

RBEM

Revista Brasileira de Educação Médica



Revista Brasileira de Educação Médica

Brazilian Journal
of Medical Education

Suplemento
Teste de
Progresso

ISSN versão on-line: 1981-5271
Volume 46 Suplemento 1, Set/Nov 2022



COLABORAÇÕES/CONTRIBUTIONS:

Os manuscritos devem ser encaminhados ao Editor da Revista através do site <https://mc04.manuscriptcentral.com/rbem-scielo> e seguir as "Instruções aos autores", disponíveis no sítio da revista.

All manuscript are received exclusively online, at <https://mc04.manuscriptcentral.com/rbem-scielo> and should comply with the "Instructions for Authors", in the site of the periodic.

INDEXAÇÃO/INDEXATION

A Revista Brasileira de Educação Médica é indexada nas seguintes bases de dados:
LILACS, SCIELO, EDUCA, PERIODICA, IRESIE, SUMARIOS, DOAJ, LATINDEX

The Brazilian Journal of Medical Education is indexed in the following Databases:

LILACS, SCIELO, EDUCA, PERIODICA, IRESIE, SUMARIOS, DOAJ, LATINDEX

Copyright: Associação Brasileira de Educação Médica
Home-page: <http://www.abem-educmed.org.br>

Revista Brasileira de Educação Médica, Associação Brasileira de Educação Médica, ABEM – v.46, supl. 1 (2022). – Brasília : Associação Brasileira de Educação Médica, 2023.

v. 46;
Quadrimestral
Trimestral a partir de 2008

ISSN 0100-5502

1. Educação Médica. Periódico I. Associação Brasileira de Educação Médica, ABEM.

CDD 610.7

A Revista Brasileira de Educação Médica é associada à Associação Brasileira de Editores Científicos.



A Revista Brasileira de Educação Médica e seu Editor são associados à World Association of Medical Editors.

PERIÓDICO DA ASSOCIAÇÃO BRASILEIRA DE EDUCAÇÃO MÉDICA
JOURNAL OF THE BRAZILIAN ASSOCIATION OF MEDICAL EDUCATION

EXERCÍCIO BOARD 2022 - 2024

DIRETOR PRESIDENTE (CHAIR DIRECTOR) – Sandro Schreiber de Oliveira
DIRETORA VICE-PRESIDENTE (VICE-CHAIR DIRECTOR) – Denise Herdy Afonso
DIRETORA TESOUREIRA (TREASURER DIRECTOR) – Aristides A. Palhares Neto
DIRETORA DE INOVAÇÃO (INOVATIO DIRECTOR) – Luciana Brandão Carreira
DIRETORA SECRETÁRIA (SECRETARY DIRECTOR) – Aída R. M. de Assunção
DIRETORA DISCENTE (STUDENT DIRECTOR) – Brenda Vieira Pinheiro
DIRETORA MÉDICA RESIDENTE – Djerly Marques Araújo da Silva

DIRETORES REGIONAIS REGIONAL DIRECTORS

CENTRO-OESTE – Ubirajara José Picanço de Miranda Junior
MINAS GERAIS – Alessandra Carla de Almeida Ribeiro
NORDESTE I – Estevão Toffolli Rodrigues
NORDESTE II – Daniela Chiesa
NORTE – Francisca Regina Oliveira Carneiro
RJ/ES – Paulo Roberto Alves de Pinho
SÃO PAU LO – Ugo Caramori
SUL I – Francisco Jorge Arsego de Oliveira
SUL II – Evelin Massae Ogatta Muraguchi

CONSELHO EDITORIAL EDITORIAL BOARD

Editor-Chefe: Rosiane Viana Zuza Diniz (UFRN)

Editores Associados: Antônio da Silva Menezes Junior (UFG), Cristiane Barelli (UFP), Daniela Chiesa (UNIFOR), Danilo Borges Paulino (UFU), Fernando Antonio de Almeida (PUC-SP), Gustavo Antonio Raimondi (UFU), Izabel Cristina Meister Martins Coelho (FPP), Jorge Carvalho Guedes (UFBA), Kristopherson Lustosa Augusto (UNICHRISUS), Maira Tiyomi Sacata Tongu Nazima (UNIFAP), Margareth Rodrigues Salerno (PUCRS), Maria Helena Senger (PUC-SP), Maurício Abreu Pinto Peixoto (UFRJ), Pedro Tadao Hamamoto Filho (UNESP), Roberto Zonato Esteves (UEM).

GERÊNCIA | MANAGER Rozane Landskron Gonçalves

NORMALIZAÇÃO | NORMALIZATION

Editoração Eletrônica: Jéssica Akemi Ychlsawa/Grifo Diagramações

Revisão: Carlos Vilarruel

Revisão Inglês: Ben Khon e Sonia Strong

Correspondência: ABEM - SCN – QD 02 – BL D – Torre A – Salas: 1021 e 1023
(LIBERTY MALL) – Asa Norte
CEP: 70.712-903 – Brasília – DF

Telefones: (61) 3024-9978 | (61) 3024-8013.

E-mail: rbem.abem@gmail.com

Homepage: <https://website.abem-educmed.org.br/>

TESTE DE PROGRESSO ABEM



O Teste de Progresso (TP) Abem é uma avaliação da qualidade do ensino médico, formativa, externa, de caráter institucional, indicada pela aquisição progressiva de competências pelo estudante de medicina, ao longo da sua formação, que oportuniza uma autoavaliação da escola sobre a adequação do seu projeto pedagógico para atendimento às necessidades de saúde da população brasileira e a autoavaliação do estudante.

Referenciado em bases técnico-científicas sólidas e reconhecidas internacionalmente, o TP Abem tem periodicidade regular e compromisso com a facilitação do acesso a todas as instituições. O teste é comprometido com o não ranqueamento, acolhe todas as escolas médicas associadas, é construído colaborativamente a partir de uma matriz orientada pelas Diretrizes Curriculares Nacionais, favorece o trabalho em redes, promove o desenvolvimento docente, incentiva a pesquisa científica em educação médica e incentiva a participação discente em todo o processo.

A Abem vem trabalhando para sua expansão e consolidação como estratégia colaborativa transformadora da educação médica brasileira, com base na sustentabilidade e no atendimento às demandas de formação.

Coordenação: Prof. Henry de Holanda Campos / Prof. Leandro Tuzzin

Sumário

EDITORIAL

Teste de Progresso: avanços e perspectivas	1
Progress Test: advances and prospects	3
<i>Aristides Augusto Palhares Neto, Nildo Alves Batista, Rosiane Viana Zuza Diniz, Sandro Schreiber de Oliveira, Valdes Roberto Bollela</i>	

ARTIGO ORIGINAL

Correlação entre o desempenho no Teste do Progresso e a aprovação na residência médica	5
<i>Lorena Freitas de França Guimarães, Daniela Chiesa, Olivia Andrea Alencar Costa Bessa</i>	
Teste de Progresso e mudança curricular: trilhas do curso de Medicina da Universidade de Brasília	12
<i>Odete Messa Torres, Iruena Moraes Kessler, Guilherme Viana Ferreira, Hudson Fernando Nunes Moura, Edinei Carvalho dos Santos, José Eduardo Baroneza</i>	
Teste de Progresso: a percepção do discente de Medicina	19
<i>Marlene Moraes Rosa Chinelato, Jose Eduardo Martinez, Gisele Regina de Azevedo</i>	
Onze anos de Teste de Progresso na Unicamp: um estudo sobre a validade do teste	26
<i>Ruy Guilherme Silveira de Souza, Angélica Maria Bicudo</i>	
Análise da adequação dos itens do Teste de Progresso em medicina	34
<i>Eldaine Faria de Moura Villela, Miguel Angelo Hyppolito, Julio Cesar Moriguti, Valdes Roberto Bollela</i>	
Desempenho cognitivo dos estudantes de Medicina no Teste de Progresso	41
<i>Érika Feitosa Queiroz, Olivia Andréa Alencar Costa Bessa, Daniela Chiesa</i>	

RELATO DE EXPERIÊNCIA

Como avaliar programas de residência a partir do Teste de Progresso?	48
<i>Gustavo Salata Romão, Cesar Eduardo Fernandes, Agnaldo Lopes da Silva Filho, Sheyla Ribeiro Rocha, Marcos Felipe Silva de Sá</i>	
Teste de Progresso para estudantes de Medicina: experiência de um consórcio na Região Nordeste do Brasil	57
<i>Taciana Barbosa Duque, Daniela Chiesa, Raquel Autran Coelho Peixoto, Carla Adriane Fonseca Leal de Araújo</i>	
O Teste de Progresso na residência médica em ginecologia e obstetrícia: a experiência nacional	64
<i>Gustavo Salata Romão, Cesar Eduardo Fernandes, Agnaldo Lopes da Silva Filho, Marcos Felipe Silva de Sá</i>	
Inserção do Teste de Progresso num curso de Medicina paulista: relato de uma experiência bem-sucedida	72
<i>Helena Landim Gonçalves Cristóvão, Júlio Cesar Massonetto, Luciane Lúcio Pereira, Alba Regina de Abreu Lima, Cíntia Leci Rodrigues, Júlio César André</i>	
Teste de Progresso como ferramenta de intervenção curricular na implementação de um curso de Medicina	78
<i>Elisângela de Mattos Silva, Livia Sissi Gonçalves Souza Piechnik, Mariana Xavier e Silva, Carlos Mauricio Santos, Izabel Cristina Meister Martins Coelho</i>	

Teste de Progresso: avanços e perspectivas

Progress Test: advances and prospects

Aristides Augusto Palhares Neto¹ aristides.palhares@unesp.br
Nildo Alves Batista² nabatista11@gmail.com
Rosiane Viana Zuza Diniz³ rosianevzdiniz@gmail.com
Sandro Schreiber de Oliveira^{4,5} sandro.schreiber@yahoo.com.br
Valdes Roberto Bollela⁶ vbollela@fmrp.usp.br

A partir de 2013, a Associação Brasileira de Educação Médica (Abem) passou a incentivar o uso do Teste de Progresso (TP) de maneira mais ampliada em diferentes regiões do Brasil, com o propósito de estimular a criação de núcleos regionais que o aplicam anualmente de modo regionalizado. Hoje se encontram constituídos 18 núcleos, dos quais participam mais de 220 escolas médicas brasileiras.

Em 2015, a Abem propôs um projeto com o objetivo de incentivar as escolas brasileiras a adotar o TP como uma estratégia de avaliação do domínio cognitivo da competência esperada para o exercício da medicina¹. Como resultado, nesse ano, foi realizada a primeira prova nacional que contou com a participação de 58 escolas e 23.065 estudantes².

O TP situa o estudante em seu processo evolutivo e promove oportunidades de reflexão para a aprendizagem (função formativa). Exames de progresso padronizados podem promover a melhoria do desempenho acadêmico com base na capacidade de autorregulação e metacognição, para além do pensamento crítico do educando³. O TP pode ser utilizado também como avaliação externa às disciplinas e aos estágios do curso, tendo como finalidade a tomada de decisão (função somativa) e permite à instituição realizar o diagnóstico de fortalezas e lacunas na estrutura curricular (função informativa ou diagnóstica)⁴. Participar da elaboração da prova constitui uma oportunidade única para o desenvolvimento docente nas boas práticas para a elaboração de testes de múltipla escolha com apenas uma alternativa correta.

A partir de 2019, a Abem retomou o processo de consolidação do TP nacional na perspectiva de institucionalizar um processo dirigido e organizado por ela e pelos consórcios regionais, com periodicidade anual, sem caráter classificatório. Em 2021, aconteceu o segundo TP nacional, em inédito formato *on-line*, que foi realizado por cerca de 50 mil estudantes oriundos de 130 escolas médicas⁵.

A Abem tem direcionado esforços para fortalecer essa iniciativa a fim de buscar o aperfeiçoamento, a expansão e a consolidação do TP. A construção do TP nacional tem permitido a construção de redes colaborativas de trabalho e pesquisa por meio da aproximação de sujeitos com interesses comuns, e ou da criação de verdadeiras comunidades de práticas.

Com base nesses dois momentos, a Abem realizou, em julho de 2022, a primeira oficina nacional sobre o TP na cidade de São Paulo. Entre os objetivos desse evento, merecem destaque o reconhecimento das diferentes iniciativas de aplicação do TP no Brasil e no mundo. Os principais resultados do segundo TP nacional, aplicado no ano anterior, foram apresentados e discutidos, com consequente avaliação das fortalezas e fragilidades do processo. Apresentou-se a plataforma do TP Abem. Em grupos de discussão e plenárias, fez-se uma série de propostas, e, ao final, alinharam-se e pactuaram-se perspectivas e expectativas para as próximas etapas do projeto. Na ocasião, estiveram presentes os conselhos diretor e de administração da Abem, e representantes da maioria dos núcleos de TP: Sul I – Gaúcho; Sul II – Núcleo de Apoio Pedagógico Interinstitucional Sul II (Napisul II) e Novo Sul; RJ/ES – RJ/ES; SP – Núcleo Interinstitucional de Estudos e Práticas de Avaliação em Educação Médica (Niepaem), Caipira, Paulista I, II e III; MG – Comissão do Consórcio Mineiro do Teste de Progresso (TEP Minas I e TEP Minas III); NE I – Alagoas e Baiano; NE II – Consórcio Interinstitucional Nordeste I (CIN I) e Consórcio Interinstitucional Nordeste II (CIN II); CO 1 – Centro-Oeste.

¹Universidade Estadual Paulista, Botucatu, São Paulo, Brasil.

²Universidade Federal de São Paulo, São Paulo, São Paulo, Brasil.

³Universidade Federal do Rio Grande do Norte, Natal, Rio Grande do Norte, Brasil.

⁴Universidade Federal do Rio Grande, Rio Grande, Rio Grande do Sul, Brasil.

⁵Universidade Católica de Pelotas, Pelotas, Rio Grande do Sul, Brasil.

⁶Universidade de São Paulo (FMRP-USP), Ribeirão Preto, São Paulo, Brasil.

Recebido em 03/11/22; Aceito em 03/11/22.

Eis alguns dos encaminhamentos da oficina: o fortalecimento do projeto por meio do apoio da Abem e de suas regionais, a valorização dos núcleos regionalizados, a inclusão e qualificação equânime dos atores que contribuíram para a prova nacional, o amplo envolvimento discente, o aperfeiçoamento da análise da prova e dos relatórios produzidos para as escolas e para os estudantes (nesse caso, de forma individual), a garantia da sustentabilidade financeira e o incentivo à pesquisa e difusão dos saberes e do conhecimento produzidos a partir dessa experiência.

A partir da oficina, a Abem decidiu abrir uma chamada pública para este suplemento da *Revista Brasileira de Educação Médica* (RBEM) que tem como tema: “Teste de Progresso: avanços e perspectivas na educação médica”. O objetivo desta publicação é divulgar as iniciativas exitosas que envolvem a realização do TP, de modo a reconhecer e valorizar as experiências realizadas especialmente no cenário nacional. Espera-se, dessa forma, subsidiar o desenvolvimento de novas iniciativas considerando os desafios, as potencialidades e os impactos do TP na formação na graduação e também na residência médica. As linhas temáticas foram: TP na graduação, TP na residência médica, impacto do TP na gestão dos cursos de graduação e residência médica, avaliação formativa mediada pelo TP e desenvolvimento docente para realização do TP. Os modelos de artigo possíveis foram de estudos originais quantitativos e qualitativos, e relatos de experiência.

Houve a submissão de 15 trabalhos para este número especial, os quais foram enviados aos revisores. De acordo com o processo editorial habitual da revista (*double blind review*), aceitaram-se onze artigos: seis relatos de experiência e cinco estudos originais. A temática abordada incluiu: impacto do TP no curso de Medicina analisado a partir do desempenho dos estudantes e da análise de seus desdobramentos na escola (dois artigos originais e um relato de experiência); percepção dos estudantes em relação ao TP (um artigo original); qualidade dos itens da prova como estratégia que suporta o desenvolvimento docente e a busca pela excelência (um artigo original); um relato de experiência de um consórcio na Região Nordeste do Brasil;

o TP como ferramenta preditiva de desempenho na seleção da residência médica e como ferramenta da certificação do título de especialistas em ginecologia e obstetrícia (um artigo original e um relato de experiência); e finalmente o uso do TP como ferramenta de gestão e avaliação de programas educacionais (mudança curricular) na graduação em Medicina e na residência de GO (um artigo original e dois relatos de experiência).

Esperamos que o material publicado neste número temático inspire e incentive o desenvolvimento de novas iniciativas relacionadas ao TP e fortaleça as já existentes nos diferentes contextos, graduação e residência médica, no âmbito nacional.

Desejamos a todas e todos uma boa leitura.

CONTRIBUIÇÃO DOS AUTORES

Todos os autores contribuíram para o desenvolvimento, a concepção, o desenvolvimento, a redação e revisão deste editorial.

CONFLITO DE INTERESSES

Declaramos não haver conflito de interesses.

FINANCIAMENTO

Declaramos não haver financiamento.

REFERÊNCIAS

1. Wrigley W, van der Vleuten CP, 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. doi: 10.3109/0142159X.2012.704437.
2. Bicudo AM, Hamamoto Filho PT, Abbade JF, Hafner MLMB, Maffei CML. Teste de Progresso em consórcios para todas as escolas médicas do Brasil. *Rev Bras Educ Med*. 2019;43(4):151-6.
3. Chang C, Colón-Berlinger M, Mavis B, Laird-Fick HS, Parker C, Solomon D. Medical student progress examination performance and its relationship with metacognition, critical thinking, and self-regulated learning strategies. *Acad Med*. 2021;96(2):278-84. doi: 10.1097/ACM.0000000000003766.
4. Wang L, Laird-Fick HS, Parker CJ, Solomon D. Using Markov chain model to evaluate medical students' trajectory on progress tests and predict USMLE step 1 scores: a retrospective cohort study in one medical school. *BMC Med Educ*. 2021;21:200. doi: 10.1186/s12909-021-02633-8.
5. de Oliveira SS, Postal EA, Afonso DH, Merse CE, Cyrino EG, de Abreu Junior AF, et al. Teste de Progresso da Abem: consolidando uma estratégia de avaliação para o ensino médico. *Rev Bras Educ Med*. 2022;46(1):e061.



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.

Progress Test: advances and prospects

Teste de Progresso: avanços e perspectivas

Aristides Augusto Palhares Neto¹  aristides.palhares@unesp.br
Nildo Alves Batista²  nabatista11@gmail.com
Rosiane Viana Zuza Diniz³  rosianevzdiniz@gmail.com
Sandro Schreiber de Oliveira^{4,5}  sandro.schreiber@yahoo.com.br
Valdes Roberto Bollela⁶  vbollela@fmrp.usp.br

As of 2013, the Brazilian Association of Medical Education (ABEM, *Associação Brasileira de Educação Médica*) started to encourage the use of the Progress Test in a more widespread way in different regions of Brazil, by promoting the creation of regional centers that apply the test annually in a regionalized way. Currently, 18 centers have been implemented, of which more than 220 Brazilian medical schools participate.

In 2015, ABEM proposed a project aiming to encourage Brazilian schools to adopt the Progress Test (PT) as a strategy to evaluate the cognitive domain of the expected competence for the practice of Medicine¹. As a result, the first national test was held in this year, with the participation of 58 schools and 23,065 students².

The PT situates students in their evolutionary process and promotes opportunities for reflection and learning (training function). The performance in standardized progress tests can promote the improvement in academic performance based on the capacity for self-regulation and metacognition, in addition to the student's critical thinking³. It can also be used as an external assessment of the disciplines and clerkships, with the purpose of decision-making (summative function), as well as allowing the institution to carry out the diagnosis of strengths and gaps in the curricular structure (informative or diagnostic function)⁴. Participating in the preparation of the test is a unique opportunity for Faculty development in good practices on item writing in the format of single best answer.

As of 2019, ABEM resumed the process of consolidating the national PT with a view to institutionalizing a process directed and organized by ABEM itself and its regional consortia, on an annual basis, without a classificatory nature. In 2021, the second national PT took place, in an unprecedented online format, which was performed by approximately 50,000 students from 130 medical schools⁵.

ABEM has directed its efforts to strengthen this initiative, seeking its improvement, expansion and consolidation. The construction of the national PT has allowed the creation of collaborative work and research networks, either by bringing together individuals with common interests or the creation of true communities of practice.

Based on these two moments, in July 2022, ABEM held the first national workshop on the PT, in the city of São Paulo. Among the objectives of this event, it is worth mentioning: acknowledgement of the different initiatives to apply the PT in Brazil and in the world. The main results of the second national PT, applied in the previous year, were presented and discussed, with the consequent assessment of the strengths and weaknesses of the process. The ABEM PT platform was introduced. In discussion groups and plenary sessions, a series of proposals were made and, at the end, perspectives and expectations for the next stages of the project were aligned and agreed. At the time, the Board of Directors and Directors of ABEM were present, as well as representatives from most of the progress test centers: Sul I – Gaúcho; Sul II – Núcleo de Apoio Pedagógico Interinstitucional Sul II (Napisul II) and Novo Sul; RJ/ES – RJ/ES; SP – Núcleo Interinstitucional de Estudos e Práticas de Avaliação em Educação Médica (Niepaem), Caipira, Paulista I, II e III; MG – Comissão do Consórcio Mineiro do Teste de Progresso (TEP Minas I and TEP Minas III); NE I – Alagoas and Baiano; NE II – Consórcio Interinstitucional Nordeste I (CIN I) and Consórcio Interinstitucional Nordeste II (CIN II); CO 1 – Centro-Oeste.

Some of the directions of the workshop were the strengthening of the project through the support of ABEM and its regional offices, the valorization of the regionalized centers, the inclusion and equal qualification of the actors that contributed

¹Universidade Estadual Paulista, Botucatu, São Paulo, Brazil.

²Universidade Federal de São Paulo, São Paulo, São Paulo, Brazil.

³Universidade Federal do Rio Grande do Norte, Natal, Rio Grande do Norte, Brazil.

⁴Universidade Federal do Rio Grande, Rio Grande, Rio Grande do Sul, Brazil.

⁵Universidade Católica de Pelotas, Pelotas, Rio Grande do Sul, Brazil.

⁶Universidade de São Paulo (FMRP-USP), Ribeirão Preto, São Paulo, Brazil.

to the national test, broad student involvement, improvement of the test analysis and the reports produced for schools and individually for the students, ensuring financial sustainability, and encouraging research and dissemination of knowledges and information produced from this experience.

Based on the workshop, ABEM decided to open a public call for this supplement by the Brazilian Journal of Medical Education (RBEM, *Revista Brasileira de Educação Médica*), whose topic is: Progress Test: advances and prospects in Medical Education. The objective of this publication is to publicize successful initiatives involving the performance of the PT, recognizing and valuing the experiences, especially in the national scenario. Hence, we expect to support the development of new initiatives considering the challenges, potentialities and impacts of the PT on undergraduate training and also on medical residency. The thematic lines were: the PT in undergraduate school; the PT in medical residency; impact of the PT on the management of undergraduate and residency courses; formative assessment mediated by the PT and teacher development for carrying out the PT. The possible article models included original quantitative and qualitative studies and experience reports.

A total of 15 articles were submitted for this special issue and sent to the reviewers, of which 11 were accepted, following the usual editorial process of the journal (double-blind review). Of these, six are experience reports and the other five original studies. The addressed topics included: the impact of PT on the medical course, analyzed from the performance of students and the analysis of its effects at school (two original articles and one experience report); the students' perception of the PT (an original article); the quality of the test items as a strategy that supports teacher development and the search for excellence (an original article); one experience report of a consortium in the northeast region of Brazil; the PT as a predictive tool of performance in medical residency selection and as a tool for the board certification of the title of specialist in Gynecology and

Obstetrics (one original article and one experience report); and finally, the use of the PT as a management and evaluation tool for educational programs (curricular change) in undergraduate medical school and in GO residency (one original article and two experience reports).

We hope that the material published in this thematic issue will inspire and encourage the development of new initiatives related to the PT and strengthen the existing ones in the different contexts, medical undergraduate school and medical residency, at the national level.

We wish you all a good read.

AUTHORS' CONTRIBUTIONS

All authors contributed to the development, conception, planning, writing and review of this editorial.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

SOURCES OF FUNDING

The authors declare no sources of funding.

REFERENCES

1. Wrigley W, van der Vleuten CP, 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. doi: 10.3109/0142159X.2012.704437.
2. Bicudo AM, Hamamoto Filho PT, Abbade JF, Hafner MLMB, Maffei CML. Teste de Progresso em consórcios para todas as escolas médicas do Brasil. *Rev Bras Educ Med*. 2019;43(4):151-6.
3. Chang C, Colón-Berlinger M, Mavis B, Laird-Fick HS, Parker C, Solomon D. Medical student progress examination performance and its relationship with metacognition, critical thinking, and self-regulated learning strategies. *Acad Med*. 2021;96(2):278-84. doi: 10.1097/ACM.0000000000003766.
4. Wang L, Laird-Fick HS, Parker CJ, Solomon D. Using Markov chain model to evaluate medical students' trajectory on progress tests and predict USMLE step 1 scores: a retrospective cohort study in one medical school. *BMC Med Educ*. 2021;21:200. doi: 10.1186/s12909-021-02633-8.
5. de Oliveira SS, Postal EA, Afonso DH, Merss CE, Cyrino EG, de Abreu Junior AF, et al. Teste de Progresso da Abem: consolidando uma estratégia de avaliação para o ensino médico. *Rev Bras Educ Med*. 2022;46(1):e061.



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.

Correlação entre o desempenho no Teste do Progresso e a aprovação na residência médica

Correlation between performance in the Progress Test and approval in medical residency

Lorena Freitas de França Guimarães¹  lorenafreitasfranca@gmail.com
Daniela Chiesa¹  danic2305@gmail.com
Olivia Andrea Alencar Costa Bessa¹  oliviabessa@unifor.br

RESUMO

Introdução: O Teste do Progresso (TP) permite uma avaliação acadêmica seriada com diversos benefícios para o aluno e a instituição de ensino. Trata-se de uma ferramenta complementar de avaliação. Tem-se tentado cada vez mais associar os resultados no TP com aqueles obtidos no processo seletivo da residência médica (RM), em razão de ambos medirem competências cognitivas.

Objetivo: Este estudo teve como objetivo correlacionar o desempenho no TP de egressos do curso de Medicina com o resultado obtido por eles no processo seletivo para admissão na RM.

Método: Trata-se de estudo transversal, quantitativo, realizado no período de julho de 2021 a março de 2022 com 143 alunos do curso de Medicina da Universidade de Fortaleza, que realizaram anualmente o TP no período da sua graduação e participaram do Processo de Seleção Unificada para Residência Médica do estado do Ceará.

Resultado: Foi encontrada uma correlação significativa positiva ($\rho = 0,257^{**}$; $p < 0,001$) entre a média de pontuação de todos os anos do TP e a aprovação na primeira fase da residência, bem como entre as notas do TP do sexto ano com a aprovação na primeira fase da residência ($\rho = 0,354^{**}$; $p < 0,001$). Também houve correlação positiva entre as médias do TP e as notas do TP do sexto ano com a aprovação na segunda e última fase da RM ($\rho = 0,226^{**}$; $p < 0,001$ e $\rho = 0,265^{**}$; $p < 0,001$, respectivamente).

Conclusão: Como o TP mostrou ser um excelente preditor de resultados em exames somativos de ampla concorrência, ele pode ser considerado uma ferramenta importante para o sucesso nas seleções de RM. Por conta disso, sugere-se que o TP seja cada vez mais valorizado tanto pelos alunos quanto pelas instituições que o aplicam.

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

ABSTRACT

Introduction: Considered a complementary assessment tool, the Progress Test (PT) allows for a serial academic evaluation, which benefits both the student and educational institution in several ways. There has been an increasing effort to ascertain a link between the PT and the results of medical residency (MR), as both are measurements of cognitive competences.

Objective: To correlate medical students' performance in the PT with MR admissions.

Method: Cross-sectional study conducted between November 2018 and March 2021 involving 143 students from the Universidade de Fortaleza medicine course who underwent the PT and unified MR selection process in the state.

Result: A significant positive correlation ($\rho = 0,257^{**}$; $p < 0,001$) was found between the average score for all years of the PT and passing the first phase of the MR, as well as between the PT scores achieved in the sixth year and passing the first phase of MR ($\rho = 0,354^{**}$; $p < 0,001$). There was also a positive correlation found between the average and sixth year PT scores and passing the second and final phase of MR ($\rho = 0,226^{**}$; $p < 0,001$ and $\rho = 0,265^{**}$; $p < 0,001$).

Conclusion: The PT proved to be an excellent predictor of results in summative exams with broad participation and can be considered an important tool for success in medical residency selection, suggesting that it should be increasingly valued, both by students and by the institutions that apply it.

Keywords: Medical Education; Medical Residency; Academic Performance; Educational Measurement.

¹Universidade de Fortaleza, Fortaleza, Ceará, Brasil.

Editor: Aristides Augusto Palhares Neto.

Recebido em 03/09/22; Aceito em 02/10/22.

Avaliado pelo processo de *double blind review*.

INTRODUÇÃO

O Teste do Progresso (TP) avalia o desempenho longitudinal e seriado das competências cognitivas almejadas ao final do curso, de modo a permitir avaliação individual e evolutiva dos estudantes¹. É elaborado a partir de um banco de questões produzidos pelos membros de consórcios interinstitucionais, com representantes das instituições participantes². Essas questões contemplam todo o conteúdo desejável à formação médica, abrangem as grandes áreas médicas e estimulam o raciocínio e a decisão crítica com base na Teoria de Resposta ao Item^{3,4}. É necessário que as questões sejam replicáveis e revisáveis por diversos docentes. Esse processo tem como objetivos tornar a avaliação multicêntrica e garantir a validade do processo^{5,6}. Dessa forma, diminui-se o impacto da endogenia acadêmica, que pode acontecer quando as questões das provas sofrem influência dos professores que as elaboraram⁷. O teste é aplicado no mesmo momento a todos os alunos de todos os semestres de cada escola médica participante, e anualmente um novo teste é elaborado e aplicado de forma seriada, com intervalos periódicos^{8,9}.

O TP também possibilita a análise do ensino ofertado pela instituição a partir da investigação comparativa e não ranqueadora dos resultados obtidos entre as escolas participantes de provas interinstitucionais¹⁰. Permite ainda à instituição identificar as carências da base curricular e aprimorar as estratégias de gestão do curso, para aperfeiçoar as disciplinas com menor desempenho^{6,11}. Pode ser ainda uma estratégia de avaliação institucional consistente à disposição da comunidade acadêmica e da comunidade em geral, em um momento em que a necessidade de avaliação se torna mais relevante, diante da rápida expansão do número de vagas e de escolas médicas^{12,13}.

As instituições de ensino superior (IES) idealmente devem contemplar diversas ferramentas avaliativas, de forma conjugada, a fim de facilitar o processo de ensino e aprendizagem. O TP é uma avaliação considerada formativa, por acontecer de maneira longitudinal, acompanhando a construção do aprendizado. Existem ainda as avaliações diagnósticas, bastante utilizadas para compreender quais conhecimentos os alunos já possuem, que guiam os processos de ensino dos professores, e há também as avaliações somativas, que contemplam as provas tradicionalmente aplicadas ao fim de cada ciclo, visando avaliar os objetivos expressos no currículo^{14,15}.

As avaliações somativas, dependendo das implicações dos resultados obtidos pelo estudante, podem ser consideradas de baixo impacto (*low-stakes*), quando têm implicações de menor importância, pois resultados adversos podem ser compensados de outras formas (provas parciais, por exemplo),

ou de alto impacto (*high-stakes*), quando o desempenho determina aprovação, reprovação ou classificação (provas finais, concurso vestibular, título de especialista ou seleção para residência médica – RM)¹⁶.

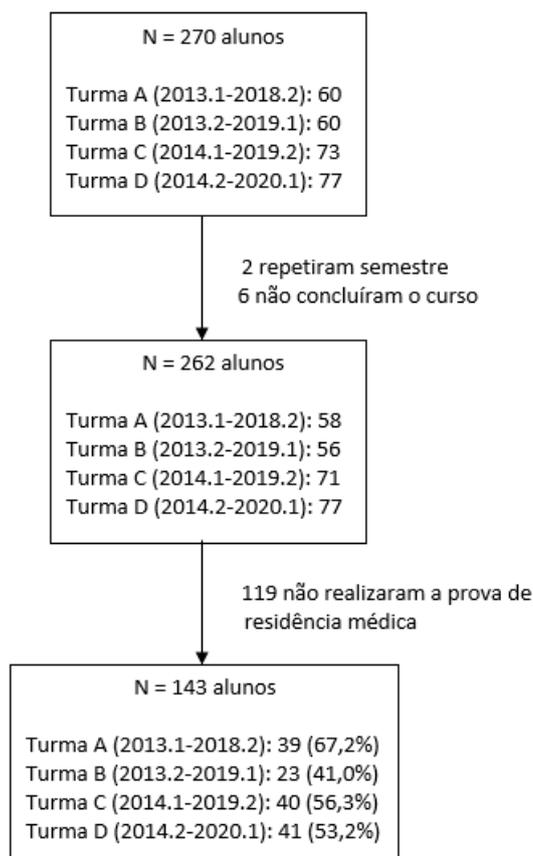
Tem-se tentado cada vez mais associar os resultados do TP com aqueles obtidos no processo seletivo da RM^{12,17}. Tais associações podem ser comparadas porque ambos medem competências cognitivas desejadas ao final do curso. Como o TP é uma avaliação seriada, ele pode extrapolar os resultados para os alunos e os cursos, de modo a relacioná-los com os objetivos destes¹⁸. O presente estudo irá abordar a correlação do desempenho dos estudantes de Medicina no TP da graduação com o resultado obtido no processo seletivo para admissão na RM.

MÉTODO

Foi realizado um estudo observacional, transversal, analítico, com abordagem quantitativa, no período de julho de 2021 a março de 2022 com os alunos do curso de Medicina da Universidade de Fortaleza (Unifor), que realizaram anualmente o TP no período da sua graduação e participaram do Processo de Seleção Unificada para Residência Médica do estado do Ceará. A amostra inicial foi constituída pelos 270 estudantes ingressantes em 2013 e 2014, que realizaram o TP do primeiro ao sexto ano do curso e fizeram concurso para RM nos anos de 2018 a 2020.

Os alunos foram categorizados em quatro turmas, de acordo com o semestre de início do curso. Como o TP é anual, todas as quatro turmas fizeram um teste por ano, totalizando seis testes ao longo da graduação. A turma A compreendeu os alunos ingressantes no primeiro semestre de 2013 (de 2013.1 até 2018.2). Como o teste é feito sempre no segundo semestre de cada ano, realizaram-se os TP nos semestres pares: segundo, quarto, sexto, oitavo, décimo e décimo segundo. A turma B ingressou no segundo semestre de 2013 (de 2013.2 até 2019.1) e realizou o TP nos semestres ímpares: primeiro, terceiro, quinto, sétimo, nono e décimo primeiro. A turma C ingressou no primeiro semestre de 2014 (de 2014.1 até 2019.2), com realização do TP nos semestres pares, e a turma D ingressou no segundo semestre de 2014 (de 2014.2 até 2020.1) e realizou o TP nos semestres ímpares. Foram excluídos os alunos que repetiram algum semestre, os que não concluíram o curso por qualquer motivo e aqueles que não prestaram prova para o processo de RM no ano de sua graduação, para evitar viés de confundimento. Desse modo, a amostra foi constituída dos resultados de 143 participantes, com a adesão ao programa de RM variando de 41% a 67,2%, a depender de cada turma. O Fluxograma 1 descreve a distribuição dos alunos recrutados para o trabalho, subdivididos nas quatro turmas (A, B, C e D), de acordo com o período de ingresso e conclusão (ano e semestre) do curso de Medicina da Unifor, considerando os critérios de exclusão do estudo.

Fluxograma 1. Distribuição dos alunos por turmas, de acordo com o ano/semestre de ingresso e de graduação



Fonte: Elaborado pelas autoras.

As informações acerca do desempenho dos estudantes no TP seriado foram coletadas no banco de dados da Unifor. Analisaram-se os dados de desempenho geral, o que inclui a nota obtida nas especialidades de clínica médica, cirurgia geral, pediatria, ginecologia-obstetrícia e saúde coletiva, totalizando 100 questões. Não se avaliaram os desempenhos na chamada área básica, que somariam 20 questões a mais, uma vez que a prova de RM não contempla esse conteúdo, o que impossibilitaria a comparação.

O desempenho no processo seletivo de acesso direto à RM foi avaliado em dois momentos, de acordo com as fases do processo. O processo seletivo da RM – Processo de Seleção Unificado para Residência Médica do estado do Ceará – PSU-RESMED/CE – é composto de duas etapas. A primeira etapa consiste na prova objetiva, com 100 questões de múltipla escolha, formuladas com igual número de questões nas cinco grandes áreas médicas: clínica médica, cirurgia geral, pediatria, ginecologia-obstetrícia e saúde coletiva. Nessa etapa, avaliou-se o resultado dos alunos, se aprovados ou reprovados para a segunda etapa. A segunda etapa consiste na análise curricular do participante classificado que obtiver 50% de acertos no exame escrito e que for selecionado dentro do quantitativo

de quatro vezes do número de vagas em cada programa. A classificação final considera a pontuação das duas etapas, em que a prova escrita tem peso 9 (90%) e a avaliação curricular peso 1 (10%), e classifica os participantes de acordo com a nota obtida. Os resultados foram computados após a segunda etapa, avaliando se os alunos foram aprovados ou reprovados de acordo com a nota de corte para cada programa.

Os dados foram obtidos de forma *on-line*, no site oficial do certame responsável pela prova da RM unificada do estado, com a anuência da coordenação do certame.

Na análise dos dados, utilizou-se o pacote estatístico Statistical Package for Social Science (SPSS), versão 25 para Windows. Realizaram-se as seguintes análises: estatísticas descritivas das pontuações no TP e da aprovação na RM; e correlações entre as pontuações do TP e a aprovação da RM e comparações das pontuações do TP em função da aprovação na residência, tanto na primeira fase como na aprovação final. Foram realizadas duas análises de regressão linear múltipla. No primeiro modelo de regressão, a nota do último ano no TP e a média geral, considerando todos os anos do teste, foram elencadas como variáveis independentes e a aprovação na primeira fase da residência como variável dependente.

As correlações entre o TP e a aprovação na RM apresentaram distribuição diferente da normal, indicando o uso de testes não paramétricos (teste de Mann-Whitney). Nas correlações, a “não aprovação” foi considerada como 0 e a “aprovação” como 1, possibilitando uma correlação ponto-bisserial. Utilizaram-se os coeficientes de correlação de Spearman (ρ) para correlações entre variáveis não quantitativas e o coeficiente de determinação (R^2) para as análises de regressão. Para possibilitar um aumento da amostra e uma análise geral, agruparam-se todas as turmas e seus dados anexados em um único banco de dados, considerando que o objetivo seja identificar ou não correlação positiva.

O projeto de pesquisa foi aprovado pelos Comitê de Ética da Unifor: Certificado de Apresentação para Apreciação Ética nº 55858521.1.0000.5052.

RESULTADOS

Na Tabela 1, são descritas as pontuações no TP de cada turma, para determinar o desempenho específico e seriado no

teste. Observamos que, ao longo dos anos, há uma progressão na média das pontuações do TP.

Os dados referentes à aprovação dos alunos por turma na primeira e segunda etapas da RM são descritos nas tabelas 2 e 3. Na primeira etapa (prova escrita), constatou-se que 38,50% (N = 55) foram reprovados, e 61,50% (N = 88), aprovados. Quanto ao final do processo, 64,30% (N = 92) foram reprovados, e 35,70% (N = 51), aprovados na residência média (Tabela 2).

Buscou-se verificar a correlação entre as médias de pontuação entre todos os anos do TP e a aprovação na primeira fase da residência. Para essa análise, a correlação foi positiva e significativa ($\rho = 0,257$; $p < 0,001$). Quando se correlacionaram as notas do último ano do TP com a aprovação na primeira fase da residência, foi possível identificar também uma correlação positiva e significativa ($\rho = 0,354$; $p < 0,001$). Entre a média de pontuação no TP e a aprovação final no processo de seleção da residência, foi encontrada uma correlação positiva e significativa ($\rho = 0,226$; $p < 0,001$). E entre as notas do TP do sexto ano e a aprovação final da RM, a correlação também foi positiva e significativa ($\rho = 0,265$; $p < 0,001$).

Tabela 1. Pontuações gerais no TP por turma

Turma	Ano	Mínimo	Máximo	Média	Desvio padrão
Turma A	1 (2013)	14	44	34,33	6,79
	2 (2014)	38	68	52,26	6,58
	3 (2015)	32	67	44,67	8,26
	4 (2016)	42	67	57,05	6,17
	5 (2017)	46	76	62,86	7,16
	6 (2018)	38	86	72,00	10,6
Turma B	1 (2013)	26	47	36,08	5,76
	2 (2014)	41	61	50,86	6,13
	3 (2015)	32	64	44,86	6,70
	4 (2016)	44	75	55,47	7,37
	5 (2017)	51	74	58,60	6,54
	6 (2018)	47	78	68,50	6,71
Turma C	1 (2014)	29	51	41,60	6,22
	2 (2015)	27	46	37,32	4,65
	3 (2016)	43	63	52,28	5,31
	4 (2017)	46	69	57,82	5,81
	5 (2018)	48	83	65,69	7,17
	6 (2019)	62	90	74,92	6,08
Turma D	1 (2014)	27	52	39,63	6,10
	2 (2015)	27	46	36,27	5,57
	3 (2016)	35	68	48,12	7,02
	4 (2017)	45	69	53,08	4,75
	5 (2018)	46	78	63,15	6,17
	6 (2019)	38	85	71,18	7,94

Fonte: Banco de dados do TP de 2013 a 2019.

Tabela 2. Análise de frequência de aprovação e reprovação na primeira e segundas etapas da RM

	Aprovados na 1ª fase	N	%	Aprovados na 2ª fase	N	%
Geral	Não	55	38,50	Não	92	64,30
	Sim	88	61,50	Sim	51	35,70
Turma A	Não	12	30,80	Não	21	53,84
	Sim	27	69,20	Sim	18	46,15
Turma B	Não	8	34,80	Não	13	56,50
	Sim	15	65,20	Sim	10	43,50
Turma C	Não	7	17,50	Não	22	55,00
	Sim	33	82,50	Sim	18	45,00
Turma D	Não	28	68,30	Não	36	87,80
	Sim	13	31,70	Sim	5	12,20

Fonte: Banco de dados do TP de 2013 a 2019 e do PSU-RESMED/CE de 2018 a 2020.

A partir de análises de regressão linear múltipla, observa-se que maiores pontuações na média do TP estão associadas à maior probabilidade de aprovação no processo da residência, seja na primeira fase [$F(2,133) = 6,390$; $p < 0,05$] ou na segunda fase da seleção [$F(2,133) = 5,313$; $p < 0,05$]. Isso indica que a média do TP é responsável por 8,80% (R^2) da aprovação na primeira fase da seleção da residência e por 7,40% da aprovação final da RM. A nota do TP do sexto ano, por si só, não se apresentou estatisticamente significativa nesse modelo de regressão, indicando que o processo como um todo é mais relevante para prever a aprovação na RM.

Com o intuito de reafirmar os resultados encontrados, buscou-se analisar apenas os 88 aprovados na primeira fase da RM e os 51 aprovados na segunda fase, a fim de comparar as pontuações que obtiveram no TP, seja na média de todos os anos ou na média da nota obtida no sexto e último ano. Verificou-se diferença estatisticamente significativa ($U = 1651,000$; $p < 0,001$) na pontuação média do TP entre os participantes que foram aprovados na primeira fase e os que não foram. Também se observou diferença estatisticamente significativa nos pontos do TP do último ano avaliado entre os participantes que foram aprovados na primeira fase e os que não foram ($U = 1278,500$; $p < 0,001$). Em ambos os casos, os participantes que foram aprovados apresentam maiores pontuações no TP do que os que não foram aprovados (Tabela 3).

Por fim, foram comparados os dados de pontuações média e do último ano no TP em função da aprovação na segunda fase da RM. Tanto para a pontuação média ($U = 1689,500$) quanto para a pontuação do último ano ($U = 1437,000$; $p < 0,05$), encontraram-se diferenças estatisticamente significativas. Em ambos os casos, os participantes aprovados no processo apresentaram maiores pontuações no TP (Tabela 4).

