

Teste de Progresso na Escola Médica: uma Revisão Sistemática acerca da Literatura

Progress Test in Medical School: a Systematic Review of the Literature

Ademir Garcia Reberti^I 

Nayme Hechem Monfredini^I 

Olavo Franco Ferreira Filho^{II} 

Dalton Francisco de Andrade^{III} 

Carlos Eduardo Andrade Pinheiro^{III} 

Jean Carl Silva^I 

RESUMO

O Teste de Progresso é uma avaliação objetiva, estruturada com 60 a 150 questões de múltipla escolha, elaborada com o objetivo de promover uma avaliação das competências cognitivas esperadas no final do curso de graduação. Esse teste é aplicado a todos os discentes no mesmo dia, de modo que seja possível comparar os resultados entre as séries e analisar a performance evolutiva do conhecimento no decorrer do curso. Este trabalho teve como objetivo realizar uma revisão sistemática e literária acerca do Teste de Progresso nas escolas médicas no Brasil e no mundo, compreendendo os benefícios de sua implantação para o desenvolvimento do aprendizado, tanto para o aluno quanto para o docente e para a instituição. A pesquisa foi realizada no período de julho de 2018 a abril de 2019 e abordou artigos publicados no período de janeiro de 2002 a março de 2019. Utilizaram-se os descritores “Teste de Progresso nas escolas médicas” e “Teoria de Resposta ao Item em Medicina” nas plataformas PubMed, SciELO e Lilacs. Não houve limitação de idioma na seleção dos artigos, porém a pesquisa foi realizada em inglês. Foram encontrados 192.026 artigos, e, após a aplicação de filtros de busca avançada, incluíram-se 11 artigos no estudo. O Teste de Progresso (TP) vem sendo aplicado nas escolas médicas, de forma isolada ou em grupos de escolas parceiras, desde o final da década de 1990. Os resultados do teste constroem curvas de desempenho dos acadêmicos, o que permite identificar fragilidades e qualidades dos estudantes nas diversas áreas do conhecimento relacionadas ao curso. O Teste de Progresso não é um instrumento exclusivo de avaliação do desempenho dos estudantes, mas assume também um significado como ferramenta avaliativa para uso na gestão acadêmica, e para isso é fundamental que as instituições assumam papel ativo na elaboração e na análise dos dados dessa avaliação. Os exames desenvolvidos para testar a competência clínica em estudantes de Medicina necessitam ser válidos e confiáveis. Para que o método avaliativo seja válido, é necessário que o assunto seja amplamente revisado e estudado, visando a melhorias e adequações na execução dos testes.

PALAVRAS-CHAVE

- Teste de Progresso.
- Estatística.
- Educação Médica.
- Avaliação Educacional.

^IUniversidade da Região de Joinville, Joinville, Santa Catarina, Brasil.

^{II}Universidade Estadual de Londrina, Londrina, Paraná, Brasil.

^{III}Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, Brasil.

KEY-WORDS

- Progress Test.
- Statistics.
- Medical Education.
- Educational Evaluation.

ABSTRACT

Progress Test is an objective assessment, consisting of 60 to 150 multiple-choice questions, designed to promote an assessment of the cognitive skills expected at the end of undergraduate school. This test is applied to all students on the same day, so that it is possible to compare the results between grades and analyze the development of knowledge performance throughout the course. This study aimed to carry out a systematic and literary review about Progress Test in medical schools in Brazil and around the world, understanding the benefits of its implementation for the development of learning for the student, the teacher and the institution. The study was carried out from July 2018 to April 2019, which addressed articles published from January 2002 to March 2019. The keywords used were: "Progress Test in Medical Schools" and "Item Response Theory in Medicine" in the PubMed, Scielo, and Lilacs platforms. There was no language limitation in article selection, but the research was carried out in English. A total of 192,026 articles were identified, and after applying advanced search filters, 11 articles were included in the study. The Progress Test (PT) has been applied in medical schools, either alone or in groups of partner schools, since the late 1990s. The test results build the students' performance curves, which allow us to identify weaknesses and strengths of the students in the several areas of knowledge related to the course. The Progress Test is not an exclusive instrument for assessing student performance, but it is also important as an assessment tool for academic management use and thus, it is crucial that institutions take an active role in the preparation and analysis of this assessment data. Assessments designed to test clinical competence in medical students need to be valid and reliable. For the evaluative method to be valid it is necessary that the subject be extensively reviewed and studied, aiming at improvements and adjustments in test performance.

Recebido em 19/9/19

Aceito em 24/10/19

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.

A técnica do Teste de Progresso (TP) foi inicialmente desenvolvida, no início dos anos 1970, na *University Missouri-Kansas City School of Medicine*, nos Estados Unidos, e na Universidade de Limburg, na Holanda².

As questões de múltipla escolha constituem um dos métodos mais utilizados em todo o mundo em exames destinados a avaliar habilidades cognitivas. O processo de avaliação busca compreender a adequação entre o desempenho do estudante e a fase em que se encontra matriculado, além de fomentar o aprendizado e embasar decisões que terão implicações em seu progresso, contribuindo para o controle da qualidade dos programas educacionais³.

O TP consiste em uma avaliação objetiva, estruturada com 60 a 150 questões de múltipla escolha, com o objetivo de promover uma avaliação das competências cognitivas esperadas no final do curso. Esse teste é aplicado a todos os discentes em uma mesma data, de modo que se possam comparar os resultados entre as séries e analisar a *performance* evolutiva do conhecimento no decorrer do curso⁴. A comparação dos resultados é possível com a equalização dos testes⁶.

O tempo de prova máximo é de quatro horas, sendo o tempo mínimo para a entrega da prova de uma hora. Dependendo da instituição, a presença dos alunos pode ou não ser obrigatória para realização do teste⁵.

O desenvolvimento cognitivo é uma importante dimensão na formação do médico, sendo um processo contínuo de aquisição e consolidação de um conjunto de componentes necessários para o domínio do conhecimento em uma ou mais áreas de desempenho⁶.

Os exames desenvolvidos para testar a competência clínica em estudantes de Medicina necessitam ser válidos e confiáveis. É reconhecido

pelo comportamento de aprendizagem do estudante de Medicina que a avaliação geralmente conduz ao aprendizado. Portanto, se os alunos estão aprendendo o que está sendo avaliado, é vital que o conteúdo da avaliação reflita os objetivos da aprendizagem⁷.

