

Teste Progresso: a importância na educação médica

Progress Test: the importance in medical education

Ana Marcia Iunes Salles Gaudard¹
Sulani Silva de Souza²

¹Farmacêutica e Bioquímica. Doutora em Ciências pelo Programa de Pós-graduação em Saúde Baseada em Evidências da Universidade Federal de São Paulo (UNIFESP). Professora do Centro Universitário de Brasília (CEUB).

²Médica. Doutorado em Tocoginecologia pelo Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto. Docente da Escola Superior de Ciências da Saúde (ESCS).

Correspondência: Sulani Silva de Souza.
E-mail: sulaniamazon@gmail.com

RESUMO

Este trabalho se propõe a discutir a crescente importância do Teste de Progresso na educação médica. No mundo, o teste é apresentado com diferentes formatos, visando refletir a variedade dos currículos médicos e diferentes natureza de avaliação, ou seja, formativa ou somativa. Ele tem sido incorporado nas escolas médicas no Brasil e na Escola Superior de Ciências da Saúde. É uma ferramenta de controle de qualidade para a melhoria do processo de aprendizagem; orientado para o estudante, o docente e a instituição. Os testes avaliativos desenvolvidos devem ser válidos e confiáveis. Isto é um desafio sendo necessário que se utilize para tanto de conhecimentos de psicometria para de forma sistemática avaliar essas propriedades.

Palavras-chave: Educação médica; Avaliação educacional; Educação

ABSTRACT

This paper aims to discuss the growing importance of Progress Tests in medical education. In the world, it is presented in different formats to reflect the variety of medical curricula and different nature of assessment, that is, formative or summative. It has been incorporated in medical schools in Brazil and in the Superior School for Health Sciences. It is a quality control tool for the improvement of the learning process, oriented for both, student and teacher, and the institution. The evaluative tests developed need to be valid and reliable. It is a challenge making necessary to use psychometric knowledge to systematically evaluate these properties.

Keywords: Medical education; Educational evaluation, Education

INTRODUÇÃO

A avaliação é um componente importante no processo educacional, frequentemente tem natureza somativa, ou seja, visa atribuir menção ao desempenho do estudante com o objetivo de avaliar a sua progressão para o semestre ou unidade subsequente. Este enfoque tem mudado, uma vez que a avaliação passa a ser mediadora do processo de aprendizagem não tendo apenas finalidade probatória, mas sendo incorporada no ato de ensinar, integrada na ação de formação dos estudantes. Neste contexto surge o Teste de Progresso (TP) como uma importante ferramenta de ensino e aprendizagem na educação médica.

Existem várias definições do que é o TP na literatura, a maioria delas converge quanto às suas principais características: ser uma avaliação longitudinal do desenvolvimento cognitivo, ou seja, aplicado múltiplas vezes no decorrer do tempo e equivalente. Assim, todas as provas são compostas para ter nível semelhante de dificuldade¹⁻³. Além disso, são comumente aplicadas a todos os estudantes, independente do ano que se encontram no curso⁴.

A criação do TP se deu ao final de 1970 nas universidades de Maastricht na Holanda e de Missouri nos EUA no mesmo período, de forma independente, como um formato natural da evolução da avaliação do conhecimento^{4,5}. Em Maastricht, era aplicada a nova metodologia *de Problem Based Learning (PBL)*, ou aprendizado baseado em problemas, que surgiu na Universidade de McMaster no Canadá⁴. Nessa metodologia, o estudo é estimulado pela formação de grupos e tutorias, em que o aluno é um participante ativo do processo de discussão e aquisição de conteúdo, mudando o foco da relação professor e aluno, em que o professor expõe seu conhecimento e os estudantes são apenas observadores e questionadores. Na época era necessário validar a metodologia PBL, portanto o TP serviria como forma de avaliar o desenvolvimento dos estudantes e comparar com a metodologia tradicional, pois a mesma prova poderia ser aplicada a ambos^{4,5}. Em Missouri, era necessário um melhor acompanhamento do desenvolvimento de seus estudantes, uma nova forma de avaliação era necessária^{4,6}. O TP é uma importante ferramenta de ensino e aprendizagem médica, pois faz parte dos seus objetivos estimular o estudante a lidar com problemas complexos, aplicar conhecimento a problemas novos e entender conceitos essenciais^{2,7}.