DISCUSSÃO

Foi confirmada a progressão das notas em cada uma das turmas analisadas, sem diferença estatística entre elas, mostrando que as turmas mantiveram padrão semelhante de evolução dos resultados. Em 2019, o estudo de Bicudo et al.² que realizou uma análise do desempenho de 23.065 estudantes no TP nacional em 2015, distribuídos por ano da graduação, mostrou uma evolução também ascendente, variando de 32,38% no primeiro ano a 62,18% no sexto ano. Quando se comparam os resultados deste estudo com os obtidos na análise de 2015, percebe-se que as quatro turmas estudadas tiveram bons rendimentos, com resultados acima da média nacional verificada no trabalho de Bicudo et al.

As associações positivas de bons resultados no TP com aprovação na RM observadas no nosso estudo indicam que maiores pontuações no TP, seja na média entre todos os anos ou somente na nota do último ano, estão relacionadas a maiores possibilidades de aprovação no processo da RM. Entretanto, foi observado que o processo do TP como um todo é mais relevante para prever a aprovação na RM do que a nota isolada do TP no sexto ano.

Observou-se ainda que um bom desempenho no TP impactou positivamente a aprovação tanto na primeira (prova cognitiva) quanto na segunda etapa da seleção da RM (avaliação curricular), embora o TP não guarde nenhuma associação com a análise curricular, não havendo correlação entre os resultados do TP e o desempenho na avaliação do currículo do estudante. O TP é uma ferramenta de avaliação cumulativa de conhecimento e é utilizado de forma a guiar a forma de estudo dos alunos³. Essa análise longitudinal permite uma boa predição de desempenho futuro e contribui para que o aluno tenha uma avaliação crítica do seu aprendizado cognitivo ao longo da graduação. Espera-se que o TP seja utilizado como

Tabela 3. Comparação entre o resultado da primeira fase da RM e os resultados do TP

Resultado da primeira fase da RM	Pontuação da média do TP	Mediana do TP	Ranking do TP	Pontuação média da nota do sexto ano do TP	Mediana	Ranking do TP
Aprovado	53,33	54,00	79,74	73,43	75,00	79,60
Reprovado	49,95	51,33	58,07	69,64	71,00	51,12

Fonte: Banco de dados do TP de 2013 a 2019 e do PSU-RESMED/CE de 2018 a 2020.

Tabela 4. Comparação entre o resultado da segunda fase da RM e os resultados do TP

Resultado da segunda fase da RM	Pontuação da média do TP	Mediana do TP	Ranking do TP	Pontuação média da nota do sexto ano do TP	Mediana	Ranking do TP
Aprovado	54,03	54,33	83,87	74,36	76,00	82,56
Reprovado	50,90	51,66	64,57	70,64	71,50	60,83

Fonte: Banco de dados do TP de 2013 a 2019 e do PSU-RESMED/CE de 2018 a 2020.

uma ferramenta de autoavaliação, e, considerando que mais avaliações devem ser feitas com o intuito final de garantir um melhor desenvolvimento das competências dos alunos, os escores obtidos sinalizam o desempenho em relação ao universo (turma/ano/semestre) em que estão inseridos.

Quando se correlaciona o desempenho do TP com os resultados positivos na prova da RM, confirma-se a vantagem de testes de alta taxonomia para avaliação de competência cognitiva, que pode prover, além do *feedback* aos estudantes, o acompanhamento curricular dos cursos^{16,18}. O estudo de Hamamoto et al.¹² com 212 estudantes da Universidade Estadual de Campinas (Unicamp) também evidenciou correlação positiva entre o desempenho no TP e a aprovação na seleção da RM.

Karay et al.¹⁸ observaram que o ganho de conhecimento progressivo verificado no TP estava associado a resultados favoráveis em exames nacionais de proficiência médica ao final da graduação de Medicina na Alemanha, o que também evidenciado foi por Andrade et al.¹⁵. Van Der Vleuten et al.¹⁰ foram além e afirmaram que o TP tem conseguido definir com precisão o desempenho dos acadêmicos de Medicina e sugerem que exames somativos após a graduação, como testes de licenciamento médico realizados em alguns países, poderiam ser eliminados, caso a escola assumisse um papel ativo no desdobramento dos resultados para seu currículo e seus estudantes. De acordo com os autores, o TP atesta o conhecimento adquirido ao longo do curso¹⁹.

Um dado relevante observado no presente estudo foi a baixa adesão dos concludentes ao processo seletivo de RM no ano da sua graduação. Vários fatores podem estar implicados nessa decisão, como programas de incentivo do governo

para trabalhar em regiões remotas ou a necessidade de obter experiência após a graduação. Não foi objetivo do estudo identificar essas causalidades, e mais estudos são necessários para verificar essas associações.

Uma das limitações deste estudo foi não considerar a concorrência e dificuldade de acesso ao programa de RM, dependente da especialidade escolhida. É notório que altas médias no TP nem sempre são suficientes para a aprovação de um aluno que optou por uma especialidade cuja concorrência por vaga estivesse bastante elevada²⁰. O índice de discriminação e o grau de dificuldade das provas não foram analisados no nosso estudo, entretanto não houve discrepâncias de resultados entre as turmas, indicando que as tais variáveis poderiam não ter influenciado neles.

CONCLUSÃO

O TP mostrou ser uma importante ferramenta para o bom desempenho no processo seletivo da RM. Desse modo, constitui-se uma forma de acompanhamento de aquisição cognitiva dos estudantes ao longo da graduação. Mesmo se tratando de uma ferramenta formativa, comprovou-se ser um excelente preditor de exames somativos de ampla concorrência, bastante almejados pelos alunos, devendo ser cada vez mais valorizado pelos graduandos e pelas instituições.

CONTRIBUIÇÃO DAS AUTORAS

Lorena Freitas de França Guimarães participou do planejamento e da escrita do projeto, da aplicação dos Termos de Consentimento Livre e Esclarecido e das cartas de anuência, da análise dos dados e da escrita do manuscrito. Daniela Chiesa participou do planejamento e da revisão do projeto, da

compilação dos dados, da edição do banco de dados para forma anônima a ser enviado para análise, da revisão da análise de dados e da revisão do manuscrito. Olivia Andrea Alencar Costa Bessa participou do planejamento e da revisão do projeto, da análise dos dados e da revisão do manuscrito.

CONFLITO DE INTERESSES

Declaramos não haver conflito de interesses.

FINANCIAMENTO

Declaramos não haver financiamento.

REFERÊNCIAS

- Schuwirth LWT, van der Vleuten CPM. The use of progress testing. *Perspect Med Educ*. 2012 Mar;1(1):24-30.
- Bicudo AM, Hamamoto Filho PT, Abbade JF, Hafner MLMB, Maffei CML. Teste de Progresso em consórcios para todas as escolas médicas do Brasil. *Rev Bras Educ Med*. 2019;43(4):151-6.
- Reberti AG, Monfredini NH, Ferreira Filho OF, de Andrade DF, Pinheiro CEA, Silva JC. Teste de Progresso na escola médica: uma revisão sistemática acerca da literatura. *Rev Bras Educ Med*. 2020;44(1):2-7.
- de Andrade MC, Strufaldi MWL, Ferreira RG, do Prado GF, Puccini RF, dos Santos AMN. Factors associated with student performance on the medical residency test. *Rev Assoc Med Bras*. 2020;66(10):1376-82.
- Wrigley W, Vleuten CPVD, Freeman A, Muijtjens A. A systemic framework for the progress test: strengths, constraints and issues. *Med Teach*. 2012;34(9):683-97.
- Sakai MH, Ferreira Filho OF, de 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.
- Pelegri T, França MTA. Endogenia acadêmica: insights sobre a pesquisa brasileira. *Estudos Econômicos*. 2020;50(4):573-610 [acesso em 24 ago 2022]. Disponível em: <https://www.revistas.usp.br/ee/article/view/165272/165873>.
- McHarg J, Bradley P, Chamberlain S, Ricketts C, Searle J, McLachlan JC. Assessment of progress tests. *Medical Education* [Internet]. Fev 2005 [citado 8 mai 2022];39(2):221-7. Disponível em: <https://doi.org/10.1111/j.1365-2929.2004.02060.x>.
- Chiesa D, Pinheiro VG, Leitão M, Cymrot M, Duque T. Constituição de consórcio interinstitucional de escolas médicas para o teste do progresso: relato da experiência de aprendizado e ação colaborativa. In: *Anais do 7th Encontro de Práticas Docentes* [Internet]; 2015 Oct 2-3; Fortaleza, CE. Fortaleza: Universidade de Fortaleza; 2015 [acesso em 1º set 2022]. Disponível em: <http://uol.unifor.br/oul/conteudosite/?cdConteudo=6133095>.
- Sakai MH, Ferreira Filho OF, Matsuo T. Avaliação do crescimento cognitivo do estudante de Medicina: aplicação do teste de equalização no Teste de Progresso. *Revista Brasileira de Educação Médica* [Internet]. 2011;35(4):493-501 [acesso em 3 jun 22]. Disponível em: <https://doi.org/10.1590/s0100-55022011000400008>.
- Van Der Vleuten CPM, Schuwirth LWT, Muijtjens AMM, Thoben AJNM, Cohen-Schotanus J, Van Boven CPA. Cross institutional collaboration in assessment: a case on progress testing. *Med Teach*. 2004 Dec;26(8):719-25.
- Hamamoto Filho PT, Lourenção PLTA, do Valle AP, Abbade JF, Bicudo AM. The correlation between students' Progress Testing scores and their performance in a residency selection process. *Med Sci Educ*. 2019 Oct 11;29(4):1071-5.
- Bica RBS, Kornis GEM. Avaliação global do ensino médico brasileiro: interesses dos atores envolvidos. *Estud Aval Educ*. 2021;32:e07592 [acesso em 1º set 2022]. Disponível em: <https://publicacoes.fcc.org.br/eae/article/view/7592>.
- Zimmermann MH, Silveira RMCF, Gomes RZ. O professor e a arte de avaliar no ensino médico de uma universidade no Brasil. *Rev Bras Educ Med*. 2019;43(3):5-15.
- Borges MC, Miranda CH, Santana RC, Bollela VR. Avaliação formativa e feedback como ferramenta de aprendizado na formação de profissionais da saúde. *Medicina (Ribeirão Preto)*. 2014;47(3):324-31.
- Bollela VR, Borges MC, Troncon LEA. Avaliação somativa de habilidades cognitivas: experiência envolvendo boas práticas para a elaboração de testes de múltipla escolha e a composição de exames. *Rev Bras Educ Med*. 2018;42(4):74-85.
- Cecilio-Fernandes D, Bicudo AM, Hamamoto Filho PT. Progress testing as a pattern of excellence for the assessment of medical students' knowledge: concepts, history, and perspective. *Medicina (Ribeirão Preto)*. 2021;54(1):e173770.
- Karay Y, Schaub SK. A validity argument for progress testing: Examining the relation between growth trajectories obtained by progress tests and national licensing examinations using a latent growth curve approach. *Medical Teacher* [Internet]. 2018;40(11):1123-9 [acesso em 5 jul 2022]. Disponível em: <https://doi.org/10.1080/0142159x.2018.1472370>.
- Van der Vleuten C, Freeman A, Collares CF. Progress test utopia. *Perspect Med Educ*. 2018 Mar 9;7(2):136-8.
- Pereira DVR, Fernandes DLR, Mari JF, Lage ALF, Fernandes APPC. Cartografia das escolas médicas: a distribuição de cursos e vagas nos municípios brasileiros em 2020. *Rev Bras Educ Med*. 2021;45(1) e005, 2021.



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.

Teste de Progresso e mudança curricular: trilhas do curso de Medicina da Universidade de Brasília

Progress Test and curricular change: paths followed by the University of Brasília Medical School

Odete Messa Torres¹  | odete.torres@unb.br
Iruena Moraes Kessler¹  | iruena@gmail.com
Guilherme Viana Ferreira¹  | guilhermeviana@unb.br
Hudson Fernando Nunes Moura¹  | HUDSONMOURA@gmail.com
Edinei Carvalho dos Santos¹  | edineicarvalho@unb.br
José Eduardo Baroneza¹  | jbaroneza@gmail.com

RESUMO

Introdução: O Teste de Progresso (TP) é um instrumento de avaliação aplicado anualmente e coordenado pela Associação Brasileira de Educação Médica (Abem). A Faculdade de Medicina da Universidade de Brasília (UnB) participa do Teste de Progresso da Região Centro-Oeste (TP-CO) desde o ano de 2013 e utiliza os dados oriundos da avaliação para diagnóstico interno e discussões sobre a formação médica.

Objetivo: Este estudo descreve os resultados obtidos pelos discentes no TP e analisa possíveis relações entre mudança do PPC e desempenho dos estudantes.

Método: Trata-se de um estudo transversal analítico, realizado a partir dos resultados obtidos pelos discentes do curso em comparação ao universo das instituições do consórcio TP-CO entre 2013 e 2021. Nesse período, participaram do teste estudantes inseridos em duas matrizes curriculares: currículo antigo (CA) e currículo novo (CN). A fim de verificar a evolução da nota média da UnB em cada período em comparação à nota média do consórcio, foi aplicado o teste de proporção para as diferenças em cada par de notas da UnB e do consórcio, em cada ano da série histórica e período dos estudantes avaliados.

Resultado: Entre 2013 e 2015, vigência do CA, não se observaram diferenças estatisticamente significativas nas médias entre os estudantes da UnB de todos os anos do curso comparados ao consórcio TP-CO. O mesmo comportamento foi verificado entre 2016 e 2018, período de transição composto por estudantes das matrizes CA e CN. Contudo, entre 2019 e 2021, vigência da nova matriz curricular, observaram-se diferenças estatisticamente significativas entre os estudantes da UnB e do consórcio.

Conclusão: O TP possibilita um olhar sobre o desempenho da escola e a identificação de lacunas que geram reflexões na busca por uma educação médica de excelência. Com a experiência apresentada, foi possível identificar as trilhas percorridas pelo curso da UnB ao longo de sua participação no consórcio, considerando as mudanças curriculares.

Palavras-chave: Educação Médica; Teste de Progresso; Currículo.

ABSTRACT

Introduction: The Progress Test (TP) is an assessment instrument applied annually and coordinated by the Brazilian Association of Medical Education (ABEM). The UnB School of Medicine has participated in the TP-CO since 2013 and uses the data from the assessment for internal diagnosis and discussions about medical education.

Objective: describe the results obtained by students in the TP and analyze possible relationships between changes in the PPC and student performance.

Method: This is an analytical cross-sectional study, based on the results obtained by the students of the course compared to the universe of institutions in the TP-CO consortium between 2013 and 2021. In this period, students enrolled in two curricular matrices sat the test: old curriculum (AC) and new curriculum (CN). In order to verify the evolution of the average score of UnB in each period compared to the average score of the consortium, the ratio test was applied to the differences in each pair of scores of UnB and consortium, in each year of the historical series and period of the students evaluated.

Results: Between the years 2013 and 2015, when the CA was in effect, no statistically significant differences were observed in the averages between the UnB students of all course years compared to the TP-CO Consortium. The same behavior was verified between 2016 and 2018, a transition period composed of CA and CN matrix students. However, between 2019 and 2021, when the new curricular matrix was in effect, statistically significant differences were observed between the UnB and consortium students.

Conclusion: The TP gives insight into the performance of the school and identifies gaps that generate reflections in the search for excellence in medical education. With the experience presented it was possible to identify the paths taken by the UnB course throughout its participation in the consortium, considering the curricular changes.

Keywords: Medical Education; Progress Test; Curriculum.

¹Universidade de Brasília, Brasília, Distrito Federal, Brasil.

Editor: Aristides Augusto Palhares Neto.

Recebido em 06/09/22; Aceito em 05/10/22.

Avaliado pelo processo de *double blind review*.

INTRODUÇÃO

As avaliações na educação médica auxiliam discentes, docentes e gestores educacionais a identificar lacunas no conhecimento dos estudantes, refletir sobre os métodos de ensino e aprendizagem, e planejar mudanças no currículo e no projeto pedagógico dos cursos. Uma das formas de avaliação aplicada anualmente a alunos de Medicina no Brasil é o Teste de Progresso (TP).

OTP foi desenvolvido no início dos anos 1970, na Kansas City School of Medicine¹ e na Universidade de Limburg², e corresponde a um tipo de avaliação repetitiva e longitudinal, aplicada em intervalos regulares a estudantes de Medicina no decorrer da formação³ e constituída por questões de múltipla escolha, com foco nos domínios de conhecimento esperados na profissão⁴.

No Brasil, o TP foi aplicado inicialmente pela Comissão Interinstitucional Nacional de Avaliação do Ensino Médico (Cinaem) em 1999, em caráter ampliado, em 60 escolas médicas⁵. Atualmente, o teste é coordenado pela Associação Brasileira de Educação Médica (Abem) e adotado por instituições conveniadas, organizadas em consórcios interinstitucionais⁶. O teste, realizado por estudantes inseridos em todos os anos da graduação, tem por finalidade avaliar a instituição e o desempenho cognitivo dos discentes, sem intenção de seleção ou classificação destes⁶, uma vez que a aplicação de instrumentos de avaliação em larga escala não é capaz de cobrir todas as dimensões necessárias para aferir a qualidade da educação médica⁷.

Ao fornecer dados longitudinais acerca do progresso dos estudantes de Medicina às instituições de ensino superior (IES), o TP permite monitorar o impacto das mudanças curriculares e planejar ajustes⁸. Ao reconhecerem o TP como instrumento de avaliação na educação médica, as escolas institucionalizam uma estratégia de avaliação de desempenho discente e da instituição, de modo a possibilitar indicadores para as mudanças curriculares.

O curso de Medicina da Universidade de Brasília (UnB) iniciou suas atividades em 1966, quatro anos após a criação da instituição. Desde então, com base nas transformações políticas, econômicas e sociais brasileiras, e também nas especificidades da demografia, da epidemiologia e do contexto da saúde do Distrito Federal, foram implantados três projetos pedagógicos de curso (PPC) – em 1966, 1988 e 2016 –, os quais são periodicamente revisados e atualizados, considerando as novas possibilidades de organização curricular⁹. Durante o século XX, os currículos que nortearam o curso de Medicina da UnB tiveram inegável influência da reforma universitária de 1968¹⁰ e do modelo norte-americano flexneriano¹¹.

Apoiadas na modernização do ensino médico, as Diretrizes Curriculares Nacionais (DCN) de 2014 instruíram, entre outras orientações, que a estrutura dos cursos de graduação em Medicina deveria privilegiar os métodos ativos de ensino, incluir as dimensões ética e humanística, incentivar o uso de tecnologias da informação e da comunicação, promover a integração entre os eixos curriculares e a interdisciplinaridade, inserir o aluno nas redes de atenção em saúde do Sistema Único de Saúde (SUS), desde as séries iniciais, além de vincular a formação do médico às necessidades sociais e de saúde. Quanto ao projeto pedagógico, as referidas DCN orientam para a centralidade no estudante como sujeito de aprendizagem, de modo que caberia ao professor utilizar métodos e tecnologias adequados para facilitar e mediar os processos de ensino-aprendizagem¹².

Na UnB, com o intuito de atender às DCN, a partir de 2016 foi implantado um novo PPC que valoriza a interdisciplinaridade, a integração curricular horizontal e vertical, os métodos ativos de ensino, a abordagem dos problemas de saúde não somente no contexto individual, mas também familiar e grupal, o enfoque na promoção de ambientes e estilos de vida saudáveis, e a ampliação dos cenários de atenção para além do hospital e dos consultórios, com foco na atenção primária à saúde, incluindo as unidades básicas de saúde, escolas, creches, equipamentos comunitários, locais de habitação e ambientes de trabalho⁹.

A Faculdade de Medicina da UnB (FM/UnB) participa do TP-CO desde o ano de 2013 e utiliza os dados oriundos da avaliação dos estudantes para diagnóstico interno e para discussões sobre os eventuais problemas na formação. Ao longo da participação no consórcio, a UnB colaborou com o desenvolvimento das questões e com a validação delas por representação em reuniões da regional Brasília, além de aplicar e bonificar o estudante que participa do teste.

As reflexões acerca do diagnóstico institucional no TP-CO são realizadas pelos integrantes do Grupo de Trabalho de Avaliação do Processo de Ensino e Aprendizagem, vinculado ao Núcleo Docente Estruturante (NDE) do curso de Medicina da UnB e pelo Colegiado de Graduação. Nesse processo de transição curricular, analisar a participação dos discentes do curso de Medicina da UnB no consórcio TP-CO se faz imperativo para depreender os caminhos percorridos ante aqueles que se delongam na jornada de formar médicos comprometidos com a realidade social e defensores do SUS. Este artigo tem como objetivo descrever os resultados obtidos pelos discentes do curso de Medicina da UnB no TP e analisar possíveis relações entre a mudança do PPC implementada no ano de 2016 com o desempenho dos estudantes.

MÉTODOS

Trata-se de um estudo transversal analítico, realizado a partir dos resultados do curso de Medicina da UnB e do universo das instituições associadas ao consórcio TP-CO entre 2013 e 2021. Nesse período, participaram do teste estudantes da UnB inseridos em duas matrizes curriculares distintas: currículo antigo (CA) e currículo novo (CN).

Considerando que o CN teve início a partir de 2016, entre 2013 e 2015 participaram do TP apenas estudantes inseridos no CA. Entre 2016 e 2019, entretanto, participaram do TP estudantes de ambos os currículos, uma vez que aqueles que ingressaram no CA continuaram neste, embora os ingressantes após 2016 já tivessem sido inseridos no CN. Por fim, no ano de 2021 participaram somente discentes matriculados em componentes do CN.

Os dados apresentados neste trabalho foram retirados dos relatórios do Sistema de Administração do Teste de Progresso (Versão: 2.4.1) em agosto de 2022, que permite o acesso às médias dos estudantes da UnB e do universo das instituições associadas em cada ano de realização da prova. As informações foram tabuladas no Microsoft Excel 2016.

Com o objetivo de verificar a evolução da nota média da UnB em cada período em comparação à nota média do consórcio, foi aplicado o teste de proporção para as diferenças em cada par de notas da UnB e do consórcio, em cada ano da série histórica e período dos estudantes avaliados. Dessa forma, testaram-se 48 hipóteses de diferença das médias, cujos resultados serão discutidos na próxima seção. Os testes estatísticos foram aplicados utilizando o *software* R, com o uso da função *prop.test*, do pacote *stats*.

Além disso, de forma a propiciar a visualização da comparação das notas médias da UnB e do consórcio, com a informação de CN e CA, elaboraram-se dois gráficos no Excel que explicitam a evolução nos anos analisados. As legendas

especificam, no caso da UnB, qual currículo de cada turma está sendo analisado em cada ano do TP-CO.

Este trabalho foi aprovado pelo Comitê de Ética em Pesquisa da Faculdade de Medicina da UnB com o número 4.933.262.

RESULTADOS E DISCUSSÃO

Desde o início do consórcio, entre 2013 e 2021, o curso de Medicina da UnB participou de todas as edições do TP-CO. O número de participantes em cada ano de prova e em cada ano de curso variou de acordo com os números apresentados na Tabela 1.

No ano de 2020, não ocorreu o TP-CO por causa da emergência sanitária da *coronavirus disease 2019* (Covid-19). Quando se analisa o desempenho dos estudantes do primeiro ao sexto ano no TP-CO, entre 2013 e 2021 (Tabela 2), é possível inferir que entre 2013 e 2015, período de vigência da matriz do CA, não se observaram diferenças estatisticamente significativas entre os estudantes da UnB de todos os anos do curso comparados ao consórcio TP-CO. O mesmo comportamento foi verificado entre 2016 e 2018, período de transição composto por estudantes da matriz do CA e do CN.

Entre 2019 e 2021, com vigência da nova matriz curricular, observaram-se diferenças estatisticamente significativas entre os estudantes da UnB e do consórcio. Em 2019, somente os estudantes do quarto ano da UnB tiveram um desempenho superior ao do consórcio (p -valor $< 0,10$). Em 2021, os estudantes do quarto, quinto e sexto anos da UnB tiveram um desempenho significativamente superior ao do consórcio com p -valores menores que 0,10, 0,05 e 0,01, respectivamente (Tabela 2 e Gráfico 2).

Na apresentação gráfica dos resultados, optou-se pela leitura deles em dois blocos: evolução do primeiro ao terceiro ano e evolução do quarto ao sexto ano.

Tabela 1. Participação de estudantes da UnB no TP-CO entre 2013 e 2021

Ano	Número de alunos participantes de acordo com o ano do curso						Total 1 (T1) de alunos participantes por ano	Total 2 (T2) de alunos matriculados por ano	Percentual de T1/T2
	1º	2º	3º	4º	5º	6º			
2013	60	38	37	51	41	6	233	548	42,5%
2014	14	51	29	29	31	1	155	555	27,9%
2015	47	44	59	51	44	19	264	570	46,3%
2016	50	65	54	65	39	45	318	581	54,7%
2017	51	60	68	48	67	46	340	593	57,3%
2018	63	60	58	66	49	66	362	604	59,9%
2019	62	67	63	51	61	48	352	590	59,7%
2020	28	62	73	63	64	24	314	645	48,7%

Fonte: Elaborada pelos autores.

Tabela 2. Desempenho dos estudantes de Medicina da UnB e do consórcio TP-CO do primeiro ao sexto ano, entre 2013 e 2021

Ano	2013		2014		2015		2016		2017		2018		2019		2021	
	UnB	CO	UnB	CO	UnB	CO	UnB	CO	UnB	CO	UnB	CO	UnB	CO	UnB	CO
Primeiro	33,3	35,6	32,6	31,8	32,9	34,2	34,2	30,6	33,3	32,1	35,3	32,4	32,6	31,1	35,3	32,3
Segundo	38,0	39,6	37,3	38,1	43,7	39,9	34,4	34,7	38,1	36,9	39,4	37,5	36,7	35,8	40,9	35,5
Terceiro	43,3	46,2	44,2	44,2	45,2	46,1	40,9	41,3	44,3	43,9	45,4	42,7	47,6	41,2	45,8	38,9
Quarto	51,3	51,9	52,6	49,4	51,9	54,2	47,6	46,4	52,1	51,1	52,4	48,3	56,5*	47,1*	52,2*	42,7*
Quinto	63,2	56,5	57,8	54,5	61,0	57,9	53,1	51,8	56,7	54,3	57,5	53,3	59,9	52,1	57,5**	47,1**
Sexto	66,3	58,8	55****	59,9	69,1	62,3	61,0	56,7	64,8	60,7	62,5	56,9	61,5	58,0	69,3***	52,2***

Para a obtenção do p-valor, utilizou-se o teste de proporções para as diferenças. * p-valor < 0,10; ** p-valor < 0,05; *** p-valor < 0,01; **** o valor não corresponde a uma média, mas à nota do único aluno que realizou o teste.

Fonte: Elaborada pelos autores.

Evolução do primeiro ao terceiro ano

O Gráfico 1 apresenta o desempenho dos estudantes dos três primeiros anos do curso de Medicina da UnB no TP-CO, bem como os resultados do universo de instituições participantes, considerando as médias das notas que os discentes alcançaram nas provas. Para cada ano em que o teste foi realizado, consideraram-se as médias dos alunos matriculados no primeiro, segundo e terceiro anos do curso de Medicina, períodos em que normalmente se valoriza no currículo sobre-tudo a abordagem de assuntos pertinentes à formação básica em ciências biológicas e da saúde.

No Gráfico 1, para cada curva representando os resultados da UnB, consideraram-se, em tonalidade mais clara, a média dos acertos dos alunos inseridos no CA e, em tons mais escuros, a média dos acertos dos alunos inseridos no CN, delimitando, visualmente, a implementação da mudança curricular vigente na UnB.

A partir da análise do gráfico, é possível constatar que as médias de acertos dos estudantes do primeiro ano do curso de Medicina nos testes de 2013, 2014 e 2015, período anterior à implementação da mudança curricular, está em geral abaixo da média do consórcio da Região Centro-Oeste, com desvio acima da média geral apenas nas turmas de segundo ano no teste de 2015. Após a mudança curricular, nos testes de 2016, 2017, 2018, 2019 e 2021, observa-se que a média de desempenho dos estudantes da instituição supera em todos os anos, a partir de 2017, a média de desempenho do consórcio. Observa-se ainda que há uma oscilação tanto no desempenho dos estudantes da UnB como do consórcio, perceptíveis nos intervalos entre os anos com curvas em queda ou ascensão, ou seja, não há um desempenho progressivo ao longo do curso, conforme prevê a literatura.

O teste de 2013 apresenta o menor desempenho da UnB dos três primeiros anos do curso, todos abaixo da média do consórcio. Em relação ao desempenho, o primeiro ano da UnB supera a média do consórcio em 2014, seguido do terceiro ano que se equipara ao consórcio no mesmo ano e da manutenção do desempenho negativo do segundo ano da UnB em relação ao consórcio na prova de 2014. Esse desempenho segue inferior ao teste do consórcio de 2015 nas turmas de primeiro e terceiro anos, e as turmas de segundos anos se destacam em relação do desempenho do consórcio em 2015. No ano de 2016, quando o CN é implementado, a UnB se destaca em relação ao consórcio com o maior desempenho das turmas de primeiro ano, ao passo que mantém a média do consórcio no desempenho das turmas de terceiro e quarto anos. Observa-se que a partir de 2017, consecutivamente até 2021, as turmas de primeiros, segundos e terceiros anos superam em desempenho as médias do consórcio TP-CO. As oscilações perceptíveis no gráfico, em geral, são acompanhadas pela UnB e pelo consórcio, o que permite inferir que remetem ao estilo de prova aplica nos anos analisados, à exceção do teste de 2019, que evidencia queda de desempenho das turmas de terceiro ano do consórcio, ao passo que os estudantes dos mesmos períodos de curso na UnB apresentam desempenho progressivo em 2019 com queda em 2021, a qual ainda assim supera as médias do TP-CO.

Evolução do quarto ao sexto ano

O Gráfico 2 apresenta o desempenho dos estudantes de Medicina da UnB e do consórcio da Região Centro-Oeste dos três últimos anos (do quarto ao sexto ano), considerando as médias das notas que os discentes alcançaram nas avaliações. Para cada ano em que o teste foi realizado,

consideraram-se as médias dos alunos matriculados no quarto, quinto e sexto anos do curso de Medicina, períodos em que normalmente se valorizam os saberes clínicos e o internato médico.

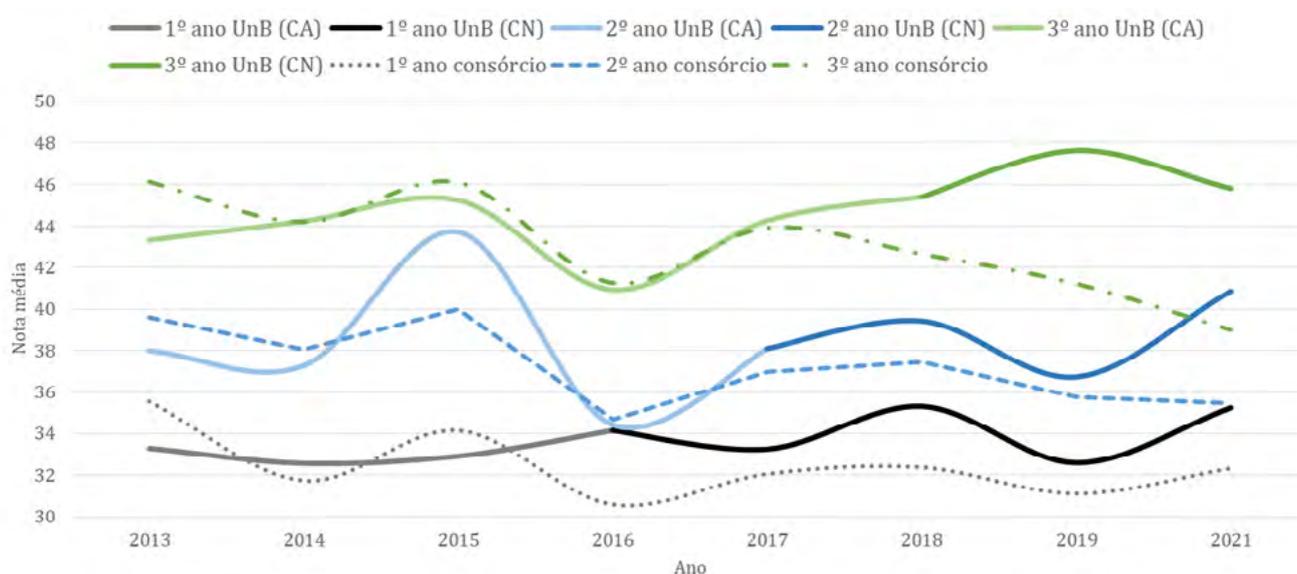
Pela visualização do gráfico, de 2013 a 2018 os estudantes dos três últimos anos do curso eram oriundos do CA. No teste de 2019, apenas os estudantes do quarto ano cursam o CN. Os estudantes dos quintos e sextos anos que realizaram os testes no período de 2013 a 2019 são oriundos do CA. O TP-CO de 2021 é o primeiro teste em que todos

os estudantes da UnB estão cursando o CN por conta da migração de turmas anteriormente apresentada.

A partir da análise do gráfico, é possível depreender que as turmas do quinto ano do curso de Medicina da UnB mantiveram melhor desempenho em relação às médias do consórcio TP-CO em todos os anos do teste, e, ainda, que as oscilações das curvas do mesmo período de formação ocorrem próximas às oscilações do consórcio.

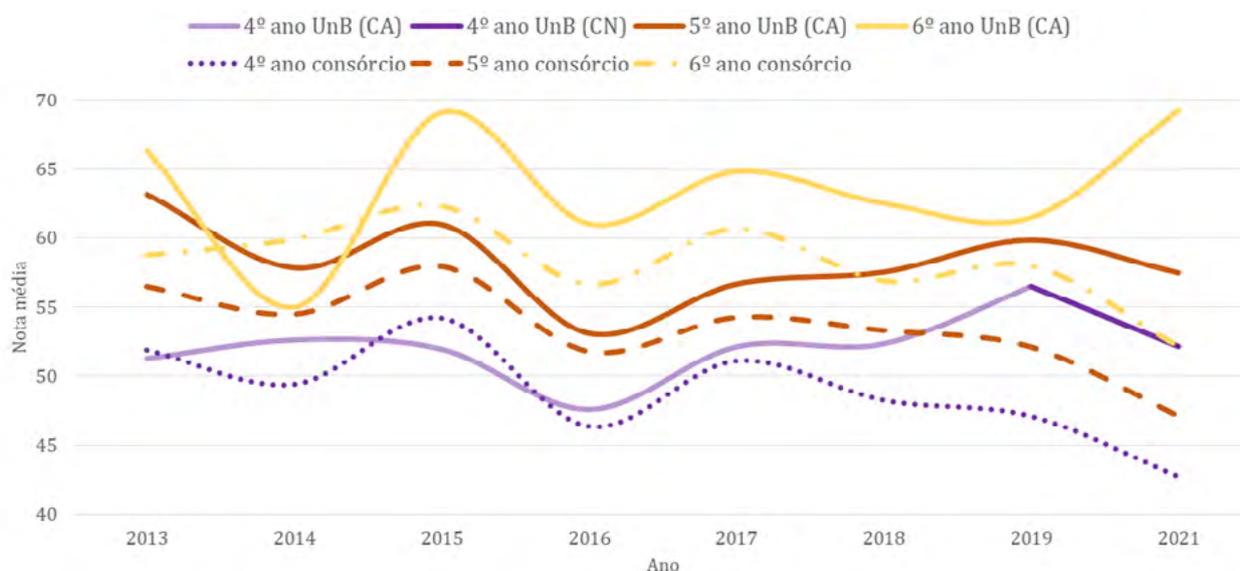
Os anos de 2013 a 2015 apresentam desempenhos aquém ao TP-CO: nas turmas de quarto ano em 2013; quinto e

Gráfico 1. Desempenho dos estudantes de Medicina da UnB e do consórcio da Região Centro-Oeste (do primeiro ao terceiro ano)



Fonte: Elaborado pelos autores.

Gráfico 2. Desempenho dos estudantes de Medicina da UnB e do consórcio da Região Centro-Oeste (do quarto ao sexto ano)



Fonte: Elaborado pelos autores.

sexto anos em 2014; e quarto ano em 2015. Maior desempenho em relação ao TP-CO: nas turmas de quinto e sexto anos em 2013; quartos e quintos anos em 2014; quintos e sextos anos em 2016. Após o teste de 2016, todas as turmas de quarto, quinto e sexto anos do curso de Medicina da UnB apresentaram maior desempenho quando comparadas às médias do consórcio TP-CO no mesmo período. Os anos de 2018, 2019 e 2020 evidenciam um destaque ao desempenho da UnB nos três últimos anos de curso, com maiores intervalos de desempenho da UnB no ano de 2021.

Entre 2013 e 2015, os estudantes da FM/UnB cursavam a matriz curricular 7111/-3 ou CA, que durava em média seis anos para ser cumprida, composta por ciclo básico (do primeiro ao quarto semestre), fase clínica (do quinto ao nono semestre) e internato (do nono ao 12º semestre). O ciclo básico era centrado em conhecimentos teóricos básicos de anatomia, embriologia, histologia bioquímica e farmacologia. A fase clínica era compartimentalizada em especialidades, como pediatria, ginecologia e semiótica, havendo uma pequena carga de práticas. Somente no internato a prática médica implementava-se de forma efetiva.

A mudança do CA para o CN em 2016 foi significativa. Houve redução expressiva na carga de optativas, e o curso de graduação em Medicina organizou-se a partir da consolidação de cinco eixos temáticos obrigatórios e um eixo temático optativo: 1. Estudo morfofuncional do ser humano; 2. processos do adoecimento do ser humano; 3. atenção à saúde do ser humano; 4. saúde coletiva e saúde da família e comunidade; 5. conhecimento de si e do outro; e 6. optativo.

Apesar da mudança significativa entre o CN e o CA, não houve interferência no desempenho dos alunos da UnB e daqueles do consórcio até 2019, exceto os do quarto ano. Observa-se um equilíbrio entre o desempenho dos alunos do consórcio e o desempenho dos discentes da UnB, o que pode estar relacionado à homogeneidade no aprendizado entre as IES, sem prejuízos aparentes ou diferencial acadêmico entre os participantes. Contudo, entre 2019 e 2021, destaca-se uma *performance* significativamente melhor da UnB em relação aos demais alunos do consórcio, o que sugere uma interferência positiva do CN em relação ao CA.

Ao avaliar a instituição, o TP permite analisar a relação entre o conteúdo e a estrutura curricular do curso de graduação em Medicina e o desenvolvimento individual e conjunto das turmas de estudantes. Ao avaliar o estudante, o TP possibilita contribuir com a avaliação formativa discente, a partir da evolução do desempenho cognitivo do estudante nas diversas áreas do curso. Com base nos resultados do TP, tanto o estudante quanto o curso podem se utilizar de seus resultados na tomada de decisão para a qualificação do discente e do curso em análise. Ainda, o TP serve como treinamento aos processos de

seleção da carreira médica, como os concursos para os cursos de residência médica⁵.

Por fim, pode-se destacar que a análise dos intervalos de confiança de cada edição do TP, em cada série avaliada, comparando os números da UnB e do consórcio, seria de grande importância para a discussão aqui levantada, até pela baixa participação, em alguns anos, de algumas turmas. No entanto, dada a natureza da disponibilização dos dados, que já vieram consolidados, não foi possível efetuar o cálculo de medidas de dispersão adequadas para auxiliar a análise, como o desvio padrão ou intervalo de confiança.

CONCLUSÕES

O TP é um poderoso instrumento de avaliação que pode auxiliar no monitoramento da *performance* acadêmica dos estudantes após a modificação de matrizes curriculares. O presente estudo relacionou os resultados dos discentes da FM/UnB no TP-CO entre 2013 e 2021 com os resultados dos discentes das instituições conveniadas no consórcio. Concluímos que as notas dos discentes da UnB evoluíram satisfatoriamente e acima da média das instituições conveniadas ao consórcio TP-CO, considerando sobretudo os dados do universo de alunos inseridos no CN, adotado a partir do ano de 2016.

O movimento gerado nas escolas médicas pelo teste por si é desencadeador de mudanças na instituição, no corpo docente e no corpo discente. O TP possibilita um olhar sobre o desempenho da escola, comparado ao contexto de formação médica, regional ou nacional, em um mesmo período e com a aplicação da mesma avaliação seriada. Ao identificar lacunas, as escolas possuem escolhas a percorrer na busca pela educação médica de excelência e da garantia de compromisso social, orientadores das mudanças curriculares, segundo as DCN. Além das escolas, a comunidade é beneficiada ao participar do TP. O corpo docente, ao se envolver na construção de questões e na validação delas dentro do consórcio, apreende novos olhares sobre os temas e modos de avaliar que desenvolve em sala de aula. O docente amplia a sua capacidade de formulação de questões com distratores que valorizam o pensamento crítico e não apenas a apreensão de conteúdo pelo estudante. Este, por sua vez, para quem a formação deve se valer, ao participar do TP acompanha sua evolução no curso e dimensiona sua avaliação dentro do contexto da escola e na relação desta com os âmbitos regional ou nacional. Paralelamente, as escolas precisam se preparar para os *feedbacks* e mecanismos de apoio individual, priorizando aqueles estudantes cujo

desempenho se coloca abaixo da média da escola e/ou da região/país. O auxílio psicopedagógico deve ser proposto pelas escolas e hoje está sendo debatido na UnB com o NDE e o Núcleo de Apoio Psicopedagógico e Bem-Estar do Estudante de Medicina (Napem).

Com a experiência apresentada, foi possível identificar as trilhas percorridas pelo curso de Medicina da UnB ao longo de sua participação no consórcio da Região Centro-Oeste do TP, bem como aquelas por cujos caminhos o currículo trespassou.

CONTRIBUIÇÃO DOS AUTORES

Todos os autores participaram da concepção e do *design* do trabalho, do preparo e da revisão do manuscrito.

CONFLITO DE INTERESSES

Declaramos não haver conflito de interesses.

FINANCIAMENTO

Declaramos não haver financiamento.

REFERÊNCIAS

1. Arnold L, Willoughby TL. The quarterly profile examination. *Acad Med.* 1990; 65(8):221-7.
2. Van der Vleuten CPM, Verwijnen GM, Wijnen WHFW. Fifteen years of experience with progress testing in McMaster University's problem-based medical curriculum: psychometric properties and effect on learning. *Acad Med.* 1996; 71(9):1002-7.
3. Blake JM, Norman GR, Keane DR, Mueller CB, Cunningham J, Didyk N. Introducing progress testing in McMaster University's problem-based medical curriculum: psychometric properties and effect on learning. *Acad Med.* 1996;71(9):1002-7 [acesso em 16 ago 2022]. Disponível em: https://journals.lww.com/academicmedicine/Abstract/1996/09000/Introducing_progress_testing_in_McMaster.16.aspx.
4. Neeley SM, Ulman CA, Sydelko BS, Borges NJ. The value of progress testing in undergraduate medical education: a systematic review of literature. *MedSci Educ.* 2016; 26:617-22.
5. Sakai MH, Ferreira Filho OF, Almeida JM, Mashima DA, Marchese MC. Teste de Progresso e avaliação do curso: dez anos de experiência da medicina da Universidade Estadual de Londrina. *RevBrasEducMed.* 2008;32(2):254-63.
6. Bicudo AM, Hamamoto Filho PT, Abbade JF, Hafner MLMB, Maffei CML. Teste de Progresso em consórcios para todas as escolas médicas do Brasil. *RevBrasEduc Med.* 2019;43(4):151-6.
7. Batista NA, Vilela RQB, Batista SHSS. Educação médica no Brasil. São Paulo: Cortez;2015.
8. Coombes L, Ricketts C, Freeman A, Stratford J. Beyond assessment: feedback for individuals and institutions based on the progress test. *MedTeach.* 2010;32:486-90.
9. Universidade de Brasília. Projeto pedagógico do curso. 2015 [acesso em 16 ago 2022]. Disponível em: <http://fm.unb.br/graduacao/projeto-pedagogico-do-curso>.
10. Brasil. Lei nº 5.540, de 28 de novembro de 1968. Fixa normas de organização e funcionamento do ensino superior e sua organização com a escola médica, e dá outras providências. *Diário Oficial da União*;1968.
11. Almeida Filho N. Reconhecer Flexner: inquérito sobre produção de mitos na educação médica no Brasil contemporâneo. *CadSaude Publica.* 2010;26(12):2234-49.
12. Brasil. Resolução CNE/CES nº 3, de 20 de junho de 2014. Institui Diretrizes Curriculares Nacionais do Curso de Graduação em Medicina e dá outras providências. *Diário Oficial da União*; 2014.



