

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

Agnaldo 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

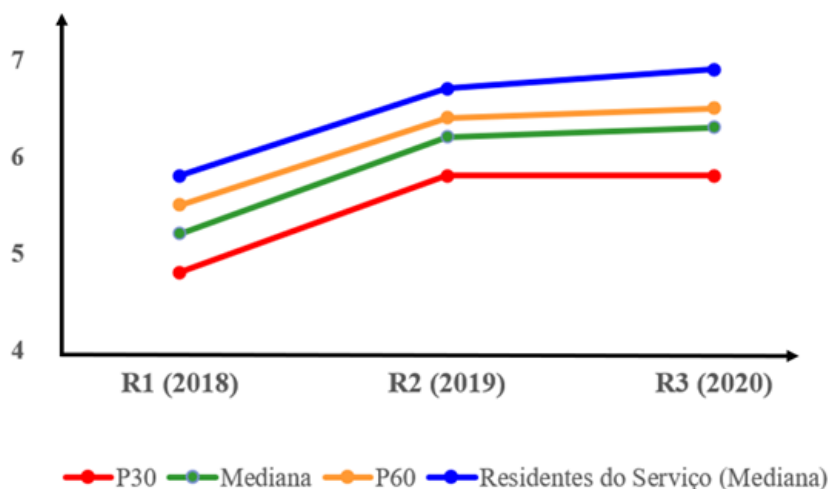
(R2-R1) e dos R3 e R2 (R3-R2), que fornecem informações sobre o ganho cognitivo dos residentes em diferentes períodos do treinamento no PRM; 4. a mediana de desempenho dos R3, que reflete o nível cognitivo dos concluintes desse PRM. Esses parâmetros foram analisados em relação à linha de tendência da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

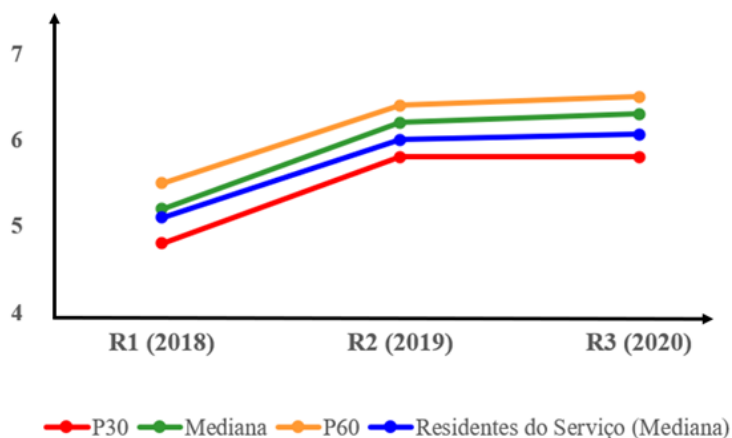
Os gráficos 1, 2, 3, 4 e 5 apresentam os cinco perfis identificados. Em cada gráfico, constam a mediana do desempenho dos residentes do PRM (azul) e a mediana (laranja), o P30 (vermelho) e o P60 (verde) do desempenho geral dos candidatos em nível nacional.

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



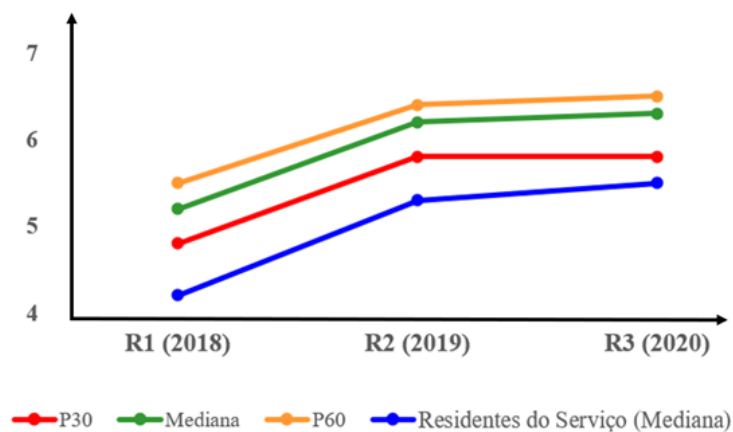
Fonte: Elaborado pelos autores.

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)



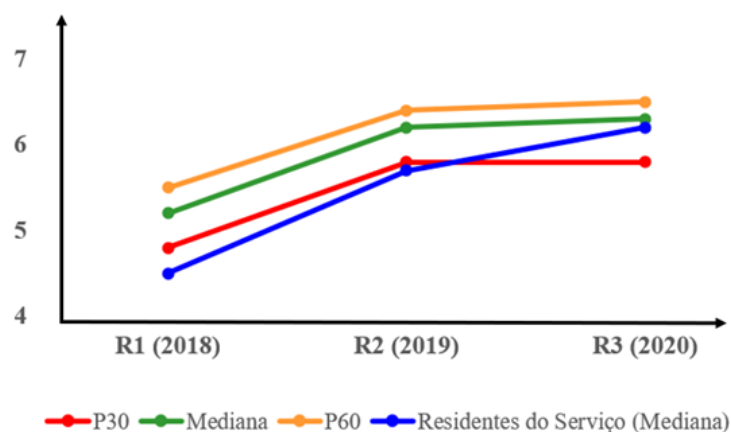
Fonte: Elaborado pelos autores.

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



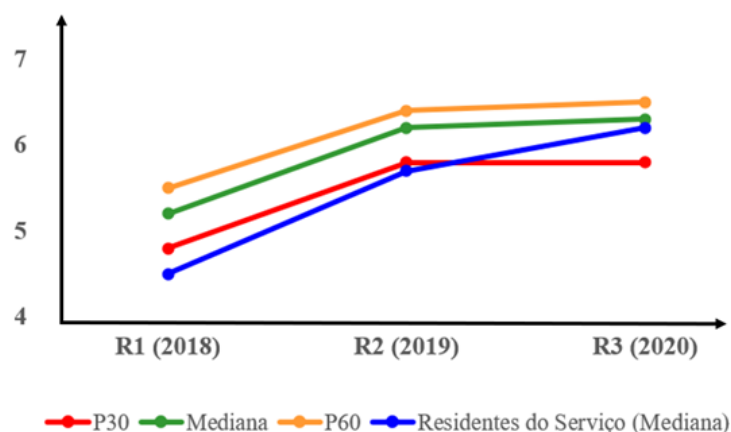
Fonte: Elaborado pelos autores.

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.

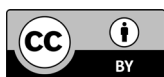
FINANCIAMENTO

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, Bollela 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.