No Brasil a partir das Diretrizes Curriculares Nacionais do Curso de Graduação em Medicina (DCN), publicadas pela Resolução do Conselho Nacional de Educação / Câmara de Educação Superior nº 4, de 7 de novembro de 2008 a Associação Brasileira de Educação Médica (ABEM) protagoniza a implementação do das DCN por meio de quatro subprojetos – “Avaliação Teste do Progresso”, “Avaliação Institucional”, “Urgência e Emergência” e “Internato”. Dentro do subprojeto “Avaliação Teste de Progresso” foi proposto como objetivo geral instituir o Teste do Progresso Interinstitucional para todas as escolas do Brasil como parte do Eixo de Avaliação da ABEM⁹. Para tanto, as regionais da ABEM promoveram reuniões convidando as escolas que fazem parte da sua área de abrangência para as oficinas de esclarecimento e formação dos Núcleos Interinstitucionais do TP. Na formação dos núcleos foi considerado as regiões geográficas do país, visando facilitar a organização dos encontros entre as escolas participantes. Foi ainda orientado que cada núcleo deveria ter representantes responsáveis de cada escola, que responderia a um coordenador escolhido pelo núcleo⁹.

A implantação dos Núcleos Interinstitucionais de TP objetivou:

1. Capacitar núcleos de escolas para a elaboração da prova do TP colaborativa;
2. Realizar oficinas sobre o TP nas oito regionais da ABEM – Norte, Nordeste, Centro-Oeste, Minas Gerais, São Paulo, Rio de Janeiro/Espírito Santo, Sul I e Sul II;
3. Auxiliar na operacionalização da elaboração e aplicação do TP para as escolas que solicitarem;
4. Realizar oficinas para apresentar os resultados, facilidades e dificuldades na aplicação da prova dentro de núcleos colaborativos, garantindo o não ranqueamento entre os cursos;
5. Elaborar prova única com questões de excelência (testadas e “aprovadas”) de todos os núcleos do Brasil.

AVALIAÇÃO SOMATIVA E FORMATIVA

O TP pode servir como um conjunto de avaliações de natureza somativa ou formativa^{1,2}. Os TPs de natureza somativa requerem que os estudantes alcancem nota mínima pré-determinada para

passarem aos próximos níveis do curso^{1-3,10,11}. Os testes somativos têm como objetivo aprovar ou reprovar estudantes, a nota final é definida pela combinação dos testes daquele ano^{1,2,10}. Pela equivalência dos testes, a cada ano a nota de corte torna-se maior^{2,10,12}. Avaliações somativas tendem a estimular os estudantes a memorização de curto prazo devido à necessidade iminente de aprovação que é frequentemente associada com sobrecarga de conteúdo. Apesar disso, mesmo em TP de natureza somativa, a melhor preparação é o estudo contínuo, já que o TP tem conteúdo extenso (todo currículo médico) e caráter longitudinal^{1,2,13}. A modalidade formativa não tem caráter de seleção ou classificação, seu principal objetivo é fornecer informações ao estudante, o *feedback*, para que ele identifique suas áreas de fragilidade para alcançar o conhecimento necessário na sua formação de médica^{1,2,10,14}.

Já foi demonstrado que o TP formativo está associado a melhores desfechos em exames de competência médica¹⁵. Para o bom funcionamento dessa metodologia, é importante que os estudantes tenham ciência do seu desempenho progressivo para usá-lo como uma ferramenta efetiva de aprendizado¹⁶.

A modalidade formativa naturalmente precisa ser acompanhada de avaliações somativas². Independente da modalidade, ambas estimulam os estudantes continuamente em várias áreas médicas, ambas têm em seu princípio o aprendizado contínuo pela estimulação de estudo autorregulado^{2,17-19}. O TP pode ser visto como uma mudança cultural, tanto na forma de se avaliar, como na forma de estimular hábitos de estudos permanentes uma vez que o estudante na realização da prova utiliza os seus conhecimentos prévios¹.

Considerando que o TP consiste em avaliações sequenciais e equivalentes com o objetivo de avaliar o desenvolvimento do conhecimento médico nos estudantes de medicina¹⁻³, é de grande importância que cada prova contemple as competências desenvolvidas pelos estudantes de medicina ao final do curso. Assim sendo, cada teste contém o conhecimento necessário para atuação profissional do médico recém-formado²⁰. Essa característica é de suma importância, pois é possível situar o aluno no seu conhecimento e naquilo que ele deve alcançar, verificando a evolução de seu desempenho cognitivo nas diferentes áreas e fases do curso^{1,21}.