A validade de uma avaliação será garantida quando for possível demonstrar que o método escolhido avalia exatamente aquilo que se pretendia que o estudante aprendesse. Depende, então, da adequação do método à natureza do domínio que se pretende avaliar³.

Visando à melhoria da qualidade das provas, são estudados métodos psicométricos de avaliações, entre eles: Teoria Clássica dos Testes (TCT) e Teoria de Resposta ao Item (TRI). A TCT busca explicar o resultado final total, considerando a soma das respostas dadas ao teste. Esse modelo visa analisar dois parâmetros do item: os graus de dificuldade e de discriminação. A análise se baseia no escore total de acertos e erros, não considerando as habilidades latentes dos estudantes e o acerto casual. Sendo assim, avalia os testes de qualidade e a correlação entre o acerto em determinado item e o número total de acertos na prova⁶.

Já por sua vez, o modelo de TRI não está interessado no escore total de um teste, mas sim no resultado obtido por cada um dos itens que compõem o teste. Tem como objetivo determinar a probabilidade de o estudante acertar ou errar um item de determinado teste e os fatores que o levam a tal resultado. São considerados três parâmetros do item: grau de dificuldade, capacidade discriminativa e probabilidade de acerto casual. Também considera as habilidades latentes do indivíduo, procurando analisar itens de qualidade⁶.

De acordo com Embretson⁸ e Andrade⁹, a TRI é o modelo mais recomendado para ser aplicado nas avaliações dos testes objetivos, em comparação com a TCT, por considerar as habilidades latentes dos estudantes e analisar os parâmetros por item e não a prova como um todo. A TRI é também a metodologia que melhor estima a capacidade de determinado item ter sido acertado ao acaso. No TP, esse é um fator de

grande importância, pois, por tratar-se de um teste elaborado em nível de sexto ano que envolve diversas competências para a resolução dos itens, tal como o raciocínio clínico, é provável que os estudantes iniciantes venham a “chutar” grande parte dos itens⁶.

O método de avaliação cumulativa tem sido utilizado como ferramenta para orientar o comportamento de estudo dos alunos, uma vez que incentiva o tempo de autoestudo de maneira mais uniforme ao longo do curso¹⁰. Independentemente da postura filosófica da avaliação, há um eixo comum entre as diversas concepções de avaliação educacional: associam-se a um processo interpretativo, supondo juízo de valor, qualidade e/ou mérito que tem por alvo diagnosticar e verificar a abrangência dos objetivos propostos no processo ensino-aprendizagem¹¹. Apesar das diferenças na formulação das questões, a essência das teorias é melhorar o sistema de ensino e o aprendizado do aluno.

OBJETIVOS

A pesquisa teve como objetivo compreender os benefícios da implantação do TP na graduação de Medicina para o desenvolvimento do aprendizado, tanto para o aluno quanto para o docente e a instituição.

METODOLOGIA

Estratégia de busca

Foi realizada uma revisão sistemática e literária acerca do TP nas escolas médicas do Brasil e do mundo. No estudo, não houve limitação sobre o idioma dos artigos. Utilizaram-se métodos de busca com o uso de descritores (DeCs Terminologia) de saúde indexados à base de dados

do Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde (Bireme). Foram consultadas as bases de dados Literatura Latino-Americana e do Caribe em Ciências de Saúde (Lilacs), Scientific Electronic Library Online (SciELO) e MEDLINE/PubMed (US National Library of Medicine of National Institute of Health).

Foram lidos todos os artigos encontrados quando o número era inferior a 20 resultados. Consideraram-se artigos repetidos aqueles que já haviam sido incluídos no trabalho em uma busca anterior.

Descritores

Na pesquisa, utilizaram-se os seguintes descritores: “Teste de Progresso nas Escolas Médicas” e “Teoria de Resposta ao Item; Medicina”. Todos os termos foram pesquisados em inglês: “*Progress test in medical school*” e “*Item response theory; medical*”.

Período de busca

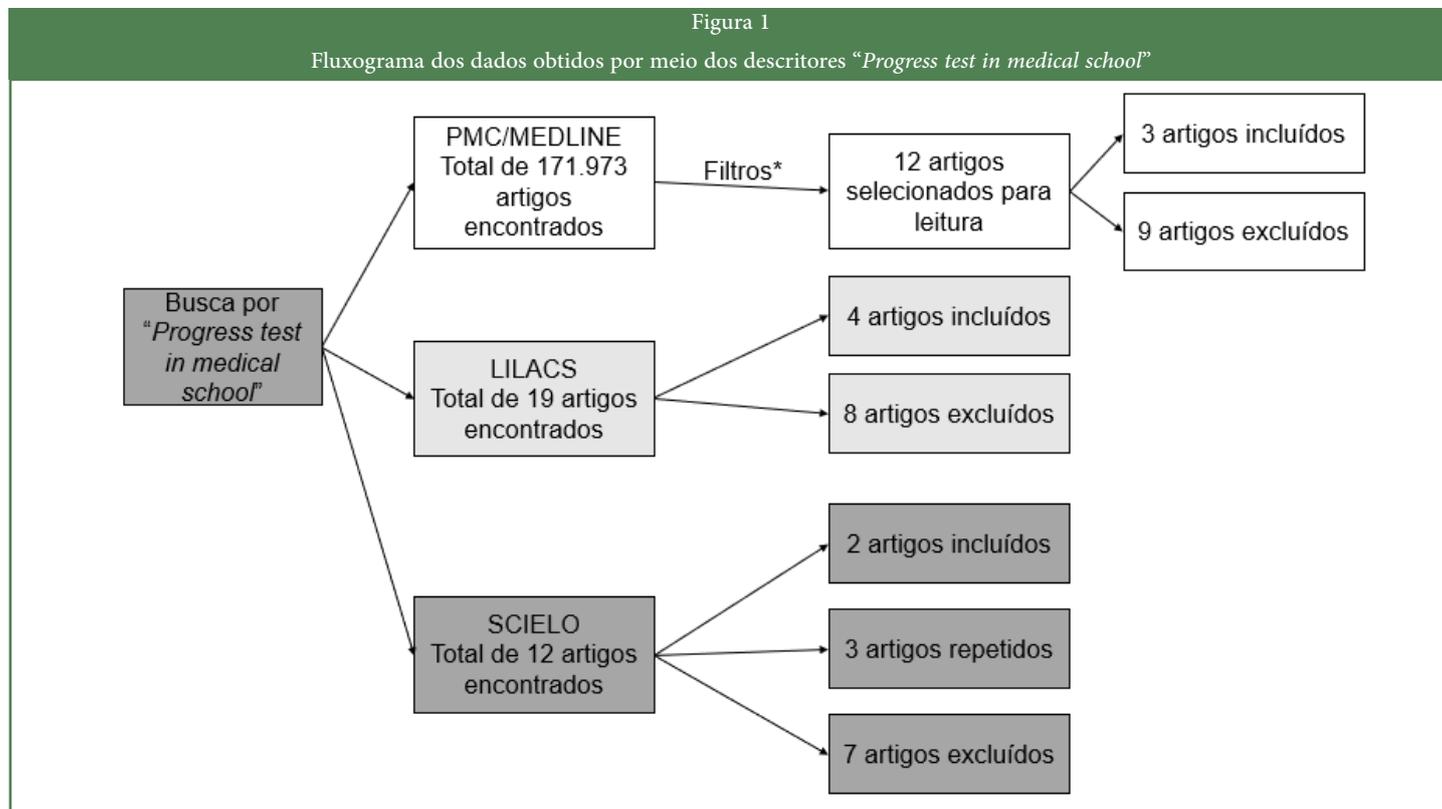
As pesquisas foram realizadas de julho de 2018 a abril de 2019, contemplando artigos que tenham sido publicados no período de janeiro de 2002 a março de 2019. Determinou-se a data de publicação dos artigos antes do início das buscas.

