiente de prática da disciplina do qual se origina a nota que embasará o CR. Historicamente, o desempenho geral do estudante de Medicina, analisado pelo CR, é constituído de altas notas e muitas delas influenciadas por notas de conceito, que é um instrumento subjetivo de avaliação da postura, do relacionamento e da atitude do estudante durante seus estágios práticos no curso médico.

Entretanto, as notas obtidas por meio da prova prática com pacientes simulados não possuem correlação com os respectivos CRs, o que nos permite sugerir novos estudos que avaliem fatores que podem ser indicativos desses dados, como o estado emocional do estudante perante o paciente simulado e um modelo padronizado de avaliação, mais objetivo e menos real para ele, por mais que se busque a reprodução da realidade.

É importante considerar, diante desses dados obtidos, que todas as atividades práticas que compõem as notas das disciplinas que geram o CR do estudante, no curso em que este estudo foi realizado, são desenvolvidas praticamente em pacientes reais e, portanto, não padronizados como ocorre em atividades de simulação.

Embora as questões da prova prática tenham sido elaboradas pelos professores que ministram a disciplina, é fundamental que a qualidade delas seja criteriosamente analisada, pois muitas vezes observamos que os conceitos abordados em avaliações nem sempre são aqueles mais enfatizados nas aulas, o que pode gerar um aumento ou uma diminuição da dedicação do aluno ao tema a ser estudado.

Outro fator que sugere novos estudos é a análise sobre a composição das notas nas disciplinas que geram o CR do estudante. É importante observar o sistema de avaliação de todas as disciplinas que compõem a nota do CR, bem como analisar se os instrumentos utilizados em cada disciplina são adequados para cada perfil.

Essas análises são importantes se considerarmos que a educação médica possui diversos estudos demonstrando a efetividade e os benefícios de se trabalhar a avaliação com pacientes simulados.

Uma das limitações deste estudo foi a impossibilidade de identificar outros fatores que poderiam influenciar no desempenho dos estudantes, tais como qualidade de vida, saúde mental e acesso à informação, pois, neste trabalho, utilizaram-se dados referentes apenas ao desempenho mensurável do estudante no sistema de avaliação da aprendizagem, e não houve a aplicação de questionários de resposta individual.

CONCLUSÃO

Este estudo teve como objetivos estimular uma análise mais profunda sobre os métodos e o instrumento de avaliação utilizados no curso de Medicina e estabelecer alternativas para solucionar problemas, especialmente no que se refere às avaliações práticas que são imprescindíveis para o estabelecimento de diagnósticos relacionados às competências e habilidades clínicas do estudante em formação.

A mudança de metodologia de avaliação neste estudo proporcionou maior homogeneidade nos critérios avaliados, trouxe maior objetividade ao exame e estimulou uma reflexão sobre a avaliação em nosso curso, de forma a buscar um método e um instrumento que melhor representem o real desempenho de nossos estudantes para subsidiar ações de melhoria na formação médica.

AGRADECIMENTOS

Agradecemos ao Programa Institucional de Bolsas de Iniciação Científica do Conselho Nacional de Desenvolvimento Científico e Tecnológico (Pibic/CNPq) e ao Núcleo de Avaliação e Pesquisa em Educação Médica da Faculdade de Ciências Médicas da Universidade Estadual de Campinas (Unicamp).

CONTRIBUIÇÃO DOS AUTORES

Durante a confecção deste texto, todos os autores trabalharam ativamente na concepção do estudo, na revisão de literatura disponível para realização do projeto, na análise estatística, na discussão dos resultados obtidos, na elaboração e na revisão do artigo.

CONFLITO DE INTERESSES

Os autores declaram não haver conflito de interesse neste estudo.

FINANCIAMENTO

O trabalho recebeu apoio financeiro do Programa Institucional de Bolsas de Iniciação Científica do Conselho Nacional de Desenvolvimento Científico e Tecnológico (Pibic/CNPq).

REFERÊNCIAS

- Borges DR, Stella RCR. Avaliação do ensino de Medicina na Universidade Federal de São Paulo. *Rev Bras Educ Med.* 1999;23(1):7-11.
- Troncon LE, Figueiredo JF, Rodrigues ML, Peres LC, Cianflone AR, Picinato CE, et al. Im-plantação de um programa de avaliação terminal do desempenho dos graduandos para estimar a eficácia do currículo na Faculdade de Medicina de Ribeirão Preto. *Rev Assoc Med Bras.* 1999;45(3):217-24.
- Amaral E, Domingues RCL, Bicudo-Zeferino AM. Avaliando competência clínica: o método de avaliação estruturada observacional. *Rev Bras Educ Med.* 2007;31(3):287-90.
- De Souza MPG, Rangel M. Avaliação: um impasse na educação médica. *Rev Bras Educ Med.* 2003;27(3):213-22.
- Demo P. *Mitologias da avaliação: de como ignorar, em vez de enfrentar problemas.* 3rd ed. Campinas; 2010. 96 p.
- Harden RM, Gleeson FA. Assessment of clinical using an objective structured clinical ex-amination (OSCE). *Med Educ.* 1979;13(1):41-54.
- Passeri SMRR, Mazur E. Peer instruction-based feedback sessions improve the retention of knowledge in medical students. *Rev Bras Educ Med.* 2019;43(3):155-62.
- Mookher-jee S, Pheatt L, Ranji SR, Chou CL. Physical examination education in graduate medical education – a systematic review of the literature. *J Gen Intern Med.* 2013;28(8):1090-9.
- Elder A, Japp A, Verghese A. How valuable is psysical examination of the cardio vascular system?. *BMJ.* 2016;354.
- Mintz M, Southern DA, Ghali WA, Ma IW. Validation of the 25-Item Stan-ford Faculty Development Program Tool on Clinical Teaching Effectiveness. *Teach Learn Med.* 2015;27(2):174-81.
- Hsu H. *Schaum's outline of probability, random variables, and random processes.* 2nd ed. New York: McGraw-Hill; 2011.
- Elliot DL, Hickam DH. Evaluation of physical examination skills. Reliability of facul-ty observers and patient instructors. *JAMA.* 1987;258(23):3405-8.
- Noel GL, Herbers JE, Caplow MP, Cooper GS, Pangaro LN, Harvey J. How well do internal medicine faculty members evaluate the clinical skills of residents? *Ann Intern Med.* 1992;117(9):757-65.
- Newble D. Techniques for measuring clinical competence: objective structured clinical examinations. *Med Educ.* 2004;38(2):199-203.
- Mazor KM, Zanetti ML, Alper EJ, Hatem D, Barrett SV, Meterko V, et al. Assessing professionalism in the context of an objective structured clinical examination: an in-depth study of the rating process. *Med Educ.* 2007;41(4):331-40.
- Harden MCG, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using Objective Structured Examination. *Br Med J.* 1975;1(5955):447-51.
- Wilkinson TJ, Frampton CM. Comprehensive undergraduate medical assessments improve prediction of clinical performance. *Med Educ.* 2004;38(10):1111-6.
- Miller GE. The assessment of clinical skills/competence/performance. *Acad Med.* 1990;65(9):S63-S67.
- Troncon LE. Avaliação do estudante de medicina. *Medicina (Ribeirão Preto).* 1996;29(4):429-39 [acesso em 20 maio 2020]. Disponível em: <http://www.journals.usp.br/rmrp/article/view/786>.



This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.