A natureza longitudinal do TP permite que o estudante avalie seu desempenho em múltiplas áreas de conhecimento médico ao apresentar sua evolução para comparação (curvas de aprendizagem), além disso permite a comparação do seu conhecimento com os dos demais estudantes, serve de preparação para as provas de residência médica e pode ser utilizado como preditivo de desempenho futuro (concursos de residência médica)^{1,14,15}. Assim, o TP fornece informações detalhadas sobre o desenvolvimento estudantil²². A grande vantagem do TP nesse quesito é que avaliações pontuais, como as avaliações de natureza somativa únicas (p. ex. avaliações de final módulo), podem variar substancialmente para mais e para menos em situar o conhecimento do aluno²³. A quantidade de vezes (número de questões por teste) sobre o mesmo conteúdo com diferentes abordagens fornece ao estudante a oportunidade de identificar o que é importante para a prática futura; e ao teste melhor validade e confiabilidade permitindo aos avaliadores situar o aluno em seu nível de conhecimento^{1,3}. O TP é importante não só para detectar e ajudar na solução de problemas de desenvolvimento cognitivo dos estudantes, mas também para avaliar o currículo da instituição^{1,2,24}.

Comumente os TPs são aplicados uma a duas vezes ao ano, com quantidade de itens variando desde 100 até 250². Entretanto, conforme o engajamento do consórcio e seus objetivos, podem ser aplicados mais vezes, como no consórcio de escolas holandesas, que é aplicado 4 vezes por ano de forma somativa^{2,25}. A distribuição temporal do TP é importante para melhorar a confiabilidade dos testes, visto que o espaçamento de itens que abordam o mesmo tema no decorrer do tempo fornece informações mais confiáveis sobre o desenvolvimento do estudante do que uma avaliação pontual extensa¹. A aplicação pode ocorrer por meio físico ou meio eletrônico, o que amplia a aplicabilidade do TP, sendo importante que sejam feitos ao mesmo tempo para evitar irregularidades por parte dos estudantes², principalmente no caso de avaliações somativas/probatórias.

Atualmente o TP está amplamente distribuído pelo mundo. Devido às suas características, abrange todos os continentes, desde Américas, Europa, Ásia, África e Oceania, restrito somente na Antártida². A perpetuação do TP é entendível, visto que avaliações longitudinais são intuitivamente mais lógicas ao combinar resultados e melhor avaliar os estudantes¹. Desde sua implementação como

método de ensino em instituições únicas ou em consórcios internacionais, o TP é continuamente reavaliado e sempre que possível é melhorado¹.

VALIDADE E CONFIABILIDADE

No panorama geral, a utilidade de uma avaliação depende de suas características. Van der Vleuten propôs que a utilidade varia conforme a validade, confiabilidade, impacto educacional, aceitação e custos²⁵. Confiabilidade é uma questão primariamente estatística² e como tal estará sujeita às mesmas regras do tratamento quantitativo de dados: como ao medirmos o desempenho humano estaremos diante de variáveis com instabilidades inerentes, será frequentemente necessário repetir a medida de diversas formas para se obter algum grau de precisão. Essa precisão estará, portanto, na dependência da amostragem utilizada no teste. As principais amostragens para garantir a estabilidade de um teste dizem respeito à amostragem de conteúdo, de tempo e de examinado. E para uma boa validade o conteúdo precisa ser significativo e adequado às necessidades futuras do aprendiz. Conteúdo é, portanto, uma das principais considerações na validade de um teste². Como validade é um conceito específico para um determinado uso ou interpretação, deveríamos avaliar somente conteúdos relevantes com a futura prática profissional².

A validade nesse sentido se refere no quão preciso um teste avalia o que se propõe. A confiabilidade é comumente avaliada pela medida do alfa de Cronbach, este tem por finalidade medir a consistência interna do teste²⁰, sendo que os valores variam de 0 a 1, no qual 0,8 é um valor geralmente aceitável². De uma forma geral, o coeficiente de Cronbach mede a correlação entre as respostas em um teste através da análise do perfil das respostas dadas pelos respondentes. Matematicamente representa o somatório da variância dos itens individuais e da soma da variância de cada avaliador^{25,26}.

Visando ainda garantir a validade da prova devem ser constituídos comitês para avaliar cada item do TP. No TP deve ser garantido a qualidade dos itens e que eles representem a melhor evidência disponível na literatura atual²⁵. Os itens escritos sem critérios pré-definidos tendem a comprometer a validade da prova, pois criadores de itens e revisores diferem quanto ao que consideram relevante²⁵. Assim sendo, critérios devem ser estabelecidos para garantir a qualidade³.

Na universidade de Maastricht por exemplo²⁵, cinco critérios mínimos devem ser preenchidos em cada item para garantir qualidade: conhecimento médico, conhecimento de rápida evocação (requisito para a prática clínica), conhecimento relevante para situações de alto risco e conhecimento base para um ou mais conceitos médicos²⁵. Critérios gerais como estes garantem o julgamento correto da relevância de cada item. Vale ressaltar que envolver os estudantes no processo de criação de itens pode ser uma importante fonte de comentários em relação ao desenvolvimento do TP, além de evidenciar áreas de fragilidades acadêmicas dos estudantes, melhorando suas motivações e entendimento do teste³. Itens criados por estudantes podem ter a mesma qualidade de itens criados por especialistas²⁵.