Crítérios de inclusão e exclusão

Consideraram-se os seguintes critérios de inclusão: 1. artigos originais; 2. artigos que possuíssem o tema abordado sem fuga de ideias; 3. artigos que tivessem sido publicados dentro do período preestabelecido; 4. artigos com contribuição científica acerca do tema. Excluíram-se 1. os estudos que envolveram mais de uma escola e não forneceram estimativas de prevalência individualizadas para cada uma delas; 2. estudos conduzidos

Figura 1

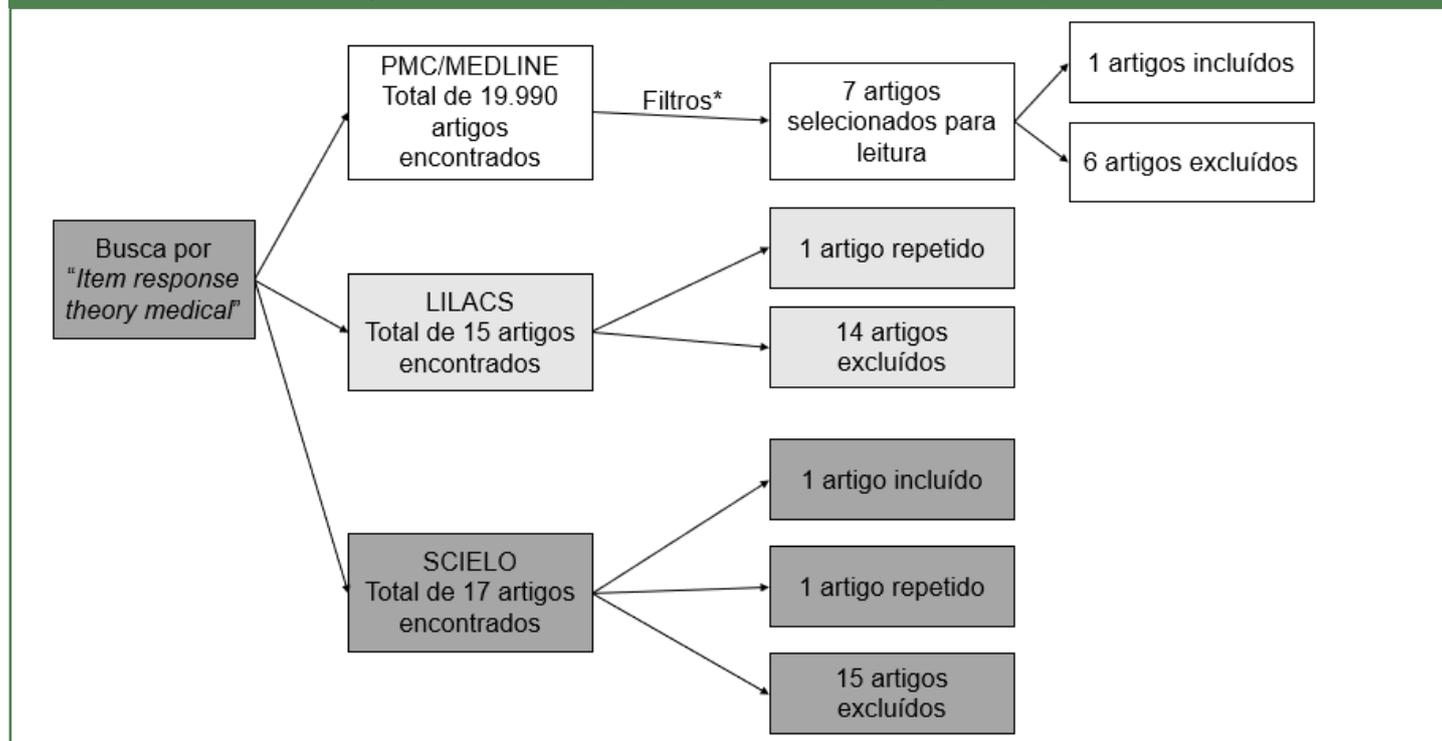
Fluxograma dos dados obtidos por meio dos descritores “*Progress test in medical school*”



*Filtros: período de publicação de 1º de janeiro de 2002 a 31 de março de 2019, com acesso aberto e sendo do *MEDLINE journals*. Em seguida, aplicou-se o filtro *Search fields* para que fossem selecionados artigos que contivessem os descritores no corpo do trabalho.

Fonte: Elaborada pelos autores.

Figura 2

Fluxograma dos dados obtidos por meio dos descritores “*Item response theory; medical*”

*Filtros: período de publicação de 1º de janeiro de 2002 a 31 de março de 2019, com acesso aberto e sendo do *MEDLINE journals*. Em seguida, aplicou-se o filtro *Search fields* para que fossem selecionados artigos que contivessem os descritores no título do trabalho.