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.

Teste de Progresso: a percepção do discente de Medicina

Progress Test: the perception of the medical student

Marlene Moraes Rosa Chinelato^{1,2}

marlenechinelato@fev.edu.br

Jose Eduardo Martinez²

jemartinez@pucsp.br

Gisele Regina de Azevedo²

grazevedo@pucsp.br

RESUMO

Introdução: O Teste de Progresso (TP) como instrumento de avaliação na educação médica constitui-se em relevante subsídio para avaliar a eficiência do programa. Contudo, a percepção do aluno quanto ao seu desempenho e o impacto do TP na aprendizagem podem variar de acordo com contextos pessoais, educacionais, sociais e culturais.

Objetivo: Esta pesquisa descritiva de abordagem qualitativa objetivou analisar a percepção dos estudantes do curso de graduação em Medicina de um centro universitário do noroeste paulista sobre o seu desempenho no TP, bem como o impacto dessa percepção em curto prazo sobre suas estratégias de estudo.

Método: A amostra de conveniência foi constituída por 20 participantes após aprovação da pesquisa no Comitê de Ética em Pesquisa. Utilizou-se a técnica de grupos focais em dois momentos distintos para a coleta de dados: um grupo foi constituído por dez estudantes do quinto período, e o outro, por dez estudantes do oitavo período. A análise de dados fundamentou-se na análise de conteúdo temática descrita por Bardin et al.

Resultado: Identificou-se que os estudantes entrevistados consideraram: 1. as condições de realização do TP inadequadas; 2. o TP uma ferramenta pedagógica relevante que permite a autoavaliação e a correção das lacunas de aprendizagem, mas sugerem que seja aprimorado; 3. a participação no TP determinou sentimentos contraditórios para os estudantes, conforme o período que estão cursando.

Conclusão: O TP é considerado pelos estudantes uma “ferramenta pedagógica” relevante. Entretanto, o impacto dos resultados de desempenho obtidos e o feedback recebidos não promoveram em curto prazo mudanças no plano de estudo dos estudantes entrevistados. A prática da autoavaliação ainda não se constitui cultura na vida acadêmica. É preciso considerar novas estratégias para a entrega do feedback formativo que permita a discussão das questões e dos resultados do curso/das turmas de forma reflexiva, de modo a potencializar o processo ensino-aprendizagem.

Palavras-chave: Educação Médica; Aprendizagem; Avaliação Educacional; Estudantes de Medicina; Percepção.

ABSTRACT

Introduction: *The Progress Test (PT) as an assessment tool in medical education is an important tool to evaluate the efficiency of the program. However, students' perception of their performance and the impact of the PT on learning may vary according to personal, educational, social and cultural contexts.*

Objective: *This descriptive research with a qualitative approach aimed to analyze the perception of undergraduate medical students of a University Center in northwestern São Paulo about their performance in the PT, as well as the impact of this perception in the short term on their study strategies.*

Method: *The convenience sample was composed of twenty participants, after approval of the research by the Research Ethics Committee. The focus group technique was used in two different moments for data collection; one group consisted of ten fifth-period students and the other, of ten eighth-period students. Data analysis was based on content analysis, thematic modality, as described by Bardin (2011).*

Results: *It was identified that the interviewed students consider: 1. the PT performance conditions as inadequate; 2. the PT a relevant pedagogical tool, which allows self-assessment and the correction of learning gaps, but they suggest that it should be improved; 3. sitting the PT determined contradictory feelings for the students, according to their period of study.*

Conclusion: *The Progress Test is considered by students a relevant “pedagogical tool”. However, the impact of the performance results obtained and the feedback received did not promote changes in the study plan of the interviewed students in the short term. The practice of self-assessment is not yet a culture in academic life. It is necessary to consider new strategies for the delivery of formative feedback, allowing for reflective discussion of the issues and the course/class results in order to enhance the teaching-learning process.*

Keywords: *Medical Education; Learning; Educational Measurement; Students Medical; Perception.*

¹ Centro Universitário de Votuporanga, Votuporanga, São Paulo, Brasil.

² Pontifícia Universidade Católica de São Paulo, Sorocaba, São Paulo, Brasil.

Editor: Aristides Augusto Palhares Neto.

Avaliado pelo processo de double blind review.

Recebido em 05/09/22. Aceito em 04/10/22.

INTRODUÇÃO

O Teste de Progresso (TP) tem sido aplicado desde os anos 1970, com a finalidade de avaliar longitudinalmente o desenvolvimento cognitivo dos estudantes de Medicina. Esse processo não tem caráter de aprovação, reprovação ou classificação, mas possibilita a tomada de decisões curriculares assertivas e a autoavaliação pelo estudante^{1,2}. Essa avaliação, habitualmente, utiliza questões objetivas de múltipla escolha construídas conforme os conteúdos que compõem a matriz curricular do curso. É aplicada, simultaneamente, com a mesma composição para todos os estudantes do primeiro ao último período curricular, refletindo o conteúdo final do curso^{3,4}.

O TP, como instrumento de avaliação na educação médica, apresenta características singulares⁵. Ao estudante, possibilita uma análise crítica do seu desempenho no curso, identificando fragilidades e potencialidades na sua formação a cada etapa em que o teste é realizado⁶. Para os docentes, o resultado apresenta a média da turma em relação à etapa, possibilitando modificações nos processos e adequações nos conteúdos. À gestão acadêmica, permite a identificação de potenciais problemas, possibilitando a tomada de decisão assertiva e implementação de ações de correção^{2,3}.

Com os objetivos de avaliar o processo ensino-aprendizagem e garantir a qualidade e execução dos objetivos curriculares, o curso de Medicina do Centro Universitário de Votuporanga (Unifev) instituiu em 2015 acordo de cooperação com outras duas instituições de educação médica visando à realização do TP interinstitucional^{1,6}.

O TP realizado pelo consórcio interinstitucional é composto por 120 questões de múltipla escolha com quatro alternativas, construídas no contexto da vivência prática, e visa à resolução de problemas por meio da aplicação de princípios que requeiram raciocínio e reflexão, de modo a evitar a simples memorização ou retenção da informação^{7,8}.

Os discentes que realizam o TP recebem individualmente o resultado da prova com a análise de desempenho por categoria das questões, o índice de dificuldade e o gabarito processado, em que se destacam as respostas corretas e a bibliografia correspondente. Os estudantes que efetuaram dois ou mais testes recebem um segundo relatório analítico que indica a evolução de seu desempenho a cada teste realizado⁹.

O primeiro TP elaborado pelo consórcio interinstitucional ocorreu em outubro de 2015. Na Unifev, foi aplicado aos estudantes do primeiro, terceiro, quinto e sétimo períodos do curso de Medicina. Em 2016, o segundo TP elaborado pelo consórcio foi aplicado também aos estudantes do nono período do curso de Medicina da Unifev.

A participação dos alunos é voluntária e sem caráter de promoção ou recompensas.

Para a gestão do curso de Medicina, o TP constitui-se em relevante subsídio para avaliar a eficiência do programa. Contudo, a percepção do aluno quanto ao seu desempenho e o impacto do TP na aprendizagem podem variar de acordo com contextos pessoais, educacionais, sociais e culturais^{10,11}.

O objetivo deste estudo foi analisar a percepção do aluno sobre seu desempenho no TP, bem como o impacto dela em curto prazo sobre suas estratégias de estudo.

METODOLOGIA

Como público-alvo do presente estudo, elegeram-se os estudantes do curso de Graduação em Medicina da Unifev que realizaram o TP interinstitucional e cursavam o quinto e o sétimo período no momento da aprovação do projeto de pesquisa. O Comitê de Ética em Pesquisa (CEP) da Unifev aprovou o estudo em 10 de outubro de 2016 – Certificado de Apresentação para Apreciação Ética (CAAE) nº 60884116.0.0000.0078. Todos os participantes assinaram o Termo de Consentimento Livre e Esclarecido (TCLE).

Os alunos participaram de forma voluntária, e a seleção foi aleatória, conforme a manifestação do desejo de cada discente. Excluíram-se os alunos que apresentavam doença infectocontagiosa ou impossibilidade de comunicar-se na data agendada para a realização dos grupos focais (GF).

Trata-se de pesquisa qualitativa com análise de conteúdo, e os dados foram coletados em entrevistas por meio de GF.

A técnica GF possibilita discutir e avaliar um tema proposto de modo flexível, dinâmico com a compreensão de contraposições, contradições, diferenças e divergências¹².

Composição e realização dos grupos focais

Realizaram-se dois GF distintos com dez participantes em cada grupo: um grupo composto por discentes do quinto período (G1) e outro composto por discentes do oitavo período (G2). Os estudantes foram agrupados de acordo com o período que estavam cursando, por conveniência de horário livre, o que permitiu manter a uniformidade do grupo e potencializar as reflexões acerca de experiências comuns¹³.

A data para a realização dos GF foi previamente agendada, para que não houvesse prejuízo acadêmico aos discentes. As entrevistas foram realizadas em uma sala de sessão tutorial climatizada, com uma mesa retangular e 13 cadeiras, de modo a acomodar confortavelmente todos os participantes e possibilitar a visualização de todos pelo moderador e dos participantes entre si.

Forma de condução dos grupos focais

Ofereceram-se aos participantes, antes do início da entrevista, sucos e lanches para evitar interrupções

posteriores, proporcionar descontração e facilitar a interação entre os membros¹⁴.

A equipe de trabalho na realização do GF foi composta de uma moderadora, um observador e uma relatora ou auxiliar de pesquisa. No primeiro GF, realizado com alunos do quinto período, a moderadora foi uma docente externa, com grande experiência na condução de GF¹⁵; na qualidade de observador, participou um docente da Unifev; e a pesquisadora, na qualidade de auxiliar de pesquisa ou relatora.

No segundo GF, realizado com alunos que cursavam o oitavo período, os papéis da moderadora e da relatora foram desenvolvidos por duas docentes da Unifev, ambas com experiência na condução de GF, e a pesquisadora participou na qualidade de observadora.

Após a apresentação do moderador e dos auxiliares (relator e observador) e depois da explicação dos objetivos que norteiam o trabalho do grupo, foi definido que os participantes seriam identificados por numerais de 1 a 10 (de N.1 a N.10) para garantir o anonimato de todos.

Na sequência, a moderadora fez a leitura do TCLE, o qual foi preenchido e assinado pelos sujeitos da pesquisa; distribuíram-se os crachás numerados; e reafirmou-se a importância do sigilo para garantir a liberdade de expressão.

O moderador é responsável pela condução e pelos direcionamentos necessários, e deve facilitar a interação entre os participantes para que se aprofundem as discussões¹⁶. O papel do observador é perceber a linguagem não verbal e registrar as impressões relevantes¹⁵.

O desenvolvimento da entrevista foi orientado por questões norteadoras. O relator registrou a intenção de participação de cada aluno e coordenou a sequência de fala dos participantes. As entrevistas foram gravadas para posterior transcrição.

No encerramento dos grupos, ao ocorrer a saturação do tema, as moderadoras realizaram uma síntese da discussão e deixaram o assunto em aberto para que todos realizassem suas considerações finais sobre o tema.

As falas dos participantes foram registradas e transcritas na íntegra pela pesquisadora e por uma assistente de pesquisa.

Questões norteadoras e procedimento de análise de dados

Utilizaram-se as seguintes perguntas norteadoras para auxiliar o desenvolvimento da discussão:

- O que você realmente acha do TP? Qual é a sua percepção acerca do TP?
- O resultado do TP mudou a sua vida estudantil? De que forma?

Os dados coletados por meio do GF são de natureza

qualitativa. De acordo com Gatti¹², o GF “permite emergir uma multiplicidade de pontos de vista e processos emocionais, pelo próprio contexto de interação e significados”.

O referencial teórico utilizado para a exploração e análise dos dados coletados foi a análise de conteúdo temática, segundo os conceitos de Bardin¹⁷.

RESULTADOS E DISCUSSÃO

O primeiro GP a colaborar foi composto por estudantes que cursavam o quinto período, com idade média de 23,42 anos (de 22 a 32 anos e desvio padrão de 3,08 anos). Participaram do segundo GP estudantes que cursavam o oitavo período, com idade média de 24,57 anos (de 22 a 29 anos e desvio padrão de 1,96 ano). Em ambos, havia sete estudantes do gênero feminino e três do gênero masculino.

Categorização do discurso do grupo focal

Organizaram-se os registros com base na contextualização das falas dos participantes, como foco na divergência e na semelhança dos relatos. Realizou-se a posterior categorização das mensagens e/ou dos fragmentos de falas por meio de vinculações entre os agrupamentos com características comuns.

Foram identificadas três categorias a partir da síntese interpretativa desses agrupamentos:

1. As condições de realização do TP são inadequadas.
2. O TP é uma ferramenta pedagógica relevante e deve ser aprimorada.
3. A participação no TP determina sentimentos contraditórios.

Essas categorias são resultantes da associação de sete subcategorias descritas no Quadro 1.

As subcategorias “frustração e segurança” que compõem a terceira categoria – “sentimentos contraditórios” – resultaram de cinco unidades de contexto descritas no Quadro 2.

Quadro 1. Categorias e subcategorias resultantes dos grupos focais.

Categorias	Subcategorias
As condições são inadequadas.	a. A prova é muito extensa.
	b. O tempo é insuficiente.
	c. A época da aplicação é inadequada.
O TP é uma ferramenta pedagógica relevante e deve ser aprimorada.	a. Realizar com mais frequência
	b. Disponibilizar feedback com especificidade.
Determina sentimentos contraditórios.	a. Frustração.
	b. Segurança.

Fonte: Elaborado pelos autores.

Quadro 2. Subcategorias e unidades de contexto da terceira categoria.

Subcategorias	Unidades de Contexto
3.a – Frustração	Falta de compreensão do TP.
	Insegurança sobre os resultados do TP.
3.b – Segurança	Comprovação do desenvolvimento dos conteúdos curriculares.
	Evidência do “monitoramento” do processo de ensino-aprendizagem pelo curso.
	Desenvolvimento da autoavaliação.

Fonte: Elaborado pelos autores.

Compreendendo as categorias identificadas**As condições de realização do TP são inadequadas**

Os discentes participantes da pesquisa consideram as circunstâncias em que ocorre a aplicação do TP inadequadas para avaliar o seu desenvolvimento cognitivo durante o curso. Os dois grupos de alunos entrevistados, do quinto e do oitavo período, questionaram o número de questões, o tempo concedido para fazer a prova e a época da sua realização.

A prova é muito extensa

A primeira citação do GF realizado com alunos do quinto período enfatiza que a prova é muito cansativa em razão do número de itens – 120: “acho que é uma prova assim muito extensa [...] ficou uma prova muito cansativa, são 120 questões, num dia só, então eu acho que isso pesa [...]” (G1 – N.7).

Sucessivamente, nos dois GF, foi referido pelos estudantes que o número de questões e o tamanho dos enunciados são fatores limitantes ao bom desempenho na prova.

[...] também acho que a prova é muito extensa, as questões muito longas [...]. A gente sai da prova muito cansado, muitas vezes já chega no fim da prova sem nem conseguir ler as questões [...] (G1 – N.8).

[...] é um teste muito cansativo, são 120 questões [...] a primeira parte, as 60 primeiras, 70 você faz concentrado, quando vai chegando nas últimas, você já está totalmente cansado, exausto, e aí meio que você não consegue prestar muita atenção (G2 – N.6).

Em uma revisão crítica sobre as principais características do TP, Albanese et al.⁹ citam que os TP geralmente são longos e variam de 180 itens a 400 itens de múltipla escolha. Testes muito curtos podem refletir um conteúdo sub-representado e comprometer a validade e a confiabilidade da avaliação⁷.

O tempo é insuficiente

No segundo GF, realizado com os discentes do oitavo período, além de considerarem a avaliação estafante pela extensão de cada item, mencionaram que o tempo disponibilizado para a realização do TP é insuficiente para o número total de questões aplicadas.

[...] o tempo a gente percebe que não é o adequado pra fazer todas as questões, então tem perguntas que a questão tem, sei lá, cinco, seis linhas pra você pensar e raciocinar [...] (G2 – N.6).

[...] chega no fim da prova, e a gente não consegue responder às últimas questões, [...] pelo tempo e por estar cansado, desse jeito umas 20 questões deixam de ser respondidas, no meu caso [...] (G2 – N.10).

No AMEE Guide nº 71,⁷ os autores referem que a experiência holandesa na aplicação do TP evidenciou que são necessários aproximadamente de 75 a 85 segundos para responder a um único item.

O TP do consórcio interinstitucional tem duração de quatro horas, que correspondem a 120 segundos para cada questão. Apesar de ser considerado pelos alunos como insuficiente, o tempo disponibilizado para fazer a prova é teoricamente adequado e maior que a média internacional.

As questões de múltipla escolha com enunciados extensos e/ou alternativas muito longas demandam maior tempo para a leitura e podem comprometer o desempenho dos alunos, especialmente em avaliações longas como o TP. A construção de bons itens, que concentrem somente informações importantes, pode otimizar o tempo necessário para a resolução da prova e diminuir o desgaste físico relatado pelos alunos¹⁸.

Época da aplicação é inadequada

O período letivo em que se aplicaram os TP, próximo ao das avaliações somativas bimestrais do curso, foi citado pelos entrevistados do quinto período como um fator prejudicial ao bom desempenho acadêmico, pois encontram-se excessivamente cansados, o que, somado ao fato de a participação ser voluntária, pode ser um dos fatores para justificar as ausências: “por estar sempre coincidindo perto do nosso horário de provas, a gente está sempre mais cansado para realizar a prova” (G1 – N.8).

A possibilidade de falharem nas avaliações somativas e a ansiedade determinada pelo risco de não progredirem na graduação induzem os discentes a intensificar as horas de estudo na véspera das provas. Assim como a expectativa por boas notas, com vistas ao sucesso em curto prazo, podem corroborar o estresse associado a essa percepção. Para Al-Kadri¹⁹, as avaliações somativas são fatores de ansiedade e a experiência mais estressante para o aluno.

Teste de Progresso como ferramenta pedagógica relevante

Os alunos entrevistados nos dois GF consideram o TP uma ferramenta útil e importante para monitorar o seu processo de ensino-aprendizagem, pois lhes permite reconhecer as dificuldades e o progresso no decorrer do curso.

[...] eu acho a prova de progresso uma coisa muito boa porque você começa a se avaliar... a intenção da prova eu acho muito boa [...] (G1 – N.3).

Eu acho que é um teste muito bom, que avalia a nossa evolução desde o primeiro até o sexto ano [...] eu acho que é muito importante (G2 – N.6).

Entretanto, acreditam que, para potencializar o impacto dessa avaliação, é preciso aperfeiçoá-la, com mais aplicações durante o ano e fornecimento de *feedbacks* específicos para cada período.

Realizar com mais frequência

Os alunos do oitavo período entendem que o TP deveria ser aplicado com mais frequência, duas ou mais vezes ao ano, e com menos itens. Essa estratégia de acordo com os alunos contribuiria para melhorar o desempenho individual e facilitar a correção de lacunas evidenciadas.

Uma sugestão seria que a gente tivesse vários testes do progresso (G2 – N.9).

A sugestão de separar ela em dois dias eu acho que seria a melhor [...] melhor opção mesmo, pra que esse teste seja 100% válido (G2 – N.10).

Essa percepção discente encontra sustentação em pesquisa realizada por Wade et al.²⁰ que demonstrou um impacto positivo na percepção dos estudantes submetidos ao TP quatro vezes por ano²¹. A realização do TP repetidamente com o objetivo de avaliar o conhecimento estimula a revisão dos conteúdos por meio do *feedback* e pode promover a aprendizagem profunda⁷.

Disponibilizar feedback com especificidade

Na percepção dos alunos do quinto período, o *feedback* realizado pelo curso é insuficiente para a autoavaliação. Acreditam que o *feedback* deveria ser específico para cada período do curso, com a descrição dos conteúdos mais importantes de cada item para o efetivo aprendizado.

[...] eu queria ter um retorno tipo: "Tal questão vocês já viram isso, só que vocês erraram, o que que aconteceu?". É assim... assim e assim. [...] aqueles gráficos, mas não têm tanta utilidade mesmo, porque, por exemplo, você não sabe se aquilo que você acertou tá bom [...] (G1 – N.1).

[...] tinha que ter esse direcionamento, [...] a gente não sabe se era pra ter visto, se não era pra ter visto, e aí vai deixando e acumulando mesmo (G1 – N.6).

Em currículos integrados com ênfase na aprendizagem centrada no aluno, espera-se que o crescimento do conhecimento ocorra progressivamente ao longo do tempo. O TP é uma avaliação longitudinal e abrangente, orientada para os objetivos finais do curso, e, portanto, está desacoplado das especificidades do que os alunos aprenderam ou estão aprendendo em qualquer momento⁹.

A dificuldade de interpretar os resultados (*feedback*) disponibilizados pode estar relacionada a uma análise pontual equivocada e/ou não realizada de modo reflexivo. Wade et al.²⁰, ao avaliarem a percepção dos alunos de duas escolas médicas do Reino Unido sobre o TP, constataram que em ambas os alunos sentiam que o *feedback* que recebiam era insuficiente.

Sentimentos contraditórios

A percepção da magnitude do impacto, positivo ou negativo, sobre a realização do TP aparenta ter correlação com o período que os estudantes estão cursando. Os estudantes do oitavo período valorizam e consideram que a realização do TP tem um impacto positivo em seu processo de ensino-aprendizagem e proporciona sentimentos de segurança por constatarem o seu desenvolvimento ao longo do curso.

Para os alunos do quinto período, os sentimentos são opostos; referem que a realização do TP provoca frustração por não compreenderem a forma como é elaborada a prova, dificuldade em interpretar o *feedback* fornecido e insegurança quanto à validade dos resultados.

Frustração

A motivação pessoal para obter altas notas pode contribuir para os estudantes do quinto período apresentarem sentimentos de frustração com o fraco desempenho inicial e demonstrarem dificuldade para compreender o TP como uma avaliação diagnóstica e formativa do desempenho cognitivo longitudinal.

[...] a prova eu acho que é até interessante, [...] como ela avalia os seis anos, ela acaba prejudicando quem está muito no começo (G1 – N.7)

A gente fica meio assim: "Ah, eu sei que fui mal em fisiologia, mas e agora como faço para melhorar?". Será que isso a culpa é da faculdade ou será que é culpa minha [...]. A gente fica meio na dúvida [...] (G1 – N.8).

Os estudantes argumentaram que, pelo fato de a avaliação ser realizada em cooperação com outras escolas médicas, sentem insegurança quanto à validade dos resultados, pois o desempenho deles é comparado ao de alunos de

escolas médicas que utilizam diferentes matrizes curriculares e estratégias de ensino-aprendizagem: “a gente tem [...] uma faculdade com um tema no segundo ano, que a outra viu no terceiro, e a gente é comparado com as outras faculdades, então esse resultado acaba não sendo tão fidedigno, né?” (G1 – N.2).

Segurança

Os participantes do oitavo período evidenciam efetivamente no TP que os conteúdos solicitados foram estudados e conseguem avaliar se o seu desempenho está adequado ou se há lacunas de aprendizagem a serem remediadas, além de perceberem que o processo ensino-aprendizagem do curso está sendo monitorado.

[...] eu consigo enxergar muito bem as matérias que a gente viu, consigo lembrar e elencar as dificuldades que eu tive naquela matéria já vista, e consigo também ver aquilo que nem passou pela minha frente, então eu acho bastante proveitoso (G2 – N.8).

[...] o Teste de Progresso não serve só pra ver o nosso progresso, serve pra ver... é eu tenho certeza que o NDE – Núcleo Docente Estruturante, tá olhando pra essa prova que a gente faz e vê onde que a gente tá falhando [...]. (G2 – N.10).

O *feedback* na percepção dos estudantes estimula a autoavaliação e pode determinar mudanças no modo de estudar. Heeneman et al.²² concluíram que incorporação do TP em programa de avaliação programática de maneira formativa afeta positivamente as pontuações do TP e a aprendizagem significativa.

[...] é um teste muito importante pro nosso desempenho durante todo o passar do curso, e até mesmo pra avaliar se está sendo realmente aproveitado o nosso estudo, e até mesmo modificar a maneira com que a gente estuda [...]. (G2 – N.2).

[...] a autocrítica, né?, que a gente acaba desenvolvendo, não que eu tenha mudado a minha forma de estudar, mas a visão que a gente tem do nosso estudo pode acabar diferenciando, você vê se você melhorou ou não, de um ano pro outro [...]. (G2 – N.1).

CONCLUSÃO

Os estudantes entrevistados consideram o TP uma “ferramenta pedagógica” relevante, que permite a autoavaliação e a correção das lacunas de aprendizagem. Porém, acreditam que as características estruturais da prova – número de questões, abrangência do conteúdo e tempo para realização – afetam negativamente o desempenho deles. E propõem mudanças: aumentar a frequência de realização, diminuir o número de itens e fornecer um *feedback* orientado para o período que estão cursando.

Os resultados de desempenho do TP e o *feedback* recebido determinaram sentimentos contraditórios de acordo com o período que estavam cursando: de frustração para os alunos do quinto período e de segurança para os do oitavo período.

O impacto determinado pelo desempenho no TP, em curto prazo, não promoveu mudanças no plano de estudo dos discentes entrevistados.

É necessário promover discussões sobre os objetivos, os benefícios e as desvantagens da aplicação do TP e dos seus resultados com docentes e discentes, e implementar estratégias para a entrega do *feedback*, de modo dirigido e reflexivo, que permita a autoavaliação e a melhoria contínua do processo de ensino-aprendizagem.

CONSIDERAÇÕES FINAIS

A percepção e a prática de autoavaliação dos alunos são habilidades instáveis que podem ser afetadas por diferentes variáveis, e o impacto que determinam pode ter magnitude positiva ou negativa²³. É fundamental oferecer um *feedback* ao aprendiz, para que o processo avaliativo desempenhe uma dimensão orientadora^{24,25}. Embora os alunos recebam um *feedback* detalhado do seu desempenho no TP, ele é subutilizado.

O TP é uma ferramenta valiosa que pode ser utilizado como uma avaliação formativa que monitore o crescimento do conhecimento e associe aspectos de avaliação somativa. É uma rica fonte de informações que precisa ser mais bem aproveitada pelos alunos, pelos professores e pelas instituições de ensino^{22,26}.

CONTRIBUIÇÃO DOS AUTORES

Marlene Moraes Rosa Chinelato participou como pesquisadora principal no projeto de pesquisa, realizou a coleta, análise e interpretação dos dados, a redação do manuscrito e a revisão crítica do estudo. Gisele Regina de Azevedo colaborou com a concepção metodológica do projeto, análise e interpretação dos dados. José Eduardo Martinez participou da concepção, orientação e supervisão do estudo, e da revisão final do artigo.

CONFLITO DE INTERESSES

Declaramos não haver conflito de interesses.

FINANCIAMENTO

Bolsa do Programa Institucional de Qualificação e Capacitação Docente (PIQCD) do Centro Universitário de Votuporanga.

REFERÊNCIAS

1. Van der Vleuten CPM, Verdwijnen GM. Fifteen years of experience with progress testing in a problem-based learning curriculum. *Med Teach*. 1996;18(2):103-9.

2. Sakai MH, Ferreira Filho OF, de Almeida MJ, Mashima DA, Marchese MDC. 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.
3. Cecilio-fernandes D, Kerdijk W, Jaarsma ADDC, Tio RA. Development of cognitive processing and judgments of knowledge in medical students : analysis of progress test results. *Med Teach*. 2016;38:1125-9.
4. Plessas A. Validity of progress testing in healthcare education. *Int J Humanit Soc Sci Educ*. 2015;2(8):23-33.
5. Vantini I, Benini L. Models of learning, training and progress evaluation of medical students. *Clin Chim Acta*. 2008;393(1):13-6.
6. Van Der Vleuten CPM, Schuwirth LWT. Assessing professional competence: from methods to programmes. *Med Educ*. 2005;39(3):309-17.
7. 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.
8. Lampert JB, Bicudo AM, organizadoras. 10 anos das Diretrizes Curriculares Nacionais do Curso de Graduação em Medicina. Rio de Janeiro: Associação Brasileira de Educação Médica; 2014. p. 80.
9. Albanese M, Case SM. Progress testing: critical analysis and suggested practices. *Adv Health Sci Educ Theory Pract*. 2016;21(1):221-34.
10. Chen Y, Henning M, Yelder J, Jones R, Wearn A, Weller J. Progress testing in the medical curriculum: students' approaches to learning and perceived stress. *BMC Med Educ*. 2015;15(147):30-38.
11. Cilliers FJ, Schuwirth LW, Adendorff HJ. The mechanism of impact of summative assessment on medical students' learning. *Adv Health Sci Educ Theory Pract*. 2010;15(5):695-715.
12. Gatti BA. Grupo focal na pesquisa em ciências sociais e humanas. Brasília: Liber Livro; 2005. p. 1-11.
13. Backes DS, Colomé JS, Erdmann RH. Grupo focal como técnica de coleta e análise de dados em pesquisas qualitativas. *Mundo Saúde*. 2011;35(4):438-42.
14. Zimmermann MH, Martins PLO. Grupo focal na pesquisa qualitativa: relato de experiência. *Anais do 8º Congresso Nacional de Educação Educere*. Curitiba: PUCPR; 2008. p. 12115-25.
15. Leopardi MT. Metodologia da pesquisa na saúde. 2a ed. Florianópolis: UFSC/Pós-Graduação em Enfermagem; 2002. 290 p.
16. Gondim SMG. Grupos focais como técnica de investigação qualitativa: desafios metodológicos. *Paidéia (Ribeirão Preto)*. 2002;12(24):149-61.
17. Bardin L, Reto LA, Pinheiro A. Análise de conteúdo. São Paulo: Edições 70; 2011. p. 229.
18. Al-Wardy NM. Assessment methods in undergraduate medical education. *SQU Med J*. 2010;10(2):203-9.
19. Al-Kadri HM. Does assessment drive students' learning? Clinical context assessment and students' approaches to learning. [Doctoral Thesis, Maastricht University]. 2012;187:93-116.
20. Wade L, Harrison C, Hollands J, Mattick K, Ricketts C, Wass V. Student perceptions of the progress test in two settings and the implications for test deployment. *Adv Heal Sci Educ Theory Pract*. 2012;17(4):573-83.
21. Al Kadri HM, Al-Moamary MS, Magzoub ME, Roberts C, van der Vleuten CPM. Students' perceptions of the impact of assessment on approaches to learning: a comparison between two medical schools with similar curricula. *Int J Med Educ*. 2011;2:44-52.
22. Heeneman S, Schut S, Donkers J, van der Vleuten C, Muijtjens A. Embedding of the progress test in an assessment program designed according to the principles of programmatic assessment. *Med Teach*. 2017;39(1):44-52.
23. Al-Kadri HM, Al-Moamary MS, Roberts C, Van der Vleuten CPM. Exploring assessment factors contributing to students' study strategies: literature review. *Med Teach*. 2012;34(Suppl 1):S42-50.
24. Panúncio-Pinto MP, Troncon LEA. Avaliação do estudante: aspectos gerais. *Med (Ribeirão Preto)*. 2014;47(3):314-23.
25. Vassar M, Wheeler DL, Davison M, Franklin J. Program evaluation in medical education: an overview of the utilization-focused approach. *J Educ Eval Health Prof*. 2010;7:1.
26. Tio RA, Schutte B, Meiboom AA, Greidanus J, Dubois EA, Bremers AJA. The progress test of medicine: the Dutch experience. *Perspect Med Educ*. 2016;5(1):51-5.



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.

Onze anos de Teste de Progresso na Unicamp: um estudo sobre a validade do teste

Eleven years of Progress Testing at Unicamp: a test validity study

Ruy Guilherme Silveira de Souza¹  | ruysouza28@gmail.com
Angélica Maria Bicudo²  | angelicabicudo@gmail.com

RESUMO

Introdução: O curso de Medicina da Universidade Estadual de Campinas (Unicamp) vem pondo à prova o aquisição cumulativa cognitiva de seus alunos por meio do Teste de Progresso (TP) há mais de uma década, de modo a possibilitar a análise da utilidade do exame como estratégia de apoio a decisões pedagógicas e apontar principais ameaças à validade dele.

Objetivo: Este estudo teve como objetivos oferecer a análise de validade do TP e explicitar as oportunidades de utilização do teste especialmente para a determinação de padrões de suficiência cognitiva para a progressão no curso e ao final deste, e a identificação de estudantes em risco.

Método: Trata-se de estudo observacional retrospectivo de uma série histórica de sucessivos testes escritos realizados para analisar o acúmulo cognitivo no período de 2006 a 2016, totalizando 11 anos e seis turmas consecutivas. Em cada momento de medida (aplicação do teste), o estudo utilizou um modelo misto, em que a exposição (realização do teste) e o desfecho (score do teste) foram avaliados no mesmo ponto de tempo, o que caracteriza um estudo transversal (cross-sectional) cujos resultados sucessivos originarão as curvas de crescimento cognitivo

Resultado: Observou-se um acúmulo cognitivo em torno de 6 pontos percentuais a cada nova testagem. Os estudantes ao completarem o sexto ano obtiveram um acerto de cerca de 65,7% ($\pm 9,1$). A cada testagem, determinou-se um "efeito piso" para identificar alunos com rendimento abaixo da média, que em geral se situou em cerca de 1,5 DP abaixo da média da respectiva turma.

Conclusão: O TP-Unicamp oferece dados confiáveis para apoiar importantes decisões pedagógicas, tais como identificação de alunos em risco acadêmico por baixa performance, critérios para progressão e desempenho cognitivo ao final do curso. Como confiabilidade sofre influência da amostragem, e o aumento do número de itens de cada teste e o aumento da frequência de testagem podem ser estratégias a serem tomadas para superar essas limitações.

Palavras-chave: Educação Médica; Avaliação de Aprendizagem; Validade; Teste de Progresso.

ABSTRACT

Introduction: The Medical School of the State University of Campinas (UNICAMP) has been testing the cumulative cognitive acquisition among its students through the Progress Test (PT) for over a decade, making it possible to analyze the utility of the test as a strategy to support pedagogical decisions and to point out the main threats to its validity.

Objective: To provide an analysis of the validity of the PT, explaining opportunities for its use, especially in determining standards of cognitive sufficiency for progression, standards of cognitive sufficiency at the end of the course, and identification of students at risk.

Method: Retrospective observational study of historical series of successive written tests performed to analyze cognitive accumulation, covering a period from 2006 to 2016, totaling 11 years and 6 consecutive classes. At each instance of measurement (test application), the study uses a mixed model where exposure (test performance) and outcome (test score) are evaluated at the same time point, characterizing a cross-sectional study, the successive results of which will generate the cognitive growth curves.

Result: A cognitive accumulation of around 6 percentage points was observed with each new test. Students, upon completing the 6th year, scored around 65.7% (± 9.1). A "floor effect" was determined for each test to identify students with below-average performance, which in general was about 1.5 SD below the average of the respective class.

Conclusion: TP-Unicamp offers reliable data to support important pedagogical decisions, such as the identification of students at academic risk for low performance, criteria for progression and cognitive performance at the end of the course. As reliability is influenced by sampling, increasing the number of items in each test and increasing the frequency of tests could represent strategies to overcome potential limitations.

Keywords: Medical Education; Learning Assessment; Validity; Progress Test.

¹Universidade Federal de Roraima, Boa Vista, Roraima, Brasil.

²Universidade Estadual de Campinas, Campinas, São Paulo, Brasil.

Editor: Aristides Augusto Palhares Neto.

Recebido em 18/09/22; Aceito em 17/10/22.

Avaliado pelo processo de *double blind review*.

INTRODUÇÃO

O Teste de Progresso (TP) vem se tornando uma das formas mais populares de avaliação do ensino médico em todo o mundo¹. A escola de Medicina da Universidade Estadual de Campinas (Unicamp) vem testando a aquisição cumulativa cognitiva de seus alunos há mais de uma década, e o volume de dados gerados já possibilita oferecer uma análise consistente da validade do exame no âmbito da formação médica na graduação. Inicialmente é necessário apresentar a justificativa epistêmica que fundamenta a visão de validade neste trabalho. Validade é o atributo mais importante para qualquer tipo de teste e sob a qual todas as demais qualidades ficam em estado de dependência²⁻⁶. No entanto, esse conceito no campo da pesquisa em avaliação de aprendizagem na educação médica ainda é confuso e fragmentado, colocando a validade no mesmo patamar de outros aspectos da testagem ou até mesmo em posição inferior, que na verdade lhe são subalternos. Uma das razões para isso está no rápido crescimento que a pesquisa em avaliação no ensino médico experimentou desde a década de 1970. Apenas para que se possa ter uma noção desse crescimento, o estudo pioneiro sobre exame clínico estruturado para avaliação de competência publicado em 1979 já foi seguido de mais de 1.800 artigos sobre o mesmo método⁷. Esse crescimento se caracterizou por um foco excessivo na testagem e em seus resultados imediatos, desviando a atenção dos pesquisadores na direção dos testes padronizados e das novas técnicas psicométricas voltadas para a obtenção de escores fidedignos fortemente ancorados em análises estatísticas cada vez mais complexas e inacessíveis em um grande número de instituições⁸, transformando a fidedignidade da medida na base para a generalização da utilização de um teste. Com isso, um teste que apresentasse por exemplo um alfa de Cronbach com uma boa consistência interna ganhava credenciais para ser replicado em diversas escolas, de diferentes contextos, sem nenhum outro tipo de reflexão. Royal⁹, ao realizar uma busca geral no banco de dados PubMed sobre estudos de avaliação do ensino em estudantes de Medicina, encontrou a expressão “instrumento confiável” mais de duas mil vezes em um período de apenas cinco anos. Souza et al.¹⁰, em uma análise bibliométrica sobre o atual conceito de validade na educação médica, concluíram que 96% dos estudos ainda apresentam um conceito fragmentado de validade, limitando-se basicamente aos resultados do processo de testagem, sem explicitar uma análise das consequências e utilidade dos testes. Essa fragmentação do conceito decorre da confusão que frequentemente acontece entre os termos “validação” e “validade”. Validação representa o processo de coleta de dados (evidências) obtidos pelos resultados de um teste³, mas isso é

somente parte do processo que só se completa quando, após a análise de todas as fontes de evidência, define-se como elas apoiam a interpretação e a utilidade de um teste; em outras palavras, sem a explicitação da utilidade e das consequências do teste, não é possível um julgamento de validade. Messick, um dos pontos de lança do movimento pela volta de um conceito unificado de validade, propunha um processo de análise de validade composto por duas etapas: a primeira, de natureza científica, deveria se preocupar com as propriedades psicométricas do teste, e a segunda, de natureza ética, deveria focar a análise extensa das consequências do teste “em termos de valores humanos”¹¹.

O futuro da educação médica, especialmente diante dos desafios criados pela pandemia da *coronavirus disease 2019* (Covid-19), aponta para novas e necessárias transformações no campo da avaliação. Faz-se necessário que a pesquisa em avaliação do ensino médico resgate a importância dessa “natureza ética” proposta por Messick, desviando um pouco o foco dos resultados numéricos para a “narrativa” do processo, de modo a possibilitar uma concepção ampla de avaliação que só termina na explicitação da utilidade e das consequências da testagem, em resumo: “uma mudança de números para palavras”¹².

Neste trabalho, propõe-se uma análise de validade do TP ancorada em uma visão unificada que explicita oportunidades de utilização do teste no dia a dia do ensino de graduação, tais como a definição de parâmetros que possibilitem a determinação de padrões de suficiência cognitiva para a progressão do estudante ao longo do curso e ao final deste, e a identificação de discentes em risco.

MÉTODO

Trata-se de um estudo retrospectivo de uma série histórica de sucessivos testes escritos de oferta anual única, realizados para analisar o acúmulo cognitivo de estudantes do curso de Medicina da Unicamp, no período de 2006 a 2016, totalizando 11 anos e seis turmas consecutivas de Medicina, computando um total de 574 alunos que concluíram sua formação médica entre 2011 e 2016. As curvas de crescimento cognitivo de cada turma foram consolidadas para uma visão global do crescimento cognitivo da escola, a fim de permitir uma melhor generalização das interpretações.

Designa-se “testagem 1” a primeira avaliação por meio do TP de cada turma testada e assim sucessivamente até a “testagem 6”. No período compreendido entre 2006 a 2016, seis turmas completaram sua formação, com seis momentos de testagem. O quadro 1 demonstra as sucessivas turmas testadas no período de 2006 a 2016.

Quadro 1. Número de testagens e turmas testadas no período de 2006 a 2016

Ano	Testagem 1	Testagem 2	Testagem 3	Testagem 4	Testagem 5	Testagem 6	Turmas
2006	95						
2007	89	81					
2008	92	89	88				
2009	100	88	94	93			
2010	98	102	91	92	92		
2011	88	96	101	96	90	89	Turma 1 (2006-2011)
2012		86	99	92	91	92	Turma 2 (2007-2012)
2013			91	98	93	89	Turma 3 (2008-2013)
2014				96	91	97	Turma 4 (2009-2014)
2015					98	90	Turma 5 (2010-2015)
2016						100	Turma 6 (2011-2016)

Fonte: Teste de Progresso Unicamp.

Como se trata de uma série histórica e, portanto, sem a possibilidade de determinação prévia de critérios para cada prova, a fim de minimizar as diferenças entre os diversos graus de dificuldade dos diferentes testes, os escores individuais de cada turma foram organizados a partir de normorreferenciamento¹³. A partir da consolidação da distribuição de frequências, foi possível determinar um “efeito teto” para alunos com rendimento de 1,5 a 2 desvios padrões acima da média, bem como um “efeito piso” para alunos com rendimento de 1,5 a 2 desvios padrões abaixo da média.

O TP-Unicamp utiliza como instrumento uma prova objetiva, do tipo múltipla escolha, composta de 120 itens. Quanto ao número de alternativas, até 2011 o TP utilizava cinco, e, a partir de 2012, cada item passou a ter quatro alternativas no formato “melhor resposta individual” e sem estratégias para acertos casuais (“chutes”). O conteúdo do exame abrange toda a área de conhecimento esperado para um aluno concluinte do curso e aplicado simultaneamente para todas as séries.

Em cada teste, obteve-se uma média global do teste geral e de cada área de conhecimento clínico (isto é, cirurgia, clínica Médica, ginecologia e obstetrícia, e pediatria).

O escore é a métrica da prova, como a soma dos pontos dos acertos, ou como a soma relativa, ou ainda como a soma média.

Por tratar-se de um estudo com medidas repetidas ao longo de uma escala de tempo, adotou-se um modelo de regressão linear de efeitos mistos para o escore relativo da prova. Para esse modelo, considerou-se como variável dependente (desfecho) o escore médio relativo de cada uma das seis turmas. As testagens (cada construção de prova, desde 2006 até 2016) foram consideradas como variáveis (ordinais) predictoras (efeito fixo).

A correlação de Pearson foi utilizada para determinar a contribuição de cada área de conhecimento analisada em relação à prova toda e nas relações interáreas. O alfa de Cronbach foi utilizado como medida de fidedignidade (confiabilidade), em termos dos domínios e do construto geral, a fim de analisá-los quanto à sua capacidade de produzir resultados precisos.

Para redução dos erros de medida na construção da curva de crescimento cognitivo, utilizou-se a variação do percentual de acertos em todo o teste, convertidos para um escore Z, definido como o valor de cada indivíduo em cada ano subtraindo a média e dividindo pelo desvio padrão da turma respectiva, como estratégia de controle da dificuldade do exame em diferentes testes e turmas.

O estudo foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Roraima (UFRR):

Certificado de Apresentação para Apreciação Ética (CAAE): nº 09314919.8.0000.5302, nº 3.206.008.