A principal forma de assegurar a validade e a confiabilidade de um teste é garantir a qualidade dos itens^{25,26}. Além disso, itens que focam em conhecimentos que integram diferentes áreas devem ser preferidos¹. Os itens devem questionar conhecimentos que são indispensáveis aos estudantes, evitando termos imprecisos, conhecimentos não essenciais, conhecimentos especializados de pouca relevância ou triviais². Para melhor definir o que os itens devem se basear, os conhecimentos devem ser classificados em grupos para a organização do TP, o *blueprint* ou matriz.

BLUEPRINT

O termo *blueprint* remete a classificação básica e fundamental dos conteúdos utilizados pelo TP. Esses grupos gerais são a primeira forma de delimitar e garantir a validade e comparabilidade entre os TP. O *blueprint* se alinha objetivamente com o que o médico recém-formado deve saber, portanto definindo as competências/ habilidades desenvolvidas pelos estudantes². Apesar do currículo das faculdades e o *blueprint* estarem alinhados com os objetivos de final de curso, esses não estão diretamente ligados²⁴. O TP é uma ferramenta que acrescenta ao desenvolvimento do estudante e não tem função de substituir as competências das faculdades. Tendo em vista a independência dos currículos das diferentes escolas o TP é um instrumento notável para compará-los^{2,6}.

No *blueprint* cada grupo de conteúdo programático recebe um peso, ou seja, grupos de maior relevância para a formação médica recebem maior quantidade de itens conforme consenso prévio: o

blueprint é definido pelos conteúdos necessários e pelo valor atribuído a esses². O conteúdo é dividido por grandes áreas temáticas de conhecimento como por exemplo: Medicina interna, Cirurgia, Ginecologia, Pediatria, Bioética, Saúde pública e Ciências básicas. Esses grupos temáticos são subdivididos para definir o valor percentual de abordagem no teste, e conseqüentemente, quantos itens cada grupo terá no TP¹.

Os itens que compõem o *blueprint* podem ser divididos em níveis cognitivos maiores e menores. A taxonomia de Bloom (Figura 1) serve para hierarquizar os processos de aquisição de conhecimento nesse sentido²⁴. Segundo Bloom, processos cognitivos de baixa ordem requerem apenas lembrar e entender (descrever ou explicar) conteúdos. Já os processos cognitivos de ordem maior requerem aplicar, analisar, julgar e, por fim, criar conhecimento, ou seja, requerem conhecimento aprofundado²⁴. Foi teorizado que no processo de aquisição de conhecimento é necessário inicialmente processos cognitivos de ordem menor para desenvolver os de nível maior²⁴. Questões de nível maior (questões contextualizadas em situações problemas) resultam em melhor retenção do conteúdo, pois requerem lembrar e raciocinar sobre as informações²⁴. Tendo em vista essas informações, o TP deve ser constituído tanto de questões de nível menor, quanto de nível maior, considerando o processo de desenvolvimento cognitivo²⁴. Além da taxonomia de Bloom, outros modelos, como a pirâmide de Miller (Figura 2), são usados com menor frequência para criar e classificar itens²⁴. Os níveis de desenvolvimento de competências clínicas descritos na pirâmide de Miller têm sido úteis para demonstrar que não se pode ficar restrito ao conhecimento teórico. O rumo da autenticidade profissional exige que o estudante saia do conhecimento meramente factual (saber), saiba aplicar esse conhecimento (saber como) e executá-lo de maneira prática em laboratórios e ambientes simulados (demonstrar) e, finalmente, aplicá-lo na realidade (fazer).

A colaboração entre instituições na criação de itens para TP é efetivo, pois cada instituição pode usar mais tempo para criação de menos itens. Os itens criados separadamente com base no *blueprint* ajudam na independência dos currículos das instituições^{2,12}. A colaboração entre universidades garante maior número de itens criados, amplificação dos comitês de revisão e repartição entre os custos³. Com esses objetivos, até alianças

internacionais podem ser criadas, como a aliança internacional envolvendo faculdades da Irlanda, Austrália, Canadá, Portugal e Índias ocidentais². Em locais que recursos são limitados, a criação de consórcios é especialmente importante. O uso de informações compartilhadas coletadas pode ser julgado e avaliado separadamente para a melhoria do TP e de outros processos acadêmicos^{1,2, 12}.