Fonte: Elaborada pelos autores.

em instituições cujos dados de interesse da pesquisa não se encontravam disponibilizados em nenhuma fonte *on-line* (site da própria instituição, do Ministério da Educação e Cultura ou de outros); 3. estudos com fuga do tema proposto.

PubMed Central/MEDLINE

Na plataforma PubMed Central (PMC), foram encontrados 171.973 artigos para os descritores “*Progress test in medical school*” com a limitação de período de publicação compreendida entre 1º de janeiro de 2002 e 31 de março de 2019, com acesso aberto e sendo do *MEDLINE journals*. Aplicou-se o filtro *Search fields* para que fossem selecionados artigos que contivessem os descritores no corpo do trabalho. Encontraram-se 12 artigos e incluíram-se três. Os outros nove foram lidos e excluídos; apesar de conterem em seu corpo o tema TP, concentravam-se em outros assuntos que não se enquadravam na revisão proposta.

Incluíram-se os seguintes artigos: “The use of progress testing”, de Schuwirth e Van Der Vleuten¹²; “Development of a competency-based formative progress test with student-generated MCQs: results from a multi-centre pilot study”, de Wagener et al.¹³; e “Progress testing in the medical curriculum: students approaches to learning and perceived stress”, de Chen et al.¹⁴.

Ainda na plataforma PMC foram encontrados 19.990 artigos para os descritores “*Item response theory; medical*” com a mesma limitação de período de publicação, acesso aberto e sendo do *MEDLINE journals*. Após aplicação dos mesmos filtros, encontraram-se 1.234 artigos. Utilizou-se então o filtro *Search fields* para que fossem selecionados os artigos que

contivessem os descritores no título do trabalho, sendo encontrados sete artigos. Entre esses, incluiu-se um artigo no estudo. Os outros seis foram lidos e excluídos, por não se enquadrarem na proposta da revisão.

O artigo incluído foi: “Using item response theory to explore the psychometric properties of extended matching questions examination in undergraduate medical education”, de Bhakta, Tennant, Horton, Lawton e Andrich⁷.

Lilacs

A plataforma Lilacs foi utilizada na busca dos descritores “*Progress test, medical education*”, e encontraram-se 19 artigos no período de publicação de 1º de janeiro de 2002 a 31 de março de 2019. Depois da leitura, incluíram-se no estudo quatro artigos. Os outros 15 artigos não tratavam do tema abordado.

Incluíram-se os seguintes artigos: “*Cumulative assessment: does it improve students knowledge acquisition and retention?*”, de Fernandes, Nagtegaal, Noordzij e Tio¹⁰; “Teste de Progresso: uma ferramenta avaliativa para a gestão acadêmica”, de Pinheiro et al.¹⁵; “Avaliação do crescimento cognitivo do estudante de Medicina: aplicação do teste de equalização no Teste de Progresso”, de Sakai, Ferreira Filho, Almeida e Matsuo⁶; e “Teste de Progresso e avaliação do curso: dez anos de experiência da medicina da Universidade Estadual de Londrina”, de Sakai, Ferreira Filho, Almeida, Mashima e Marchese².

Utilizou-se a mesma plataforma para a busca dos descritores “*Item response theory; medical*” e encontraram-se 15 artigos no mesmo período de publicação. Após a leitura dos artigos, um foi selecionado para inclusão

no estudo, no entanto ele já havia sido incluído em uma busca anterior. Os outros 14 artigos apresentaram fuga em relação ao tema proposto.

SciELO

A plataforma SciELO foi utilizada na busca dos descritores “*Progress test, medical education*”, e encontraram-se 12 artigos no período de publicação de 1º de janeiro de 2002 a 31 de março de 2019. Depois da leitura, dois artigos foram incluídos no estudo. Além destes, três artigos apareceram repetidamente. Os outros sete não tratavam do tema estudado.

Incluíram-se os seguintes artigos: “O Teste de Progresso como indicador para melhorias em curso de graduação em Medicina”, de Rosa et al.¹⁶, e “Avaliação somativa de habilidades cognitivas: experiência envolvendo boas práticas para a elaboração de testes de múltipla escolha e a composição de exames”, de Bollela, Borges e Troncon³.

Utilizou-se a mesma plataforma para a busca dos descritores “*Item response theory; medical*”, e encontraram-se 17 artigos. Depois da leitura, dois foram selecionados pelos critérios de inclusão, no entanto um deles já estava incluído no trabalho (repetido). Os outros 15 artigos foram excluídos por não se enquadrarem no tema proposto.

Incluiu-se o artigo “Avaliação critério-referenciada em Medicina e Enfermagem: diferentes concepções de docentes e estudantes de uma

escola pública de saúde de Brasília, Brasil”, de Miranda Junior, Tatsch, Braga, Spengler Junior e Novaes⁴.

RESULTADOS

Foram incluídos 11 trabalhos científicos acerca do assunto (Quadro 1).

Os trabalhos incluídos auxiliaram na compreensão do conceito de TP e do desenvolvimento cognitivo em geral, principalmente relacionado à construção do conhecimento médico.

Schuwirth e Van der Vleuten¹² realizaram uma revisão literária acerca do uso do TP. Ficou evidenciado que a abordagem de avaliação longitudinal tem efeito positivo sobre o comportamento de aprendizagem do aluno, desencorajando o aprendizado pré-provas e conseqüentemente auxiliando no aprendizado acumulativo em longo prazo, o aprendizado duradouro. Supõe-se que os alunos experimentem menos estresse na realização do TP em comparação com os testes tradicionais, visto que um único resultado ruim não pode desfazer uma série de bons resultados. Os autores ressaltam também que o método longitudinal aumenta a confiabilidade do TP.

Segundo Wagener et al.¹³, o TP fornece aos alunos um *feedback* sobre seu nível de proficiência no decorrer de seus estudos. Foi aplicado o TP a acadêmicos do primeiro ao sexto ano de Medicina em sete escolas de medicina da Alemanha, totalizando 463 estudantes participantes.