RESULTADOS

A Tabela 1 apresenta os escores relativos das seis turmas testadas ao longo de 11 anos.

A Tabela 2 apresenta os valores do alfa de Cronbach obtidos para cada momento de testagem (do primeiro ao sexto ano). Os valores de alfa de Cronbach são inferiores a 0,7, porém apresentam crescimento e ficam limítrofes após a terceira testagem.

O Gráfico 1 apresenta os perfis (dispersão) obtidos na análise longitudinal e a curva de crescimento obtida a partir dos escores relativos para cada uma das seis turmas que realizaram sua formação no período analisado.

Na Tabela 3, são apresentados os ajustes para a análise utilizando o modelo de regressão linear de efeitos mistos para

o escore relativo da prova. Com exceção das turmas de 2013 e 2014, todas as variáveis foram significativamente associadas com o escore.

Com base no modelo longitudinal linear misto, observou-se um acúmulo cognitivo em torno de 6 pontos percentuais a cada nova testagem. Pediatria e ginecologia e obstetrícia aumentaram, em média, 9 pontos a cada nova testagem, e clínica médica e cirurgia podem aumentar, em média, 8 pontos a cada nova testagem.

Os estudantes ao completarem o sexto ano obtiveram um acerto de cerca de 65,7% ($\pm 9,1$), com um intervalo de confiança de 95% de 65,6545 a 65,8206, com variações de desempenho perto de 30% de acerto até, aproximadamente de 93% de acerto.

O “efeito piso/efeito teto” é utilizado como uma ideia de um alarme para os estudantes que necessitem de remediação.

Tabela 1. Medidas descritivas do escore relativo das provas realizadas pelas seis turmas, pelo ano de formatura, durante o período de 2006 a 2016

Testagem	Ano de testagem	Ano de formatura	Média	Desvio padrão da média	Mínimo	1º Quartil	Mediana	3º Quartil	Máximo
1	2006	2011	32	4,78	23	29	32	36	50
	2007	2012	39	4,85	25	36	38	43	48
	2008	2013	33	5,32	18	28	33	37	44
	2009	2014	31	4,78	20	28	31	34	44
	2010	2015	31	4,60	21	28	31	34	44
	2011	2216	36	4,44	27	33	37	39	46
2	2007	2011	43	5,47	31	40	43	47	55
	2008	2012	33	6,29	21	28	33	38	48
	2009	2013	34	5,94	18	29	35	38	49
	2010	2014	32	5,15	20	28	32	36	43
	2011	2015	38	5,92	23	34	38	42	54
	2012	2216	34	4,48	24	31	33	36	45
3	2008	2011	42	7,16	17	38	43	46	58
	2009	2012	39	6,91	22	34	38	44	53
	2010	2013	39	6,92	22	33	38	44	56
	2011	2014	45	7,63	27	39	45	50	60
	2012	2015	39	6,01	25	36	39	43	58
	2013	2216	44	7,77	26	38	44	50	62
4	2009	2011	49	0,76	33	43	48	54	66
	2010	2012	45	0,70	26	40	45	48	61
	2011	2013	52	0,68	34	47	52	57	68
	2012	2014	44	0,73	28	39	45	50	62
	2013	2015	50	0,85	27	45	51	56	69
	2014	2216	57	0,84	36	52	57	63	71
5	2010	2011	52	6,41	26	48	53	53	68
	2011	2012	58	6,72	42	53	58	58	71
	2012	2013	53	9,28	33	46	53	53	77
	2013	2014	57	6,49	34	53	58	58	71
	2014	2015	60	7,04	38	56	61	61	77
	2015	2216	63	8,17	45	57	63	63	82
6	2011	2011	64	8,17	33	60	65	68	79
	2012	2012	58	9,81	37	50	57	67	82
	2013	2013	64	7,54	43	58	63	69	81
	2014	2014	68	6,27	48	65	68	73	81
	2015	2015	71	9,17	44	65	73	78	88
	2016	2216	69	7,54	44	64	70	74	84

Fonte: Teste de Progresso Unicamp.

Tabela 2. Medidas de confiabilidade do escore relativo dos seis momentos de testagem

Testagem	Alfa de Cronbach
1	0,4609
2	0,5810
3	0,5851
4	0,6546
5	0,6251
6	0,6904

Fonte: Teste de Progresso Unicamp.

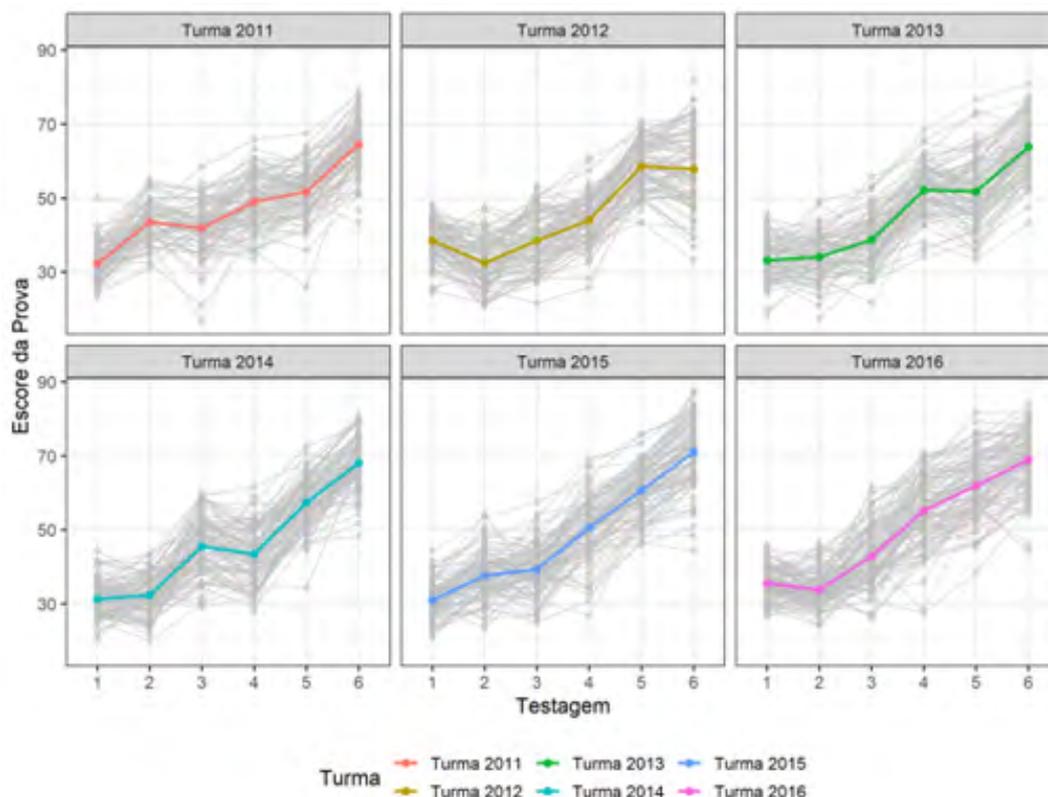
Assim sendo, observou-se uma variação de performance em torno de 18% a 28% da prova na primeira testagem; de 18% a 29% na segunda testagem; de 17% a 35% na terceira testagem; de 26% a 43% na quarta testagem; de 26% a 50% na quinta testagem; e de 33% a 58% na sexta testagem; pela tabela de normatização construída, os valores inferiores estão abaixo da média, entre o primeiro e o terceiro desvio padrão negativo.

DISCUSSÃO

A literatura sobre educação médica já apresenta amplas evidências tanto de “plausibilidade teórica” como de “viabilidade prática” do TP em medir o crescimento cognitivo

de uma turma de aprendizes em diversos modelos curriculares e em uma diversidade de cenários, oferecendo dados valiosos para o desempenho de um currículo escolar^{13,14}.

O modelo que a escola adota para seus itens é o de múltipla escolha, já amplamente referendado pela literatura como o modelo de avaliação escrita mais confiável e válido, a ponto de ter diminuído a importância dos demais formatos de testes escritos¹⁵. Mesmo o teste do consórcio holandês tido frequentemente como referência mudou seu formato original que remontava à década de 1970 para o formato de “múltipla escolha com melhor alternativa” ou “múltipla escolha tipo A”¹⁶. O modelo de teste de múltipla escolha tipo A apresenta ainda como vantagem oferecer melhor consistência interna e confiabilidade, além de exigir um grau de análise mais profunda para sua resolução¹⁷. Quanto ao número de itens, no entanto, a prova apresenta uma ameaça à consistência dos resultados. É conhecido o efeito da amostragem sobre a fidedignidade da medida, e a literatura sobre TP demonstra que o número de itens utilizados em testes com comprovada fidedignidade varia de 180 a 400¹⁸. No caso do TP-Unicamp, observou-se que a consistência interna do teste só se tornou mais adequada a partir da quarta testagem, o que em parte pode ser melhorado com o aumento do número de itens, uma estratégia a ser adotada caso a escola pretenda utilizar seus resultados para fins somativos.

Gráfico 1. Perfil do escore relativo das turmas nas seis testagens do TPI

Fonte: Teste de Progresso Unicamp.

Tabela 3. Modelo linear misto ajustado para o escore relativo das turmas nas seis testagens do TPI

	Estimativa	Erro padrão	Graus de liberdade	Valor T	Valor p
Intercepto	23,53	0,55	2767	42,51	< 0,001*
Momento	6,68	0,07	2767	96,72	< 0,001*
Turma 2012	-1,99	0,70	573	-2,85	0,005*
Turma 2013	-1,29	0,70	573	-1,84	0,066*
Turma 2014	-0,55	0,69	573	-0,79	0,432*
Turma 2015	1,43	0,70	573	2,03	0,043*
Turma 2016	2,71	0,68	573	3,98	< 0,001*

* valor p < 0,05

Fonte: Teste de Progresso Unicamp.

Outro aspecto do TP-Unicamp que merece ser analisado é a ausência de estratégia para acertos casuais (chute). Como uma das características do TP é a oferta de uma prova de conteúdo abrangente mesmo para alunos do primeiro ano, sempre foi uma preocupação a influência dos acertos casuais na confiabilidade dos resultados. Na verdade, essa é uma questão que remete ao início do século XX e à popularização dos testes padronizados, mas que também demonstra como aspectos psicométricos podem muitas vezes dominar uma discussão a ponto de obscurecer aspectos fundamentais no ensino. Muitas escolas penalizam candidatos que marcam respostas erradas para desencorajar marcações baseadas em palpites¹⁹, e outras incluem uma alternativa “não sei”, que tira a penalidade no caso de erro, e o resultado é obtido por meio da utilização de uma fórmula para o escore (*formula scoring*)¹⁷. No entanto, a literatura tem demonstrado que não somente tais estratégias apresentam menos confiabilidade²⁰, como também podem ter um efeito prejudicial sobre aspectos fundamentais do processamento de um problema por um ser humano. Bliss²³, em um estudo com crianças do ensino básico, demonstrou que a utilização da *formula scoring* penalizou alunos mais capazes, o que poderia inclusive levantar questões éticas sobre essa prática, e, da mesma forma, Muijtens et al.²⁵ demonstraram que alunos de Medicina que são menos predispostos a adivinhar e, portanto, procuram desenvolver um entendimento mais profundo da questão tendem a ter um escore mais baixo. Especialmente no caso da utilização dos testes de múltipla escolha tipo A, em que todas as alternativas são plausíveis, ao penalizarmos alunos que não escolham a opção “não sei”, negamos a possibilidade de que um conhecimento parcial possa contribuir para a resolução de problemas.

Um aspecto crítico quanto à execução do teste diz respeito ao número de testagens. Nesse sentido, o número de edições do TP-Unicamp por turma é bem inferior àqueles utilizados por outras escolas que utilizam o TP para fins

somáticos. As escolas de Medicina de McMaster e Utrecht realizam três testagens anuais; e as escolas de Maastricht e de Missouri-Kansas, quatro testagens anuais¹⁴. Sabe-se que é justamente a repetição desses testes que permite a construção da curva cognitiva. Este estudo demonstrou que os escores alfa das testagens só alcançaram valores limítrofes a 7 após a terceira testagem, o que demonstra que, caso a escola deseje utilizar os resultados como avaliação somática para certificar a progressão de seus alunos, precisará aumentar o número de edições anuais para pelo menos três testes anuais.

O crescimento cognitivo observado revelou escore médio percentual que variou de 34% para a primeira testagem no primeiro ano do curso a 65,7% para testagem ao final do curso, o que está acima da média nacional²². Para a construção da curva de crescimento cognitivo, o estudo distribuiu os dados por meio de normorreferenciamento, já que não foi possível definir um critério prévio por meio de um ponto de corte. Nesse sentido, os consórcios holandês e alemão também utilizam uma estratégia normorreferenciada, apoiados em evidências de que a determinação de um ponto de corte prévio produz resultados menos confiáveis²³. Essa estratégia seria particularmente importante em escolas sem consultores psicométricos, já que adotam estratégias estatísticas tradicionais que não necessitam de programas complexos e historicamente mais utilizadas como o escore Z²⁴.

A construção da curva também permite uma estratégia educacional formativa importante e frequentemente negligenciada em educação médica, que é o da identificação de alunos com baixo rendimento e que necessitem de suporte pedagógico (remediação). Na curva obtida no TP-Unicamp, é possível observar três padrões que podem ser combinados na tomada de decisões quanto à progressão do estudante: 1. uma expectativa anual de crescimento de 6 pontos percentuais, 2. um desempenho em relação à média não inferior a 1,5 desvio padrão e 3. um desempenho cognitivo ao final do

curso, em termos de acertos, em torno de 65%. Nenhuma decisão pedagógica de alta aposta deveria ser baseada em um único parâmetro, porém a triangulação desses indicadores pode sustentar pareceres mais fidedignos que os atualmente utilizados. Na escola de Medicina de McMaster, os alunos são identificados quando seu desempenho cai abaixo de 1,5 e 2 desvios padrão da média de sua turma¹⁴. Embora essa definição de até 2 desvios padrão seja frequentemente a mais utilizada em termos estatísticos, o mais importante é a consolidação de uma curva de crescimento consistente, a partir da qual se possam identificar variações de desempenho²⁷. No caso da escola de Medicina da Unicamp, o efeito piso ficou compreendido entre 1 e 3 desvios padrão da média, de tal maneira que, com base no presente estudo, um desempenho a partir de 1,5 desvio padrão abaixo da média poderia servir como um sinalizador para alunos que necessitem de remediação. Destaca-se então aquele que seja o aspecto de maior valor utilitário do TP: a possibilidade de oferecer estratégias formativas e somativas para apoiar decisões pedagógicas tanto na identificação e remediação de estudantes em risco (avaliação formativa) como para atestar com qualidade e confiabilidade a progressão do aprendiz ao longo do curso e definição de parâmetros para a sua conclusão (avaliação somativa).

Um aspecto necessário a ser discutido diz respeito à escolha do método utilizado para a medida dos construtos propostos. Vivemos uma época na qual a popularidade da Teoria da Resposta ao Item (TRI) cresceu a ponto de alguns estudiosos questionarem mesmo a relevância da Teoria Clássica dos Testes (TCT) e até mesmo, segundo as palavras de Zickar et al.²⁹, tratem a TRI como “uma panaceia para todos os problemas psicométricos”. No entanto, existem ainda situações em que a TCT pode ser suficiente e até mesmo preferida²⁶. Diante da conhecida complexidade dos programas computacionais e da necessidade de uma amostragem bem maior, muitas escolas podem apresentar dificuldades em desenvolver uma análise de seus testes utilizando a TRI, recorrendo a consultorias e ao mercado de banco de itens, que oferecem soluções fidedignas, mas distantes dos objetivos pedagógicos da escola. Dessa forma, como o TP é uma ferramenta que valoriza o acúmulo de conhecimento ao longo do tempo, um estudo longitudinal com normorreferenciamento e utilizando a TCT apresenta uma lógica inerente e que pode ser facilmente instituído em qualquer escola.

CONCLUSÃO

O TP no Brasil não tem sido utilizado para apoiar decisões pedagógicas de “alta aposta” (*high stakes*), ou seja, não representa uma avaliação com consequências mais significativas na vida acadêmica do aprendiz, especialmente

na forma de atestar progressão ao longo do curso. Com isso, corre o risco de se transformar em uma mera data no calendário escolar, à qual o corpo estudantil dará pouco valor. Além disso, os gestores poderão questionar o custo-benefício do TP. Dessa forma, se a escola pretender utilizar o TP como instrumento somativo válido desde o primeiro ano do ensino médico terá que promover pelo menos três testes anuais, a exemplo de outras escolas.

Mesmo com um único teste anual, o TP revelou-se um instrumento valioso para atestar a progressão do aluno ao longo do curso e principalmente para identificar os estudantes em risco e que necessitam de um programa de remediação.

A pandemia da Covid-19 trouxe para as escolas médicas inúmeros desafios, mas, acima de tudo, várias possibilidades de avanço. No atual contexto, o TP torna-se mais relevante do que nunca, já que a curva de crescimento cognitivo poderá trazer dados valiosos até mesmo sobre o impacto do ensino remoto na educação médica. O isolamento também forçou as escolas médicas a adotar estratégias de avaliação remota baseadas no computador, o que implicará significativa redução de custos, possibilitando a adequação de recursos que possibilitem um maior número de testagens.

CONTRIBUIÇÃO DOS AUTORES

Ruy Guilherme Silveira de Souza participou da pesquisa do projeto e da redação e revisão do manuscrito. Angélica Maria Bicudo participou da revisão do manuscrito.

CONFLITO DE INTERESSES

Declaramos não haver conflito de interesses.

FINANCIAMENTO

Declaramos não haver financiamento.

REFERÊNCIAS

1. Freeman A, van der Vleuten C, Nouns Z, Ricketts Z. Progress testing internationally. *Med Teach*. 2010;32:451-5.
2. Messick S. Validity. In: Linen R, editor. *Educational measurement*. 3rd ed. Washington, DC: American Council on Education; 1989. p. 13-104.
3. Newton PE, Shaw S. *Validity in educational & psychological assessment*. London: Sage; 2014.
4. Sireci S. On the validity of useless tests. *Assessment in Education: Principles, Policy & Practices*. 2016;23(2):226-35.
5. Downing S. Validity: on the meaningful interpretation of assessment data. *Med Educ*. 2003;37:830-7.
6. Downing S. Validity threats: overcoming interference with proposed interpretation of assessment data. *Med Educ*. 2004;38(3):327-33.
7. Harden R, Lilley P, Patricio M. *The definitive guide to the OSCE*. Edinburgh: Elsevier; 2016.
8. Pasquali L, Primi R. Fundamentos da teoria da resposta ao item – TRI. *Aval Psicol*. 2003;2:99-110.

9. Royal K. Four tenets of modern validity theory for medical education assessment and evaluation. *Advances in Medical Education and Practice*. 2017;8(3):567-9.
10. Souza R, Bicudo A, Costa B, Martins A. Validity concept in medical education: a bibliometric analysis. *Rev Bras Educ Med*. 2020;44(4):e166.
11. Anderson S. Social competency in young children. *Dev Psychol*. 1974;10(2):282-93.
12. Govaerts M, van der Vleuten C. Validity in work-based assessment: expanding our horizons. *Med Educ*. 2013;47(12):1164-74.
13. Glaser AN. *High-yield biostatistics*. 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 2001.
14. Tio R, Schutte B, Meiboom AA, Greidanus J, Dubois E, Bremers A, et al. Medicine, the progress test of medicine: the Dutch experience. *Perspect Med Educ*. 2016;5:51-5.
15. Albanese M, Case S. Progress testing: critical analysis and suggested practices. *Adv Health Sci Educ*. 2016;21(1):221-34.
16. Hift R. Should essays and other open-ended type questions retain a place in written summative assessment in clinical medicine? *BMC Med Educ*. 2014;14:249. doi: 10.1186/s12909-014-0249-2.
17. Schuwirth L, Bosman G, Henning R, Rinkel R, Wenink A. Collaboration on Progress Testing in medical schools in the Netherlands. *Med Teach*. 2009;32(6):476-9.
18. Wrigley W, van der Vleuten CP, Freeman A, Muijtjens A. A systemic framework for the progress test: strengths, constraints and issues. *Med Teach*. 2012;34(9):683-97.
19. Rademakers J, Ten Cate T, Bär P. Progress testing with short answer questions. *Med Teach*. 2005;27(7):578-82.
20. McHarg J, Bradley P, Chamberlain S, Ricketts C, Searle J, Mclachlan J. Assessment of progress tests. *Med Educ*. 2005;39:221-7.
21. CJ R, et al. The don't know option in Progress Testing. *Adv Health Sci Educ*. 2015;20:1325-38.
22. Muijtjens A, Mameren H, Hoogenboom R, Evers J, van der Vleuten C. The effect of a "don't know" option on test scores: number-right and formula scoring compared. *Med Educ*. 1999;33(4):267-75.
23. Bliss L. A test of Lord's assumption regarding examinee guessing behavior on multiple-choice tests using elementary school students. *J Educ Meas*. 1980;12(2):147-53.
24. Bicudo AM, Hamamoto Filho PT, Abbade JF, Hafner MLMB, Maffei CML. Teste de Progresso em consórcios para todas as escolas médicas do Brasil. *Rev Bras Educ Med*. 2019;43(4):151-6.
25. Muijtjens A, Hoogenboom R, Verwijnen G, van der Vleuten C. Relative or absolute standards in assessing medical knowledge using Progress Tests. *Adv Health Sci Educ*. 1998;3(2):81-7.
26. Langer M, Swanson D. Practical considerations in equating progress tests. *Med Teach*. 2010;32(6):509-12.
27. Callegari-Jacques SM. *Bioestística: princípios e aplicações*. São Paulo: Artmed; 2003.
28. Harvey R, Hammer A. Item Response Theory. *Couns Psychol*. 1999;27:353-83.
29. Zickar J, Broadfoot A. The partial revival of a dead horse? Comparing Classical Test Theory and Item Response Theory. In: Lance C, Vandenberg R, editors. *Statistical and methodological myths and urban legends: doctrine, verity and fable in the organizational and social sciences*. New York: Routledge; 2009. p. 38-57.



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.

Análise da adequação dos itens do Teste de Progresso em medicina

Item analysis of Progress Test in medicine

Edlaine Faria de Moura Villela¹ edlaine@alumni.usp.br
Miguel Angelo Hyppolito¹ mahyppo@fmrp.usp.br
Julio Cesar Moriguti¹ moriguti@fmrp.usp.br
Valdes Roberto Bollela¹ 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.

¹Faculdade 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*.

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 aluno¹.

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 psicométricas².

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 Medicina³.

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 médica⁶. 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 todo⁷.

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 instituição⁸.

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 existentes⁹. 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 Brasil¹⁰.

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 informativa⁷. 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 questões¹⁰.

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 Bloom¹².

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 obstetria, 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 conhecimento no período de 2013 a 2018

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

Fonte: Elaborada pelos autores.

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
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

Anos	Grandes áreas do conhecimento						Total
	Ciências básicas	Clínica médica	Cirurgia	Ginecologia e obstetrícia	Medicina social	Pediatria	
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.

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

Anos	Grandes áreas do conhecimento						Total
	Ciências básicas	Clínica médica	Cirurgia	Ginecologia e obstetrícia	Medicina social	Pediatria	
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						

Fonte: Elaborada pelos autores.

DISCUSSÃO

OTP 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 aprovação¹³. 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 profissional¹⁴.

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 risco¹⁵. 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/reprovação¹⁵. 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 internacional⁸.

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 educacional¹⁹. 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 dificuldade². A reprodutibilidade (confiabilidade) da prova indica sua consistência interna, ou se a amostragem foi adequada e se existe estabilidade na medida que foi feita¹⁹.

Além de aderir às boas práticas na elaboração de itens⁸, 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 prova²⁰.

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 nacionalmente¹⁸⁻²¹.

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 nº 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 in longitudinal 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 international context: a start for benchmarking global standardization? *Med Teach*. 2005;27(6):514-20.
15. Lahner FM, Schaubert 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.



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.

Desempenho cognitivo dos estudantes de Medicina no Teste de Progresso

Cognitive performance of medical students in the Progress Test

Érika Feitosa Queiroz¹ erikafqueiroz@gmail.com

Olivia Andréa Alencar Costa Bessa¹ oliviabessa@unifor.br

Daniela Chiesa¹ danichiesa@unifor.br

RESUMO

Introdução: O Teste de Progresso (TP) é uma avaliação seriada e formativa, com conteúdo cumulativo desejável ao final do curso. O TP fornece um feedback do desenvolvimento cognitivo do estudante. Para o curso, faz uma avaliação diagnóstica do currículo e permite identificar potencialidades e possíveis falhas ao longo do processo de formação.

Objetivo: Este estudo teve como objetivo avaliar a evolução cognitiva de graduandos de uma escola médica da Região Nordeste do país por meio do TP.

Método: Trata-se de estudo observacional, analítico, transversal, com abordagem quantitativa, em que se utilizaram os resultados dos TP das turmas de 2013 a 2016 do curso de Medicina da Universidade de Fortaleza. Foram incluídos os resultados dos TP de 2013 a 2021, descritos pela porcentagem da média de cada score, calculado o diferencial de desempenho entre o primeiro e sexto anos, pelo teste T pareado e pelo coeficiente de correlação de Pearson para os resultados do sexto ano. O estudo respeitou os aspectos éticos e foi aprovado pelo Comitê de Ética da IES.

Resultado: Foram incluídos os resultados de 361 estudantes (88,0%) de seis turmas que participaram do teste desde S1. Verificou-se um aumento progressivo no desempenho cognitivo global em cada turma ao longo dos seis anos de TP. A diferença de desempenho (DDO) da turma 2015.2 foi superior às demais ($p = 0,047$; $R = 0,73$). O grau de dificuldade das provas foi semelhante. Em relação ao desempenho global do sexto ano, houve diferença no desempenho dos alunos entre as turmas, variando de 64,3% em 2016.2 a 72,5% em 2014.2 ($p = 0,003$; $R = 0,85$). A partir do terceiro ano, a análise do desempenho mostrou-se crescente para todas as turmas. O desempenho nas áreas de cirurgia e saúde coletiva apresentou maior variação entre as turmas.

Conclusão: O TP confirmou ser uma ferramenta importante para a avaliação formativa dos estudantes e para o diagnóstico do currículo.

Palavras-chave: Teste de Progresso; Avaliação Formativa; Educação Médica.

ABSTRACT

Introduction: The progress test (PT) is a knowledge assessment tool, consisting of desirable cognitive undergraduate content, that provides feedback on the student's cognitive development. For the school, it serves as a diagnostic evaluation of the curriculum.

Objective: The aim of the study was to assess the cognitive performance in the progress test of students in the Brazilian Northeast.

Method: A cross-sectional survey was conducted with results from six classes of students that performed the PT since the first semester of the medical course. The analysis consisted of PT results from 2013 to 2021, described by means, Y6-Y1 performance difference, Student's T test and Pearson correlation at 6Y. The study was approved by a Research Ethics Committee.

Result: The results of 361 (88%) students showed a progressive increase in PT global scores from Y1 to Y6. The performance difference of the 2015.2 class was higher ($p=0.047$; $R=0.73$) than the others and the test difficulty index was similar for all. There was a difference between Y6 classes, ranging from 64.3% to 72.5% ($p=0.003$; $R=0.85$). From Y3 to Y6, all classes increased their PT scores. Higher variability between classes was reported in the areas of Surgery and Public Health.

Conclusion: The progress test proved to be an important tool for knowledge assessment and for evaluation of the medical curriculum.

Keywords: Progress Test; Assessment; Medical Education.

¹Universidade de Fortaleza, Fortaleza, Ceará, Brasil.

Editor: Valdes Roberto Bollela.

Recebido em 02/09/22; Aceito em 19/10/22.

Avaliado pelo processo de *double blind review*.

INTRODUÇÃO

A formação médica exige avaliações periódicas para que o perfil do egresso seja compatível com as necessidades dos pacientes nos diferentes serviços de saúde. Uma das habilidades cognitivas necessárias ao médico é a capacidade de acumular as informações adquiridas ao longo da graduação, de modo organizado, para a construção do raciocínio clínico. As escolas precisam estimular o aprimoramento de tal habilidade em seus alunos, avaliando o ganho de conhecimentos ao longo do curso por meio de diferentes ferramentas¹. A avaliação cumulativa pode orientar o comportamento de estudo dos estudantes ao longo do curso, incentivando o estudo autodirigido para desenvolver a aprendizagem. Cursos que adotam currículos integrados e metodologias de ensino-aprendizagem centradas no estudante valorizam o percurso percorrido no processo da aprendizagem, oportunizando evolução cognitiva dos seus graduandos².

Uma das formas de avaliar o aprendizado ao longo do curso de Medicina é o Teste de Progresso (TP). Essa ferramenta avalia individualmente e de forma seriada a aquisição de conhecimento por parte do estudante. Desde a década de 1990, as escolas brasileiras aplicam o TP isoladamente ou constituem núcleos colaborativos para aplicação conjunta, de modo a facilitar o processo de elaboração da prova e análise dos resultados, com troca de experiência, compartilhamento de recursos e otimização do custo-benefício entre elas. A parceria contribui para a confecção de questões com alta qualidade, além de vantagens educacionais globais, permitindo uma comparação dos avanços e/ou das limitações entre as instituições participantes, sem haver qualquer tipo de classificação ou ranqueamento³⁻⁵. Em 2012, a Associação Brasileira de Educação Médica (Abem) estimulou a criação de núcleos interinstitucionais para aplicação do TP, com o objetivo de desenvolvimento de parcerias entre escolas e de ações que impulsionassem a melhoria da formação do médico. Em 2015 e 2021, foram realizadas provas nacionais, e o TP tem sido a ferramenta mais utilizada para quantificar o desenvolvimento cognitivo, em virtude das vantagens de ser aplicado em larga escala⁶⁻⁹.

O TP caracteriza-se por ser uma avaliação cognitiva, formativa, com conteúdo cumulativo desejável ao final do curso. Além disso, fornece um *feedback* do desenvolvimento cognitivo do estudante. Possibilita ao estudante uma análise crítica de seu desempenho ao longo do curso, destacando as potencialidades e identificando as fragilidades nas áreas avaliadas. Os resultados do teste objetivam direcionar os esforços dos estudantes para contemplar os conhecimentos desejáveis em todas as grandes áreas do currículo^{8,10}. Para o curso, o TP é uma forma

de avaliação diagnóstica do currículo que permite identificar as potencialidades possíveis falhas ao longo do processo de formação. O conteúdo do teste não segue um modelo curricular específico e, portanto, avalia os objetivos finais do currículo médico em sua totalidade, independentemente da proposta pedagógica adotada¹¹. É constituído de um número variável de questões, contemplando de forma ampla todas as grandes áreas da medicina: ciências básicas, saúde coletiva, clínica médica, pediatria, ginecologia/obstetrícia, cirurgia, urgência/emergência, ética médica e humanidades. Sugere-se que as questões sejam de múltipla escolha e priorizem alta taxonomia, já que o processo de aprendizado envolve competências complexas^{2,12}. A prova é aplicada a todos os estudantes em um mesmo momento, pelo menos uma vez ao ano, e permite comparar os resultados e analisar a evolução do desempenho do discente ao longo da graduação. Para cada ano, é esperado um escore médio progressivo, variando de acordo com a estrutura curricular do curso e a dificuldade do teste^{13,14}. Como o TP representa um importante instrumento para avaliação de estudantes e do currículo nas escolas médicas, o objetivo deste estudo foi verificar a evolução cognitiva de graduandos de uma escola médica da Região Nordeste do país por meio desse instrumento.

MÉTODO

Trata-se de um estudo observacional, analítico, transversal, com abordagem quantitativa, em que se utilizaram os resultados dos TP das turmas de 2013 a 2016 do curso de Medicina da Universidade de Fortaleza (Unifor). A Unifor faz parte de um consórcio interinstitucional para o TP com mais sete escolas e realiza TP desde 2013. O curso tem ingresso semestral, e o TP é realizado no segundo semestre do ano, entre setembro e outubro. Elegeram-se como população do estudo as oito turmas que realizaram seis TP ao longo da graduação, e incluíram-se os resultados dos testes de progresso de 2013 a 2021. Foi comparado os desempenhos nos seis anos do curso, global, por períodos e por grandes áreas da medicina. Extraíram-se os dados do banco de dados do TP, no qual estão presentes as notas de todos os estudantes que participaram dos TP, descritas em escore de 0 a 10, porcentagem, média e desvio padrão. O TP interinstitucional é uma prova objetiva, com 120 questões de múltipla escolha, com quatro alternativas (gabarito e quatro distratores), sendo 20 questões por cada grande área: Ciências Básicas, Cirurgia, Clínica Médica, Ginecologia/Obstetrícia, Pediatria e Saúde Coletiva. Os conteúdos de Urgência/Emergência, Ética Médica, Bioética e Humanidades estão inseridos longitudinalmente nas grandes áreas. A matriz da prova apresenta a distribuição das questões com base nos objetivos de aprendizagem dos temas considerados relevantes

após validação pelas escolas do consórcio e contempla cenários dos três níveis de atenção à saúde. As questões são elaboradas por docentes de todas as escolas do consórcio, e as questões do banco são, na sua maioria, de nível 3 ou mais da taxonomia de Bloom. As questões são analisadas pela Teoria Clássica dos Testes (TCT) para comparar o grau de dificuldade e de discriminação das provas. Os resultados de acerto casual foram excluídos da análise.

Os resultados do sexto ano foram escolhidos para análise comparativa por serem concludentes. Incluíram-se todos os resultados dos TP aplicados de 2013 a 2021. Excluíram-se os resultados de estudantes com escores sinalizados como acerto casual (respostas em apenas uma letra ou padrão gráfico de respostas), que não realizaram o TP no ano de conclusão do curso ou que não compareceram a duas ou mais provas. As turmas iniciadas em 2015.1 e 2016.1 foram excluídas da análise do sexto ano, pois não realizaram o TP no S12 por causa da colação de grau antecipada, não tendo resultados completos para comparação.

As turmas foram assim denominadas: turma 2013.1 (TP de 2013 a 2018), turma 2013.2 (TP de 2013 a 2018), turma 2014.1 (TP de 2014 a 2019), turma 2014.2 (TP de 2014 a 2019), turma 2015.2 (TP de 2015 a 2020) e turma 2016.2 (TP de 2016 a 2021). As turmas ingressantes no primeiro semestre (ano.1) realizaram TP nos semestres pares do curso (de S2 a S12), e as ingressantes no segundo semestre (ano.2) realizaram TP nos semestres ímpares do curso (de S1 a S11).

Os dados foram tabulados em planilhas do software Microsoft Excel®, analisados por meio de estatística descritiva e descritos pela porcentagem da média de cada escore. Calculou-se o diferencial de desempenho observado (DDO), subtraindo o escore do sexto ano do escore do primeiro ano, expresso em porcentagem. Para as comparações, aplicaram-se o teste T pareado e a análise de variância para medidas repetidas, para as médias entre as turmas e entre os semestres de uma mesma

turma, com intervalo de confiança de 95%, sendo o valor de $p < 0,05$ considerado estatisticamente significativo. Aplicou-se o coeficiente de correlação de Pearson. As comparações entre as turmas respeitaram a equivalência do semestre de ingresso em dois blocos: 2013.1 e 2014.1 e 2013.2, 2014.2, 2015.2 e 2016.2.

O estudo respeitou os aspectos éticos e foi aprovado pelo Comitê de Ética da Universidade de Fortaleza sob o número 1.498.070/2016.

RESULTADOS E DISCUSSÃO

Entre 2013.1 e 2016.2, 552 (100%) estudantes de primeiro ano, pertencentes às oito turmas de ingressantes nesses semestres, foram inscritos para realizar o TP. Excluíram-se os 142 concludentes das turmas de 2015.1 e 2016.1 que colaram grau antes da data da realização do TP no sexto ano, respectivamente em 2020 e 2021. Dos 410 estudantes das demais seis turmas, incluíram-se os resultados de 361 (88,0%), que correspondem aos estudantes que realizaram os seis TP ao longo do curso.

Verificou-se um aumento progressivo no desempenho cognitivo global dos estudantes em cada turma, ao longo dos seis anos de TP (Tabela 1). A diferença de desempenho entre o sexto e primeiro ano (DDO) das turmas 2013.1 e 2014.1 foi semelhante ($p = 0,235$), e a turma 2015.2 teve DDO superior às demais turmas de ingresso de segundo semestre ($p = 0,047$; $R = 0,73$). Isso ocorreu porque essa turma teve um desempenho inferior no TP do S1 (29% de acertos), em relação às demais turmas, sendo significativa a diferença em relação à turma 2014.2 ($p = 0,026$). Não houve diferença significativa em relação às turmas 2013.2 ($p = 0,186$) e 2016.2 ($p = 0,368$). É esperado que os acertos no teste (em percentuais) tenham crescimento do primeiro ano para o sexto ano. O grau de dificuldade das provas analisadas foi semelhante, com cerca de 90% a 95% de questões com grau de dificuldade fácil e médio, e 50% a 60% com questões com boa discriminação (considerando a TCT do ano 6), conforme demonstrado na Tabela 2.

Tabela 1. Desempenho geral dos estudantes em média de acertos (%)

Turmas	TP 2013	TP 2014	TP 2015	TP 2016	TP 2017	TP 2018	TP 2019	TP 2020	TP 2021	DDO*
2013.1	36,1%	52,2%	44,6%	56,8%	61,1%	71,8%	-	-	-	35,7%
2013.2	35,7%	49,5%	41,8%	52,8%	57,2%	66,7%	-	-	-	31%
2014.1	-	40,7%	37,1%	50,9%	56,5%	64%	72,5%	-	-	31,8%
2014.2	-	39,8%	36,4%	49,1%	53,4%	62,9%	71,3%	-	-	31,5%
2015.2	-	-	29,8%	41,0%	49,3%	59%	66,2%	69,9%	-	40,1%
2016.2	-	-	-	32,6%	42,5%	51,1%	59,8%	61,8%	64,3%	31,7%

* Diferencial de desempenho observado.

Fonte: Banco de dados dos TP de 2013 a 2021.

Quando se comparou o desempenho global do sexto ano (TP S11) nas turmas ingressantes no segundo semestre, houve diferença no desempenho dos alunos entre as turmas, variando de 64,3% em 2016.2 a 72,5% em 2014.2 ($p = 0,003$; $R = 0,85$). As turmas 2013.2 e 2014.2 tiveram desempenhos semelhantes no TP do S12. A variação observada provavelmente foi mais influenciada pelas características específicas de cada turma, já que as provas tiveram graus de dificuldade semelhantes (Tabela 2).

A partir do terceiro ano, a análise do desempenho das turmas ao longo dos seis testes da graduação mostrou-se crescente para todas as turmas (Gráfico 1). Em relação à diferença de desempenho (DDO) entre o primeiro e o último

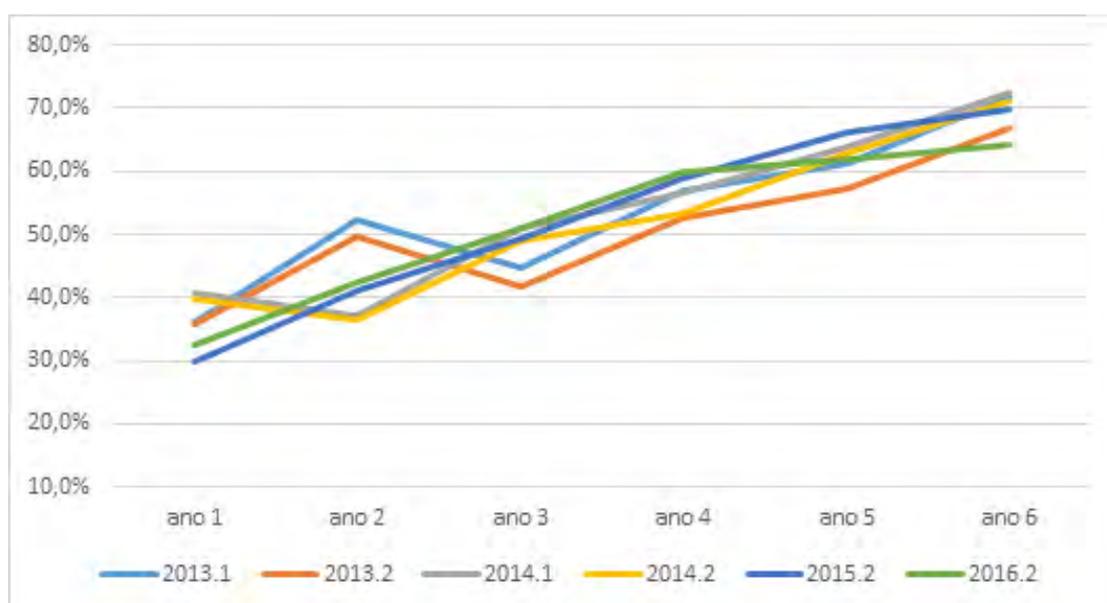
ano das turmas, verificou-se que quase todos os valores estavam dentro da faixa esperada (25%-35%)¹, com exceção do valor para a turma 2015.2, que foi superior aos demais. Tal diferença foi observada por um escore inferior no teste do S1, quando comparada às demais turmas, mantendo o padrão de crescimento para os demais anos. Em 2015, a prova realizada foi TP nacional, em que as questões estavam divididas em cinco áreas, com o conteúdo das ciências básicas distribuído entre as demais áreas⁷, o que pode em parte explicar tal diferença de desempenho. Ao mesmo tempo, esse teste continha questões elaboradas por escolas de todo o Brasil, o que também é um fator para explicar tal desempenho, pela menor familiaridade com o tipo de questão.

Tabela 2. Análise das questões do TP pela Teoria Clássica dos Testes (%)

	2018	2019	2020	2021
<i>Dificuldade</i>				
Fácil	30%	75%	45%	50%
Médio	65%	20%	45%	45%
Difícil	5%	5%	10%	5%
<i>Discriminação</i>				
Deficiente	20%	35%	10%	5%
Regular	25%	15%	30%	10%
Boa	55%	50%	60%	85%

Fonte: Banco de dados dos TP de 2018 a 2021.

Gráfico 1. Evolução de desempenho ao longo de seis anos da graduação (% média de acertos)



Fonte: Banco de dados dos TP de 2013 a 2021.

O Gráfico 2 mostra o desempenho do sexto ano nas grandes áreas do teste da medicina. Observa-se um número médio de acertos similar para Ciências Básicas, Clínica Médica, Pediatria e Ginecologia/Obstetrícia, com pequenas variações por turma. A área de Ciências Básicas manteve-se com média de acertos elevada nas turmas do último ano da graduação, apesar de constituir uma área com conteúdos ministrados nos primeiros semestres da graduação.