Ainda nesse sentido, as questões podem ser divididas em dois tipos: questões básicas e aquelas baseadas em casos clínicos. As questões básicas não necessitam de informações relacionadas a pacientes e tendem a apresentar níveis cognitivos menores. Elas são mais bem aplicadas a estudantes dos primeiros anos do curso. As questões de casos clínicos representam o tipo de raciocínio que se utiliza na prática clínica. Questões de nível cognitivo maior tendem a ser difíceis para estudantes das séries iniciais. Portanto, já que o TP abrange vários níveis de estudantes ao mesmo tempo, é necessário que questões de nível maior e menor estejam presentes em consonância (comumente metade para cada nível)². Esse balanço se alinha com a metodologia formativa e repetitiva do TP. Dessa forma o TP garante que o conhecimento seja aplicado de forma válida a todos estudantes. É recomendado que os itens na sua criação estejam de acordo com o nível cognitivo (maior ou menor), e não somente de acordo com o tipo (básico ou clínico) melhor definindo a dificuldade de cada teste².

TIPOS DE QUESTÕES

Em relação aos tipos de itens, os mais usados são questões de múltipla escolha ou verdadeiro ou falso. No TP geralmente se opta pelas questões de múltipla escolha com escolha da melhor resposta, visto que apresentam alta validade e confiabilidade²⁷. Essas questões podem ter quantidade variável de opções (distratores) sendo importante não incluir alternativas que são facilmente reconhecidas como erradas. Também apresentam mais possibilidades e delimitam melhor se o estudante avaliado detém o conhecimento necessário para acertá-la.

A construção de questões de múltipla escolha, atualmente, para o TP do Consórcio do Centro Oeste parte dos conteúdos e habilidades definidas no *Blueprint*. Posteriormente são elaborados os casos clínicos que serão utilizados nas situações-problemas seguidos do comando para resposta, que deve indicar de forma clara e objetiva a tarefa

a ser realizada em conexão com a habilidade que se pretende avaliar na questão.

O terceiro passo é a construção de quatro alternativas de respostas, que devem ser elaboradas tendo-se em vista a produção de informações relevantes sobre o processo de construção da habilidade avaliada. Isso significa que a resposta correta – o gabarito – deve validar a capacidade do estudante em relação à determinada habilidade cognitiva. As demais alternativas, os distratores, produzem informações importantes para a avaliação, na medida em que apontam possíveis caminhos de raciocínio dos estudantes, delimitando a etapa do desenvolvimento da aprendizagem em que ele se encontra.

ANÁLISE DAS QUESTÕES

A análise das questões envolve a avaliação da dificuldade, discriminação e qualidade dos distratores.

A dificuldade do TP deve ser equivalente em todas suas edições para permitir a comparação e assegurar sua validade. Uma forma usual de avaliar a dificuldade dos testes é pelo cálculo do valor p . Para isso, divide-se a quantidade de indivíduos que acertaram a questão pela quantidade total de indivíduos que responderam²⁵. A dificuldade geral do teste é calculada pela média dos valores de p ²⁵. A consistência na dificuldade é o que torna o TP importante para acompanhar o desenvolvimento estudantil. A dificuldade varia de 0 a 1, sendo que quanto mais próximo da unidade mais fácil será o item e quanto mais próximo de zero mais difícil.

A discriminação pode ser representada pela diferença do percentual de acertos do grupo de alto e baixo desempenho. Estes grupos são definidos a partir dos 27% que tiveram maior e menor acerto no teste, respectivamente. Outra forma é o cálculo da correlação item-total, mais comum, que informa a correlação de determinado item em relação ao resultado total do teste. A discriminação ideal para os itens construídos é de 0,30²⁶.

A qualidade do distrator é verificada considerando o quanto os estudantes optaram por determinada opção. Se ela não for selecionada ou selecionada em percentual inferior a cinco será considerada um distrator não funcional. Assim, questões com quatro opções de resposta podem ficar com zero, um, dois ou três distratores, conforme a seleção dos estudantes²⁷.