Quadro 1

Artigos incluídos no estudo: autores, país de realização, ano de publicação e tipo de estudo

Título	Autores/país	Ano de publicação	Tipo de estudo
“The use of progress testing”	Schuwirth e Van Der Vleuten / Holanda	2012	Revisão literária
“Development of a competency-based formative progress test with student-generated MCQs: results from a multi-centre pilot study”	Wagener et al. / Alemanha	2015	Coorte
“Progress testing in the medical curriculum: students approaches to learning and perceived stress”	Chen et al. / Nova Zelândia	2015	Estudo-piloto multicêntrico
“Using item response theory to explore the psychometric properties of extended matching questions examination in undergraduate medical education”	Bhakta, Tennant, Horton, Lawton e Andrich / Reino Unido	2005	Coorte
“Cumulative assessment: does it improve students knowledge acquisition and retention?”	Fernandes, Nagtegaal, Noordzij e Tio / Holanda	2018	Coorte
“Teste de Progresso: uma ferramenta avaliativa para a gestão acadêmica”	Pinheiro et al. / Brasil	2015	Coorte transversal
“Avaliação do crescimento cognitivo do estudante de medicina: aplicação do teste de equalização no Teste de Progresso”	Sakai, Ferreira Filho, Almeida e Matsuo / Brasil	2011	Pesquisa-ação
“Teste de progresso e avaliação do curso: dez anos de experiência da medicina da Universidade Estadual de Londrina”	Sakai, Ferreira Filho, Almeida, Mashima e Marchese / Brasil	2008	Agregado-observacional-transversal
“O Teste de Progresso como indicador para melhorias em curso de graduação em Medicina”	Rosa et al. / Brasil	2017	Coorte transversal
“Avaliação somativa de habilidades cognitivas: experiência envolvendo boas práticas para a elaboração de testes de múltipla escolha e a composição de exames”	Bollela, Borges e Troncon / Brasil	2018	Revisão literária
“Avaliação critério-referenciada em Medicina e Enfermagem: diferentes concepções de docentes e estudantes de uma escola pública de saúde de Brasília, Brasil”	Miranda Junior, Tatsch, Braga, Spengler Junior e Novaes / Brasil	2018	Estudo qualitativo, descritivo, com delineamento transversal

Fonte: Elaborado pelos autores.

Destes 61,3% eram do sexo feminino, 35% do sexo masculino e 3,7% sem indicação de sexo. A média de idade dos participantes foi de 24,56 anos. O estudo observou que o número de respostas corretas aumentou constantemente com o avanço do nível do semestre letivo em que o aluno encontrava-se matriculado.

Chen et al.¹⁴ exploraram o efeito do TP em comparação aos testes tradicionais sobre o estresse percebido pelos acadêmicos. O estudo foi realizado em dois tempos, avaliando a mudança em relação ao tempo de decorrido. Houve um total de 864 alunos participantes. Ao longo dos dois momentos estudados, o TP pode ter reduzido o estresse para os alunos, pois aqueles que participaram do grupo tradicional experimentaram aumentos significativos no estresse quando fizeram os exames tradicionais de final de ano, enquanto os níveis de estresse não aumentaram significativamente para os alunos do grupo do TP.

Bhakta, Tennant, Horton, Lawton e Andrich⁷ analisaram os resultados do Exame de Questões de Correspondência Estendida realizado por estudantes de Medicina do quarto ano (entre 2001 e 2002). A análise de Rasch foi usada para examinar se o conjunto de perguntas utilizadas no exame correspondia a uma escala unidimensional. Observou-se o grau de dificuldade de perguntas intra e interespecialidades médicas e cirúrgicas, além do padrão de respostas nas perguntas individuais para avaliar o impacto das opções do distrator. Com o estudo foi possível entender como os alunos usam as opções e as informações fornecidas, juntamente com o conhecimento especializado para a compreensão das perguntas e decisão acerca da resposta correta. Essas informações são fundamentais para que os redatores melhorem a qualidade das perguntas a serem criadas.

Fernandes, Nagtegaal, Noordzij e Tio¹⁰ estudaram a avaliação cumulativa como ferramenta para orientar o comportamento de estudo dos alunos, comparando o aumento do conhecimento entre estudantes que participaram com avaliação cumulativa e aqueles que permaneceram no método tradicional de avaliação no final dos ciclos de ensino. Foram utilizados dados dos primeiros quatro TPs interuniversitários holandeses anteriores ao experimento. Participaram 62 estudantes, dos quais 37 estavam na condição de avaliação tradicional e 25 na de avaliação cumulativa. Demonstrou-se que houve aumento significativo no conhecimento dos estudantes nos quatro TPs no decorrer do tempo decorrido de curso. Não foi possível mostrar diferença entre os dois grupos, indicando que os dois métodos de avaliação são semelhantes em relação ao crescimento do conhecimento dos alunos.

Pinheiro et al.¹⁵ averiguaram as potencialidades do TP no auxílio à gestão acadêmica. Foram analisados os desempenhos dos estudantes, do primeiro ao sexto ano do curso de Medicina, no TP aplicado em 2008 e reaplicado sem modificações em 2011. Os resultados demonstraram que, nos dois momentos de aplicação do teste, houve aquisição de conhecimento cognitivo dos estudantes no decorrer do curso de Medicina. Entretanto, o comportamento de aprendizado cumulativo mostrou diferenças nos dois momentos de aplicação. Em 2008, a progressão ocorreu a cada dois anos de curso, e, em 2011, essa progressão iniciou-se somente a partir do terceiro ano de curso. Esse estudo foi fundamental para compreensão dos gestores acerca da necessidade de maiores reflexões no ensino ofertado e da qualidade das questões criadas para a avaliação.

Sakai, Ferreira Filho, Almeida e Matsuo⁶ realizaram, em uma universidade estadual brasileira, uma avaliação do crescimento cognitivo do estudante de Medicina a partir da aplicação do Teste de Equalização

no TP. Analisaram-se todos os resultados dos estudantes do primeiro ao sexto ano acadêmico no período de 2004 a 2007. Em um primeiro momento, buscou-se explicar o sentido que têm as respostas dadas pelos estudantes a uma série de itens (questões). Foram utilizados dois modelos psicométricos: TCT e TRI. Para a análise com a TCT, considerou-se um item de boa qualidade aquele que tinha correlação bisserial (entre o acerto em determinado item e o número total de acertos na prova) acima de 0,40. Para a TRI, foram utilizados três parâmetros para itens de múltipla escolha: (a) capacidade discriminativa do item; (b) grau de dificuldade do item; (c) acerto casual.