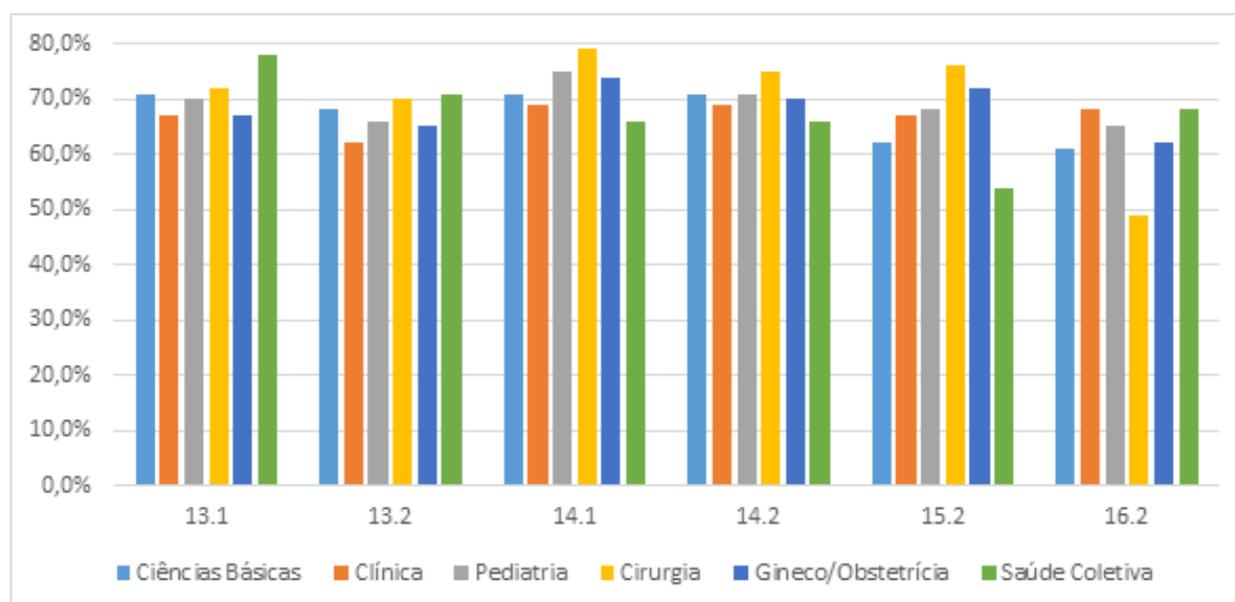
O desempenho nas áreas de Cirurgia e Saúde Coletiva apresentou maior variação entre as turmas. A turma de 2016.2 apresentou o menor desempenho do TP do S11, provavelmente relacionado ao escore na prova de cirurgia (49%), com diferença significativa nessa área para as demais turmas ($p = 0,012$). Em Saúde Coletiva, a turma 2015.2 teve desempenho inferior não significativo no TP S11 em relação às demais ($p = 0,070$). A turma de 2016.2 foi a que apresentou o menor desempenho global no TP do sexto ano. Essa turma realizou a prova durante a pandemia do novo coronavírus e de forma virtual, o que pode ter impactado negativamente os resultados por conta das necessidades de isolamento impostas, com prejuízo às atividades do internato no último ano do curso¹⁵.

O impacto da pandemia pode ser observado também nos resultados da turma 2015.2 no TP S11, na área de Saúde Coletiva. Essa turma em especial fez um semestre de internato adaptado para os serviços que tinham vagas disponíveis, em razão das necessidades de remanejamento causadas pela pandemia. Quase todos os estágios foram em serviços hospitalares de atendimento à doença pelo coronavírus 19 (Covid-19), já que

os serviços ambulatoriais eletivos estavam fechados. Em nossa cidade, a maioria dos serviços de Atenção Primária à Saúde (APS) e Medicina de Família e Comunidade, além do internato rural realizado no interior, suspendeu as atividades de ensino, entre abril e outubro de 2020, por causa dos desafios impostos pela pandemia e da falta de equipamentos de proteção individual para atender de forma adequada os estudantes. Com isso, alguns internos não completaram seus rodízios de APS antes do TP S11, o que pode ter influenciado no desempenho da turma.

Em relação à Cirurgia Geral, foi a área com menor desempenho no sexto ano no primeiro TP em 2013 – 58% de acertos (média da prova = 69,7% de acertos), confirmando uma tendência observada: menor inserção desse conhecimento em relação às demais áreas na matriz curricular do curso. Toda a inserção curricular de temas cirúrgicos foi revista utilizando os resultados do TP como ferramenta de avaliação para definir em quais momentos as competências em cirurgia deveriam ser reforçadas ou inseridas. As atividades foram reorganizadas aumentando a inserção em todos os semestres e redefinindo os rodízios de internato na clínica cirúrgica. Os avanços nessa área culminaram em maiores médias de acertos ao final do sexto ano (variando de 70% a 79%) nas turmas que ingressaram até 2015.2 e provavelmente são consequência dos ajustes realizados no currículo. A pandemia da Covid-19, com os desafios impostos aos serviços de saúde, pode ter impactado o menor desempenho em Cirurgia Geral no TP S11 pela turma 2016.2. Essa turma ingressou no internato em 2020.2, e os

Gráfico 2. Desempenho por área no sexto ano (% média de acertos)



Fonte: Banco de dados dos TP de 2013 a 2021.

rodízios de cirurgia foram suspensos na primeira onda (abril a outubro/2020), e novamente foram impactados na segunda onda, que se iniciou em janeiro de 2021. Em Fortaleza, os leitos destinados às cirurgias eletivas foram direcionados para o atendimento clínico de pacientes com Covid-19, e essa turma ficou praticamente três semestres do internato sem ter oportunidade de manter os rodízios previstos nessa área.

O curso de Medicina da Unifor utiliza metodologias ativas de ensino-aprendizagem, tendo a aprendizagem baseada em problemas como orientação pedagógica do S1 ao S8. O TP é aplicado desde 2013, e a escola participou de nove edições, e nesse período seis turmas fizeram o TP anualmente, do primeiro ao sexto ano. A participação dos estudantes é voluntária. Em 2020 e 2021, a prova foi aplicada em plataforma virtual, de forma síncrona, em razão da pandemia pelo novo coronavírus.

Ainda há a necessidade do incremento de estímulos para garantir a adesão da totalidade dos estudantes ao TP, de modo que eles entendam que se trata de uma ferramenta de estudo com potencial para a melhoria na aquisição de conhecimento durante a graduação. Observamos que desempenho dos estudantes no sexto ano foi acima da média observada em alguns estudos, que registraram uma média percentual de acertos para o sexto ano entre 50% e 60% em todo o teste^{7,9,13,14}. A variação observada na nesta amostra (de 64,3 a 72,5% de acertos) provavelmente foi mais influenciada pelas características específicas de cada turma, já que as provas tiveram graus de dificuldade semelhantes. O curso possibilita experiências de ensino-aprendizagem em um currículo consolidado, o que pode ser evidenciado por desempenhos similares entre diferentes turmas ao longo dos anos de aplicação do TP aos alunos.

A evolução do desempenho ao longo dos seis anos do curso manteve o padrão de crescimento a partir do terceiro ano para ambas as turmas, o que condiz com a aquisição satisfatória de conhecimento ao longo do curso, estando de acordo com as exigências curriculares e com o observado em estudos semelhantes^{7,9,13,14}.

CONCLUSÃO

O TP confirmou ser uma ferramenta importante para a avaliação formativa dos estudantes e para diagnóstico do currículo, de modo a evidenciar suas potencialidades e fragilidades, e permitir subsidiar ajustes no processo pedagógico. Para além do ganho em desempenho, coexiste uma melhor apropriação dos conteúdos transmitidos e, conseqüentemente, um desenvolvimento de competências essenciais, com aprimoramento da formação profissional.

CONTRIBUIÇÃO DAS AUTORAS

Érika Feitosa Queiroz participou do planejamento e da escrita do projeto, da aplicação dos Termos de Consentimento Livre e Esclarecido, da análise dos dados e da escrita do manuscrito. Olivia Andréa Alencar Costa Bessa participou do planejamento e da revisão do projeto, da análise dos dados e da revisão do manuscrito. Daniela Chiesa participou do planejamento e da revisão do projeto, da compilação dos dados, da edição do banco de dados para forma anônima para ser enviado para análise, da revisão da análise de dados e da revisão do manuscrito.

CONFLITOS DE INTERESSE

Érika Feitosa Queiroz era bolsista de iniciação científica quando o projeto foi iniciado e realizou o TP como estudante. Entretanto, não teve acesso ao banco de dados identificados, recebendo os dados anonimizados. Analisaram-se os dados, e o artigo foi elaborado após a sua colação de grau. Olivia Bessa e Daniela Chiesa são professoras do curso de Medicina da Unifor e estavam na gestão acadêmica do curso quando foi implantado o TP. Daniela Chiesa teve acesso ao banco de dados completo do TP e foi responsável por sua edição para posterior envio à bolsista Érika.

FINANCIAMENTO

Declaramos não haver financiamento.

REFERÊNCIAS

1. 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.
2. Sakai MH, Ferreira Filho OF, Matsuo T. Avaliação do crescimento cognitivo do estudante de Medicina: aplicação do teste de equalização no Teste de Progresso. *Rev Bras Educ Med.* 2011;35(4):493-501.
3. van der Vleuten CPM, Schuwirth LWT, Muijtjens AMM, Thoben AJNM, Cohen-Schotanus J, Van Boven CPA. Cross institutional collaboration in assessment: a case on progress testing. *Med Teach.* 2004;26:719-25.
4. Muijtjens A, Schuwirth L, Cohen-Schotanus J, van der Vleuten CPM. Differences in knowledge development exposed by multi-curricular progress test data. *Adv in Health Sci Educ.* 2008;13:593-605.
5. Wrigley W, van der Vleuten CP, 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.
6. Lampert JB, Bicudo AM. 10 anos de Diretrizes Curriculares Nacionais dos cursos de graduação em Medicina. Rio de Janeiro: Associação Brasileira de Educação Médica; 2014.
7. Bicudo AM, Hamamoto PT, Abbade JF, Hafner MDLMB, Maffei CML. Teste de Progresso em consórcios para todas as escolas médicas do Brasil. *Rev Bras Educ Med.* 2019;43(4):151-6.
8. Zeferino AMB, Domingues RCL, Amaral E. Feedback como estratégia de aprendizado no ensino médico. *Rev Bras Educ Med.* 2007;31(2):176-9.
9. Schreiber SO, Postal EA, Afonso DH, Merse CE, Cyrino EG, Abreu Junior AF, et al. Teste de Progresso da Abem: consolidando uma estratégia de avaliação para o ensino médico. *Rev Bras Educ Med.* 2022;46(1):e061.
10. Pinheiro OL, Spadellal MA, Moreira HM, Ribeiro ZMT, Guimarães 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.

11. Sartor LB, Rosa LL, Madeira K, Uggioni MLR, Ferreira Filho OF, Rosa MI. Undergraduate Medical student's perception about the Progress Testing. *Rev Bras Educ Med.* 2020;44(2):e062.
12. Reberti AG, Monfredini NH, Ferreira Filho OF, Andrade DF, Pinheiro CEA, Silva JC. Progress Test in medical school: a systematic review of the literature. *Rev Bras Educ Med.* 2020;44(1):e014.
13. Tomic ER, Martins MA, Lotufo PA, Benseñor IM. Progress testing: evaluation of four years of application in the school of medicine, University of São Paulo. *Clinics* 2005;60(5):389-96.
14. van der Vleuten CPM, Verwijnew GM, Wijnen WHFW. Fifteen years of experience with progress testing in a problem-based learning curriculum. *Med Teach.* 1996;18(2):103-9.
15. Coelho BM, Meira EB, Araujo LS, Araujo PFF, Tenisi SS, Santos ACSR, et al. O impacto da pandemia da Covid-19 na formação médica: uma revisão integrativa. *Revista Ibero-Americana De Humanidades, Ciências E Educação.* 2021;7(12):522-45.



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.

Como avaliar programas de residência a partir do Teste de Progresso?

How can residency programs be evaluated based on the Progress Test?

Gustavo Salata Romão¹

gsalata@uol.com.br

Cesar Eduardo Fernandes²

cesarefernandes_dircient@febrasgo.org.br

Aginaldo Lopes da Silva Filho³

agnaldo.ufmg@gmail.com

Sheyla Ribeiro Rocha⁴

sheyla@ufscar.br

Marcos Felipe Silva de Sá⁵

marcosfelipe@fmrp.usp.br

RESUMO

Introdução: Embora o conhecimento especializado seja um elemento fundamental para a prática médica qualificada, não há, na maioria das especialidades, uma avaliação cognitiva unificada dos médicos residentes, e, conseqüentemente, não é possível verificar o conhecimento agregado durante o treinamento pelos programas de residência médica (PRM). O Teste de Progresso (TP) oferece uma oportunidade para avaliação dos PRM a partir do desempenho dos seus residentes. Em 2018, a Febrasgo implementou o Teste de Progresso Individual do Residente em Ginecologia e Obstetrícia (TPI-GO), que tem sido aplicado em todo o Brasil.

Relato de experiência: Este estudo descritivo se refere ao acompanhamento longitudinal dos residentes que iniciaram a participação no TPI-GO em 2018 como R1 (n = 497) e concluíram a participação em 2020 como R3 (n = 314). O desempenho desses residentes no TPI-GO serviu de base para analisar o perfil de 32 PRM localizados nas Regiões Sul (28,1%), Sudeste (68,8%) e Centro-Oeste (3,1%), sendo identificados cinco diferentes perfis de PRM em relação ao desempenho dos residentes iniciantes, diferenças de desempenho entre R3 e R1 e desempenho dos concluintes.

Discussão: No Brasil, não são oferecidas avaliações abrangentes e unificadas de conhecimento aos médicos residentes na maioria das especialidades, e conseqüentemente ainda não é possível incorporar essas informações na avaliação dos PRM. No modelo aqui apresentado, o desempenho dos residentes no TP possibilita inferir sobre o processo seletivo, o conhecimento agregado pelo PRM ao longo do treinamento e o nível de conhecimento dos concluintes, sendo reconhecidos PRM qualificados (tipo 1) e PRM que necessitam de melhorias (tipos 2, 3, 4 e 5).

Conclusão: O TP oferece uma oportunidade para avaliação dos PRM a partir do desempenho dos seus residentes. Por meio do modelo aqui apresentado, é possível obter informações para subsidiar decisões institucionais que promovam melhorias dos PRM e do seu processo de formação na especialidade.

Palavras-chave: Educação Médica; Residência Médica; Avaliação de Programas e Projetos de Saúde; Teste de Progresso; Ginecologia e Obstetrícia.

ABSTRACT

Introduction: Although specialized knowledge is a fundamental element for qualified medical practice, in most specialties, there is no unified cognitive assessment of resident physicians, and it is impossible to verify the knowledge gained during training in Medical Residency Programs (MRPs). The Progress Test (PT) provides an opportunity to evaluate MRPs based on the performance of the residents. In 2018, Febrasgo applied the Resident Progress Test in Gynecology and Obstetrics (TPI-GO) throughout Brazil.

Experience report: This descriptive study refers to the longitudinal follow-up of residents who started participating in the TPI-GO in 2018 as R1 (n=497) and completed their participation in 2020 as R3 (n = 314). The performance of these residents in the TPI-GO served as a basis for analyzing the profile of 32 MRPs located in the South (28.1%), Southeast (68.8%), and Central-West (3.1%) regions of Brazil, with five different identified PRM profiles in relation to the performance of beginner residents, differences in performance between R3 and R1 and the performance of residency graduates.

Discussion: In Brazil, comprehensive and unified assessments of knowledge are not offered to resident physicians in most specialties, and consequently, it is not yet possible to incorporate this information into the assessment of MRPs. In the model presented here, the performance of residents in the PT enables one to make inferences about the selection process, the knowledge added by the MRP throughout the training, and the level of knowledge of the graduates, being recognized as qualified MRPs (type 1) and MRPs that need improvement (types 2, 3, 4 and 5).

Conclusion: The PT offers an opportunity to evaluate MRPs based on the performance of their residents. Through the model presented here, it is possible to obtain information to support institutional decisions that promote improvements in MRPs and their training process in the specialty.

Keywords: Medical Education; Medical Residency; Program Evaluation; Progress Test; Gynecology and Obstetrics.

¹Universidade de Ribeirão Preto, Ribeirão Preto, São Paulo, Brasil.

²Centro Universitário Faculdade de Medicina do ABC, Santo André, São Paulo, Brasil.

³Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brasil.

⁴Universidade Federal de São Carlos, São Carlos, São Paulo, Brasil.

⁵Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo, Ribeirão Preto, São Paulo, Brasil.

Editor: Valdes Roberto Bollela.

Recebido em 09/09/22; Aceito em 02/10/22.

Avaliado pelo processo de *double blind review*.

INTRODUÇÃO

A residência médica é uma modalidade de pós-graduação *lato sensu*, sendo reconhecida como padrão ouro no treinamento de especialidades médicas. O modelo proposto por William Halsted serviu de base estruturante para a maioria dos programas de residência médica (PRM) em todo o mundo. Segundo esses preceitos, a especialização médica requer um treinamento que proporcione experiência prática com pacientes reais durante um período fixo preestabelecido, responsabilização crescente pelo cuidado e supervisão por médicos especialistas mais experientes¹.

Entretanto, diversas publicações apontam para o papel central do conhecimento técnico-científico para a formação e a atividade médica qualificada^{2,3}. Nesse sentido, a avaliação e o acompanhamento da aquisição de conhecimento na especialidade pelos médicos residentes ao longo do treinamento se tornam essenciais.

De acordo com o consenso para uma boa avaliação proposto por Norcini et al.⁴, uma avaliação de desempenho individual adequada deve atender a uma série de quesitos, que incluem a validade, a confiabilidade, a equivalência, a viabilidade, a aceitabilidade e os efeitos educacional e catalisador. O Teste de Progresso (TP) consiste em uma avaliação de conhecimento abrangente, seriada e longitudinal, com níveis satisfatórios de validade, confiabilidade e equivalência, que permitem subsidiar decisões de médio e alto impactos, como progressão, habilitação ou certificação. Por sua vez, a viabilidade, a aceitabilidade e os efeitos educacional e catalisador atribuídos ao TP o credenciam também como método apropriado para avaliação formativa⁵.

Quando comparado a outras modalidades de avaliação cognitiva, o TP apresenta uma série de vantagens para o aprendiz, os supervisores, os gestores de programas educacionais e a sociedade em geral. Sob a perspectiva dos PRM, o TP fornece informações confiáveis sobre o nível de conhecimento dos seus residentes e permite a identificação de desvios (*outliers*) que necessitam de maior apoio, supervisão, remediação ou progressão acelerada^{6,7}. O desempenho de um grupo de aprendizes de um mesmo PRM no TP permite a avaliação indireta da qualidade dos programas e estágios curriculares, identificando áreas de bom desempenho e setores e estágios com necessidades de melhoria^{8,9}.

Na graduação em Medicina, o TP foi introduzido pela primeira vez na década de 1970, pelas Universidades de Kansas City nos Estados Unidos e Maastrich na Holanda,

sendo atualmente utilizado em todo o mundo⁶. Na residência médica, o seu uso tem sido descrito desde 1999, quando foi implementado na Holanda para avaliação dos médicos residentes em ginecologia e obstetrícia⁵.

No Brasil, embora TP seja amplamente utilizado na graduação em Medicina¹⁰, seu uso na residência tem sido bastante restrito, sendo oferecido por poucas especialidades médicas⁸. Diante dessa constatação, a Federação Brasileira das Associações de Ginecologia e Obstetrícia (Febrasgo) desenvolveu o Teste de Progresso Individual do Residente em Ginecologia e Obstetrícia (TPI-GO) e ofereceu essa modalidade de avaliação a todos os médicos residentes nessa especialidade no Brasil a partir de 2018⁸. Este artigo tem por objetivo compartilhar a experiência da Febrasgo na utilização de um modelo de avaliação dos PRM a partir do desempenho dos residentes no TPI-GO.

RELATO DE EXPERIÊNCIA

Este é um estudo descritivo que relata a experiência da Febrasgo no acompanhamento longitudinal de candidatos que iniciaram o TPI-GO em 2018 como R1 e terminaram em 2020 como R3, correspondendo à primeira coorte de médicos residentes que participaram dessa avaliação. O desempenho dos aprendizes nas avaliações individuais tem sido amplamente utilizado como componente na avaliação de programas educacionais¹¹. Amparada nos preceitos da avaliação orientada por objetivos¹², a Febrasgo tem utilizado o desempenho do grupo de residentes de cada programa no TP como base para compor um modelo de avaliação do conhecimento agregado pelos PRM durante o treinamento.

Este relato de experiência utilizou um banco de dados, cujas informações são agregadas, sem possibilidade de identificação individual e, por isso, está dispensado de ser registrado ou avaliado pelo sistema do Comitê de Ética em Pesquisa e da Comissão Nacional de Ética em Pesquisa (CEP/Conep), conforme Resolução nº 510, de 7 de abril de 2016, do Conselho Nacional de Saúde¹³.

Teste de Progresso Individual do Residente em Ginecologia e Obstetrícia (TPI-GO)

O TPI-GO foi implementado pela Febrasgo em 2018⁹. A prova teórica para obtenção do Título de Especialista em Ginecologia e Obstetrícia (Tego) foi escolhida como base para o TPI-GO, por tratar-se de uma avaliação abrangente e pautada no conteúdo da Matriz de Competências em Ginecologia e Obstetrícia^{14,15}. Essa prova contém 100 testes

de múltipla escolha com quatro alternativas, sendo 50 testes de ginecologia e 50 de obstetrícia⁹. A matriz de prova e as questões são elaboradas e revisadas pela Comissão Nacional do Tego (CN-Tego), composta por especialistas de diferentes regiões brasileiras⁸. Ao final de cada aplicação do TPI-GO, realiza-se a análise psicométrica dos itens quanto ao grau de dificuldade e poder discriminativo, fornecendo bases para os trabalhos da comissão de revisores no ano subsequente¹⁶.

Desde a sua implementação, o TPI-GO é oferecido anualmente a todos residentes do primeiro, segundo e terceiro anos de treinamento na especialidade (R1, R2 e R3, respectivamente), em nível nacional. A participação na prova é voluntária (não obrigatória), e, para estimular a adesão dos residentes, foram estabelecidos critérios que possibilitam a dispensa ou bonificação na prova teórica do Tego no ano subsequente à conclusão do PRM⁸. De acordo com esses critérios, serão dispensados da prova teórica do Tego os candidatos que participarem do TPI-GO como R1, R2 e R3, apresentarem um desempenho igual ou superior ao percentil 60 (P60) em relação aos seus pares em pelo menos duas versões da prova (incluindo necessariamente o desempenho como R3) e obtiverem uma percentagem de acertos de pelo menos 60% nas questões de ginecologia e de obstetrícia na última participação (como R3). Serão bonificados na prova teórica do Tego os candidatos que participarem do TPI-GO como R1, R2 e R3, apresentarem um desempenho superior ao percentil 30 (P30) em relação aos seus pares em todas as versões da prova e obtiverem uma percentagem de acertos de pelo menos 50% nas questões de ginecologia e de obstetrícia, na última participação (como R3). Nesta última condição, os candidatos terão um bônus de 10% nas questões de ginecologia e de obstetrícia na prova teórica do Tego no ano subsequente⁸. Segundo o Guia da Associação Europeia de Educação Médica para o TP (Association for Medical Education in Europe – Amee – Guide nº 71)¹⁷, por causa da variação nos níveis de dificuldade entre as diferentes versões do TP, a análise do desempenho baseada em valores fixos da nota (*cut-off*) torna-se mais precária, e recomendam-se nesse caso critérios referenciados por norma, que dependem do desempenho total dos candidatos^{17,18}. Por essa razão, o desempenho referenciado por norma foi adotado pelo TPI-GO⁸. Os percentis 30 e 60 (P30 e P60) foram definidos como pontos de corte para dispensa e bonificação dos candidatos na prova teórica do Tego pela equipe de implementação do TPI-GO, por meio do método de Angoff modificado¹⁹.

A pontuação no TPI-GO reflete diretamente o número de acertos das questões. Cada candidato recebe as informações de forma sigilosa sobre seus desempenhos

absoluto e relativo em relação à mediana, ao percentil 30 (P30) e ao percentil 60 (P60) de seus pares em nível nacional, além do seu progresso referente aos anos anteriores. Para os R3, também é fornecido o laudo final sobre a dispensa ou bonificação na prova teórica do Tego do ano subsequente. O consolidado do desempenho dos residentes de um mesmo serviço é apresentado ao supervisor responsável pelo PRM. Para preservar o sigilo em relação ao desempenho individual dos residentes, somente estão habilitados a receber esses resultados os supervisores de PRM em que pelo menos três médicos residentes de cada categoria tenham se aplicado ao TPI-GO. Para evitar ranqueamento e exposições indevidas dos programas, somente o supervisor de cada PRM recebe os dados sobre o desempenho consolidado dos seus residentes^{8,9}. A devolutiva é enviada aos residentes e aos PRM por *e-mail* até três meses após a realização da prova.

Em 2018 e 2019, o TPI-GO foi aplicado presencialmente ao final do primeiro semestre. Em 2020, o TP foi adaptado para o modelo *on-line* em razão da pandemia e aplicado ao final do segundo semestre^{8,20}. Na modalidade *on-line*, o candidato pode realizar o TPI-GO a partir do domicílio, utilizando seu próprio equipamento. Para garantir a segurança, o candidato instala um programa em que bloqueia o acesso a navegadores de internet e permite a monitorização utilizando a câmera e o microfone do próprio computador. Além disso, são designados fiscais de prova que monitoram os candidatos, sendo possível detectar sons, movimentos corporais e desvios do olhar, bem como a presença de outras pessoas ou artefactos suspeitos no ambiente de prova. Havendo infrações ou tentativas de fraude, os candidatos são advertidos, e, em casos persistentes, a prova é interrompida²⁰.

Descrição da coorte de residentes participantes no estudo

Este relato de experiência se refere ao acompanhamento longitudinal da primeira coorte de residentes do TPI-GO, em que 497 médicos residentes iniciaram a participação em 2018 como R1, dos quais 457 continuaram em 2019 como R2 e 314 concluíram a participação em 2020 como R3, o que representa uma aderência de 63% entre iniciantes e concluintes. A evolução da coorte de residentes acompanhados entre 2018 e 2020 em nível nacional mostrou um aumento progressivo de desempenho no TPI-GO durante esse período, sendo o P30 = 4,8, 5,8 e 5,8, a mediana = 5,2, 6,2 e 6,3, e o P60 = 5,5, 6,4 e 6,5, respectivamente para o R1 (2018), o R2 (2019) e o R3 (2020). Dentre os candidatos inscritos como R1 em 2018, a maioria foi composta por mulheres (86%). Todos os dados analisados foram anonimizados.

Análise dos PRM a partir do desempenho dos seus residentes

A experiência aqui apresentada incluiu somente os programas em que pelo menos três médicos residentes de cada categoria (R1, R2 e R3) participaram do TPI-GO durante todo o período de observação (de 2018 a 2020). Foram excluídos dessa análise os PRM que apresentaram menos de três residentes em cada categoria no TPI-GO. Entre os 32 PRM incluídos, nove (28,1%) estão localizados na Região Sul; 22 (68,8%), no Sudeste; e um (3,1%); no Centro-Oeste.

A análise dos PRM foi realizada pela Comissão de Residência Médica da Febrasgo, em que se comparou a mediana de desempenho dos residentes desses programas ao longo dos três anos consecutivos de participação na prova (como R1, R2 e R3) com a curva de desempenho nacional, que incluiu a mediana, o P30 e o P60 do desempenho geral. Para o cálculo desses valores, utilizou-se o pacote de análise estatística do Excel com seus diversos recursos disponíveis. Esse mesmo *software* foi utilizado na elaboração dos gráficos do desempenho dos residentes de cada PRM. A partir desses gráficos, os PRM foram classificados com base nos seguintes parâmetros: 1. a mediana do desempenho dos R1 de um PRM, que reflete o nível de exigência do processo seletivo e a qualidade dos ingressantes; 2. a diferença entre as medianas de desempenho dos concluintes (R3) e dos ingressantes (R1) (R3-R1), que reflete o ganho de conhecimento proporcionado pelo PRM aos seus residentes durante todo o treinamento; 3. as diferenças intermediárias entre as medianas de desempenho dos R2 e R1

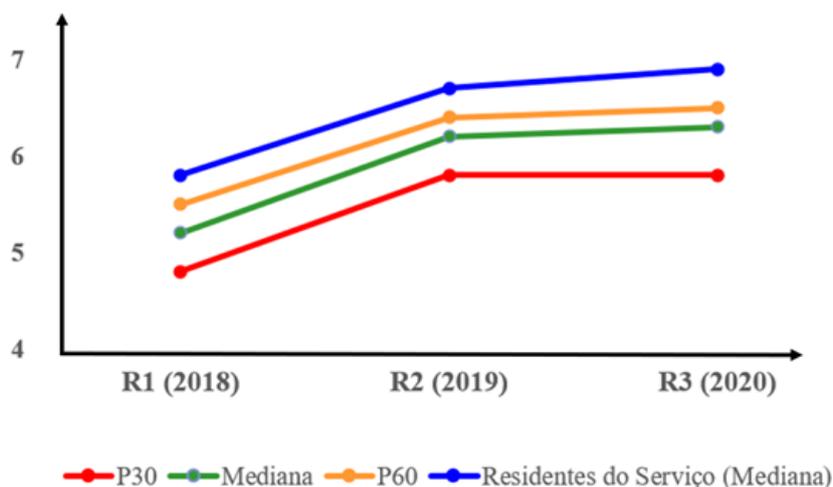
(R2-R1) e dos R3 e R2 (R3-R2), que fornecem informações sobre o ganho cognitivo dos residentes em diferentes períodos do treinamento no PRM; 4. a mediana de desempenho dos R3, que reflete o nível cognitivo dos concluintes desse PRM. Esses parâmetros foram analisados em relação à linha de tendência da mediana de desempenho dos residentes de cada PRM, não sendo considerada a significância estatística das diferenças entre os valores.

Identificaram-se cinco diferentes perfis ou tipos de PRM:

- Programa do tipo 1: o desempenho dos residentes é superior ao P60 nacional em todas as categorias (R1, R2 e R3).
- Programa do tipo 2: o desempenho dos residentes é intermediário e está entre o P30 e o P60 nacional em todas as categorias (R1, R2 e R3).
- Programa do tipo 3: o desempenho dos residentes está abaixo do P30 nacional em todas as categorias (R1, R2 e R3).
- Programa do tipo 4: o desempenho dos R1 está abaixo do P30 nacional, enquanto o desempenho dos R3 está entre o P30 e o P60 nacional.
- Programa do tipo 5: o desempenho dos R1 está acima do P60 nacional, enquanto o desempenho dos R3 está entre o P30 e o P60 nacional.

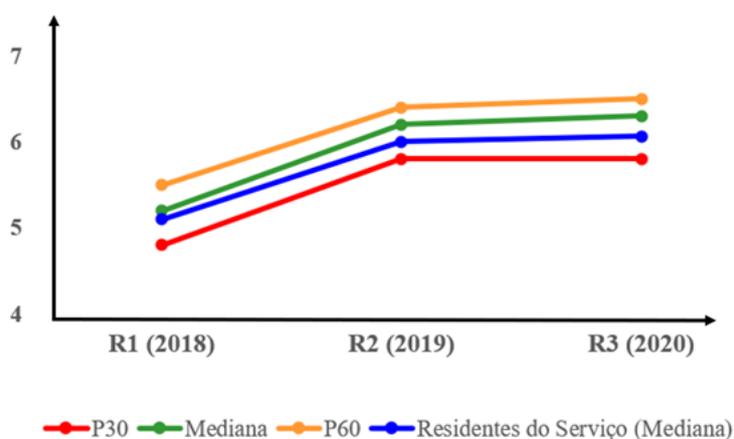
Os gráficos 1, 2, 3, 4 e 5 apresentam os cinco perfis identificados. Em cada gráfico, constam a mediana do desempenho dos residentes do PRM (azul) e a mediana (laranja), o P30 (vermelho) e o P60 (verde) do desempenho geral dos candidatos em nível nacional.

Gráfico 1. Programa do tipo 1: o desempenho dos residentes é superior ao P60 nacional em todas as categorias (R1, R2 e R3)



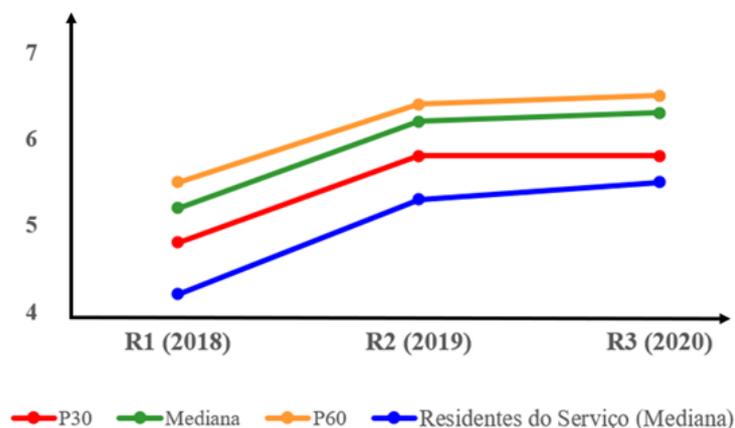
Fonte: Elaborado pelos autores.

Gráfico 2. Programa do tipo 2: o desempenho dos residentes é intermediário e está entre o P30 e o P60 nacional em todas as categorias (R1, R2 e R3)



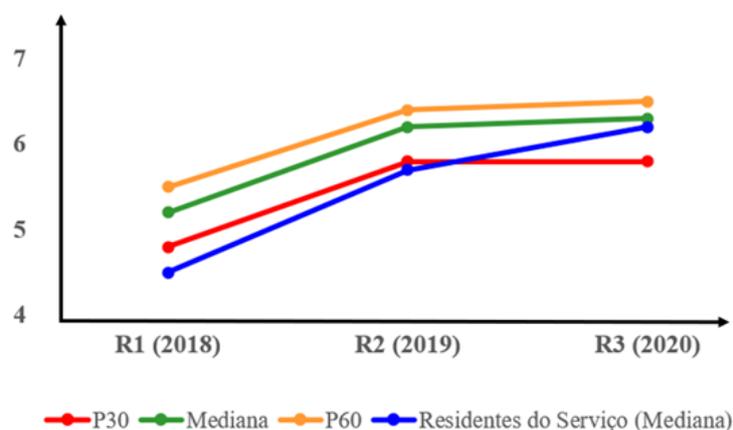
Fonte: Elaborado pelos autores.

Gráfico 3. Programa do tipo 3: o desempenho dos residentes está abaixo do P30 nacional em todas as categorias (R1, R2 e R3)



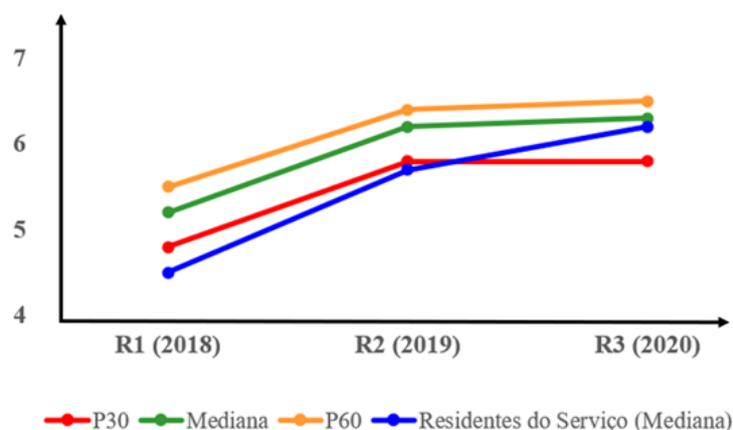
Fonte: Elaborado pelos autores.

Gráfico 4. Programa do tipo 4: o desempenho dos R1 está abaixo do P30 nacional, enquanto o desempenho dos R3 está entre o P30 e o P60 nacional



Fonte: Elaborado pelos autores.

Gráfico 5. Programa do tipo 5: o desempenho dos R1 está acima do P60 nacional, enquanto o desempenho dos R3 está entre o P30 e o P60 nacional



Fonte: Elaborado pelos autores.

DISCUSSÃO

A experiência aqui retratada apresenta um modelo de avaliação dos PRM a partir do desempenho dos seus residentes no TPI-GO. A análise desses resultados identificou cinco diferentes perfis de PRM, o que possibilita inferir a magnitude do conhecimento agregado pelos seus residentes ao longo do treinamento, bem como identificar fortalezas e fragilidades do PRM em promover o desenvolvimento cognitivo desses aprendizes.

A relevância do conhecimento médico para a prática profissional na especialidade vem sendo debatida desde o século passado, quando Abraham Flexner destacou a importância da ciência como base para a formação médica²¹. Por sua vez, William Osler defendia um modelo de formação do especialista mais voltado para a prática, em que o ensino à beira do leito se sobreponha ao aprendizado nos anfiteatros⁵. A visão de Osler foi inspiradora para alguns educadores que implementaram, em alguns currículos canadenses e norte-americanos, a aprendizagem baseada em problemas, na década de 1970, priorizando o treinamento voltado para a solução dos desafios profissionais²². Porém, trabalhos posteriores demonstraram que a mera imersão na prática não é suficiente para desenvolver a *expertise* médica e que a resolução de problemas por especialistas requer uma base cognitiva bem estruturada e conhecimento especializado. Dessa forma, o papel central do conhecimento para a *expertise* médica voltou a ser reconhecido^{2,3}.

No âmbito da residência médica, entretanto, nem sempre a avaliação do conhecimento adquirido é priorizada. No Brasil e em muitos outros países, não há uniformidade na avaliação cognitiva dos médicos residentes, e muitos PRM

seguem realizar avaliações regulares de seus aprendizes⁵. Mais recentemente, novos métodos de avaliação em ambientes de prática têm sido utilizados, tais como o Miniexercício Clínico Avaliativo (Miniex), a *Direct Observation of Procedural Skills* (DOPS) e o *Multisource Feedback*. Mas esses métodos priorizam o desempenho na prática em detrimento do conhecimento técnico-científico, sendo insuficientes para uma avaliação consistente do componente cognitivo das competências⁵.

O TP se apresenta como um método de grande valia para preencher as lacunas de avaliação cognitiva na residência médica. Trata-se de uma avaliação abrangente e confiável que permite o acompanhamento dos residentes, desde o seu ingresso até a conclusão do PRM. Por meio dos seus resultados, é possível reconhecer o domínio cognitivo dos residentes em algumas áreas do conhecimento na especialidade e identificar lacunas e necessidades de aprendizagem em outras, fornecendo subsídios para a autoavaliação e avaliação formativa dos aprendizes⁵. O TP oferece também informações valiosas para a autoavaliação dos PRM a partir do desempenho de um grupo de residentes matriculados em um mesmo programa.

Nos programas do tipo 1, pode-se presumir que o processo seletivo é exigente, já que esse PRM seleciona ingressantes com elevado nível de conhecimento na área. O conhecimento agregado ao longo do programa é relevante, e o nível dos concluintes é superior à maioria dos egressos. Pode-se inferir que esse perfil corresponde a um PRM bem estruturado e que promove um desenvolvimento substancial do componente cognitivo de seus residentes ao longo do treinamento.

Nos programas do tipo 2, pode-se presumir que a exigência no processo seletivo é moderada, uma vez que os R1 ingressantes apresentam um perfil de desempenho entre o P30

e o P60. O conhecimento agregado ao longo do programa é relevante, e o nível dos concluintes também se encontra entre o P30 e o P60 do desempenho nacional. Esse perfil corresponde aos programas razoavelmente estruturados que oferecem condições aceitáveis de treinamento, mas que podem se beneficiar do aprimoramento dos estágios e das atividades teórico-práticas.

Nos programas do tipo 3, pode-se presumir que o processo seletivo é pouco exigente, pois possibilita o ingresso de médicos que apresentam deficiências importantes na formação e no conhecimento essencial. Embora exista um certo acréscimo de conhecimento ao longo do treinamento, a mediana do desempenho dos concluintes fica abaixo do P30, sugerindo um despreparo dos egressos para atuação na especialidade. Essa hipótese deveria ser verificada por meio da triangulação de informações com o desempenho dos residentes em outras modalidades de avaliação, como as observações diretas e indiretas do desempenho clínico, as discussões de casos e o *Multisource Feedback*. Pode-se inferir que esse perfil corresponde aos PRM que apresentam maiores deficiências em termos de estrutura pedagógica, estágios, supervisão e oportunidades de treinamento. O perfil dos ingressantes pode ser melhorado pelo aprimoramento do processo seletivo. A aquisição de conhecimento pode ser promovida pela reestruturação e qualificação dos estágios e das atividades teórico-práticas.

Nos programas do tipo 4, pode-se presumir que, embora o processo seletivo seja pouco exigente, existe um ganho substancial de conhecimento dos residentes entre o R1 e o R3, haja vista que a mediana de desempenho dos R3 se encontra entre o P30 e o P60 do desempenho nacional. Os concluintes desses programas demonstram níveis aceitáveis de conhecimento adquirido ao final do treinamento. Pode-se inferir que esse perfil corresponde aos PRM que oferecem estrutura pedagógica, estágios e supervisão que impulsionam e promovem o aprendizado dos seus residentes e, mesmo diante das lacunas e deficiências dos seus ingressantes, conseguem agregar conhecimento substancial. Em alguns casos, é possível melhorar ainda mais os resultados pelo aprimoramento do processo seletivo, da qualidade dos estágios e das atividades teórico-práticas.

Nos programas do tipo 5, pode-se presumir que o processo seletivo é exigente, pois seleciona ingressantes com nível intermediário ou superior de conhecimento na especialidade. Entretanto, quando a mediana do desempenho dos residentes do serviço é comparada com os valores nacionais, verifica-se uma queda ou um crescimento relativamente menor no componente cognitivo ao longo do treinamento, sugerindo uma agregação de conhecimento insatisfatória ou abaixo

das expectativas. Pode-se inferir que esse perfil corresponde a programas que sofreram desajustes em sua estrutura pedagógica ou assistencial durante o período de observação. Considerando a pandemia da Covid-19 como o mais recente evento que impactou o treinamento dos residentes em níveis nacional e internacional, é plausível a hipótese de que esses programas foram os que mais sofreram essas consequências²³. É recomendável que eles reorganizem seus estágios, a supervisão e as atividades teórico-práticas, de modo a minimizar as perdas decorrentes do período da pandemia.

Uma avaliação adequada dos programas educacionais requer a coleta sistemática de informações sobre as atividades realizadas e os resultados alcançados. Tais informações são determinantes para reafirmar valores e identificar fragilidades, visando à melhoria de qualidade¹¹. Diferentes modelos foram descritos na literatura e podem ser utilizados para a avaliação dos programas de acordo com a necessidade. Entre os modelos mais utilizados na educação médica, merecem destaque o modelo de avaliação em quatro níveis de Kirkpatrick, que se concentra apenas nos resultados da avaliação dos aprendizes²⁴; o Modelo Lógico, que incorpora o contexto do programa em seu processo de avaliação²⁵; e o Modelo Contexto/Entrada (*Input*)/Processo/Produto (CIPP), que avalia o contexto e os produtos em uma abordagem mais flexível e considera também a complexidade e o dinamismo dos programas educacionais¹¹. Cada um desses modelos se baseia em diferentes teorias, mas todos guardam em comum a avaliação dos resultados que inclui, necessariamente, o desempenho dos aprendizes.

No Brasil, o credenciamento, a supervisão e a avaliação dos programas de residência são atribuições da Comissão Nacional de Residência Médica (CNRM) do Ministério da Educação (MEC), que conta com o apoio das Comissões Estaduais de Residência Médica (Cerem) na realização dessas tarefas. Durante o processo de avaliação dos programas, são realizadas visitas *in loco* por avaliadores que verificam as dependências destinadas à assistência e ao ensino, e realizam reuniões com o presidente da Comissão de Residência Médica (Coreme) local, com o supervisor responsável pelo programa, com os preceptores e com os médicos residentes. Nesse processo de avaliação, também são verificadas e conferidas as documentações, atas e reuniões da Coreme. Entretanto, informações sobre o desempenho dos residentes em avaliações cognitivas ou práticas não são consideradas no processo avaliativo dos PRM²⁶.

O TP pode oferecer aos PRM informações valiosas sobre o desempenho de seus residentes em testes de conhecimento essencial na especialidade. Por meio dessas informações, é possível verificar o conhecimento agregado aos seus residentes durante o treinamento e identificar necessidades de melhoria,

contribuindo assim para a qualificação desses programas. O modelo aqui apresentado é fruto da análise do perfil de todos os 32 PRM cujos residentes se aplicaram ao TPI-GO entre 2018 e 2020, em número superior ou igual a três por categoria. Por meio desse modelo, é possível interpretar as curvas de desempenho dos médicos residentes de um PRM no TP de forma simples e objetiva, e avaliar indiretamente os PRM. Sendo assim, o TP se apresenta como uma estratégia de avaliação externa tanto do desempenho dos médicos residentes quanto dos PRM, sanando a falta de dados válidos e confiáveis que permitam o monitoramento do processo de formação dos especialistas brasileiros. Outros estudos serão necessários para confirmar esses achados.

As limitações deste estudo incluem o número limitado de PRM que apresentaram dados suficientes para a análise e o fato de ter analisado os PRM apenas à luz do desempenho cognitivo dos seus residentes no TPI-GO. Sob a ótica da formação por competências, a avaliação dos residentes deverá contemplar também os outros componentes, como as habilidades e o profissionalismo¹⁵. Além disso, uma avaliação mais completa dos PRM requer informações sobre a sua estrutura física e pedagógica, incluindo o projeto de ensino, os campos de estágio, as oportunidades de treinamento e avaliação, a qualidade da supervisão e o apoio aos residentes^{11,26}.