FEEDBACK

A característica fundamental do TP é o feedback, que consiste em estatísticas de acertos e erros encaminhadas aos estudantes, geralmente de forma online^{2,3}. No *feedback* o estudante pode avaliar seus pontos de dificuldade e saber em quais aspectos necessita aprimorar-se. A função primordial do TP é que o estudante tenha consciência do próprio conhecimento e possa melhorar¹⁴. Os tópicos mais comuns de *feedback* são:

1. Nota
2. Percentual de acerto
3. Notas das avaliações progressas (progressão)
4. Acertos por categorias
5. Média da turma
6. Média geral

A efetividade do processo de aprendizagem do *feedback* depende dos conteúdos e da forma como ele é apresentado²⁷. É necessário entendimento completo do conceito de avaliações formativas e feedback por parte dos estudantes. Quando estes são esclarecidos sobre essas características, a motivação e confiança são maiores. Os objetivos do TP e a sua relevância precisam ficar claro para os estudantes, do contrário esses tendem a não fazer a prova com seriedade, pois não percebem seu benefício^{2,16}. Quanto maior o uso das plataformas de *feedback*, maior o crescimento do conhecimento⁶. A maioria dos estudantes percebem o TP como uma forma de situar seu nível de conhecimento e não como um instrumento de aprendizagem^{14,16}, ou seja, não percebem o TP como uma forma de se autoavaliar e escolher quais competências e habilidades focar seus estudos. Estudantes com baixa performance no TP são os que mais se beneficiam do *feedback*, entretanto mesmo entre os que mais acertam, ele é importante para demonstrar lacunas de conhecimento^{3,10,16}. Para a instituição, o *feedback* tem especial importância para identificar estudantes com problemas no progresso e que necessitam de intervenção. O impacto formativo do TP se dá no *feedback*, fundamental atenção deve ser dada a esse tópico pelos educadores e estudantes para que seu máximo desempenho seja estabelecido na educação médica. Por fim, discussões sobre as questões do TP podem associar-se à melhor entendimento e resultados por parte dos estudantes numa tentativa de melhorar a relação com o planejamento de estudos individuais^{10,14}, principalmente nos primeiros anos. Pode-se progressivamente mudar

o foco para indicação de apostilas/referências para estimular a autorregulação do estudo^{2,4}.

TP NA ESCOLA SUPERIOR DE CIÊNCIAS DA SAÚDE E NO CENTRO-OESTE.

O TP da ESCS foi instituído em 2001 e considerado obrigatório desde essa data, sendo atualmente aplicado uma vez ao ano, mas em alguns anos letivos tivemos duas aplicações. No primeiro ano (2001) de aplicação, o teste foi fornecido pela Faculdade de Medicina e Enfermagem de Marília (FAMEMA), que na ocasião tinha um contrato firmado de consultoria. O TP permitiu durante todos esses anos monitoramento do progresso ou crescimento no alcance de objetivos previstos ao final do currículo como um todo.

As questões desde 2001 integram várias disciplinas e são elaboradas na modalidade de resposta escolhida. A elaboração das questões TP na modalidade de resposta escolhida sempre exigiu da Gerência de Avaliação (GA) da ESCS um rigor técnico partindo inicialmente da seleção de conteúdo a ser examinado em cada teste. Os professores responsáveis da GA elaboram um *blueprint* em que as questões na sua maioria são situações problemas. Elas atendem uma distribuição percentual pelas grandes áreas de conhecimento e as competências esperadas respectivamente para cada área. As situações problemas também contemplam uma distribuição percentual por diferentes faixas etária e contextos (cenários).

Na região Centro-Oeste, por iniciativa da ESCS, foi criado em 2013 um Consórcio de Escolas Médicas Públicas e Privadas do Centro-Oeste com a participação inicial das seguintes escolas – Universidade de Brasília (UnB), Universidade Federal de Goiás (UFG), Universidade Federal de Mato Grosso do Sul (UFMS), Universidade Federal do Mato Grosso (UFMT), Universidade Federal do Grande Dourados (UFGD), Escola Superior de Ciências da Saúde (ESCS), Universidade Católica de Brasília (UCB), Universidade Evangélica de Goiás (UniEvangélica), Universidade para o Desenvolvimento do Estado e da Região do Pantanal (UNIDERP) e Universidade de Cuiabá (UNIC). O Consórcio de Cursos de Medicina da Região Centro Oeste, atualmente conta com a participação de 38 cursos.

O TP do consórcio do Centro-Oeste é um instrumento de avaliação de caráter formativo, constituído de 100 questões de múltipla escolha, elaboradas para promover uma avaliação das

competências cognitivas consideradas pertinentes ao final do curso de graduação, com base no perfil desejado do médico que a escola busca alcançar. As desigualdades, diferenças e a diversidade social, cultural e econômica e de situação de saúde existentes no Brasil exigem, portanto, flexibilidade curricular. Estas características estão expressas nas discussões do Consórcio e na definição das habilidades e competências selecionados para integrar o *blueprint* (matriz) do TP.