Ainda acerca do estudo de Sakai, Ferreira Filho, Almeida e Matsuo⁶, não foi possível seguir adiante na etapa de aplicação da TRI, pois o número de itens com correlação bisserial acima de 0,40 ficou abaixo do recomendado, além da insuficiência do número necessário de respondentes por grupo. Entretanto, foi possível definir as questões-âncora que seriam utilizadas no momento seguinte. A média de acertos dos estudantes, em todos os TP, variou de um mínimo de 33,5% na primeira série até um máximo de 66,4% na sexta série, nos resultados sem equalização. Já os escores médios nos TP com equalização variaram de 31,0% até 73%, para a primeira e sexta série, respectivamente. Foi sugerido que escolas médicas formem parcerias para compor uma base maior de questões e experiências com critérios preestabelecidos, aumentando o número de itens com alto grau de discriminação e de respondentes, o que permitiria usar a TRI para a equalização dos teste.

Em 2008, Sakai, Ferreira Filho, Almeida, Mashima e Marchese² realizaram uma avaliação do curso de Medicina e do uso do TP em um decorrer de dez anos de experiência em uma universidade estadual de Medicina no Sul do Brasil. O TP foi utilizado como instrumento de avaliação do curso e como resultado desse período: houve aumento na participação dos estudantes para realizar o teste; o desempenho cognitivo deles apresentou aumento de uma série para outra em cada teste; a média de acertos nas áreas de clínica médica foi menor na quinta série, no oitavo e nono TP; na saúde coletiva, houve alto percentual de acerto nas duas primeiras séries. Tais resultados refletem a estrutura curricular, bem como as potencialidades e fragilidades do curso. Esse estudo foi um bom indicador do processo de autoavaliação do curso, mas sugeriu-se que ainda necessita de aprofundamento nos estudos de técnicas de análise dos resultados, para permitir estimar o crescimento cognitivo dos estudantes.

Rosa et al.¹⁶ utilizaram o TP como indicador para melhorias no curso de graduação em Medicina. Foram realizados três estudos transversais institucionais durante os três anos da aplicação desse teste no curso de Medicina em uma universidade na Região Sul do Brasil. Participaram estudantes de graduação em Medicina que haviam feito o TP em 2011, 2012 e 2013. A análise estatística foi realizada com um nível de confiança de 95%. A adesão média ao longo dos três anos variou de 91,8% a 100%. Concluiu-se que o TP é um excelente indicador para os gestores, pois pode ser usado para desenvolver intervenções para melhorar a qualidade dos cursos. Após a aplicação do primeiro teste, realizaram-se mudanças no curso de Medicina da universidade. Testes posteriores demonstraram a eficácia dessas mudanças.

Bollela, Borges e Troncon³ utilizaram uma avaliação somativa de habilidades cognitivas para compreender mais sobre o uso de boas práticas para a elaboração de testes (questões) de múltipla escolha e a composição de exames. De acordo com os autores, para a avaliação

de habilidades referentes ao domínio cognitivo, utilizaram-se três níveis referenciais para elaboração de questões: 1. básico, que exige conhecimento factual, envolvendo basicamente a memorização de fatos e sua recuperação; 2. intermediário, que envolve não só o conhecimento adquirido, mas também as capacidades de compreensão, interpretação e aplicação daquele conhecimento para resolver problemas mais simples; 3. avançado, que envolve os níveis de análise, síntese e avaliação para propor soluções para problemas mais complexos como os que habitualmente se apresentam ao profissional da área da saúde. O estudo concluiu que as questões de múltipla escolha com somente uma alternativa correta têm inúmeras vantagens, mas, para a aplicabilidade adequada, é necessária atenção às boas práticas de confecção de questões e de construção dos exames, envolvendo recomendações relativas ao conteúdo e à forma dos testes, bem como à composição da matriz de avaliação ou tabela de especificação.

Miranda Junior, Tatsch, Braga, Spengler Junior e Novaes⁴ estudaram diferentes concepções de docentes e estudantes de uma escola pública de saúde de Brasília sobre a avaliação critério-referenciada em Medicina e Enfermagem. Esse tipo de avaliação analisa o desempenho acadêmico sem comparar um estudante com outro, diferindo da avaliação normorreferenciada que objetiva determinar os melhores e piores desempenhos. Obteve-se uma amostra de 413 participantes, totalizando 54 docentes e 344 estudantes. Observou-se fragilidade do conceito de avaliação critério-referenciada em todo o corpo acadêmico e insatisfação quanto aos critérios avaliativos adotados na escola (323 ou 79,6%). Entre os motivos apontados, destacam-se a subjetividade na correção dos exames empregados e dificuldades ao concorrer em concursos de residência. Outros autores relatam que o método critério-referenciado tende a inflacionar as menções dos estudantes, embora os avalie de maneira mais integral. Por isso, recomendam a adoção dos sistemas critério e normorreferenciado para, respectivamente, aferir melhor o desempenho dos estudantes e compará-los entre si. Os autores concluíram que, diante disso, é necessária a adoção de ambos os modelos (normo e critério-referenciado), com a aplicação de padrões de desempenho. Os instrumentos avaliativos também necessitam de critérios mais objetivos para uma correção de testes mais uniforme e justa.

DISCUSSÃO

O TP tem como finalidade a avaliação do desempenho cognitivo dos estudantes durante o curso de Medicina, além de ser uma ferramenta que auxilia na compreensão dos acertos e das falhas do próprio curso. Um único teste é desenvolvido e aplicado em todos os semestres/anos do curso em um mesmo momento². Ao ser aplicado anualmente, possibilita verificar a evolução do acadêmico no decorrer do curso.

É um tipo de avaliação longitudinal do desenvolvimento cognitivo dos estudantes no decorrer do curso. Os testes aplicados pelas diversas escolas médicas apresentam diferenças, tais como: norma referenciada ou critério referenciado e as questões podem ser do tipo verdadeiro/falso ou múltipla escolha¹⁷.

No cenário de educação médica no Brasil, o TP vem sendo aplicado nas escolas médicas, de forma isolada ou em parceria, desde o final da década de 1990⁶.