CONCLUSÃO

Embora o conhecimento seja um elemento fundamental para a prática médica qualificada, na maioria das especialidades não há uma avaliação cognitiva unificada dos médicos residentes, e conseqüentemente não se pode verificar apropriadamente o conhecimento agregado pelos PRM a esses aprendizes durante o treinamento.

O TP oferece uma oportunidade para avaliação dos PRM a partir do desempenho dos seus residentes. O modelo aqui apresentado possibilita inferências sobre o processo seletivo, o nível cognitivo dos ingressantes, o conhecimento proporcionado pelo PRM a esses residentes durante o treinamento e o nível de conhecimento dos egressos. Essas informações podem subsidiar decisões institucionais que promovam melhorias na qualidade dos estágios, na supervisão e nas atividades teórico-práticas, contribuindo para o processo de formação na especialidade.

AGRADECIMENTOS

Aos membros da Comissão de Residência Médica da Febrasgo e da Comissão do Título de Especialista, à Secretaria da Febrasgo, aos preceptores e supervisores de programas de residência médica, e aos médicos residentes que se aplicaram ao TPI-GO.

CONTRIBUIÇÃO DOS AUTORES

Gustavo Salata Romão participou da concepção do projeto, do delineamento metodológico, da coleta e análise estatística dos dados, da interpretação dos resultados e da redação do manuscrito. Cesar Eduardo Fernandes, Agnaldo Lopes da Silva Filho e Marcos Felipe Silva de Sá participaram da concepção do projeto, do delineamento metodológico, da interpretação dos resultados e da redação do manuscrito. Sheyla Ribeiro Rocha participou da análise estatística dos dados, da interpretação dos resultados e da redação e aprovação da versão final do artigo.

CONFLITO DE INTERESSES

Declaramos não haver conflito de interesses.

FINANCIAMENTO

Projeto financiado pela Federação Brasileira das Associações de Ginecologia e Obstetrícia (Febrasgo).

REFERÊNCIAS

1. Ten Cate O. Competency-based postgraduate medical education: past, present and future. *GMS J Med Educ.* 2017;34(5):Doc69. doi: 10.3205/zma001146.
2. Schmidt H, Norman G, Boshuizen H. A cognitive perspective on medical expertise: theory and implications. *Acad Med.* 1990;65(10):611-21.
3. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med.* 2004;79(10 Suppl):S70-81.
4. Norcini J, Anderson MB, Bollella V, Burch V, Costa MJ, Duvivier R et al. Consensus framework for good assessment. *Med Teach.* 2018; 40(11): 1102-1109. doi: 10.1080/0142159x.2018.1500016.
5. Dijksterhuis MG, Scheele F, Schuwirth LW, Essed GG, Nijhuis JG, Braat DD. Progress testing in postgraduate medical education. *Med Teach.* 2009;31(10):e464-8. doi: 10.3109/01421590902849545.
6. Freeman A, Van Der Vleuten C, Nouns Z, Ricketts C. Progress testing internationally. *Med Teach.* 2010;32(6):451-5. doi: 10.3109/0142159x.2010.485231.
7. Freeman AC, Ricketts C. Choosing and designing knowledge assessments: experience at a new medical school. *Med Teach.* 2010;32(7):578-81. doi: 10.3109/01421591003614858.
8. de Sá MFS, Romão GS, Fernandes CE, da Silva Filho AL. The Individual Progress Test of Gynecology and Obstetrics Residents (TPI-GO): the Brazilian experience by FEBRASGO. *Rev Bras Ginecol Obstet.* 2021;43(6):425-8. doi:10.1055/s-0041-1731803.
9. Romão GS, Fernandes CE, de Sá MFS. Teste de Progresso Individual do residente em GO: relato da experiência do primeiro ano de implantação no Brasil. *Femina.* 2019; 47(5): 282-287.
10. Bicudo AM, Hamamoto PT, Abbade JF, Hafner MdLMB, Maffei CML. Teste de Progresso em consórcios para todas as escolas médicas do Brasil. *Rev Bras Educ Med.* 2019;43:151-6.
11. Frye AW, Hemmer PA. Program evaluation models and related theories: AMEE guide no. 67. *Med Teach.* 2012;34(5):e288-99. doi: 10.3109/0142159x.2012.668637.
12. Cook DA. Twelve tips for evaluating educational programs. *Med Teach.* 2010;32(4):296-301.
13. Brasil. Resolução nº 510, de 7 de abril de 2016. Diário Oficial da República Federativa do Brasil; 2016. p. 44.

14. Romão GS, dos Reis FJC, Cavalli RC, de Sá MFS. Matriz de competência em ginecologia e obstetria: um novo referencial para os programas de residência médica no Brasil. *Femina*. 2017;45(3):172-7.
15. Romão GS, de Sá MFS. Competency-based training and the competency framework in gynecology and obstetrics in Brazil. *Rev Bras Ginecol Obstet*. 2020;42(5):272-88. doi: 10.1055/s-0040-1708887.
16. Coombes L, Roberts M, Zahra D, Burr S. Twelve tips for assessment psychometrics. *Med Teach*. 2016;38(3):250-4. doi: 10.3109/0142159x.2015.1060306.
17. Wrigley W, van der Vleuten CP, 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. doi: 10.3109/0142159x.2012.704437.
18. Muijtjens AMM, Hoogenboom RJ, Verwijnen GM, Van der Vleuten CPM. Relative or absolute standards in assessing medical knowledge using progress tests. *Adv Health Sci Educ*. 1998;3(2):81-7.
19. Bandaranayake RC. Setting and maintaining standards in multiple choice examinations: AMEE Guide No. 37. *Med Teach*. 2008;30(9-10):836-45.
20. Salata Romão G, Sá MFS, Fernandes CE, da Silva Filho AL. Comment on "progress testing anytime, anywhere – does a mobile-learning approach enhance the utility of a large-scale formative assessment tool?". *Med Teach*. 2021; 43(8): 976-977. doi: 10.1080/0142159x.2020.1854453.
21. Flexner A. Medical education in the United States and Canada. *Science*. 1910;32(810):41-50.
22. Patel VL, Evans DA, Groen GJ. Biomedical knowledge in clinical reasoning. In: Evans DA, Patel VL (eds). *Cognitive Science in Medicine: Biomedical Modeling*. Cambridge: MIT Press, 1989; p. 53-112.
23. Brito LGO, Romão GS, Fernandes CE, Silva-Filho AL. Impact of Covid-19 on Brazilian medical residencies in obstetrics and gynecology. *Int J Gynaecol Obstet*. 2020;150(3):411-2. doi: 10.1002/ijgo.13283.
24. Smidt A, Balandin S, Sigafos J, Reed VA. The Kirkpatrick model: a useful tool for evaluating training outcomes. *J Intellect Dev Disabi*. 2009;34(3):266-74.
25. Frechtling JA. *Logic modeling methods in program evaluation*. John Wiley & Sons. San Francisco; 2007.
26. Brasil. Resolução CNRM nº 02/2006, de 17 de maio de 2006. Dispõe sobre requisitos mínimos dos Programas de Residência Médica e dá outras providências. *Diário Oficial da União*; 2006.



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.

Teste de Progresso para estudantes de Medicina: experiência de um consórcio na Região Nordeste do Brasil

Progress Test for medical students: experience of a consortium in Northeastern Brazil

Taciana Barbosa Duque¹

tacianaduque@fps.edu.br

Daniela Chiesa²

danic2305@gmail.com

Raquel Autran Coelho Peixoto³

raquelautrancp@gmail.com

Carla Adriane Fonseca Leal de Araújo¹

carla.leal@fps.edu.br

RESUMO

Introdução: A colaboração entre escolas médicas para a realização do Teste de Progresso (TP) vem sendo ampliada com o propósito de tornar-se um modelo de prática educacional.

Relato de experiência: O núcleo de escolas médicas CIN1 foi instituído no ano de 2013, tendo na sua organização um coordenador-geral e coordenadores de cada escola participante. Realiza dois testes ao ano, variando o número de escolas participantes. O teste possui 120 questões distribuídas nas seis áreas (ciências básicas aplicada, saúde coletiva, pediatria, clínica médica, cirurgia/urgência e emergência, e ginecologia e obstetria) e tem seu blueprint baseado em matriz previamente validada. O teste é aplicado a todos os estudantes das escolas participantes, no mesmo dia e com duração de quatro horas. Aconteceu on-line em 2020 e 2021, por causa da pandemia da Covid-19. A pontuação total do estudante é calculada pelo número de questões corretas, sem a opção “não sei”, nem penalidades para incorretas. A análise é realizada por meio da teoria clássica do item. Os estudantes recebem a análise do desempenho comparativamente com a média obtida pelos discentes no mesmo ano e a progressão em relação ao teste anterior. As escolas recebem o desempenho dos respectivos estudantes, a análise da dificuldade e discriminação geral dos itens, e uma avaliação global do teste.

Discussão: A implantação do CIN1 representou um grande avanço para as escolas envolvidas. A colaboração não se deu apenas sobre o teste, mas também para o desenvolvimento docente de forma conjunta. Impedimentos de regimento interno e custos podem justificar a variação na participação das escolas.

Conclusão: A organização de um núcleo de escolas para o TP traz benefícios e muitos desafios. O aperfeiçoamento dos critérios de análises e feedback e a inclusão de estudantes na avaliação de questões são etapas a serem discutidas e implantadas. A utilização do ambiente virtual, de mais baixo custo, pode ser um facilitador para ampliar a participação das escolas.

Palavras-chave: Avaliação Educacional; Educação de Graduação em Medicina; Aprendizagem.

ABSTRACT

Introduction: Collaboration between medical schools to hold the PT has expanded, constituting a model of educational practice.

Experience report: The CIN 1 group of medical schools was established in 2013, organized with a general coordinator and coordinators from each participating school. It holds two tests a year, with a varying number of participant schools. The test has 120 questions distributed in 6 areas (Applied Basic Sciences, Public Health, Pediatrics, Internal Medicine, Surgery/Urgent and Emergency, Gynecology and Obstetrics) and its blueprint is based on a previously validated Matrix. The 4-hour test is applied to all students from participating schools, on the same day and at the same time. It was administered online in 2020 and 2021 due to the COVID-19 pandemic. The student's total score is calculated by the number of correct questions, excluding the “don't know” option, or penalties for incorrect ones, and is analyzed by classical item theory. Students receive performance analysis compared to the average attained by students in the same year, and the progression in relation to the previous test. The schools receive the performance of the respective students and analysis of the difficulty and general breakdown of the items and an overall evaluation of the test.

Discussion: The implementation of CIN 1 represented a great advance for the schools involved. The collaboration was not only about the test, but also for the teacher development in a joint way. Bylaw impediments and costs may justify the variation in school participation.

Conclusion: The organization of a group of schools for the TP brings benefits and many challenges. Improving analysis and feedback criteria and including students in the assessment of questions are steps to be discussed and implemented. The use of the lower cost virtual environment may facilitate increased participation of schools.

Keywords: Educational Measurement; Education Medical Undergraduate; Learning.

¹Faculdade Pernambucana de Saúde, Recife, Pernambuco, Brasil.

²Universidade de Fortaleza, Fortaleza, Ceará, Brasil.

³Centro Universitário Christus, Fortaleza, Ceará, Brasil.

Editor: Valdes Roberto Bollela

Recebido em 05/09/22; Aceito em 07/10/22.

Avaliado pelo processo de *double blind review*.

INTRODUÇÃO

Muitos estudos com aplicação do Teste de Progresso (TP) têm demonstrado a importância dessa avaliação no acompanhamento da aquisição de conhecimento durante o período de formação. O formato longitudinal abrangente do TP, baseado no conhecimento esperado ao final do curso e com a participação simultânea de estudantes de todos os anos do curso, permite revisar as estratégias de aprendizagem e contribuir para a atualização do currículo e para o acompanhamento por parte dos discentes das suas curvas de aprendizado na dimensão conhecimento¹⁻⁴. Tem sido demonstrado também o papel preditor do TP em exames para residência médica, além de poder fornecer informações mais confiáveis para decisões sobre a progressão do estudante no curso⁵⁻⁸.

O TP foi aplicado pela primeira vez nas Universidades de Maastricht e Missouri no final da década de 1970 e no Brasil, em 1998, na Universidade Estadual de Londrina. A partir daí o TP vem sendo amplamente utilizado principalmente em cursos de Medicina, seja em construção individual ou por núcleos de escolas médicas^{1,2,9,10}.

A colaboração entre escolas médicas para a realização do TP vem sendo estimulada e ampliada, contribui para a redução de custos, o compartilhamento de banco de itens e o desenvolvimento docente, além de constituir um modelo de prática educacional de qualidade para acompanhamento da aprendizagem¹¹.

Apesar da sua larga utilização em diversos países e em regiões do Brasil, que atualmente possui 17 núcleos de escolas médicas para o TP¹², observa-se uma variação na sua forma de aplicação, repetição, *blueprint*, número de itens, entre outros. Alguns desses fatores podem dever-se a uma acomodação das condições locais para a efetiva aplicação do TP^{11,13}.

Apresentamos o relato da experiência e organização de um núcleo de colaboração de escolas médicas na Região Nordeste do Brasil para o TP, que recebeu a denominação de Consórcio Interinstitucional Nordeste 1 (CIN1), aqui descrito como o núcleo CIN1.

RELATO DE EXPERIÊNCIA

O presente relato é composto pelos históricos da criação do núcleo CIN1 e da participação das escolas, seguidos do relato da organização e do fluxo, tendo como base o modelo proposto por Wrigley et al.¹¹ que apresenta um arranjo sistêmico com interação de quatro grandes eixos: construção, administração, análise dos resultados e *feedback*.

Histórico de participação das escolas

O núcleo CIN1 foi instituído no ano de 2013, por estímulo e adesão das escolas participantes ao projeto da Associação Brasileira

de Educação Médica (Abem) denominado “Abem 50 anos – 10 anos de DCN”. O projeto da Abem tinha como um dos subprojetos “Avaliação do estudante: contribuição do Teste de Progresso”, que incluiu, entre as ações para qualificação do ensino médico, a ampliação e o desenvolvimento de núcleos interinstitucionais, os chamados consórcios, para realização do TP.

O CIN1 tem na sua organização um coordenador do núcleo e coordenadores de cada escola participante. Os coordenadores, responsáveis pela gestão interna do teste, respondem pelas decisões relacionadas à prova e pela interlocução com as demais escolas e a coordenação do CIN1. O processo de trabalho dos coordenadores de escola e do TP ocorre presencialmente e a distância, por meio de *e-mail* e videoconferência. A qualificação do ensino por meio do desenvolvimento docente também faz parte dos objetivos da criação do consórcio, sendo realizadas oficinas para os docentes das escolas sobre elaboração de itens e discussão de matriz. A coordenação do núcleo e os coordenadores de escolas, no momento da implantação do consórcio, realizaram estudo de validação de uma matriz para o teste, baseando-se nas competências esperadas ao final do curso de Medicina definidas pelas Diretrizes Curriculares Nacionais para o curso de Medicina e detalhadas na matriz para o teste de revalidação de diploma no Brasil. O conteúdo para a matriz foi selecionado e organizado por listas de temas das grandes áreas da medicina, em escala do tipo Likert de 5 pontos, e submetido ao julgamento de professores das diversas instituições, em que se utilizou a técnica Delphi, tendo como critério de consenso para compor a matriz de referência para o teste o item com *ranking* médio acima de 3.

A matriz de referência para o teste resultante desse estudo possui seis áreas – ciências básicas, saúde coletiva, pediatria, ginecologia/obstetrícia, clínica médica e cirurgia/urgência e emergência –, com distribuição das questões dentro de objetivos de aprendizagem e de cenários, contemplando os diferentes níveis de complexidade da atenção em saúde, balanceados de acordo com as diretrizes do Sistema Único de Saúde. Após conclusão da matriz, houve validação pelas escolas participantes do consórcio.

O núcleo CIN1 realizou o seu primeiro teste em outubro de 2013 com a participação de sete escolas médicas: seis do estado do Ceará – Centro Universitário Christus (Unichristus), Universidade de Fortaleza (Unifor), Universidade Estadual do Ceará (Uece), Universidade Federal do Ceará (UFC), *campus* Fortaleza e *campus* Sobral, e Universidade Federal do Cariri (Ufca) – e uma do estado de Pernambuco: Faculdade Pernambucana de Saúde (FPS). Participavam do núcleo tanto escolas que já realizavam o TP com os seus estudantes de forma individual quanto as que estavam introduzindo essa

modalidade de avaliação após a participação no núcleo. Em 2019, a Estácio Faculdade de Medicina de Juazeiro (Estácio FMJ) e a Universidade Federal do Rio Grande do Norte (UFRN) foram agregadas ao grupo.

O CIN1 desde a sua criação realiza dois testes ao ano, nos meses de maio e outubro. A adesão aos dois testes se deu inicialmente por duas escolas (FPS e Unichristus), e a partir do ano de 2021 a Estácio FMJ também passou a realizar os dois testes anuais. Nos anos de 2015 e 2021, foram realizados testes nacionais com a participação de 57 e 130 escolas, respectivamente¹². Por causa da urgência sanitária da pandemia da *coronavirus disease* 2019 (Covid-19), nos anos de 2020 e 2021 os testes foram realizados na modalidade on-line e de forma remota. Em 2020.1, apenas uma escola, a FPS, realizou o TP, não havendo participação das demais por conta das dificuldades com a aplicação do teste na modalidade *on-line*.

As escolas participantes do CIN1 e o número de estudantes envolvidos são apresentados na Tabela 1.

Construção do teste

O teste contém 120 questões distribuídas nas seis áreas definidas na matriz (ciências básicas aplicada, saúde coletiva, pediatria, clínica médica, cirurgia/urgência e emergência, e ginecologia e obstetrícia), cada uma com 20 questões. A construção do teste começa com o plano do teste (*blueprint*) que se baseia na matriz previamente validada. O plano define o número de questões por competência em cada área da matriz e permite o acompanhamento da distribuição dos conteúdos para essas competências nos sucessivos testes.

Com base no plano do teste, são construídas as encomendas das questões. Essas encomendas contêm por área a competência geral, o conhecimento a ser avaliado dentro daquela competência, o cenário da questão a ser elaborada, a idade e o sexo e/ou gênero da paciente quando cabível e a escola responsável pela elaboração. Essas encomendas são organizadas pela coordenação do consórcio e encaminhadas aos coordenadores de cada escola membro do consórcio que irá participar do teste. Idealmente, utiliza-se a estratégia

Tabela 1. Núcleo CIN1 de escolas médicas para o Teste de Progresso – escolas participantes, número de estudantes, adesão e modalidade do teste, de 2013 a 2022

Ano/ semestre	Número de escolas	Modalidade do teste	Número de estudantes	Adesão dos estudantes (%)
2013.2	7*	Presencial	2.654	64,3
2014.1	2**	Presencial	1.317	83,9
2014.2	7*	Presencial	3.616	73,3
2015.1	2**	Presencial	1.590	88,9
2015.2	58 ^N	Presencial	24.313	76,4
2016.1	2**	Presencial	1.496	91,4
2016.2	3***	Presencial	2.236	91,0
2017.1	2**	Presencial	1.496	91,4
2017.2	3***	Presencial	2.393	90,8
2018.1	2**	Presencial	1.709	91,9
2018.2	3***	Presencial	2.768	90,2
2019.1	2**	Presencial	1.877	91,4
2019.2	3***	Presencial	2.919	90,8
2020.2	5****	<i>On-line</i>	2.446	80,5
2021.1	3*****	<i>On-line</i>	2.746	90,8
2021.2	130 ^N	<i>On-line</i>	48.946	72,1
2022.1	3*****	<i>On-line</i>	2.779	89,9

* FPS, Unichristus, Unifor, Uece, UFC *campus* Fortaleza, UFC *campus* Sobral, UFC *campus* Cariri.

** FPS, Unichristus.

*** FPS, Unichristus, Unifor.

**** FPS, Uece, UFRN, Unichristus, Unifor.

***** FPS, Unichristus, Estácio FMJ.

^N Teste de Progresso Nacional.

Fonte: Elaborada pelas autoras.

de “escolas espelhos”, em que cada duas escolas recebem encomendas iguais. As escolas têm o prazo de 30 dias para concluir a etapa de elaboração da questão. As questões elaboradas são organizadas por área para revisão. Após a etapa de revisão, são selecionadas as questões que cumprem com os requisitos recomendados para uma questão de múltipla escolha, devendo predominar fortemente no teste questões com a taxonomia de compreensão e resolução de problemas. As questões selecionadas são encaminhadas para a elaboração do teste, a depender da modalidade definida, se impresso ou ambiente *on-line*. As questões são arquivadas de acordo com o índice de dificuldade e discriminação após serem testadas.

Administração do teste

De acordo com a modalidade de aplicação do teste, se impresso ou *on-line*, existe um fluxo de organização. Desde a formação do núcleo CIN1, os testes são realizados no formato impresso, presencial, exceto no período de emergência sanitária por causa da pandemia Covid-19. A formatação, diagramação e impressão das provas, a correção das folhas de resposta, a análise e divulgação dos resultados para os estudantes e as escolas, e a emissão de certificado de participação aos estudantes são centralizadas, realizadas por empresa contratada pelas escolas com experiência em trabalhar com núcleos para o TP em outras regionais.

Após o consenso de data da prova, cada escola responsabiliza-se pelo envio das listagens de alunos à empresa e pela divulgação do TP aos seus estudantes. É pactuado o horário de início da prova entre todas as escolas, e a divisão dos estudantes nas salas segue a ordem alfabética ou os semestres do curso, sendo uma opção de cada escola. A prova é aplicada a todos os estudantes do primeiro ao sexto ano das escolas do núcleo, no mesmo dia e com duração de quatro horas. Em 2020 e 2021, por conta da pandemia da Covid-19, o processo de aplicação da prova foi virtual. O propósito do teste, se exclusivamente formativo ou somativo, fica a critério de cada escola.

Análise dos resultados

A pontuação total do estudante é calculada pelo número de questões corretas. No teste, não há a opção “não sei”, nem penalidades para opções incorretas. O teste é analisado globalmente pela teoria clássica do item. O processo de análise psicométrica dos itens é iniciado pela ordenação do escore total dos examinandos. Constituem-se dois subgrupos extremos: um superior, composto de 27% dos que obtiveram os maiores escores, e grupo inferior, organizado com 27% daqueles que alcançaram os menores escores. A análise obtida fornece, além do percentual de acertos dos estudantes, o

índice de dificuldade e discriminação dos itens, o coeficiente de fidedignidade e a correlação ponto bisserial.

Feedback aos estudantes

O *feedback* detalhado é disponibilizado aos estudantes inicialmente por acesso ao gabarito comentado com as referências bibliográficas e a análise dos seus resultados em ambiente virtual. Os estudantes recebem uma análise do seu desempenho de forma individual comparativamente com a média obtida pelo grupo de escolas no mesmo ano de curso, demonstrando também a sua progressão em relação ao teste anterior (Figura 1).

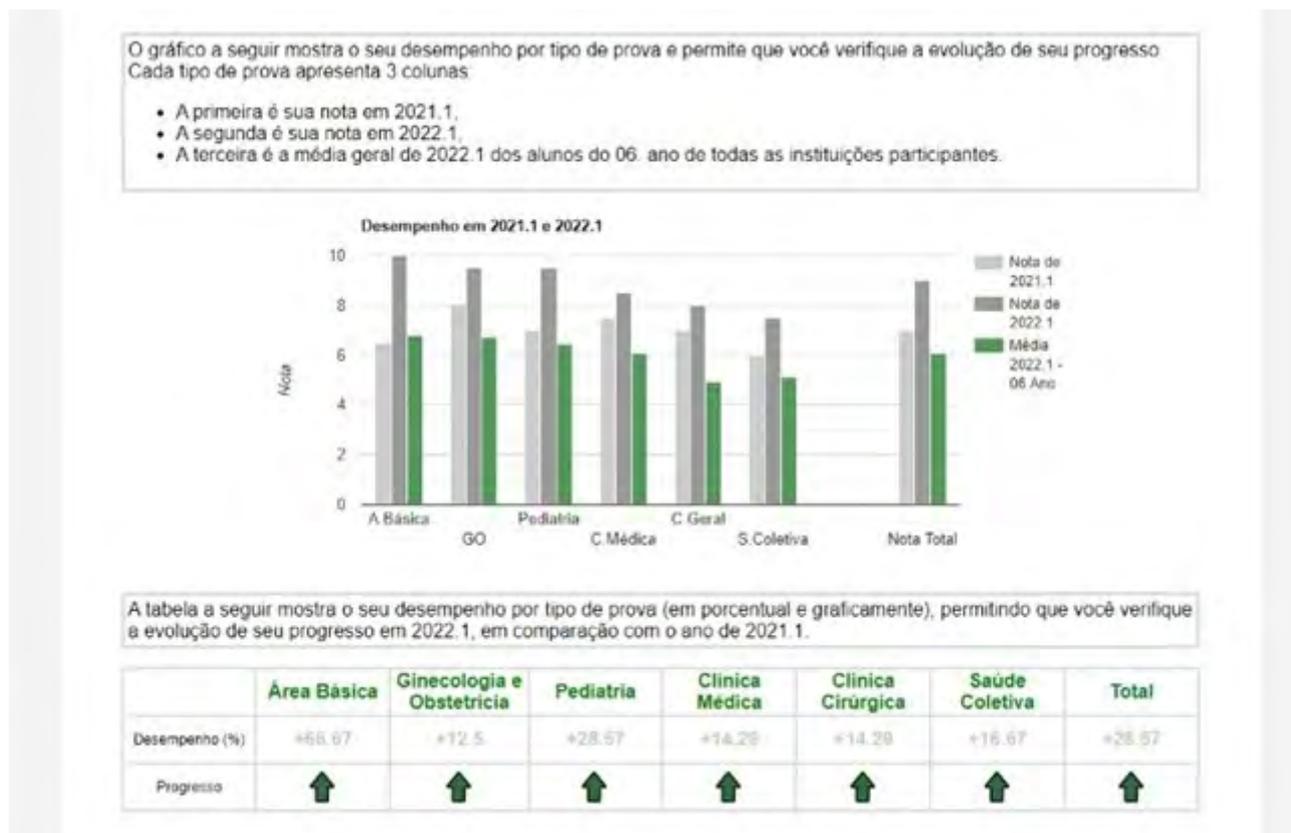
Feedback à instituição

Após liberação dos resultados, o TP possibilita que o curso reveja os próprios conteúdos e o projeto pedagógico. A aplicação de psicométrica às avaliações, por meio de análise dos graus de dificuldade e de discriminação geral e por áreas de conhecimento, permite avaliar a qualidade das questões e a correlação entre o acerto em determinado item e o número total de acertos na prova. Essas informações de erros e acertos, principalmente nas questões de melhor discriminação, trazem devolutivas valiosas aos docentes envolvidos nos conteúdos envolvidos.

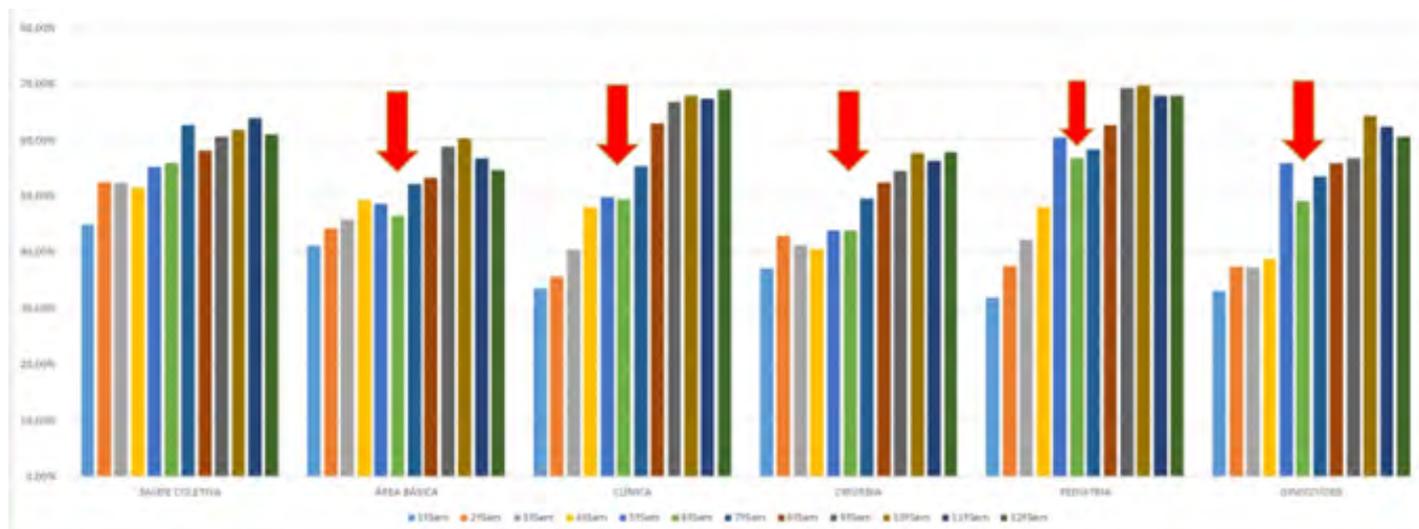
Tem sido observada variação nas dificuldades médias do TP por área de conhecimento ao longo dos anos, quando o número de acertos deveria ser equivalente em suas edições para ampliar o poder de comparação e assegurar sua validade.

A devolutiva aos docentes a respeito das questões formuladas possibilita ainda desenvolvimento docente para a elaboração de itens de boa qualidade, de modo a buscar processos cognitivos mais bem preparados segundo a taxonomia de Bloom, por meio de questões contextualizadas em situações-problema que contribuem para a melhor retenção do conteúdo. Outra vantagem tem sido o banco de itens gerado nos consórcios.

As instituições fornecem *feedback* aos coordenadores e professores, o que motiva reflexões internas quanto à necessidade de melhorias. A Figura 2 mostra um exemplo de sinalização de uma turma (semestre 6) com alguma dificuldade observada em todas as grandes áreas, evidenciado por uma das escolas desse consórcio em uma aplicação de TP. O apoio institucional e o comprometimento dos diretores, coordenadores e professores são fundamentais nesse processo, tanto na fase de elaboração de itens como na participação ativa na discussão dos resultados apresentados.

Figura 1. Exemplo de *feedback* eletrônico recebido pelos estudantes após a realização do teste

Fonte: Edudata Informática.

Figura 2. Exemplo de acompanhamento da evolução do desempenho em TP ao longo dos semestres do curso, por área de conhecimento

Fonte: Edudata Informática.

DISCUSSÃO

A implantação do CIN1, com processo colaborativo de realização da prova, representou um grande avanço para as escolas envolvidas. As escolas iniciaram um processo de colaboração não apenas para o TP, mas também para a realização de atividades de desenvolvimento docente, com

intercâmbio de informações e experiências. Foram realizados eventos sobre outros temas de educação médica a partir da constituição do consórcio.

A variação na participação das escolas ao longo dos anos, observada no presente relato, pode ser atribuída em sua maioria a dificuldades com os custos para realização do TP,

mesmo de forma compartilhada. O ambiente *on-line* oferece essa possibilidade com mais baixo custo, mas as dificuldades técnicas apresentadas no período de pandemia não permitiram observar com precisão esse potencial benefício.

A aplicação e análise dos resultados proporcionam aos cursos e aos estudantes uma avaliação externa, pois a elaboração das questões é distribuída uniformemente entre as escolas, o que reduz o viés da endogenia (questões feitas por professores da própria escola)¹⁴. Isso permite melhoria na qualidade da prova, em razão da troca de experiências entre as instituições de ensino e por conta da montagem de um banco de itens com questões testadas com alto grau de discriminação¹⁵. A colaboração entre faculdades garante maior número de itens criados, amplificação dos comitês de revisão e repartição entre os custos¹⁶. Em locais em que os recursos são limitados, a criação de consórcios é especialmente importante para a melhoria do teste, assim como de outros processos acadêmicos^{11,17,18,19}.

É fornecido aos estudantes o desenvolvimento do aprendizado a partir da avaliação, pois eles têm acesso ao próprio desempenho e ao gabarito comentado com referências bibliográficas. A participação dos estudantes do CIN1 vem se mantendo elevada, apesar de ter oscilado na pandemia após a mudança para a modalidade da prova *on-line*. A longitudinalidade dessa modalidade de avaliação formativa fornece uma medida única e demonstrável da progressão cognitiva dos estudantes¹¹. Após cada avaliação, é possível reafirmar e consolidar conhecimentos e identificar lacunas de aprendizagem e pontos a serem melhorados¹⁹. A participação recomendada dos estudantes na elaboração ou análise das questões ainda não é uma realidade na organização do CIN, o que está programado para implantação futura.

A avaliação dos professores envolvidos na coordenação e organização do teste nas suas escolas é positiva. Há intercâmbio entre os outros consórcios constituídos no país, com realização de fóruns e oficinas nos congressos de educação médica. Um dos desafios é aumentar a participação dos estudantes de algumas escolas para garantir maior homogeneidade na avaliação.

O TP possibilita que os cursos revejam suas matrizes curriculares e mesmo seus projetos pedagógicos por meio da análise dos conteúdos (gerais e por área)²⁰. Fornece subsídios à instituição de ensino para a avaliação do modelo curricular utilizado²¹, por meio da identificação de como os conhecimentos dos alunos progridem durante a formação médica⁶. Para isso, é fundamental que as instituições assumam papel ativo na elaboração e na análise dos dados dessa avaliação^{10,14,22}.

CONCLUSÃO

O presente relato apresentou a experiência de implantação e organização de um núcleo de escolas médicas para o TP. A organização desse núcleo trouxe benefícios e muitos desafios. O aperfeiçoamento dos critérios de análises e *feedback* e a inclusão dos estudantes na avaliação de questões são etapas a serem discutidas e implantadas. Há ainda outro desafio: a ampliação e manutenção do número de escolas. O custo ainda é um fator limitante à participação das escolas de forma mais regular, e a experiência com a utilização do ambiente virtual, de mais baixo custo, aponta um dos caminhos para superar essa dificuldade.

CONTRIBUIÇÃO DAS AUTORAS

Taciana Barbosa Duque participou do planejamento do estudo, da revisão da literatura e da redação e revisão do manuscrito. Daniela Chiesa, Raquel Autran Coelho Peixoto e Carla Adriane Fonseca Leal de Araújo participaram do planejamento estudo, da revisão da literatura e da redação do manuscrito.

CONFLITO DE INTERESSES

Declaramos não haver conflito de interesses.

FINANCIAMENTO

Declaramos não haver financiamento.

REFERÊNCIAS

1. Neeley SM, Ulman CA, Sydelko BS, Borges NJ. The value of Progress Testing in undergraduate medical education: a systematic review of the literature. *Med Sci Educ*. 2016;26:617-22.
2. Van der Vleuten C, Freeman A, Collares CF. Progress test utopia. *Perspect Med Educ*. 2018;7(2):136-8.
3. Albanese M, Case SM. Progress Testing: critical analysis and suggested practices. *Adv Health Sci Educ*. 2016;21(1):221-34.
4. Blake JM, Norman GR, Keane DR, Mueller B, Cunnington J, Didyk N. Introducing progress testing in McMaster University's problem-based curriculum: psychometric properties and effect of learning. *Acad Med*. 1996;71:1002-7.
5. Norman G, Neville A, Blake J, Mueller B. Assessment steers learning down the right road: impact of progress testing on licensing examination performance. *Med Teach*. 2010;32:496-9.
6. Karay Y, Schaubert SK. A validity argument for progress testing: examining the relation between growth trajectories obtained by progress tests and national licensing examinations using a latent growth curve approach. *Med Teac*. 2018;28:1-7.
7. Hamamoto Filho PT, Lourenção APLT, do Valle AP, Abbade JF, Bicudo AM. The correlation between students' Progress Testing scores and their performance in a residency selection process. *Med Sci Educ*. 2019;29(1):1071-5.
8. Guaraná CV, Duque TB, Melo e Lima TR. Desempenho no teste de progresso e coeficiente de rendimento final de curso são preditores de aprovação na residência médica? *Medicina (Ribeirão Preto)*. 2022;55(2):e-18211.
9. 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.

10. Bicudo AM, Hamamoto Filho PT, Abbade JF, Hafner MLMB, Maffei CM. Teste de Progresso em consórcios para todas as escolas médicas do Brasil. *Rev Bras Educ Med.* 2019;43(4):151-6.
11. 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.
12. de Oliveira SS, Postal EA, Afonso DH, Meress CE, Cyrino EG, de Abreu Junior AF, et al. Teste de Progresso da Abem: consolidando uma estratégia de avaliação para o ensino médico. *Rev Bras Educ Med.* 2022;46(1):e061.
13. Ricketts C, Freeman A, Pagliuga G, Coombes I, Archer J. Difficult decisions for progress testing: how much and how often? *Med Teach.* 2010;32:513-5.
14. Pinheiro OL, Spadella MA, Moreira HM, Ribeiro ZMT, Guimarães 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.
15. Sakai MH, Ferreira Filho OF, Almeida MJ, Matsuo T. Avaliação do crescimento cognitivo do estudante de Medicina: aplicação do teste de equalização no Teste de Progresso. *Rev Bras Educ Med.* 2011;35(4):493-501.
16. Heeneman S, Schut S, Donkers J, Van der Vleuten C, Muijtjens A. Embedding of the progress test in an assessment program designed according to the principles of programmatic assessment. *Med Teach.* 2017;15:1-9.
17. Schuwirth LW, van der Vleuten CP. The use of progress testing. *Perspectives on Medical Education.* 2012;1(1):2430.
18. Muijtjens AMM, Schuwirth LWT, CohenSchotanus J, van der Vleuten CPM. Differences in knowledge development exposed by multicurricular progress test data. *Adv Health Sci Educ.* 2008;13:593-605.
19. Dijksterhuis MG, Scheele F, Schuwirth LW, Essed GG, Nijhuis JG, Braat DD. Progress testing in postgraduate medical education. *Med Teach.* 2009;31(10):e464-e468.
20. Reberti AG, Monfredini NH, Ferreira Filho OF, Andrade DFD, Pinheiro CEA, Silva JC. Progress Test in medical school: a systematic review of the literature. *Rev Bras Educ Med.* 2020;44(1):e015.
21. Swanson D, Holtzman K, Butler A, Langer MM, Nelson MV, Chow JW, et al. Collaboration across the pond: the multi-school progress testing project. *Med Teach.* 2010;32:480-5.
22. Rosa MI, Isoppo CC, Cattaneo HD, Madeira K, Adami F, Ferreira Filho OF. O Teste de Progresso como indicador para melhorias em curso de graduação em Medicina. *Rev Bras Educ Med.* 2017;41(1):58-68.



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.

O Teste de Progresso na residência médica em ginecologia e obstetrícia: a experiência nacional

The Progress Test in gynecology and obstetrics medical residency: the national experience

Gustavo Salata Romão¹ | gsalata@uol.com.br

Cesar Eduardo Fernandes² | cesarefernandes_dircient@febrasgo.org.br

Aginaldo Lopes da Silva Filho³ | agnaldo.ufmg@gmail.com

Marcos Felipe Silva de Sá⁴ | marcosfelipe@fmrp.usp.br

RESUMO

Introdução: O Teste de Progresso (TP) é uma avaliação cognitiva abrangente e longitudinal com vantagens para o aprendiz, os programas educacionais e a sociedade. No Brasil, embora seja amplamente utilizado na graduação, existe pouca experiência do seu uso na residência médica. Este estudo visa retratar a experiência da Federação Brasileira das Associações de Ginecologia e Obstetrícia (Febrasgo) nos quatro primeiros anos de implementação do TP Individual do Residente em Ginecologia e Obstetrícia (TPI-GO).

Relato da experiência: O TPI-GO foi implementado em 2018, sendo oferecido anualmente aos residentes brasileiros. Nos dois últimos anos, o TPI-GO passou a ser oferecido no formato on-line. Para incentivar a participação, estabeleceram-se critérios que possibilitam dispensa ou bonificação na prova teórica para a obtenção do Título de Especialista em Ginecologia e Obstetrícia (Tego) para os candidatos com melhor desempenho. Os resultados são fornecidos de maneira sigilosa a cada candidato. Os resultados do desempenho dos residentes de um mesmo serviço são fornecidos ao supervisor do programa sem a identificação dos seus residentes, de modo a oferecer subsídios para autoavaliação da qualidade e indicar pontos de melhoria.

Discussão: Desde a implementação do TP, verificou-se um número crescente de candidatos inscritos com elevada taxa de aderência à prova. As medianas de desempenho do grupo de residentes evidenciaram evolução cognitiva entre os iniciantes (R1) e os concluintes (R3) dos programas.

Conclusão: A implementação do TP na residência é factível e aceitável, apresentando vantagens e benefícios. No entanto, requer preparo e envolvimento da equipe organizadora, além de apoio logístico e financeiro. O sigilo na divulgação dos resultados é recomendável e preserva o candidato. A possibilidade de bonificação e dispensa na prova teórica do Tego é um grande atrativo para a adesão e aderência dos médicos residentes ao TP. A aplicação da prova em ambiente on-line amplia o acesso, com níveis aceitáveis de segurança.

Palavras-chave: Educação Médica; Residência Médica; Avaliação Educacional; Teste de Progresso; Ginecologia e Obstetrícia.

ABSTRACT

Introduction: *The Progress Test (PT) is a comprehensive, longitudinal cognitive assessment with benefits for the learner, educational programs, and society. Although it is widely used in undergraduate courses in Brazil, there is little experience regarding its use in medical residency. This study aims to portray Febrasgo's experience in the first four years of implementing the Resident's Individual PT in ObGyn (TPI-GO).*

Experience report: *The TPI-GO was implemented in 2018, being offered annually to Brazilian residents. In the last two years, the TPI-GO has been provided online. Febrasgo established criteria that allow for exemption or bonus in the theoretical test to encourage the trainees' participation in the TPI-GO to obtain the Specialist Certification in Gynecology and Obstetrics (TEGO). Results are provided confidentially to candidates and allow self-assessment. The results of the residents' performance from the same service are provided to the program supervisor, offering subsidies for self-assessment of quality and indicating points for improvement.*

Discussion: *Since it was implemented, there has been an increasing number of candidates enrolled in the TPI-GO and a high adherence rate. The medians of performance of the two cohorts of residents studied showed cognitive evolution between beginners (R1) and graduates (R3), with differences attributed in part to the adverse effects of the Pandemic on training.*

Conclusion: *The implementation of the PT in medical residency is feasible, acceptable, and shows advantages and benefits, but it requires the commitment of the organizing team, besides logistical and financial support. The results confidentiality is recommended and preserves the candidate. The possibility of granting bonuses and exemptions in the TEGO theoretical test is an excellent attraction aiming at the adherence and adherence of residents to the PT. The TPI-GO online increases access with acceptable levels of security.*

Keywords: *Medical Education; Medical Residency; Educational Assessment; Progress Test; Obstetrics and Gynecology.*

¹Universidade de Ribeirão Preto, Ribeirão Preto, São Paulo, Brasil.

²Centro Universitário Faculdade de Medicina do ABC, Santo André, São Paulo, Brasil.

³Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brasil.

⁴Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo, Ribeirão Preto, São Paulo, Brasil.

Editor: Aristides Augusto Palhares Neto.

Recebido em 07/09/22; Aceito em 04/10/22.

Avaliado pelo processo de *double blind review*.

INTRODUÇÃO

O ensino e a avaliação do conhecimento clínico são fundamentais durante a formação médica. Entre as diferentes modalidades de avaliação cognitiva, o Teste de Progresso (TP) se destaca por ser um método abrangente, longitudinal e seriado, em que um mesmo conjunto de itens é aplicado a aprendizes que se encontram em diferentes fases ou períodos de treinamento¹. Quando comparado a outras modalidades de avaliação cognitiva, apresenta vantagens como promover a autoavaliação e avaliação formativa, permitindo reafirmar saberes prévios e identificar lacunas de aprendizagem². Por causa de sua abrangência, representa um estímulo à aprendizagem contínua e um desestímulo ao estudo de véspera de prova, conscientizando e engajando o aprendiz em seu processo metacognitivo³. Dessa forma, o TP exerce um efeito educacional favorável não somente em termos de aprendizagem funcional, como também em relação à retenção do conhecimento em longo prazo⁴.