As questões atualmente são classificadas nas seguintes áreas do conhecimento médico: aspectos morfofuncionais, aspectos fisiológicos e terapêuticos, habilidades médicas, saúde da mulher e da criança, saúde do adulto, saúde mental, saúde da comunidade e bioética. As questões são construídas de modo a focalizar os comportamentos mais elaborados (compreensão, aplicação, análise, síntese e avaliação) da taxonomia de Bloom. O resultado individual é sigiloso e de conhecimento exclusivo do estudante. Os desempenhos individuais, identificados somente por senha, são encaminhados a cada estudante, por carta fechada.

Para o estudante, os propósitos da aplicação do TP incluem:

1. Uso como instrumento de autoavaliação, para propiciar ao estudante o acompanhamento da sua progressão no curso de medicina.
2. Treinamento do estudante na realização de teste de múltipla escolha, uma vez que esse é ainda o método tradicional de avaliação dos exames de residência médica, de admissão e de título de especialização.

Para a Escola, os resultados não-individualizados de cada série são conhecidos e utilizados para a avaliação de fragilidades específicas e para a elaboração de planos de melhoria.

Por conseguinte, considerando um currículo baseado em problemas, os objetivos do teste incluem os seguintes:

1. Analisar o grau de aquisição de conhecimentos, geral e por área, dos estudantes em cada série.
2. Avaliar o grau de progressão do conhecimento médico nas séries do Curso de Medicina.
3. Analisar as questões de múltipla escolha quanto à taxonomia de Bloom.

4. Determinar a progressão dos estudantes na resolução de testes de múltipla escolha, conforme o grau de taxonomia das questões de múltipla escolha.

5. Determinar estratégias de regulação para a superação de fragilidades, eventualmente observadas.

Os TP referentes aos anos de 2014, 2015, 2016, 2017, 2018 e 2019 foram aplicados, no mesmo dia, para todos os estudantes de todos os semestres do Curso de Medicina que compõem o consórcio. A confiabilidade do TP do consórcio nestes anos foi sempre superior a 0,80 pelo alfa de Cronbach. A Universidade Federal do Mato Grosso do Sul é a responsável pela consolidação dos dados do TP referente ao consórcio e de manter o *link* que orquestra toda a administração e envio de relatórios individuais para os estudantes. A realização do TP foi considerada obrigatória para a maioria dos estudantes das escolas consorciadas.

CONSIDERAÇÕES FINAIS

O Teste de Progresso na educação médica surgiu como uma forma de medir o progresso dos estudantes em relação ao tempo decorrido de curso. O TP na educação médica no Brasil tem natureza formativa. Os estudantes devem ser esclarecidos dos objetivos formativos do TP e de que se trata de uma avaliação longitudinal do desenvolvimento cognitivo dos estudantes. Além disso, o teste permite, em conjunto com outros indicadores avaliativos, identificar prováveis deficiências nas unidades curriculares desenvolvidas no curso, visando aprimorar e intervir nos processos de ensino.

Os estudantes se beneficiam ao quantificar seu progresso de aprendizagem por meio das curvas de desempenho cognitivo, que são construídas a partir dos seus resultados no teste ao longo do tempo. As curvas permitem identificar as fragilidades e potencialidades dos estudantes sendo que as fragilidades identificadas podem ser melhoradas em processo. O TP é uma ferramenta que tem grande importância na formação médica e nas demais áreas formadoras de profissionais para a área da saúde.