Visto o interesse das escolas médicas em evoluir o processo educacional, parcerias foram criadas para compartilhamento de informações e aprendizados, possibilitando a melhoria da aplicação e

utilidade do teste na educação médica. Visando a esse objetivo, o Núcleo de Apoio Pedagógico Interinstitucional Sul II (Napisul II) foi estabelecido em 2010 por iniciativa de oito escolas médicas com apoio da Associação Brasileira de Educação Médica (Abem), com o propósito de formular, implementar e analisar o TP¹⁶.

Além disso, em 2004 o Ministério da Educação introduziu um novo modelo de Exame Nacional de Desempenho dos Estudantes (Enade), cujos objetivos são semelhantes ao do TP, mas aplicado apenas aos estudantes ingressantes e concluintes dos cursos¹⁸. O TP é uma avaliação de caráter institucional que não visa formar um *ranking* das faculdades de Medicina, e, por isso, as datas das provas são estipuladas por cada instituição, por núcleos ou consórcios¹⁹.

As questões do TP são elaboradas por uma equipe multidisciplinar, constituída por professores das próprias instituições, com diferentes graus de dificuldade. Cada questão tem de quatro a cinco alternativas possíveis nas respostas. Além disso, são divididas em sete áreas do conhecimento: clínica geral, pediatria, ginecologia e obstetrícia, cirurgia, saúde coletiva, ética médica e ciências básicas. As questões podem ser contextualizadas em casos clínicos, de modo que os enunciados não exijam apenas a memorização do aluno, mas também raciocínio lógico¹⁹.

Os resultados do teste constroem curvas de desempenho dos acadêmicos, o que permite identificar fragilidades e qualidades dos estudantes nas diversas áreas do conhecimento relacionadas ao curso².

Os métodos de avaliação podem servir como ferramenta de melhoria do ensino. A avaliação formativa pode ser definida como a avaliação de forma contínua, de caráter não probatório, que tenha por objetivo a melhoria do aprendizado. O processo de avaliar atua como regulador, mostrando falhas e possíveis soluções a obstáculos apresentados pelos estudantes, podendo gerar melhorias nas ferramentas didáticas¹⁹.

A TRI passou a ser conhecida, principalmente, a partir do ano de 1968 com o trabalho de Lord e Novick intitulado *Statistical theories of mental tests scores*. Desde então, inúmeras aplicações têm sido exploradas para a teoria: criação de bancos de itens, avaliação adaptativa computadorizada, equalização de provas, avaliação de mudança cognitiva, entre outras. A TRI tem como unidade de análise o item, formalizando a relação que existe entre a probabilidade de acertar o item e a capacidade latente requerida na resolução dele. Quanto maior for a capacidade de um sujeito (traço latente), maior será a probabilidade de ele acertar um determinado item que meça esse construto²⁰.

A TRI tem sido considerada um marco para a psicometria moderna, apresentando vantagens em relação à TCT, considerando a virtual invariância dos parâmetros dos itens em relação à amostra, a estimação mais precisa e interpretável nível de habilidade dos indivíduos e os procedimentos de equalização de testes mais eficientes²⁰.

O TP também pode ser uma ferramenta auxiliar no treinamento dos estudantes para os processos seletivos dos quais participarão no decorrer de sua vida profissional, como os concursos para residência médica².

De acordo com o grau de dificuldade, a questão será considerada fácil quando obtiver taxa de acerto igual ou superior a 80%, intermediária quando a taxa de acerto estiver entre 20% e 80% e difícil quando tiver taxa de acerto igual ou inferior a 20%⁶.

A tendência de as escolas médicas formarem parcerias para elaborar e aplicar os testes aponta uma melhoria na qualidade da prova, pois isso permite a troca de experiências entre elas, a constituição de grupos de

trabalho com as especialidades envolvidas e a montagem de um banco de itens com questões testadas e avaliadas. Além disso, aumentaria o número de itens com alto grau de discriminação e de respondentes, o que permitiria usar a TRI para a equalização dos testes⁶.

A avaliação do desempenho dos estudantes depende fundamentalmente da qualidade dos itens da prova. Assim sendo, a TCT e a TRI constituem métodos estatísticos que podem contribuir para essa finalidade²¹.

O TP não é um instrumento exclusivo de avaliação do desempenho dos estudantes, mas assume também um significado como ferramenta avaliativa para uso na gestão acadêmica, e para isso é fundamental que as instituições assumam papel ativo na elaboração e na análise dos dados dessa avaliação¹⁵.

Outra vantagem assumida é que a coleta de dados longitudinais é mais preditiva de competência/desempenho futuro do que medições pontuais.

Entre as vantagens da aplicação do teste, destaca-se a validade preditiva que é elevada à medida que aumenta a confiabilidade, e esta se dá pela combinação longitudinal de resultados. A coleta de dados longitudinais é mais preditiva de competência/desempenho futuro em comparação a medições pontuais¹².

A possibilidade de determinar as características de dificuldade, discriminação e acerto ao acaso incentiva novas formas de avaliação, e TP adaptativos computadorizados estão sendo desenvolvidos^{23,24}.

O TP possibilita que o curso reveja o seu projeto pedagógico e os conteúdos tendo por base análises gerais e por área. Com o uso rotineiro, a análise de dados pode proporcionar aos cursos e aos estudantes uma avaliação externa cognitiva, na medida em que deixa de haver endogenia resultante de questões feitas sempre pelos mesmos professores²⁵.

As mudanças educacionais começam a pequenos passos: a união de faculdades de Medicina pode auxiliar, implementar e otimizar uma ferramenta de avaliação econômica capaz de promover melhorias na aprendizagem. O TP, além de ser estratégico para o aprendizado e a gestão, é uma ferramenta de baixo custo²².

Além disso, foi estudada a correlação entre a pontuação dos TP dos alunos e seu desempenho em um processo de seleção de residência, o qual demonstrou que existe correlação entre o desempenho dos estudantes no TP e na prova de residência. Esses dados podem ser explicados pelo fato de que ambos os instrumentos medem competências e conhecimentos cognitivos²⁶.

CONSIDERAÇÕES FINAIS

Este estudo trouxe como principal contribuição a discussão acerca do conceito do aprendizado durante o tempo de passagem dos acadêmicos pela escola de Medicina.