Sob a perspectiva dos professores e preceptores, o TP fornece informações valiosas sobre o desempenho dos aprendizes e permite a identificação daqueles que necessitam de maior apoio, remediação ou progressão acelerada. Adicionalmente o TP provê maior estabilidade ao processo de avaliação cognitiva, não sendo necessário adaptá-lo às eventuais mudanças curriculares de um programa¹.

O uso do TP na residência médica tem sido descrito desde 1999, quando foi implementado na Holanda para avaliação dos médicos residentes em ginecologia e obstetrícia⁵. No Brasil, embora o TP seja amplamente utilizado na graduação em Medicina, o uso na residência tem sido bastante restrito, já que é oferecido por poucas especialidades médicas.

Em resposta aos anseios e às expectativas da sociedade por uma prática médica mais qualificada e segura, o TP apresenta maior validade preditiva da qualidade dos futuros profissionais quando comparado a outras modalidades de avaliação cognitiva pontual⁶. Por esses e outros motivos, a Federação Brasileira das Associações de Ginecologia e Obstetrícia (Febrasgo) desenvolveu o TP Individual do Residente em Ginecologia e Obstetrícia (TPI-GO) para aplicação em todo o Brasil a partir de 2018⁷. Este artigo tem por objetivo retratar a experiência dos quatro primeiros anos de implantação do TPI-GO.

RELATO DE EXPERIÊNCIA

Este é um estudo descritivo que relata a experiência da Febrasgo na implantação do TPI-GO e no acompanhamento longitudinal de duas coortes de médicos residentes em ginecologia e obstetrícia. Cada coorte representa um mesmo grupo de residentes ao longo dos três anos consecutivos de participação no TPI-GO. A primeira coorte refere-se

aos residentes que iniciaram a participação no TPI-GO em 2018 como R1, participaram da prova em 2019 como R2 e concluíram a participação em 2020 como R3. Já a segunda coorte corresponde aos residentes que iniciaram a participação no TPI-GO em 2019 como R1, participaram da prova em 2020 como R2 e concluíram a participação em 2021 como R3.

Este relato de experiência utilizou um banco de dados que não foi produzido com a finalidade de pesquisa e cujas informações são agregadas, sem possibilidade de identificação individual dos sujeitos. Sendo assim, nesse caso, este relato fica dispensado de ser registrado ou avaliado pelo sistema de Comitês de Ética em Pesquisa da Comissão Nacional de Ética em Pesquisa (CEP/Conep), conforme Resolução nº 510, de 7 de abril de 2016, do Conselho Nacional de Saúde⁸.

Implementação do TPI-GO

Em 2018, o TPI-GO foi implementado pela Comissão de Residência Médica (Coreme) da Febrasgo e disponibilizado a todos os residentes regularmente matriculados em programas nessa especialidade credenciados pelo Ministério da Educação (MEC)⁷.

A prova teórica para obtenção do Título de Especialista em Ginecologia e Obstetrícia (Tego) concedido pela Febrasgo é a base para o TPI-GO. A matriz da prova é baseada na Matriz de Competências em Ginecologia e Obstetrícia^{9,10} e assegura a distribuição balanceada entre as subáreas de atuação do ginecologista e obstetra. A prova contém 100 questões de múltipla escolha com quatro alternativas e apenas uma opção correta, sendo 50 de ginecologia e 50 de obstetrícia. As questões são elaboradas e revisadas pela Comissão Nacional do Tego (CN-Tego)¹¹.

O TPI-GO é oferecido anualmente aos residentes do primeiro (R1), segundo (R2) e terceiro (R3) anos de treinamento. A participação é voluntária, e, para estimular a adesão, foram estabelecidos critérios que possibilitam a dispensa ou bonificação na prova teórica do Tego no ano subsequente à conclusão do programa de residência médica (PRM). De acordo com esses critérios, serão dispensados da prova teórica do Tego os candidatos que participarem do TPI-GO como R1, R2 e R3, apresentarem um desempenho igual ou superior ao percentil 60 (P60) em relação aos seus pares em pelo menos duas versões da prova (incluindo necessariamente o desempenho como R3) e obtiverem uma percentagem de acertos de pelo menos 60% nas questões de ginecologia e 60% de obstetrícia na última aplicação (como R3). Serão bonificados na prova teórica do Tego os candidatos que participarem do TPI-GO como R1, R2 e R3, apresentarem um desempenho superior ao percentil 30 (P30) em relação aos seus pares em todas as versões da prova e obtiverem uma percentagem de acertos de pelo menos 50%

nas questões de ginecologia e de obstetrícia na última aplicação (como R3). Nesta última condição, os candidatos terão um bônus de 10% nas questões de ginecologia e de obstetrícia na prova teórica do Tego no ano subsequente¹¹. A adoção dos percentis 30 e 60 (P30 e P60) como pontos de corte na prova teórica do Tego foi estabelecida pela equipe que implementou a avaliação por meio do método de Angoff modificado¹². Essa escolha se baseou no pressuposto de que pontos de corte referenciados por norma são mais recomendados do que valores fixos (*cut-off*) por conta da variação nos níveis de dificuldade entre as diferentes versões do TP^{11,13,14}.

A aplicação do TPI-GO ocorre simultaneamente à prova teórica do Tego, sob a coordenação da Coreme-Febrasgo. Inicialmente, o TPI-GO foi oferecido presencialmente em 11 cidades: Porto Alegre (RS), Curitiba (PR), São Paulo (SP), Ribeirão Preto (SP), Rio de Janeiro (RJ), Belo Horizonte (MG), Salvador (BA), Recife (PE), Fortaleza (CE), Brasília (DF) e Belém (PA)¹¹.

A partir de 2020, passou a ser oferecido na modalidade *on-line*, o que permitiu ao candidato realizar a prova em seu domicílio utilizando computador próprio. Para garantir a segurança, os candidatos instalam um aplicativo que bloqueia o acesso ao navegador de internet e permite a monitorização contínua e automática, utilizando a câmera e o microfone do próprio computador. Além disso, os candidatos são monitorados remotamente por fiscais, sendo possível detectar sons, movimentos corporais e desvios do olhar, bem como a presença de outras pessoas ou artefatos suspeitos no ambiente. Quando há infrações ou tentativas de fraude, os candidatos são advertidos, e, em casos de persistência, a prova é interrompida¹⁵.

A divulgação dos resultados do desempenho individual é feita somente ao próprio candidato e de maneira sigilosa. São fornecidas informações sobre o seu desempenho geral e nas questões de ginecologia e de obstetrícia, e o desempenho progressivo em relação aos anos anteriores. Também são disponibilizados gráficos comparativos em relação ao desempenho geral de seus pares (R1, R2 ou R3) em nível nacional, utilizando como parâmetros a mediana, o percentil 30 (P30) e o percentil 60 (P60). Para os R3, o laudo inclui informações sobre a dispensa ou bonificação na prova teórica do Tego¹¹. Para o cálculo desses valores e a identificação dos candidatos que preenchem os critérios para bonificação ou dispensa na prova teórica do Tego no ano subsequente, utilizou-se o pacote de análise estatística do Excel com seus diversos recursos

disponíveis. Esse mesmo *software* foi utilizado na elaboração dos gráficos do desempenho dos residentes em relação ao P30, à mediana e ao P60 do desempenho geral em nível nacional. A análise estatística das diferenças de desempenho entre R1 e R2 e entre R2 e R3 em cada coorte foi realizada por meio do teste de Mann-Whitney considerando a não homogeneidade da distribuição de notas dos residentes. Em todas as análises, consideraram-se significativas as diferenças para $p < 0,05$. A consistência interna das provas de ginecologia e de obstetrícia passou a ser analisada a partir de 2020 por meio do coeficiente alfa de Cronbach, considerando-se adequados os valores de alfa acima de 0,64¹⁶. Para as análises foi utilizado o *software* IBM SPSS Statistics – V22.

O consolidado do desempenho dos residentes de um mesmo serviço é enviado ao supervisor do PRM. Com base nesses resultados, é possível avaliar a evolução do grupo ao longo do PRM e compará-lo com as curvas de desempenho nacional em cada categoria (R1, R2 e R3). Para preservar o sigilo, somente poderão receber esses resultados os supervisores de PRM em que pelo menos três residentes de cada categoria se aplicaram ao TPI-GO¹¹. Os residentes não são identificados no relatório.

Ao final de cada aplicação do TPI-GO, realiza-se a análise psicométrica dos itens quanto ao grau de dificuldade e poder discriminativo, fornecendo bases para os trabalhos da comissão de revisores no ano subsequente¹⁷.

Perfil dos residentes participantes

Desde 2018, quando o TPI-GO foi implementado, tem-se observado aumento do número de participantes (Tabela 1), com predomínio de R1 em todos os anos. Quanto ao perfil, a idade dos candidatos oscila entre 24 e 52 anos, com média de 28,3 e mediana de 28,2 anos, e predomínio de mulheres (87%). Em relação ao estado civil, 84% dos candidatos são solteiros, 15% são casados e 0,8% é divorciado. Quanto à localização dos PRM de origem dos candidatos, 117 (52%) estão no Sudeste, 42 (19%) no Sul, 32 (14%) no Nordeste, 20 (9%) no Centro-oeste e 12 (5%) no Norte.

Analisamos os candidatos em duas coortes com medidas repetidas. Na primeira, foram incluídos os candidatos que realizaram as três provas no período de 2018 a 2020, e, na segunda, os candidatos que o fizeram no período de 2019 a 2021. A aderência à prova foi avaliada por meio da proporção de candidatos que se aplicaram ao TPI-GO como R3 em relação ao número de candidatos que se aplicaram à prova como R1 no início de cada coorte (Tabela 2).

Tabela 1. Médicos residentes inscritos no TPI-GO por ano

Ano	Total de inscritos	R1	R2	R3
2018	1.202	497	360	345
2019	1.314	568	457	289
2020	1.422	628	480	314
2021	1.495	634	501	360

Fonte: Elaborada pelos autores.

Tabela 2. Aderência dos candidatos ao TPI-GO

Coorte	Total de R1 em 2018	Total de R3 em 2020	Aderência
De 2018 a 2020	497	314	63%
De 2019 a 2021	568	360	63%

Fonte: Elaborada pelos autores.

Desempenho dos residentes participantes

A evolução dos candidatos no TPI-GO reflete o ganho de conhecimento adquirido ao longo do treinamento e pode ser avaliada a partir do desempenho geral dos candidatos em nível nacional, em cada coorte (gráficos 1 e 2). Na coorte 1, houve aumento estatisticamente significativo do desempenho entre os R1 e os R2 ($p < 0,001$), sem aumento estatisticamente significativo do desempenho entre os R2 e os R3 ($p = 0,584$). Na coorte 2, houve redução estatisticamente significativa do desempenho entre os R1 e os R2 ($p < 0,001$), com aumento estatisticamente significativo do desempenho entre os R2 e os R3 ($p < 0,001$).

A consistência interna das provas, medida pelo coeficiente alfa de Cronbach para as provas de ginecologia e obstetrícia, foi 0,640 e 0,560 em 2020, e 0,647 e 0,709 em 2021, respectivamente, o que se traduz em valores adequados na prova de ginecologia em 2020 e 2021, e na prova de obstetrícia em 2021, segundo os critérios aqui utilizados.

DISCUSSÃO

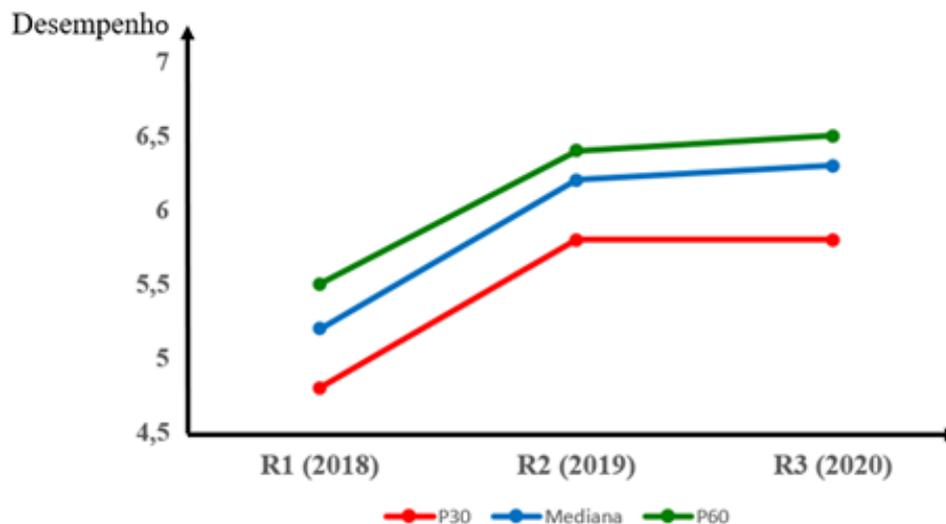
O TPI-GO é uma iniciativa da Febrasgo ante o conhecimento das vantagens e dos benefícios do teste, e diante da constatação da inexistência de uma avaliação cognitiva confiável, abrangente e unificada que pudesse ser oferecida nacionalmente a todos os médicos residentes em ginecologia e obstetrícia. O desenvolvimento e a implementação do TPI-GO envolveram um meticuloso trabalho da Coreme-Febrasgo e da CN-Tego, com o apoio da diretoria científica e da presidência da Febrasgo em duas gestões consecutivas^{7,11}.

Desde a implementação em 2018, o TPI-GO tem apresentado um número crescente e expressivo de candidatos inscritos, atingindo um total 634 candidatos iniciantes

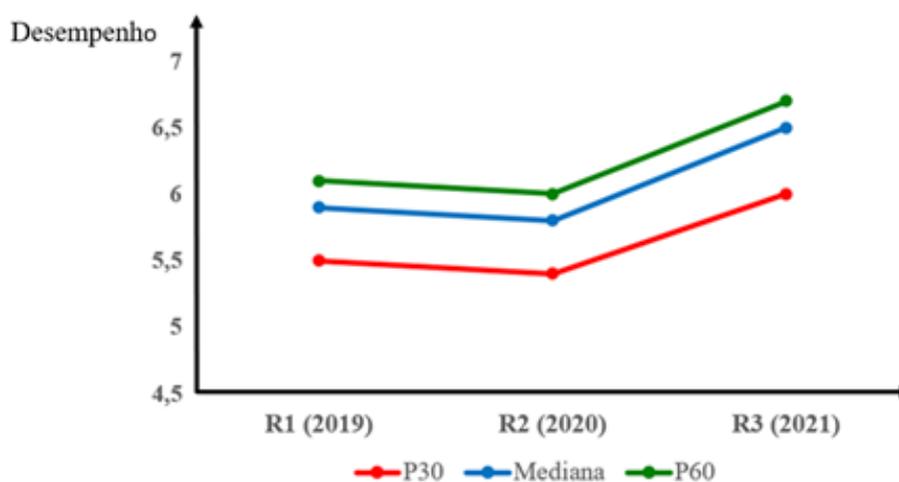
(R1) inscritos em 2021, o que corresponde a 48% do total de R1matriculados em PRM credenciados pelo MEC nessa especialidade ($n = 1.325$)¹⁸. Também são notórias as taxas de aderência dos candidatos nos três anos consecutivos de aplicação da prova, atingindo 63% nas duas coortes avaliadas. Diversas são as razões para o êxito desse modelo de avaliação. A começar pelo sigilo na divulgação dos resultados ao candidato, de modo a evitar a sua exposição perante os pares, preceptores e supervisores, e prevenir atitudes discriminativas que levem ao constrangimento, ao estresse ou à perda da autoconfiança. A avaliação formativa orientada para o aprendizado, o sigilo e o *feedback* construtivo oferecem melhores resultados quando comparados às avaliações excessivamente competitivas que expõem os aprendizes à desqualificação ou depreciação¹⁹. Diversos estudos demonstraram um ganho consistente no aprendizado e na retenção de conteúdo em longo prazo a partir da avaliação formativa e do *feedback* pós-teste²⁰⁻²².

Outro fator determinante para a adesão e aderência dos candidatos ao TPI-GO é a possibilidade de dispensa ou bonificação na prova teórica do Tego. Ser dispensado ou bonificado não significa apenas uma maior facilidade na obtenção do título, mas também o reconhecimento do seu desenvolvimento na especialidade, corroborado pelo desempenho satisfatório em uma avaliação confiável.

Também deve ser considerada a maior facilidade de acesso ao TPI-GO pela aplicação em ambiente virtual a partir de 2020. A modalidade *on-line* permite ao candidato realizar o TPI-GO utilizando seu próprio computador, com um sistema de monitorização que garante níveis aceitáveis de segurança. Dessa forma, prescinde-se da necessidade de deslocamento até os centros aplicadores, o que representa um grande desafio para os serviços que dependem da força

Gráfico 1. Evolução do desempenho dos residentes no TPI-GO na coorte de 2018 a 2020

Fonte: Elaborado pelos autores.

Gráfico 2. Evolução do desempenho dos residentes no TPI-GO na coorte de 2019 a 2021

Fonte: Elaborado pelos autores.

de trabalho dos residentes, além de reduzir os gastos com passagens e hospedagem.

No tocante ao perfil dos candidatos, verifica-se um predomínio das mulheres. Esse achado é corroborado por outros levantamentos que demonstram a inserção cada vez maior das mulheres nos PRM brasileiros¹⁸. Em relação à procedência, mais da metade é oriunda de PRM da Região Sudeste, que concentra 54,5% dos residentes em ginecologia e obstetria¹⁸.

O desempenho geral no TPI-GO é de grande valia no acompanhamento do ganho cognitivo pelos residentes ao longo do treinamento. O conhecimento é um componente essencial das competências do médico e tem impacto direto sobre o seu desempenho profissional²³. Logo, a verificação e a mensuração do conhecimento adquirido durante o PRM se

tornam essenciais. A magnitude de aquisição do conhecimento pode ser estimada pelo “tamanho do efeito”, cujo valor reflete a força da relação entre uma intervenção pedagógica e o aprendizado²⁴. Uma importante finalidade do TP é avaliar o ganho cumulativo de conhecimento e o efeito do treinamento sobre o aprendizado dos residentes²⁴.

A análise da primeira coorte de candidatos que iniciaram a aplicação no TPI-GO em 2018 como R1, participaram da prova em 2019 como R2 e completaram a participação em 2020 como R3 (coorte 2018-2020) mostra um aumento progressivo na mediana do desempenho. Entretanto, o ganho foi expressivamente maior do R1 para o R2, em que se verificou um aumento estatisticamente significativo. Já para a diferença de desempenho entre os R2 e os R3, não foi verificada significância estatística. Diversas hipóteses

poderiam justificar a estagnação do ganho cognitivo verificada entre os R2 e os R3 na primeira coorte.

Primeiramente, é preciso considerar as potenciais diferenças no grau de dificuldade da prova entre as diversas versões do TPI-GO. Considerando que a Febrasgo não dispõe de um banco de itens para elaboração da prova, não é possível assumir que não existam oscilações nesse quesito. Os consórcios internacionais recomendam o armazenamento de questões pré-testadas quanto ao poder discriminativo e ao grau de dificuldade em um banco de itens. Cada nova versão da prova deve conter pelo menos 30% das questões provenientes do banco (questões-âncoras) para garantir a estabilidade da avaliação. Para reduzir a margem de acertos por memorização de questões, os itens devem permanecer no banco por um período mínimo de três anos para que possam ser reutilizadas. No entanto, a legislação brasileira exige a divulgação das questões da prova do Tego imediatamente após a sua aplicação, visando assegurar maior transparência e *feedback* aos candidatos. Essa divulgação, embora vantajosa, dificulta a reutilização de questões e limita o valor do banco de itens. Apesar dessas dificuldades, a Febrasgo está se empenhando na construção de um banco de questões validadas após sua aplicação no Tego.

Outro aspecto a ser considerado se refere à heterogeneidade das condições de aplicação do TPI-GO, que foi presencial em 2018 e 2019 e *on-line* a partir de 2020. A utilização do formato *on-line* representou uma mudança significativa no ambiente e formato da avaliação. Algumas normas de segurança foram estabelecidas, destacando-se o fato de que o candidato precisa completar uma questão para visualizar a próxima e não pode retornar a uma questão já respondida. É possível que a primeira aplicação digital tenha representado uma fonte de estresse e insegurança aos candidatos, muito embora alguns consórcios internacionais (como o consórcio alemão) tenham verificado maior preferência dos candidatos pela modalidade *on-line* do TP²⁵.

Finalmente, é preciso considerar os efeitos adversos da pandemia Covid-19 sobre o treinamento dos médicos residentes, particularmente em 2020. Durante o período mais crítico, diversas adaptações nos PRM foram necessárias, tanto para atender às exigências sanitárias quanto para reforçar as equipes no atendimento aos pacientes acometidos²⁶. Entre diversas mudanças, destacam-se o adiamento de consultas rotineiras, a suspensão de cirurgias eletivas e o deslocamento dos residentes para unidades de pronto atendimento, enfermarias e unidades de terapia intensiva²⁷. Ademais, houve desfalque nas equipes de supervisores em decorrência dos afastamentos por doença ou por enquadramento nos grupos de alto risco²⁷. As atividades teóricas foram adaptadas para o

ensino remoto emergencial, o que representou um desafio aos preceptores e supervisores²⁷. No treinamento em assistência à saúde da mulher, embora tenham sido mantidos os serviços essenciais nas maternidades e urgências, houve um déficit significativo nas outras áreas de aprendizagem²⁷. É plausível a hipótese de que todas essas adaptações durante a pandemia tenham impactado negativamente o desempenho dos residentes no TPI-GO de 2020, repercutindo negativamente na mediana de desempenho dos R3 da primeira coorte.

Na segunda coorte, em que os candidatos iniciaram a aplicação no TPI-GO em 2019 como R1, participaram da prova em 2020 como R2 e completaram a participação em 2021 como R3 (coorte 2019-2021), observa-se uma redução estatisticamente significativa do desempenho do R1 para o R2 com aumento estatisticamente do desempenho do R2 para o R3. Um dos fatores relacionados à redução da mediana dos R1 para o R2 nessa coorte foi a pandemia da Covid-19. Como já foi discutido anteriormente, essa emergência mundial de saúde pública exerceu impacto negativo sobre o treinamento dos residentes em diversas especialidades médicas, incluindo a ginecologia e obstetrícia²⁷. Talvez esse fator seja um dos responsáveis pelo desempenho abaixo das expectativas dos R2 que se aplicaram à prova em 2020, que foi sabidamente o ano mais afetado pela pandemia. Outros possíveis fatores relacionados ao baixo desempenho dos R2 no ano de 2020 foram a transição da versão presencial para a versão *on-line* da prova e as variações no grau de dificuldade da prova entre suas diferentes versões, conforme discutido anteriormente.

Em relação à consistência interna, verificaram-se níveis adequados do coeficiente alfa de Cronbach nas provas de ginecologia em 2020 e 2021 e de obstetrícia em 2021. O alfa representa uma estimativa de confiabilidade que é explicada pela variância do teste em cada ano de aplicação. Diversos fatores podem influenciar o valor de alfa, como o número de itens de teste, a interrelação entre os itens e a dimensionalidade do construto²⁸. Na literatura, encontramos diferentes relatos sobre os valores satisfatórios de alfa que variam de 0,60 a 0,95, sendo os valores ideais entre 0,70 e 0,90^{16,28,29}. No presente estudo, optamos por considerar adequados valores de alfa acima de 0,64¹⁶ e verificamos que, para a prova de obstetrícia de 2020, o valor de alfa não alcançou a faixa esperada. Esse achado pode ser atribuído a um baixo número de perguntas, à baixa interrelação entre itens ou a construtos heterogêneos^{16,28}. Resultados semelhantes foram relatados por Dijksterhuis et al.⁵ em uma amostra de residentes de ginecologia e obstetrícia que se aplicaram ao TP na Holanda de 1999 a 2008. Os valores de alfa variaram de 0,2 a 0,81 com uma média de 0,65, que foi considerada uma confiabilidade moderada nesse estudo. Segundo os autores, a explicação mais provável para a confiabilidade encontrada

seria a grande diversidade de subdomínios abordados no TP, a qual teve como consequência um número insuficiente de itens para amostrar adequadamente o domínio completo, resultando em perda de sinal mensurável⁵. No relato de experiência aqui apresentado, é preciso considerar também as dificuldades de elaboração e validação das questões de prova durante o período da pandemia, que impossibilitou a realização de reuniões presenciais dos revisores para validação das questões.

Algumas medidas têm sido implementadas para melhorar a confiabilidade do TPI-GO, tais como o aumento do número total de questões de 100 para 120, a readequação da matriz e prova, e as reuniões presenciais de revisores para discutir a qualidade dos itens. Futuramente, pretende-se desenvolver um banco de itens com questões pré-testadas quanto ao índice discriminativo e ao grau de dificuldade, o que deverá contribuir ainda mais para o aprimoramento dessa avaliação.

No TPI-GO, os candidatos recebem um *feedback* individual e sigiloso que traz informações sobre o número e a proporção de acertos totais, a proporção de acertos nas questões de ginecologia e de obstetrícia, o desempenho relativo aos seus pares em nível nacional (em relação à mediana, ao P30 e ao P60) e a evolução do seu desempenho em relação aos anos anteriores. Os candidatos do último ano de aplicação (R3) recebem informações adicionais sobre a dispensa e bonificação na prova teórica do Tego do ano seguinte com base em seu desempenho, segundo os critérios previstos em edital. Essas informações constituem a base para a autoavaliação e avaliação formativa dos residentes, de modo a permitir a reafirmação de conhecimento prévio e a identificação de lacunas de aprendizagem, e potencializando o efeito educacional do TPI-GO.

CONCLUSÃO

A partir da experiência da Febrasgo com o TPI-GO, é possível concluir que a implementação do TP na residência é factível e aceitável, e apresenta uma série de vantagens e benefícios para os médicos residentes, para os preceptores e supervisores, para os programas e para a sociedade em geral. A implementação do teste requer preparo, envolvimento e comprometimento dos membros da equipe organizadora, além de apoio logístico e financeiro. O sigilo na divulgação dos resultados é recomendável porque preserva o candidato de uma exposição desnecessária e que lhe possa ser prejudicial. A possibilidade de bonificação e dispensa na prova teórica do Tego é um grande atrativo para a adesão e aderência. A aplicação da prova em ambiente remoto amplia o acesso com níveis aceitáveis de segurança.

O TP, quando utilizado na residência médica, oferece uma grande oportunidade para a avaliação formativa,

promovendo ganho consistente no aprendizado e retenção de conhecimento em longo prazo. Também permite acompanhar o ganho cognitivo dos residentes ao longo do treinamento.

AGRADECIMENTOS

Agradecemos aos membros da Coreme da Febrasgo, da Comissão do Título de Especialista, à secretaria da Febrasgo e aos médicos residentes, preceptores e supervisores de PRM que se aplicaram ao TP da Febrasgo.

CONTRIBUIÇÃO DOS AUTORES

Gustavo Salata Romão participou da concepção do projeto, do delineamento metodológico, da coleta e análise estatística dos dados, da interpretação dos resultados e da redação e aprovação da versão final do artigo. Marcos Felipe Silva de Sá, Cesar Eduardo Fernandes e Agnaldo Lopes da Silva Filho participaram da concepção do projeto, do delineamento metodológico, da interpretação dos resultados e da redação do manuscrito.

CONFLITO DE INTERESSES

Declaramos não haver conflito de interesses.

FINANCIAMENTO

Projeto financiado pela Febrasgo.

REFERÊNCIAS

1. Freeman A, Van Der Vleuten C, Nouns Z, Ricketts C. Progress testing internationally. *Med Teach*. 2010;32(6):451-5. doi: 10.3109/0142159x.2010.485231.
2. Dion V, St-Onge C, Bartman I, Touchie C, Pugh D. Written-based Progress Testing: a scoping review. *Acad Med*. 2022;97(5):747-57. doi: 10.1097/acm.0000000000004507.
3. Van Berkel HJM, Nuy HJP, Geerligs T. The influence of progress tests and block tests on study behaviour. *Instructional Science*. 1994;22(4):317-33. doi: 10.1007/BF00891784.
4. Ricketts C, Freeman AC, Coombes LR. Standard setting for progress tests: combining external and internal standards. *Med Educ*. 2009;43(6):589-93. doi: 10.1111/j.1365-2923.2009.03372.x.
5. Dijksterhuis MG, Scheele F, Schuwirth LW, Essed GG, Nijhuis JG, Braat DD. Progress testing in postgraduate medical education. *Med Teach*. 2009;31(10):e464-8. doi: 10.3109/01421590902849545.
6. Schuwirth LW, van der Vleuten CP. The use of progress testing. *Perspect Med Educ*. 2012;1(1):24-30. doi: 10.1007/s40037-012-0007-2.
7. Romão GS, Fernandes CE, de Sá MFS. Teste de Progresso Individual do residente em GO: relato da experiência do primeiro ano de implantação no Brasil. *Femina*. 2019;282-7.
8. Brasil. Conselho Nacional de Saúde. Resolução nº 510, de 7 de abril de 2016. *Diário Oficial da República Federativa do Brasil*; 2016.
9. Romão GS, dos Reis FJC, Cavalli RC, de Sá MFS. Matriz de competência em ginecologia e obstetrícia: um novo referencial para os programas de residência médica no Brasil. *Femina*. 2017;45(3):172-7.
10. Romão GS, de Sá MFS. Competency-based training and the competency framework in gynecology and obstetrics in Brazil. *Rev Bras Ginecol Obstet*. 2020;42(5):272-88. doi: 10.1055/s-0040-1708887.

11. de Sá MFS, Romão GS, Fernandes CE, da Silva Filho AL. The Individual Progress Test of Gynecology and Obstetrics Residents (TPI-GO): the Brazilian experience by Febrasgo. *Rev Bras Ginecol Obstet.* 2021;43(6):425-8. doi: 10.1055/s-0041-1731803.
12. Bandaranayake RC. Setting and maintaining standards in multiple choice examinations: AMEE Guide No. 37. *Med Teach.* 2008;30(9-10):836-45.
13. Wrigley W, van der Vleuten CP, 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. doi: 10.3109/0142159x.2012.704437.
14. Muijtjens AMM, Hoogenboom RJJ, Verwijnen GM, Van der Vleuten CPM. Relative or absolute standards in assessing medical knowledge using progress tests. *Adv Health Sci Educ.* 1998;3(2):81-7.
15. Romão GS, de Sá MFS, Fernandes CE, da Silva Filho AL. Comment on "progress testing anytime, anywhere – does a mobile-learning approach enhance the utility of a large-scale formative assessment tool?". *Med Teach.* 2021 Aug;43(8):976-977. doi: 10.1080/0142159x.2020.1854453. Epub 2020 Dec 2.
16. Taber KS. The use of Cronbach's alpha when developing and reporting research instruments in science education. *Res Sci Educ.* 2018;48(6):1273-96.
17. Coombes L, Roberts M, Zahra D, Burr S. Twelve tips for assessment psychometrics. *Med Teach.* 2016;38(3):250-4. doi: 10.3109/0142159x.2015.1060306.
18. Scheffer M, Cassenote A, Guerra A, Guilloux AGA, Brandão APD, Miotto BA. *Demografia médica no Brasil 2020.* São Paulo: FMUSP, CFM; 2020.
19. Butler AC, Roediger HL. Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Mem Cognit.* 2008;36(3):604-16.
20. Kromann CB, Jensen ML, Ringsted C. The effect of testing on skills learning. *Med Educ.* 2009;43(1):21-7.
21. Roediger HL, Karpicke JD. The power of testing memory: Basic research and implications for educational practice. *Perspect Psychol Sci.* 2006;1(3):181-210.
22. Larsen DP. When I say... test-enhanced learning. *Med Educ.* 2013;47(10):961.
23. Wenghofer E, Klass D, Abrahamowicz M, Dauphinee D, Jacques A, Smees S, et al. Doctor scores on national qualifying examinations predict quality of care in future practice. *Med Educ.* 2009;43(12):1166-73. doi: 10.1111/j.1365-2923.2009.03534.x.
24. Görlich D, Friederichs H. Using longitudinal progress test data to determine the effect size of learning in undergraduate medical education – a retrospective, single-center, mixed model analysis of progress testing results. *Med Educ Online.* 2021;26(1):1972505. doi: 10.1080/10872981.2021.1972505.
25. Karay Y, Reiss B, Schaubert SK. Progress testing anytime and anywhere – does a mobile-learning approach enhance the utility of a large-scale formative assessment tool? *Med Teach.* 2020;42(10):1154-62. doi: 10.1080/0142159x.2020.1798910.
26. Romão GS, Schreiner L, Laranjeiras CLS, Bella ZIKJD, Coelho RA, Simões MDCR et al. Medical residency in gynecology and obstetrics in times of Covid-19: recommendations of the National Specialized Commission on Medical Residency of Febrasgo. *Rev Bras Ginecol Obstet.* 2020;42(7):411-4. doi: 10.1055/s-0040-1715147.
27. Brito LGO, Romão GS, Fernandes CE, Silva-Filho AL. Impact of Covid-19 on Brazilian medical residencies in obstetrics and gynecology. *Int J Gynaecol Obstet.* 2020;150(3):411-2. doi: 10.1002/ijgo.13283.
28. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ.* 2011;2:53-5. doi: 10.5116/ijme.4dfb.8dfd.
29. Souza AC, Alexandre NMC, Guirardello EB. Propriedades psicométricas na avaliação de instrumentos: avaliação da confiabilidade e da validade. *Epidemiol Serv Saúde.* 2017;26:649-59.



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.

Inserção do Teste de Progresso num curso de Medicina paulista: relato de uma experiência bem-sucedida

Insertion of the Progress Test in a São Paulo Medicine course: report of a successful experience

Helena Landim Gonçalves Cristóvão¹ hcristovao@prof.unisa.br
Júlio Cesar Massonetto¹ jmassonetto@prof.unisa.br
Luciane Lúcio Pereira¹ lpereira@unisa.br
Alba Regina de Abreu Lima² alba.lima09@gmail.com
Cíntia Leci Rodrigues¹ cintialeci@prof.unisa.br
Júlio César André² julio.andre@edu.famerp.br

RESUMO

Introdução: O Teste de Progresso no curso de Medicina (TPMed) tem sido utilizado pelas escolas médicas que implantaram mudanças curriculares, com currículos baseados/orientados na comunidade, aprendizagem baseada em problemas e currículos orientados por competências, entre outros, além de alguns programas de pós-graduação ou disciplinas isoladas. Baseado nessa premissa, o presente relato se propõe a descrever a experiência no processo de implantação do TPMed em nossa instituição e a participação estudantil nele.

Relato de experiência: Dada a importância do TPMed, o curso de Medicina da Universidade Santo Amaro (Unisa) se filiou ao Consórcio São Paulo I (SP1), composto pelas seguintes escolas de Medicina: ABC, Jundiaí, Catanduva, PUC Sorocaba, PUC Goiás e Unoeste, iniciando nossa experiência em 2018. A inserção da Unisa no TPMed foi estimulada a partir da participação no Congresso Brasileiro de Educação Médica (Cobem), com a subsequente associação ao Consórcio SP1. Inúmeras iniciativas de sensibilização do corpo docente e do corpo discente da instituição, com a colaboração da diretoria da Regional São Paulo da Associação Brasileira de Educação Médica (Abem), resultaram em 96,74% de participação discente no TPMed em 2018, 97,12% em 2019, 98,85%, em 2020 e 86,10% em 2021. Em relação ao desempenho dos estudantes, encontramos um aumento progressivo de acertos ao longo dos seis anos de graduação: 32,63% como média de acerto no primeiro semestre, e 62,87%, no 12º semestre.

Discussão: A influência positiva do TPMed na aprendizagem do aluno foi a motivação da Unisa para vencer os desafios e mobilizar os estudantes no primeiro ano em que aplicamos a prova, corroborando a grande participação deles. Houve uma queda da adesão ao Teste de Progresso Nacional da Abem – 2021 dos alunos da Unisa, tendo como possível causa os problemas técnicos ocorridos na aplicação dele, que foi on-line.

Conclusão: O efeito positivo do TPMed foi observado de forma clara na Unisa. O trabalho direcionado, enfatizando a relevância do TPMed, faz-se mister no seu reconhecimento, resultando em uma forte adesão dos nossos acadêmicos a essa avaliação.

Palavras-chave: Teste de Progresso; Educação Médica; Avaliação Educacional; Indicador; Participação das Partes Interessadas.

ABSTRACT

Introduction: The Medical Course Progress Test (TPMed) has been used by medical schools that have implemented curricular changes, with, for example, community-based/oriented curricula, problem-based learning and competency-oriented curricula, in addition to some graduate programs or isolated disciplines. Based on this premise, the present report proposes to describe the experience in the TPMed implementation process in our Institution and student participation in it.

Experience report: Given the importance of the TPMed, the Medicine course at Universidade Santo Amaro (UNISA) joined the SP1 Consortium, with our experience beginning in 2018. The inclusion of UNISA in the TPMed was driven by the school's participation in the Brazilian Congress of Medical Education (COBEM), and its subsequent association with the Consortium (SP1). Numerous initiatives to raise awareness of the institution's faculty and students, with the collaboration of the São Paulo Regional Board of the Brazilian Association of Medical Education (ABEM) resulted in 96.74% student participating in the TPMed in 2018, 97.12% in 2019, 98.85% in 2020 and 86.10% in 2021. Regarding student performance, we found a progressive increase in correct answers over the six years of the undergraduate course, 32.63% as an average score in the first semester and 62.87% in the twelfth semester.

Discussion: The positive influence of the TPMed on student learning was UNISA's motivation to overcome the challenges and mobilize the students in the first year in which the test was applied, corroborating their substantial participation. There was a drop in adherence to the ABEM National Progress Test – 2021 by UNISA students, possibly due to technical problems that occurred in its online application.

Conclusion: The positive effect of the TPMed was clearly observed in UNISA. This directed work, which emphasized the relevance of the TPMed, is essential for its recognition, and resulted in a strong adhesion of our academics in this evaluation.

Keywords: Progress Test; Medical Education; Educational Assessment; Indicators; Stakeholder Participation.

¹Universidade Santo Amaro, Santo Amaro, São Paulo, Brasil.

²Faculdade de Medicina de São José do Rio Preto, São José do Rio Preto, São Paulo, Brasil.

Editor: Valdes Roberto Bollela.

Recebido em 05/10/22; Aceito em 13/10/22.

Avaliado pelo processo de *double blind review*.

INTRODUÇÃO

A avaliação é uma atividade fundamental para o planejamento e desenvolvimento da educação. Essa atividade permite acompanhar o processo de apropriação e complexificação do conhecimento, avaliando a evolução de aprendizagens relevantes e significativas para o desenvolvimento de habilidades e competências dos alunos¹.

O Teste de Progresso (TPMed) tem por finalidade avaliar o desempenho cognitivo dos estudantes durante o curso e o próprio curso. A este, permite a análise da relação entre conteúdo e estrutura curricular da graduação e o desenvolvimento dos estudantes. Ao estudante, dá a oportunidade de verificar a evolução de seu desempenho cognitivo nas diversas áreas do curso, servindo como avaliação formativa e identificando problemas potenciais. Dessa forma, possibilita implementar ações para a melhoria contínua do estudante e do curso^{2,3}.

O TPMed tem sido utilizado pelas escolas médicas que implantaram mudanças curriculares, com currículos baseados/orientados na comunidade, aprendizagem baseada em problemas e currículos orientados por competências, entre outros, além de alguns programas de pós-graduação ou disciplinas isoladas. Essa técnica foi inicialmente desenvolvida, nos anos 1970, na Universidade de Missouri, no Kansas, e na Universidade de Limburg, na Holanda⁴.

Em 2000, o projeto *Canadian Medical Education Directions for Specialists* (CanMED) descreveu sete dimensões da competência para prática profissional que deveriam ser incluídos em todos os níveis da educação médica: especialista, comunicador, colaborador, gerente, defensor da saúde, comprometido com educação continuada e profissional⁵. Apesar dessa abordagem multidimensional de competência clínica, vários estudos sustentam que, na prática, ela é bidimensional, pois verdadeiramente os alunos são avaliados por duas dimensões substancialmente correlacionadas que, seguindo a taxonomia do Accreditation Council for Graduate Medical Education (ACGME), são denominadas: competência humanista – conhecimento médico, cuidado com o paciente e atenção baseada no sistema, englobando as habilidades de comunicação interpessoal e profissionalismo – e competência técnica^{6,7}.

Todas essas discussões repercutem na prática do ensino médico. Muitas universidades e órgãos de classe estão unindo esforços para repensar maneiras de sistematizar a avaliação dessas competências, acrescentando-lhe a necessária perspectiva formativa e indutora de novas práticas pedagógicas. A avaliação, como parte integrante do processo de ensino-aprendizagem, tem como funções: estabelecer a gama de conhecimentos e habilidades que um indivíduo precisa ter atingido num certo estágio de sua carreira, estimar de forma

precisa o nível dos alunos, oferecer aos alunos informações sobre seu progresso, alterar o comportamento em relação à aprendizagem, motivar os alunos, conscientizando-os sobre a importância do que estão aprendendo, e fornecer informações sobre a adequação do currículo aos professores e diretores^{8,9}.

O TPMed foi implantado no Brasil em 2005, nas seguintes instituições: Universidade Estadual de Londrina (UEL), Universidade Estadual de Campinas (Unicamp), Universidade Federal de Santa Catarina (UFSC), Universidade Estadual Paulista (Unesp), Universidade Federal de São Paulo (Unifesp), Faculdade de Medicina da Universidade de São Paulo (FMUSP), Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (FMRP-USP) e Fundação Universidade Regional de Blumenau (Furb), com expansão para 30 escolas em 2012 e para 80 escolas em 2015, com o incentivo do projeto da Associação Brasileira de Educação Médica (Abem), a partir de 2012¹⁰.

A parceria entre escolas tem contribuído para a confecção de questões com alta qualidade, além de combinar benefícios econômicos com vantagens educacionais globais, pois tira do processo de avaliação o efeito da endogenia e permite uma comparação salutar dos avanços e das limitações entre as instituições participantes, sempre evitando qualquer tipo de ranqueamento^{11,12}.

Um aspecto importante é que o TP não fornece informações apenas para autoavaliação do estudante, mas representa também uma valiosa ferramenta para a gestão acadêmica. Um TP com ampla abrangência das áreas que compõem o currículo de um curso, associado a uma abordagem de temas-chave para a formação do estudante durante a graduação, poderá fornecer importantes subsídios para a gestão curricular. Embora tenha sido concebido no curso de Medicina, o TP pode ser usado também em outros cursos, da área da saúde ou não, graças à sua lógica de construção e aplicação¹³.

No decorrer desses anos, o TPMed se difundiu por todo o Brasil: na Região Norte, nos estados do Amazonas e Pará; na Região Nordeste, com os consórcios interinstitucional nordeste, nordestino e baiano; na Região Centro-Oeste, com o consórcio Centro-Oeste; na Região Sudeste, com os consórcios Mineiro e Regional, o Núcleo Interinstitucional e Prática de Educação Médica, o São Paulo I e o Paulista II; e, finalmente, na Região Sul, com os consórcios Núcleo de Apoio Pedagógico Interinstitucional Sul (Napisul) e Rio Grande do Sul¹⁴.

Atualmente, a Abem oferece o TP a todos os cursos de Medicina a ela associados, correspondendo, aproximadamente, a 80% de todas as escolas médicas do país, estimativa realizada a partir de dados apresentados na plataforma Diretório das Escolas de Medicina (Direm-Br)¹⁵.

Baseado nessa premissa, o presente relato se propõe a descrever a experiência no processo de implantação do TPMed em nossa instituição e a participação estudantil nele.

RELATO DE EXPERIÊNCIA

O presente estudo se caracteriza como observacional, descritivo e retrospectivo que retrata a experiência com a introdução do TPMed no curso de Medicina da Universidade Santo Amaro (Unisa).