REFERÊNCIAS

1. Schuwirth LW, van der Vleuten CP. The use of progress testing. *Perspectives on Medical Education*. 2012 Mar;1(1): 24-30. doi: 10.1007/s40037-012-0007-2.
2. William Wrigley, Cees PM Van Der Vleuten, Adrian Freeman & Arno Muijtjens (2012) A systemic framework for the progress test: Strengths, constraints and issues: AMEE Guide No. 71, *Medical Teacher*, 34:9, 683-697. doi: 10.3109/0142159X.2012.704437.
3. 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. doi: 10.1080/0142159X.2016.1230183.
4. van der Vleuten CPM, Verwijnen GM, Wijnen WHFW. 1996. Fifteen years of experience with progress testing in a problem-based learning curriculum. *Medical Teacher* 18(2):103–110. doi:10.3109/01421599609034142.
5. Arnold L, Willoughby TL. 1990. The quarterly profile examination. *Academic Medicine* 65(8):515–516. doi: 10.1097/00001888-199008000-00005.
6. Gold J, DeMuth R, Mavis B, Wagner D. Progress testing 2.0: clinical skills meets necessary science. *Med Educ Online*. 2015 May 5; 20: 27769. doi: 10.3402/meo.v20.27769.
7. Koens F, Rademakers JDJ, Ten Cate TJ. 2005. Validation of core medical knowledge by postgraduates and specialists. *Med Educ* 39:911–917. doi: 10.1111/j.1365-2929.2005.02246.x.
8. Ministério da Educação. Conselho Nacional de Educação. Resolução CNE/CES nº 4 /2001. Institui Diretrizes Curriculares Nacionais do Curso de Graduação em Medicina. [Internet]. Brasília:MEC;2001. [citado em março 2021] Disponível em: <http://portal.mec.gov.br/cne/arquivos/pdf/Med.pdf>
9. Barbosa L, Bicudo AM. 10 anos das Diretrizes Curriculares Nacionais dos Cursos de Graduação em Medicina. Rio de Janeiro: Associação Brasileira de Educação Médica, 2014 80p [citado em março 2021] Disponível em: <http://www.ufrgs.br>
10. Tio R, Schutte B, Meiboom AA, Greidanus J, Dubois EA, Bremers AJA. The progress test of medicine: the Dutch experience. *Perspect. Med Educ*. 2016; 5(1):51-55. doi: 10.1007/s40037-015-0237-1.
11. 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 Sep 11; 15:147. doi: 10.1186/s12909-015-0426-y.
12. Muijtjens AMM, Schuwirth LWT, Cohen-Schotanus J, van der Vleuten CPM. 2008. Differences in knowledge development exposed by multicurricular progress test data. *Adv Health Sci Educ* 13:593–605. doi: 10.1007/s10459-007-9066-2.
13. Wade L, Harrison C, Hollands J, Mattick K, Ricketts C, Wass L. 2011. Student perceptions of the progress test in two settings and the implications for test deployment. *Adv Health Sci Educ* 01 November 2011 ed., Springer. doi: 10.1007/s10459-011-9334-z.
14. 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. doi: 10.3205/zma000988.
15. Norman G, Neville A, Blake J, Mueller B. 2010. Assessment steers learning down the right road: Impact of progress testing on licensing examination performance. *Med Teach* 32:496–499. doi: 10.3109/0142159X.2010.486063.
16. Given K, Hannigan A, McGrath D. (2016). Red, yellow and green: What does it mean? How the progress test informs and supports student progress. *Med Teach*;38(10):1025-1032. doi: 10.3109/0142159X.2016.1147533
17. Al-Kadri HM, Al-Moamary MS, Roberts C, Van der Vleuten CP. Exploring assessment factors contributing to students' study strategies: literature review. *Med Teach*. 2012;34 Suppl 1: S42-50. doi: 10.3109/0142159X.2012.656756.

18. Koh LC. Refocusing formative feedback to enhance learning in pre-registration nurse education. *Nurse Educ Pract.* 2008 Jul;8(4):223-30. doi: 10.1016/j.nepr.2007.08.002.
19. Freeman A, Ricketts C. 2010. Choosing and designing knowledge assessments: Experience at a new medical school. *Med Teach* 32:578–581. doi: 10.3109/01421591003614858.
20. Yelder J, Wearn A, Chen Y, Henning M, Weller J, Lillis S, Mogol V, Bagg W. A qualitative exploration of student perceptions of the impact of progress tests on learning and emotional wellbeing. *BMC Med Educ.* 2017;17(1):148. doi: 10.1186/s12909-017-0984-2.
21. Dijksterhuis MGK, Scheele F, Schuwirth LWT, Essed GGM, Nijhuis JG. 2009. Progress testing in postgraduate medical education. *Med Teach* 31: e464–e468. doi: 10.3109/01421590902849545.
22. Coombes L, Ricketts C, Freeman A, Stratford J. Beyond assessment: Feedback for individuals and institutions based on the progress test. *Med Teach.* 2010; 32(6):486-490. doi: 10.3109/0142159X.2010.485652
23. Muijtjens AMM, Schuwirth LWT, Cohen-Schotanus J, Thoben AJNM, van der Vleuten CPM. Benchmarking by cross-institutional comparison of student achievement in a progress test. *Med Educ.* 2008; 42(1):82-88. doi: 10.1111/j.1365-2923.2007.02896.x.
24. Van der Vleuten CPM, Schuwirth LWT, Muijtjens AMM, Thoben A, Cohen-Schotanus J, van Boven CPA. Cross institutional collaboration in assessment: A case on progress testing. *Med Teach* 2004; 26(8):719–725. doi: 10.1080/01421590400016464
25. 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.
26. Pasquali L. *Psicometria.* Rio de Janeiro: Editora vozes; 2003.
27. Tarrant M, Ware J, Mohammed AM. An assessment of functioning and non-functioning distractors in multiple-choice questions: A descriptive analysis. *BMC Med Educ.* 2009;9(1):1–8. doi: 10.1186/1472-6920-9-40.