O interesse pelo progresso educacional proporcionou o desenvolvimento de estudos acerca do aprendizado. O Teste de Progresso em Medicina (TP) surgiu como uma forma de medir o desenvolvimento dos acadêmicos em relação ao tempo de curso decorrido. Além disso, o teste possibilita identificar prováveis deficiências nas disciplinas ativas do curso, visando melhorar os métodos de ensino e intervir neles. Os acadêmicos se beneficiam quando quantificam o seu progresso de aprendizado no decorrer do tempo, o que pode ajudá-los no preparo para as provas seletivas de residência médica. Já para a instituição e seus gestores, é possível averiguar pontos de deficiência e eficiência do curso para promover melhorias nos métodos de ensino.

Sugerem-se mais estudos de caráter quantitativo para melhor análise

estatística acerca da qualidade na criação dos testes cumulativos e determinação do seu impacto positivo e/ou negativo na construção do aprendizado.

REFERÊNCIAS

1. Ferreira ARL, Weyh CB. Tecnologia da informação no processo da avaliação como estratégia para qualificar a prática docente. In: Congresso Internacional de Educação e Tecnologias; 2018; São Carlos, Brasil. São Carlos: UFSCar; 2018. p. 1-7.
2. 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.
3. 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.
4. Miranda Junior UJP, Tatsch JFS, Braga CS, Spengler Junior SL, Novaes MRCG. Avaliação critério-referenciada em Medicina e Enfermagem: diferentes concepções de docentes e estudantes de uma escola pública de saúde de Brasília, Brasil. *Rev. bras. educ. med.* 2018; 42(3):67-77.
5. Associação Brasileira de Educação Médica. Avaliação do estudante: contribuição do teste de progresso. ABEM, 2017. Disponível em: <https://abem-educmed.org.br/acoes/avaliacao-do-estudante-contribuicao-do-teste-de-progresso/>. Acesso em: de maio de
6. 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.
7. Bhakta B, Tennant A, Horton M, Lawton G, Andrich D. Using item response theory to explore the psychometric properties of extended matching questions examination in undergraduate medical education. *BMC med. educ* 2005;5(9):1-13.
8. Embretson SE. The new rules of measurement. *Psychol. Assess.* 1996;8(4):341-9.
9. Andrade DF. Comparando desempenho de grupos de alunos por intermédio da teoria de resposta ao item. *Est. aval. educ.* 2001;23:31-69.
10. Fernandes DC, Nagtegaal M, Noordzij G, Tio RA. Cumulative assessment: does it improve students knowledge acquisition and retention? *Scientia Medical Education in Health Sciences* 2018;28(4):ID31880.
11. Rodrigues M. Proposta de análise de itens das provas do Saeb sob a perspectiva pedagógica e a psicométrica. *Est. aval. educ.* 2006;34:43-78.
12. Schuwirth LWT, Van Der Vleuten CPM. The use of progress testing. *Perspect Med Educ* 2012; 1:24-30.
13. Wagener S, Moltner A, Timbil S, Gornostayeva M, Schultz JH, Brustle P et al. Development of a competency-based formative progress test with student-generated MCQs: results from a multi-centre pilot study. *GMS Z Med Ausbild* 2015;32(4):Doc46.
14. 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 Edc* 2015; 15(147):1-8.
15. 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.

16. 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.
17. Verhoeven BH, Verwijnen GM, Scherpbier AJ, van der Vleuten CP. Growth of medical knowledge. *Med Educ* 2002;36(8):711-7.
18. Limana A, Brito MRF de. O modelo de avaliação dinâmica e o desenvolvimento de competências: algumas considerações a respeito do Enade. *Avaliação* 2005;10(2):9-32.
19. Baldim TL, Vicente CAO, Arcuri MB. O teste de progresso sob a visão do discente. *Revista da Faculdade de Medicina de Teresópolis* 2018;2(1):41-54.
20. Nunes CHSS, Primi R. Impacto do tamanho da amostra na calibração de itens e estimativa de escores por teoria de resposta ao item. *Aval. psicol.* 2005;4(2):141-153.
21. Henning E, Ramos MS, Konrath AC. Análise de itens de uma prova de raciocínio probabilístico. In: *Actas del VII CIBEM*; 2013; Montevideo, Uruguay. Colegio Seminario; 2013. p. 2025-2032.
22. Vleuten CVD, Freeman A, Collares CF. Profress test utopia. *Perspect Med Educ* 2018;7:136-8.
23. De Champlain AF. A primer on classical test theory and item response theory for assessments in medical education. *Med Educ* 2010;44(1):109-17.
24. Collares CF, Cecilio-Fernandes D. When I say...computerised adaptative testing. *Med Educ* 2019;53(2):115-6.
25. Bicudo AM, Hamamoto Filho PT, Abbade JF, Hafner MLMB, Maffei CML. Consortia of cross-institutional progress testing for all medical schools in Brazil. *Rev. bras. educ. med.* 2019; 43(4)151-6.
26. Hamamoto Filho PT, Lourenção PLTA, Valle AP do, Abbade JF, Bicudo AM. A correlação entre a pontuação dos testes de progresso dos alunos e seu desempenho em um processo de seleção de residência. *Medical Science Educator* 2019;29(1):1-5.

CONTRIBUIÇÃO DOS AUTORES

Ademir Garcia Reberti participou da pesquisa, seleção de textos, redação e correções acerca do manuscrito.

Nayme Hechem Monfredini participou da pesquisa, seleção de textos, redação e correções acerca do manuscrito.

Olavo Franco Ferreira Filho participou da pesquisa, leitura, indicação de quais artigos eram mais pertinentes para confecção deste manuscrito.

Dalton Francisco de Andrade participou da pesquisa, leitura, indicação de quais artigos eram mais pertinentes para confecção deste manuscrito, principalmente ao que se refere a estatística.

Carlos Eduardo Andrade Pinheiro participou da pesquisa, leitura, indicação de quais artigos eram mais pertinentes para confecção deste manuscrito, principalmente ao que se refere ao Teste de Progresso a nível internacional.

Jean Carl Silva participou da pesquisa, seleção de textos, redação e correções acerca do manuscrito.

CONFLITO DE INTERESSES

Os autores declaram não haver conflitos de interesse neste estudo.

ENDEREÇO PARA CORRESPONDÊNCIA

Ademir Garcia Reberti

Rua Henrique Meyer, 280, Centro, Sala 311, Joinville, Santa Catarina, Brazil. CEP: 89201-405.

E-mail: agrebert@terra.com.br



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.