O instrumento de avaliação do TPMed é constituído por testes de múltipla escolha, com o propósito de verificar o ganho de conhecimento do estudante ao longo do curso de graduação. A mesma prova é aplicada aos seis anos, cujo conteúdo compreende uma amostra do conhecimento de todas as áreas que compõem o currículo do curso de Medicina e reflete os objetivos finais deste. Oferece ainda a oportunidade de acompanhar a evolução dos alunos ao longo do curso e comparar o desempenho deles com os demais discentes, de modo a fornecer *feedback* sobre a formação dos estudantes. Por meio do *feedback*, permite a análise da relação entre conteúdo e estrutura curricular da graduação e o desenvolvimento dos estudantes.

Para os docentes e a instituição, é uma ferramenta de diagnóstico sobre o ensino. Por meio dela, é possível implementar ações para a melhoria contínua do estudante e do curso que visem ao diagnóstico, ao prognóstico, às correções e às intervenções educacionais e à oportunidade de desenvolvimento docente.

Um docente da nossa instituição esteve presente no 55º Congresso Brasileiro de Educação Médica (Cobem), realizado em Porto Alegre, de 12 a 15 de outubro de 2017. Esse docente conheceu o TPMed e constatou a competência dessa ferramenta para a gestão acadêmica em nível institucional. O referido docente levou essa informação para conhecimento das instâncias superiores e para a instância máxima da Unisa, a reitoria, que de pronto reconheceu a importância e deferiu seu aval para a implantação dessa valiosa avaliação formativa em nossa instituição.

Devido à localização e ao perfil das escolas que compunham o Consórcio São Paulo I (SP1) e com a mediação do então diretor da Regional São Paulo da Abem, o docente da nossa instituição entrou em contato com a então coordenadora do Consórcio SP1, e, após reunião com os coordenadores locais de cada instituição, fomos acolhidos pelo consórcio, de forma que, no início de 2018, começamos os trabalhos relativos à aplicação do TPMed nesse mesmo ano.

Dada a importância do TPMed, o curso de Medicina da Unisa se filiou ao Consórcio SP1, composto pelas seguintes

escolas de Medicina: Faculdade de Medicina do ABC (FMABC), Faculdade de Medicina de Jundiaí (FMJ), Faculdade de Medicina de Catanduva (Fameca), Pontifícia Universidade Católica de São Paulo (PUC-SP), *campus* de Sorocaba, Pontifícia Universidade Católica de Goiás (PUC Goiás) e Universidade Estadual do Oeste do Paraná (Unoeste).

A partir da inclusão da nossa instituição no TPSP1, iniciamos a sensibilização da coordenação do curso de Medicina, de docentes e discentes por meio de palestras que o então diretor da Regional São Paulo da Abem ministrou, primeiramente aos coordenadores e docentes, e, oportunamente, no Congresso Acadêmico Médico da Unisa (Coacme), aos discentes. Vale ressaltar que foi citada a pontuação no programa de residência médica atrelada à participação do TPMed, em várias escolas médicas.

Dando continuidade à divulgação do TPMed, no início de maio de 2018, foi encaminhado um comunicado institucional da Unisa aos coordenadores e docentes, o qual informava as peculiaridades, com todas as etapas do processo e a data da nossa primeira participação: 3 de outubro de 2018. Concomitantemente, a coordenadora institucional local percorreu as salas de aula no intuito de esclarecer a importância dessa avaliação formativa e dirimir dúvidas dos estudantes. Na segunda quinzena desse mesmo mês, enviou-se um segundo comunicado institucional que solicitava a indicação dos coordenadores de área da Unisa para as funções de coordenadores auxiliares do TPMed, de cada área, assim como o anexo da matriz do TP de 2018, orientações sobre a elaboração de itens (questões de múltipla escolha) e a bibliografia.

Com a premissa de disseminarmos de uma forma mais objetiva o TPMed, elaborou-se um *flyer*, que disponibilizamos em todos os canais de comunicação e que pode ser visualizado na Figura 1.

Em 2018, tivemos uma grande participação dos nossos alunos. A Unisa ficou em terceiro lugar (96,74% de presença), comparada com as sete instituições participantes. Em relação ao desempenho dos estudantes, encontramos um aumento progressivo de acertos ao longo dos seis anos de graduação: 32,63% como média de acerto no primeiro semestre, e 62,87%, no 12º semestre.

Em 2019, tivemos a edição paulista do TP, também com grande participação dos nossos alunos, e a Unisa ficou em 12º lugar (97,12% de presença), comparada com as 43 instituições de ensino superior (IES) participantes, sendo 38 do estado de São Paulo e cinco de outros estados da Federação (Goiás, Paraná, Minas Gerais e Santa Catarina), agremiadas em quatro consórcios, num universo de 22.396 alunos.

Figura 1. Flyer institucional utilizado para a divulgação da primeira participação do TPN da Uni-sa, 2018

UNISA
Universidade de Santo Amaro

03/10/2018
13H
CAMPUS I DA UNISA
TESTE DE PROGRESSO

O QUE É O TESTE DE PROGRESSO?
É uma avaliação **CONTINUADA** (ocorre todo ano) de caráter **FORMATIVO**, que avalia o **DESEMPENHO COGNITIVO**, mensurando o **GANHO DE CONHECIMENTO** ao longo do curso. A **PROVA**, com 120 questões de **MÚLTIPLAS ESCOLHA**, divididas em **CIÊNCIAS BÁSICAS, SAÚDE COLETIVA, CLÍNICA, CIRURGIA, GINECOLOGIA E OBSTETRÍCA** e **PEDIATRIA**, possui ainda questões de **ÉTICA E BIOÉTICA** permeando cada grande área. A mesma prova é aplicada para todos os anos/semestres do curso e é construída baseada numa **MATRIZ DE CONTEÚDOS** essenciais para o **MÉDICO GENERALISTA**.

O Curso de **MEDICINA** da UNISA realizará pela primeira vez na sua história o **TESTE DE PROGRESSO (TP)**.
A UNISA pertence hoje ao consórcio para realização do TP onde figuram as seguintes escolas médicas:
Faculdade de Medicina do ABC / Faculdade de Medicina de Catanduva / UNOESTE / PUC Sorocaba / PUC Campinas e PUC Goiás.

Não APROVA ou REPROVA ninguém.
Os resultados são **PESSOAIS: SÓ VOCÊ TEM ACESSO** a partir de um Login e Senha próprios.

Não perca essa oportunidade de verificar a EVOLUÇÃO de seu conhecimento cognitivo, nas diversas áreas do curso, e ajudando-o a identificar problemas potenciais.

1º ano 2º ano 3º ano 4º ano 5º ano 6º ano

Insucesso
Sucesso

Fonte: Elaborada pelos autores.

Em 2020, obtivemos uma porcentagem de presença na ordem de 98,85%, demonstrando um aumento crescente, nos últimos anos, da participação dos nossos alunos. A Unisa ficou em segundo lugar (98,85% de presença), comparada com as sete IES participantes.

Já em 2021, tivemos uma queda na adesão do corpo discente à participação do TP, com uma porcentagem de 86,10%, ocupando o sétimo lugar, em relação ao nosso consórcio (SP1).

DISCUSSÃO

Em 2015, foi organizado o primeiro Teste de Progresso Nacional (TPN), com a participação do grupo de coordenação do TP da Abem, como atividade de encerramento do projeto "Abem 50 anos". Isso constituiu um momento importante na construção de um teste nacional único, de modo a indicar um horizonte cada vez mais inclusivo e representativo, e apontar o caminho para uma estratégia capaz de induzir e transformar o ensino e a cultura de avaliação no país¹⁶.

A prova nacional foi realizada no dia 30 de setembro de 2015, com a participação de 58 escolas, com 23.065 estudantes que tiveram seus gabaritos analisados. A análise do desempenho

dos estudantes mostrou o esperado aumento progressivo de acertos ao longo dos seis anos de graduação, com a média inicial do primeiro ano de 32,38%, atingindo 61,28% no sexto ano¹⁵, porcentagens semelhantes às encontradas em nosso relato de experiência nas aplicações do TP em nossa instituição em 2018: 32,63% como média de acerto no primeiro semestre, e 62,87%, no 12º semestre. Quando se avaliou a progressão de conhecimentos entre o momento de entrada do estudante no curso (32%) e o término, na sexta série (54%), verificou-se, no estudo de Pinheiro et al.¹³, uma incorporação de conhecimentos de 22%. Esses testes foram aplicados aos estudantes de todas as séries do curso médico do ano de 2008 (n = 457) e 2011 (n = 446), em um mesmo dia e horário, com duração de quatro horas e permanência obrigatória de duas horas na sala. Os dados relatados também estão próximos aos encontrados em nossa instituição.

Os resultados encontrados na edição do TPN da Abem refletem o que já se identificava em relação ao processo de aprendizagem ao longo de seis anos do curso de Medicina: o contínuo aumento do desempenho dos discentes na avaliação. Verificou-se que o incremento no desempenho dos estudantes

no TPN da Abem de 2021 foi de 29,6 pontos percentuais. Resultados semelhantes foram observados na edição nacional de 2015, momento em que se identificou um incremento de 28,9 pontos percentuais na média de acertos entre os acadêmicos do primeiro ao sexto ano. Além disso, as médias de acertos em cada ano do curso no TPN da Abem de 2021 foram semelhantes ao resultado de 2015¹⁰. O efeito positivo do TPMed pode ser observado de forma clara na nossa instituição, em consonância com a literatura que descreve o crescimento do conhecimento médico nos que se abriram para acolher a proposta desse método de avaliação.

Em relação à participação dos alunos no TPMed, foram divulgados, em uma publicação do colegiado de Medicina da UEL no período de 1998 a 2002, os resultados relativos aos desempenhos gerais dos estudantes em todos os TPMed e o desempenho geral e por área de conhecimento do oitavo e nono TPMed. Verificou-se um aumento progressivo na participação dos estudantes nos TPMed, com exceção do terceiro TPMed, que teve uma participação total de 31%¹⁶.

A influência positiva do TPMed na aprendizagem do aluno e um incentivo de acréscimo de 0,5 ponto em um módulo, escolhido pelo estudante, caso participasse da prova e a realizasse por um período mínimo de três horas, foram as motivações da Unisa para vencer os desafios e mobilizar os estudantes no primeiro ano em que aplicamos a prova, corroborando a grande participação discente. Acreditamos que todas essas ações relatadas possam ter contribuído para uma maior motivação dos nossos alunos na aderência à realização dessa avaliação em 2018, com 96,74% de presença, o que vem ao encontro dos resultados do estudo de Sakai et al.¹⁶ quanto ao comparecimento dos estudantes nos TPMed, na Medicina da UEL, de 95,6%, em 2006, em que houve um aumento da participação dos alunos, com a obrigatoriedade do teste a partir de 2003 e a instituição de um prêmio no nono TPMed/2006 à série com o maior número de comparecimentos.

Em 2019, na edição paulista do TP, tivemos uma adesão ainda um pouco maior, comparada ao ano anterior, com divulgação em salas de aula sobre a importância dessa avaliação e esclarecimento das vantagens, como a oportunidade de acompanhar a evolução do aluno ao longo do curso, comparar o seu desempenho com os demais discentes e verificar a evolução de seu desempenho cognitivo nas diversas áreas do curso, servindo como avaliação formativa e identificando problemas potenciais, o que pode ter auxiliado a adesão ao TPMed.

Já em 2020 nos deparamos com a pandemia da *coronavirus disease 2019* (Covid-19), quando discutimos em nosso consórcio a possibilidade de realizarmos o TPMed 2020 *on-line*. Ele foi agendado para o dia 7 de outubro de 2020 às 13h30, porém, nessa data, às 10h30 o *site* começou a apresentar instabilidades

no acesso por motivos técnicos desconhecidos, tendo sido solicitado um parecer técnico da empresa onde o *site* estava hospedado. Recebemos, então, um comunicado de adiamento da aplicação da prova até que a falha fosse solucionada, para que não houvesse prejuízo aos alunos. Mantendo o compromisso assumido com os discentes, a prova foi realizada em novembro, com uma porcentagem de participação dos discentes maior em relação aos anos anteriores.

Tivemos esse mesmo problema com o segundo TPN de 2021, promovido pela Abem, e a prova, que inicialmente havia sido agendada para o dia 6 de outubro de 2021, foi cancelada e remarcada para o dia 4 de novembro de 2021, às 13h30. Todas as escolas de Medicina do Brasil, que aderiram ao TPMed de 2021, tiveram a oportunidade de participar dessa avaliação, realizada *on-line*. A participação dos estudantes do curso de Medicina da Unisa foi menor, se comparada aos anos anteriores. Apesar de todo o suporte da Abem aos estudantes, inclusive no dia da prova, e da Fundação para o Vestibular da Universidade Estadual Paulista (Vunesp), por meio do *chat* do sistema, telefone e *e-mail*, tivemos problemas com alguns dispositivos eletrônicos, havendo atraso no início da prova, com prorrogação do horário de acesso, para que nenhum aluno fosse prejudicado, mas ainda assim alguns estudantes referiram que não conseguiram acessar a prova. Apesar da assessoria prestada aos alunos com dificuldades de acesso, acreditamos que a queda da adesão ao TPN da Abem de 2021 dos discentes da Unisa ocorreu por conta dessas questões técnicas.

Entre o momento de cancelamento da primeira avaliação e a efetivação da prova no segundo agendamento, houve a desistência da participação de 15 escolas, motivada por desafios de conciliação com o calendário letivo, e agregaram-se duas novas escolas, que não estavam participando no primeiro momento. Assim, o TPN da Abem de 2021 se concretizou com a participação de 130 escolas médicas (37,3% do total de escolas brasileiras e 47,5% das associadas à Abem) e a inscrição de 67.871 acadêmicos, dos quais 48.946 realizaram a prova. A taxa de absenteísmo de 27,9% foi semelhante à da edição presencial de 2015, que contabilizou 23,6% de ausências¹⁰.

CONCLUSÃO

A adesão significativa dos acadêmicos ao TPMed foi ampliada pelo incentivo do acréscimo de 0,5 ponto em um módulo, escolhido pelo aluno, caso participasse da prova e a realizasse por um período mínimo de três horas. Entretanto, ainda há a necessidade de estratégias de sensibilização nos meios discente e docente para aumentar a participação dos estudantes.

O desempenho geral observado encontra-se dentro do esperado e apontado na literatura. Houve um aumento

progressivo de acertos ao longo dos seis anos de graduação, em relação à média inicial do primeiro semestre e à média atingida no 12º semestre.

O impacto do TPMed na aprendizagem dos alunos é geralmente considerado profundo, de sorte que a avaliação por meio dele pode muito bem ser uma das ferramentas mais poderosas que temos à disposição para influenciar a aprendizagem dos estudantes e garantir a construção de um exímio profissional.

O trabalho direcionado, enfatizando a relevância do TPMed, faz-se mister no seu reconhecimento, de modo a resultar em uma forte adesão dos nossos acadêmicos a essa avaliação.

A partir da experiência vivenciada, tem-se como perspectiva futura a reflexão sobre a percepção dos envolvidos, dos dirigentes institucionais, do corpo docente e do corpo discente, com vistas a levantar aspectos que possam consolidar a participação da IES no TPMed.

CONTRIBUIÇÃO DOS AUTORES

Helena Landim Gonçalves Cristóvão participou da concepção do relato, da coleta e análise dos dados, e da redação e da revisão do manuscrito. Alba Regina de Abreu Lima participou da concepção do relato e da revisão do manuscrito. Júlio César André, Júlio César Massonetto e Luciane Lúcio Pereira participaram da concepção do relato, da redação e da revisão do manuscrito. Cíntia Leci Rodrigues participou da coleta e análise dos dados, e da revisão do manuscrito.

CONFLITO DE INTERESSES

Declaramos não haver conflito de interesses.

FINANCIAMENTO

Declaramos não haver financiamento.

REFERÊNCIAS

1. Ferreira ARL, Weyh CB. Tecnologia da informação no processo da avaliação como estratégia para qualificar a prática docente. *Anais do Congresso Internacional de Educação e Tecnologias*; 2018; São Carlos. São Carlos: UFSCar; 2018. p. 1-7.
2. van der Vleuten CPM, Verwijnew GM, Wijnen WHFW. Fifteen years of experience with progress testing in a problem-based learning curriculum. *Med Teach*. 1996;18(2):103-9. doi: 10.3109/01421599609034142.
3. Ito K, Janowski E. Resultados dos testes de progresso da medicina realizados no período de 1998 a 2002. Londrina: Colegiado de Medicina da Universidade Estadual de Londrina; 2003.
4. Blake JM, Norman GR, Keane DR, Mueller CB, Cunnington J, Didyk N. Introducing progress testing in McMaster University's problem-based medical curriculum: psychometric properties and effect on learning. *Acad Med*. 1996;71(9):1002-7. doi: 10.1097/00001888-199609000-00016.
5. Prideaux D, Alexander H, Bower A, Dacre J, Haist S, Jolly B, et al. Clinical teaching: maintaining an educational role for doctors in the new health care environment. *Med Educ*. 2000;34(10):820-6. doi: 10.1046/j.1365-2923.2000.00756.x.
6. DiMatteo MR, DiNicola DD. Sources of assessment of physician performance: a study of comparative reliability and patterns of intercorrelation. *Med Care*. 1981;19(8):829-42. doi: 10.1097/00005650-198108000-00003.
7. Silber CG, Nasca TJ, Paskin DL, Eiger G, Robeson M, Veloski JJ. Do global rating forms enable program directors to assess the ACGME competencies? *Acad Med*. 2004;79(6):549-56. doi: 10.1097/00001888-200406000-00010.
8. Haydt RC. Avaliação do processo ensino-aprendizagem. São Paulo: Ática; 2002.
9. Hulsman RL, Mollema ED, Hoos AM, de Haes JC, Donnison-Speijer JD. Assessment of medical communication skills by computer: assessment method and student experiences. *Med Educ*. 2004;38(8):813-24. doi: 10.1111/j.1365-2929.2004.01900.x.
10. Oliveira SS, Postal EA, Afonso DH, Merse CE, Cyrino EG, de Abreu Junior AF, et al. Teste de Progresso da Abem: consolidando uma estratégia de avaliação para o ensino médico. *Rev Bras Educ Med*. 2022;46(1):e061. doi: 10.1590/1981-5271v46.1-EDITORIAL.
11. van der Vleuten CP, Schuwirth LW, Muijtjens AM, Thoben AJ, Cohen-Schotanus J, van Boven CP. Cross institutional collaboration in assessment: a case on progress testing. *Med Teach*. 2004;26(8):719-25. doi: 10.1080/01421590400016464.
12. Friedman-Ben DM, Hunter I, Harden R. Introduction of a progress test as a tool for defining core undergraduate curriculum. *Educ Med*. 2001;4:97-9.
13. Pinheiro OL, Spadella MA, Moreira HM, Ribeiro ZMT, Guimarães 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. doi: 10.1590/1981-52712015v39n1e02182013.
14. Hamamoto PT, Bicudo AM. Implementation of the Brazilian National Network for Practices and Research with Progress Testing – BRAZ-NPT. *Rev Bras Educ Med*. 2020;44(3):e074. doi: 10.1590/1981-5271v44.3-20200089.
15. Bicudo AM, Hamamoto PT, Abbade JF, Hafner MDLMB, Maffei CML. Teste de Progresso em consórcios para todas as escolas médicas do Brasil. *Rev Bras Educ Med*. 2019;43:151-6. doi: 10.1590/1981-52712015v43n4RB20190018.
16. 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.



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.

Teste de Progresso como ferramenta de intervenção curricular na implementação de um curso de Medicina

Progress Test as a curricular interventional tool in the implementation of a medical course

Elisângela de Mattos Silva¹ elisangela.silva@professor.edu.br

Lívia Sissi Gonçalves Souza Piechnik¹ livia.piechnik@fpp.edu.br

Mariana Xavier e Silva¹ mariana.silva@fpp.edu.br

Carlos Mauricio Santos¹ carlos.santos@fpp.professor.edu.br

Izabel Cristina Meister Martins Coelho¹ izabel.coelho@fpp.edu.br

RESUMO

Introdução: O Teste de Progresso (TP) é uma modalidade de avaliação longitudinal composta por questões de múltipla escolha que permite a avaliação do estudante com realização do feedback individual e a análise do processo de ensino-aprendizagem institucional. O presente estudo relata a experiência do uso do TP como ferramenta de gestão e aprendizado na implementação de um curso de Medicina de currículo integrado e baseado em metodologias ativas.

Relato de experiência: Desde 2014, o curso de Medicina de uma capital do país utiliza o TP como ferramenta de avaliação para estimular a autonomia do estudante, baseado em seu desempenho longitudinal, possibilitar ao docente a reflexão e ação adequadas ao processo de ensino-aprendizagem e fornecer elementos significativos para a tomada de decisão institucional relacionada à adequação curricular e à implementação do curso. Com diversas estratégias, como a formação de uma Comissão de Avaliação do Estudante e do Curso (Caec), oficinas com professores e estudantes e consórcios interinstitucionais, os atores que compõem o curso refletem seu papel na consolidação do conhecimento para o futuro profissional médico e se engajam nesse processo em diferentes momentos e de diferentes formas.

Discussão: O TP é uma ferramenta com potenciais e limites que devem considerar a segurança, a confiabilidade e validade da avaliação, utilizando-se de estratégias qualificadoras, incluindo o ambiente virtual. O TP possibilita a autodeterminação e o autoaperfeiçoamento do estudante, do docente e do curso ao estabelecer comparações entre médias de desempenho longitudinais dos discentes em diversos estágios de aprendizado e instituições.

Conclusão: A experiência demonstrou a importância do TP como ferramenta de avaliação e trouxe vantagens e benefícios na implantação do curso e orientação de intervenções para a melhoria do processo de ensino-aprendizagem.

Palavras-chave: Avaliação Educacional; Aprendizagem; Aquisição de Conhecimento.

ABSTRACT

Introduction: The Progress Test (PT) is a longitudinal assessment modality composed of multiple-choice questions and allows the assessment of the student with individual feedback and the analysis of the institutional teaching-learning process. The present study aims at reporting the experience of using the progress test as a management and learning tool in the implementation of a medical course with an integrated curriculum based on active methodologies.

Experience report: Since 2014, a medical school in a capital of the country uses the PT as an assessment tool to stimulate student autonomy, based on their longitudinal performance, allowing the teacher to reflect on and take the appropriate actions aimed at the teaching and learning process and providing significant elements for institutional decision-making related to curricular adequacy and course implementation. Using several strategies, such as the formation of a student and course evaluation committee (CAEC), workshops with teachers and students and inter-institutional consortia, the actors that make up the course reflect on their role in consolidating knowledge for the future medical professional and engage in this process at different times and in different ways.

Discussion: The progress test is a tool with potentials and limits that must consider the safety, reliability and validity of the evaluation, using qualifying strategies, including the virtual environment. The PT enables the students, the faculty and the course self-determination and self-improvement by establishing comparisons between longitudinal performance averages of students at different stages of learning and institutions.

Conclusion: The experience demonstrated the importance of the PT as an evaluation tool, brought advantages and benefits in the course implementation and guidance on interventions aimed at improving the teaching-learning process.

Keywords: Educational Evaluation; Learning; Knowledge Acquisition.

¹Faculdades Pequeno Príncipe, Curitiba, Paraná, Brasil.

Editor: Valdes Roberto Bollela.

Recebido em 19/09/22; Aceito em 15/11/22.

Avaliado pelo processo de *double blind review*.

INTRODUÇÃO

O Teste de Progresso (TP) é uma avaliação longitudinal realizada periodicamente por meio de questões de múltipla escolha (QME)¹. Foi introduzido nos anos 1970 pelas Universidades de Maastricht e Missouri num contexto de reflexão e mudança das metodologias de ensino e de avaliação^{2,3}.

Desde então, o TP foi incorporado a um número crescente de escolas em todos os continentes^{2,3}. Em 2014, a Associação Brasileira de Educação Médica (Abem) implementou ações de fortalecimento da educação médica de modo a ampliar o TP⁴. Em 2019, 40% das escolas médicas realizaram essa avaliação, e atualmente em torno de 200 instituições participam por meio de 18 núcleos ou consórcios interinstitucionais no Brasil⁵.

Realizado anualmente, o TP é composto por 120 QME divididas em cinco grandes áreas (clínica, pediatria, cirurgia, tocoginecologia e saúde pública). Trata-se de questões que integram conhecimentos básicos e aplicados, incluindo bioética³.

Dentro de uma nova filosofia educacional, a *problem-based learning* (PBL) surge como proposta de metodologia ativa que reestrutura o processo de ensino-aprendizagem em várias escolas, com foco no estudante ativo e participativo^{2,6}. Além disso, a metodologia de avaliação torna-se parte do processo e precisa ser repensada de modo a proporcionar elementos para o autoaperfeiçoamento do estudante^{5,6}.

As avaliações por meio de QME sempre foram amplamente utilizadas durante a graduação. As QME permitem a elaboração de um grande número de questões que abrangem diferentes conteúdos, podem ser administradas em um período relativamente curto de tempo e possibilitam a correção dos gabaritos com o auxílio de programas computacionais^{1,6}. Apesar de esse tipo de avaliação ser um método limitado, principalmente no que se refere a habilidades e atitudes, é um componente crítico do processo de ensino-aprendizagem e, quando usado corretamente, pode ter um importante papel para atingir os objetivos curriculares desejados^{3,6}.

Entretanto, o TP possui características distintas da avaliação habitual por meio de QME: utiliza questões relacionadas aos objetivos finais de aprendizagem do curso que são aplicadas a todos os estudantes do programa desde o primeiro ano, de forma igualitária e em intervalos regulares durante todo o curso. Com essas características, o TP permite uma avaliação mais aprofundada em relação ao ganho de conhecimento de forma longitudinal durante toda a formação e a aquisição dos objetivos de aprendizagem do curso^{2,3}.

Além disso, a forma como o *feedback* é executado também preenche as necessidades em relação à percepção de lacunas e dificuldades e serve como instrumento para o docente, o discente e a instituição estabelecerem estratégias de condução do processo de ensino-aprendizagem^{3,6}.

O TP consiste em uma ampla ferramenta que cumpre as funções diagnóstica e formativa⁵. Embora ainda não seja frequente, pode ser utilizada também como avaliação somativa, por apresentar o resultado em porcentagens ajustadas de acordo com o período do curso e as competências esperadas para o estágio⁴.

O TP vem sendo utilizado como importante ferramenta de análise do currículo, objetivos de aprendizagem, avaliação docente e institucional quando os dados são analisados comparativamente a outras instituições e de maneira contínua^{3,4}.

Objetivo

Este artigo relata a experiência do uso do TP como ferramenta de gestão e aprendizado na implementação de um curso de Medicina de currículo integrado e baseado em metodologias ativas.

METODOLOGIA

Trata-se de um relato de experiência desenvolvido com professores e alunos do curso de Medicina das Faculdades Pequeno Príncipe (FPP), na cidade de Curitiba, no Paraná.

O curso de Medicina das FPP iniciou sua primeira turma em julho de 2014. Em seu planejamento, apresentava como base o uso exclusivo de metodologias ativas de ensino (incluindo principalmente a PBL) em diferentes unidades curriculares que compunham um currículo integrado.

Os temas e as competências de aprendizagem foram definidos pelas Diretrizes Curriculares Nacionais (DCN) e distribuídos ao longo de seis anos, sendo os dois últimos dedicados ao internato (estágio curricular obrigatório para os cursos de Medicina).

Em 2015, a instituição participou pela primeira vez do TP com as três turmas (150 estudantes) existentes na ocasião. A partir disso, a instituição passou a participar todos os anos sempre com número crescente de estudantes que foram sendo incorporados até completar os seis anos do curso. Anualmente, a instituição indica diferentes professores que atuam na etapa de elaboração de QME e participam das oficinas de avaliação e seleção das questões e de qualificação do teste, enquanto os discentes são constantemente estimulados a aderir à proposta.

Em 2020, não houve TP por conta da situação pandêmica. Em 2021, a instituição participou do TP Nacional, realizado em plataforma *on-line* por causa da necessidade de manter o distanciamento social imposto pela pandemia da *coronavirus disease 2019* (Covid-19), com consórcios de todo o país.

A FPP participa do Núcleo de Apoio Pedagógico Interinstitucional Sul II (Napisul II), composto por 13 instituições, entre públicas e privadas, com metodologias tradicionais ou ativas, do Paraná e de Santa Catarina.

As atividades a serem relatadas referem-se a todas as etapas desde o início do curso, nas quais se utilizou o TP para auxiliar na gestão, implementação e avaliação do curso de Medicina.

• *Etapa 1 – Avaliações do TP como ferramenta avaliativa na Caec:* Desde o início, o processo avaliativo foi uma preocupação constante na instituição, motivo pelo qual se criou já no primeiro semestre do curso uma Comissão de Avaliação do Estudante e do Curso (Caec). O intuito da Caec é promover estudos, oficinas e educação continuada e permanente para todo o corpo docente em relação a diversas formas de avaliação, assim como estabelecer os padrões de avaliação institucional do curso. Assim, além das questões avaliativas institucionais internas, o TP também passou a ser acompanhado e interpretado por essa comissão. Os docentes pertencentes à Caec participam do processo de elaboração do TP em parceria com o consórcio, organizam a execução do teste dentro da instituição com conscientização dos estudantes e posteriormente realizam a análise dos resultados e gráficos gerados.

• *Etapa 2 – Atividades da semana pedagógica dos docentes:* Palestras, oficinas e grupos de trabalho relacionados ao TP são realizados desde então no início de cada semestre, durante a semana pedagógica, para treinamento continuado e melhoria na qualidade das diferentes formas de avaliação utilizadas pela instituição. Além disso, os docentes discutem e analisam os ajustes dos objetivos de aprendizagem e planejamento das atividades a serem desenvolvidas em cada unidade curricular.

• *Etapa 3 – Adesão dos estudantes:* A estratégia para melhorar a adesão dos discentes ao longo dos anos envolve lideranças estudantis e docentes que são convidados a participar da divulgação e sensibilização dos pares, em mídias sociais, eventos e abordagem direta em salas de aula de todos os semestres da graduação. As devolutivas dos resultados obtidos pela instituição em comparação à média das demais instituições do consórcio são fornecidas anualmente aos estudantes como forma de motivar a participação deles. A instituição passou a bonificar também com pontuação na unidade curricular vigente aqueles estudantes que participam do TP e com pontuação mais expressiva os que ficam com média acima da média de sua turma. Dessa forma, temos tido adesão significativa dos estudantes (Tabela 1), o que contribui para melhor análise dos resultados.

• *Etapa 4 – Utilização das questões do TP em outras etapas avaliativas:* Em nossa instituição, as questões do TP são utilizadas no período pré-internato, nas avaliações teóricas curriculares, em provas de múltipla escolha. Para o internato, o TP passou a ter caráter também somativo desde 2018. A avaliação teórica é realizada no segundo semestre por meio do TP Externo – promovido pelo consórcio Napisul 2 – e no primeiro semestre por meio de um TP Interno – planejado, elaborado e aplicado

por professores e supervisores que compõem o corpo docente do internato. A partir da primeira turma do internato, essa avaliação passa a fazer parte do portfólio de ferramentas que compõem a nota final do internato.

• *Etapa 5 – Ajustes curriculares baseados no TP:* A partir da análise das questões, das oficinas com docentes e das devolutivas com docentes e discentes, algumas alterações e adequações foram feitas no currículo.

RESULTADOS E DISCUSSÃO

A etapa 1 foi um importante passo para a aquisição das habilidades necessárias para que o TP pudesse efetivamente ser utilizado como ferramenta de gestão auxiliar na implantação do curso.

Nessa etapa, os resultados do TP são amplamente discutidos com estudantes em várias instâncias colegiadas e com os docentes de modo a analisar os resultados e planejar as intervenções necessárias para possibilitar o alcance dos objetivos de aprendizagem previstos pelas DCN e definidos no currículo do curso.

A análise dos resultados e gráficos é feita por equipe formada por docentes representantes do TP, pela coordenação do curso e pelos estatísticos. Por meio dessas análises dos resultados, a instituição acompanha o ganho de conhecimento dos estudantes ao longo do curso que tem se mostrado crescente. Essa análise é feita por meio de gráficos comparativos entre os períodos e entre as médias dos mesmos períodos de outras instituições do consórcio.

O modelo de currículo utilizado pela instituição proporciona um desafio e uma grande oportunidade para a elaboração do TP, já que prescinde da integração dos temas e da abordagem ampla que interligue objetivos de aprendizagem de diversos estágios do curso.

Assim, essa etapa tem se mostrado uma base de apoio importante para a gestão do curso, sendo o TP uma das ferramentas utilizadas por essa comissão.

Os grupos de trabalhos e as oficinas realizadas na etapa 2 habilitam os docentes que fazem uma análise crítica das questões e com isso conseguem fornecer um *feedback* formativo aos discentes, com explicações e esclarecimentos de dúvidas, contribuindo de forma importante para uma aprendizagem significativa. A abordagem das questões nos momentos de aula é sistematicamente utilizada como estratégia de ensino e bastante valorizada, já que são elaboradas a partir de uma perspectiva construtivista que integra conhecimentos e se alinha à proposta curricular do curso.

Nas oficinas, um *feedback* sobre as questões e os resultados também é fornecido de forma rotineira para

professores e estudantes. Esse *feedback* individual e coletivo é realizado no sentido de orientar ações que possam ajudar docentes e discentes a alcançar os objetivos de aprendizagem, de modo a reforçar temas de questões com desempenho inadequado dos estudantes e adequar os conteúdos de cada unidade curricular.

Por meio dessa experiência, o corpo docente passou a ter consciência da importância do processo avaliativo e engajou-se em treinamentos e construções com o olhar crítico para a elaboração de questões, além de dar a devida atenção aos objetivos da docência de proporcionar o alcance dos objetivos finais de aprendizagem do curso de Medicina.

Os aprendizados advindos da elaboração do TP, da participação em oficinas de formação e da qualificação de questões refletiram positivamente nas avaliações internas do curso. À medida que a experiência dos professores aumentava, tornava-se mais fácil a execução de avaliações com maior qualidade e capacidade discriminatória ao longo do semestre. Esse fato foi evidenciado na análise da qualidade das questões realizada por meio do método Angoff, que consiste em uma classificação quanto ao grau de dificuldade e distribuição na prova, e em uma classificação quanto ao índice de acerto e grau de discriminação dos itens elaborados.

Por esse motivo, as oficinas de elaboração de QME tornaram-se rotina anual para o desenvolvimento docente, nas quais se incluíram a elaboração, a discussão e a avaliação das assertivas e questões a partir de objetivos de aprendizagem, além de exercícios formativos de *feedback* e autoavaliação, baseados em evidências científicas.

Oficinas de avaliação de edições anteriores do TP também são realizadas anualmente para avaliação dos objetivos das unidades curriculares. Oficinas de *blueprinting* associadas a elas foram realizadas para análise do currículo.

Assim, na etapa 2, as oficinas, os grupos de trabalho e as palestras relacionadas ao TP serviram para consolidar e formalizar a utilização ampla do instrumento na implantação do curso.

Na etapa 3, a estratégia consistiu na ampla divulgação do TP e no envolvimento direto dos professores, de modo a sensibilizar os alunos a participar do TP, ressaltar os benefícios diretos e indiretos ao estudante e enfatizar o uso do TP como ferramenta dinâmica de construção de conhecimentos durante a formação acadêmica.

Os resultados obtidos com as abordagens realizadas na etapa 3 geraram uma adesão dos estudantes que em média ultrapassa 90%, conforme pode ser observado na Tabela 1, e conferiram uma confiabilidade tal ao TP que permite a compreensão profunda sobre a aquisição e retenção do conhecimento por parte dos estudantes, e valida a tomada de decisão que nele se baseia.

Assim, o engajamento dos estudantes reflete o compromisso institucional com o aprendizado deles. Os discentes reconhecem esse esforço e participam ativamente do processo avaliativo em parceria com os docentes, pois o TP promove a melhoria do aprendizado individual e o fortalecimento do ensino no curso.

Na etapa 4, a utilização de questões do TP em avaliações teóricas e de todo o teste como avaliação semestral no internato foi de notada importância na adesão dos estudantes.

Tabela 1. Adesão dos estudantes ao Teste de Progresso

Adesão	2017	2018	2019	2020	2021	2022
1º período	97,9	98,1	88,2	-----	88	100
2º período	100	96,2	64	-----	68	84,6
3º período	72,9	97,8	89,1	-----	94	96,2
4º período	96,6	94	98	-----	80	76,6
5º período	95,7	71,4	90,9	-----	94	100
6º período	93,6	91,5	81,3	-----	75	91,6
7º período	86,3	83,3	66	-----	90	90
8º período	-----	95,9	85,4	-----	96	60,7
9º período	-----	100	100	-----	100	100
10º período	-----	-----	100	-----	100	100
11º período	-----	-----	97,9	-----	100	100
12º período	-----	-----	-----	-----	100	98
TOTAL	91,9	92	87,3	-----	91	91,6

Fonte: Elaborada pelos autores.

Temos com isso a adesão de 100% dos estudantes do internado (quinto e sexto anos) e uma excelente adesão dos demais, o que confere maior confiabilidade dos resultados na análise institucional. Essa etapa motivou também os docentes a se familiarizar com o teste para que possam utilizar as questões em suas avaliações. Os docentes participaram ativamente dos debates referentes às melhorias do currículo com base nos resultados das questões.

Na etapa 5, realizaram-se ajustes importantes relacionados à forma e ao momento em que cada tema era abordado. Esses temas foram indicados pelas comissões e implementados nos primeiros anos do curso para sua qualificação. Por exemplo, identificou-se como falha a abordagem da dengue e das emergências oftalmológicas no currículo, e ambos os assuntos foram revistos e reforçados nas unidades curriculares pertinentes.

O desempenho dos estudantes, aquém do desejado em alguns períodos e abaixo da média do consórcio, nas áreas de pediatria e ginecologia/obstetrícia nos anos de 2016 e 2017, também foi motivo de reestruturação e revisão da linha do tempo dessas áreas. O ensino clínico de pediatria foi inserido no quinto período do curso, e reorganizaram-se alguns temas de ginecologia/obstetrícia.

Na área cirúrgica, alguns temas estavam sendo abordados apenas precocemente no currículo durante o segundo período e necessitaram de reforço dos conceitos e da aplicabilidade clínica mais no final do curso, como a classificação de feridas operatórias e cirurgias quanto ao grau de contaminação. A punção de acesso venoso central também foi um tema que ganhou atenção não somente do ponto de vista teórico, mas também o treinamento dessa habilidade em simuladores foi reforçado após discussões realizadas nas oficinas docentes.

Os bons resultados dos discentes da instituição, quando comparados com as médias das demais instituições do consórcio, comprovam a consolidação do conhecimento dos estudantes ao longo do curso e reforçam as fortalezas da metodologia e do currículo integrado em promover o alcance dos objetivos de aprendizado no curso.

Até o momento, pôde-se observar que a complexidade do processo de ensino-aprendizagem se expressa na complexidade dos currículos, dos processos de ensino e de avaliação. Portanto, é importante reforçar que o processo de avaliação deve ser capaz de diagnosticar a aquisição de competências e permitir a tomada de decisão de modo a potencializar o aprendizado. Diante do uso de metodologias ativas, a avaliação deve permitir aos atores desse contexto a

adoção de novas estratégias de ensino-aprendizagem⁵.

Embora o TP seja um modelo limitado³, é passível de automação⁶, como ocorreu em 2021 por força da pandemia da Covid-19, e carece de um cuidado ampliado em sua preparação que garanta a sua confiabilidade e validade. Oliveira et al.⁷ apresentam uma série de adaptações que foram realizadas na preparação para o TP Nacional, com especial atenção à segurança na realização do teste e na divulgação dos dados dos participantes.

Ranganath et al.⁸ estudaram a percepção dos estudantes sobre o uso das ferramentas *on-line* para a aplicação de testes de múltipla escolha e consideraram o exame suficiente para avaliação do conhecimento, desde que observados os cuidados referidos por Oliveira et al.⁷.

Entre as limitações do TP, podemos citar o fato de o processo organizacional ser complexo e depender de várias horas e atores em dedicação e agendas institucionais^{3,6}, e a importância de estar alinhado aos objetivos institucionais³ e ao currículo dos cursos². Prioritariamente as instituições devem valorizar um *Core Curriculum* para a eleição de temas relevantes de maneira integrada e comprometida com a avaliação formativa.

O TP é um instrumento de avaliação do estudante que permite aferir a retenção do conhecimento³ por meio da medida de desempenho e de questões qualificadas que discriminam o conhecimento em vários níveis de dificuldade e aprofundamento e evitam distratores⁹. Além de permitir a avaliação, é parte do processo de aprendizado que promove a autonomia do discente a partir da autodeterminação e do autoaperfeiçoamento⁵.

As oficinas de elaboração de QME devem seguir requisitos mínimos de qualificação, partindo da definição clara de quais objetivos de aprendizagem serão avaliados e de sua relevância no contexto da formação médica e na evidência científica para sua aplicação¹⁰.

Como ferramenta de gestão acadêmica, o TP oferece a possibilidade de avaliar o curso internamente correlacionando-se o currículo, as unidades curriculares, os temas das questões e o desempenho dos estudantes. Possibilita também a avaliação externa por meio de comparação com as médias das escolas do consórcio³.

Outras avaliações relacionadas à adoção de novas estratégias metodológicas para ensino já foram realizadas em comparação com metodologias tradicionais³. No estudo de Cecílio-Fernandes et al.³, a instituição que utiliza currículo integrado e metodologia ativa de ensino com exclusividade apresenta desempenho superior a outras escolas, incluindo as que priorizam metodologias tradicionais.

CONSIDERAÇÕES FINAIS

A experiência aqui descrita mostrou-se de extrema importância como ferramenta de avaliação dos discentes, do curso e da instituição, permitindo novos aprendizados e intervenções de melhorias.

O TP trouxe importantes vantagens e benefícios na implantação do curso, na formação docente e na adequação de currículo por meio da participação da instituição desde a implantação do curso.

Com o decorrer dos anos, a participação continuada permitiu avaliar o ganho de conhecimento dos estudantes e implantar ações de melhoria no currículo e nos processos avaliativos da instituição, bem como constatar resultados satisfatórios de desempenho da instituição como um todo.

CONTRIBUIÇÃO DOS AUTORES

Todos os autores participaram da experiência como atores institucionais e contribuíram na elaboração do relato de experiência.

CONFLITO DE INTERESSES

Declaramos não haver conflito de interesses.

FINANCIAMENTO

Declaramos não haver financiamento.

REFERÊNCIAS

1. Epstein RM. Assessment in medical education. *N Engl J Med.* 2007 Jan 25;356(4):387-96.
2. Van der Vleuten CPM, Verwijnen GM, Wijnen WHFW. Fifteen years of experience with progress testing in a problem-based learning curriculum. *Med Teach.* 1996;18:103-9.
3. Cecílio-Fernandes D, Bicudo AM, Hamamoto Filho PT. Progress test as a pattern of excellence for the assessment of medical students knowledge: concepts, history and perspective. *Medicina (Ribeirão Preto).* 2021;54(1):e-173770.
4. Guaraná CVPS, Duque TB, Lima TRM. Desempenho no teste de progresso e coeficiente são preditores da aprovação na residência médica? *Medicina (Ribeirão Preto).* 2022;55(2):e-182110.
5. Laluna MCMC, Ferraz CA. Finalidades e função da avaliação na formação de enfermeiros. *Rev Bras Enferm.* 2007;60(6):641-5.
6. Olayemi E. Multiple choice questions as a tool for assessment in medical education. *Ann Biomed Sci.* 2013;12(1):15-23.
7. Oliveira SS, Postal EA, Afonso DH, Meress CE, Cyrino EG, Abreu Junior AF, et al. Teste de Progresso da Abem: consolidando uma estratégia de avaliação para o ensino médico. *Rev Bras Educ Med.* 2022;46(1):e-061.
8. Ranganath R, Rajalaksmi C, Simon MA. Medical student's perceptions of e- assessment: multiple choice questions used as a tool of assessment for preclinical years. *J Med Educ.* 2017;16(1):35-43.
9. Licon-Chavez AL, Boehringer PKM, Velazquez-Liano LR. Quality assessment of a multiple choice test though psychometric properties. *MedEdPublish.* 2020;9(91):1-17.
10. Salam A, Yousuf R, Abu Bakar SM. Multiple choice questions in medical education: how to construct high quality questions. *International Journal of Human and Health Sciences.* 2020;4(2):79-88.



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.