

Avaliação do Crescimento Cognitivo do Estudante de Medicina: Aplicação do Teste de Equalização no Teste de Progresso

Assessment of the Cognitive Growth of the Medicine Student: Applying the Equalization Test to the Progress Test

Marcia Hiromi Sakai¹
Olavo Franco Ferreira Filho¹
Tiemi Matsuo¹

PALAVRAS-CHAVE

- Avaliação
- Educação Médica
- Teste de Progresso
- Psicometria
- Avaliação Educacional

KEYWORDS

- Evaluation.
- Medical Education.
- Progress Testing.
- Psychometrics.
- Educational Measurement.

Recebido em: 01/04/2011

Aprovado em: 05/09/2011

RESUMO

O desenvolvimento cognitivo é avaliado, geralmente, por meio de testes objetivos, e o Teste de Progresso (TPMed) é um dos métodos que as escolas médicas utilizam. A comparação dos resultados é possível com a equalização dos testes. Este trabalho avaliou o desenvolvimento cognitivo dos estudantes de Medicina de uma universidade estadual através da análise dos TPMed no período de 2004 a 2007, que foram equalizados para estimar o progresso dos estudantes durante o curso. A metodologia foi pesquisa-ação e aplicaram-se os métodos para equiparação dos testes, Teoria Clássica dos Testes (TCT) e Teoria de Resposta ao Item (TRI), seguidos pelo método linear por meio do teste de ancoragem para grupos não equivalentes. Observou-se uma média de acerto de 36,9% e 64,4% na primeira e sexta séries, respectivamente. A análise dos resultados baseou-se na psicometria e realizou-se a equiparação dos testes pela TCT, pela impossibilidade de aplicar a TRI. O alfa de Cronbach mostrou alta consistência interna dos testes, e os índices de dificuldade e discriminação dos itens mostraram boa evolução do nível das questões. O crescimento cognitivo dos estudantes de Medicina foi evidente em todos os TPMed.

ABSTRACT

Cognitive development is usually assessed by applying objective tests, and the Progress Test (TPMed) is one of method used by medical schools. The results can be compared through test equalization. This study assessed the cognitive growth of medicine students at a state university by analysing the TPMed tests in the period of 2004 to 2007, using equalization to estimate student progress during the course. The methodology used was action research. Moreover, Classical Test Theory (CTT) and Item Response Theory (IRT) were applied as test equalization methods and, non-equivalent group with anchor test design. It was found 36.9% and 64.4% score in the first and sixth series, respectively. The results were based on psychometric analysis and CTT test equalization was applied, due to IRT being impossible to apply. Cronbach's alpha demonstrated displayed high internal consistency of the tests and the indices of difficulty and item discrimination showed good level of the question. Cognitive growth of medical students was evident in all TPMeds.

INTRODUÇÃO

O desenvolvimento cognitivo é uma importante dimensão na formação do médico^{1,2}. É um processo contínuo de aquisição e consolidação de um conjunto de componentes necessários ao domínio do conhecimento em uma ou mais áreas de desempenho³. Esta dimensão, junto com as habilidades clínicas e de atitudes, é contemplada nos cursos de Medicina, principalmente nos que implantaram mudanças curriculares nas últimas décadas. A escolha pedagógico-metodológica facilitaria ou não a aquisição do conhecimento pelo estudante durante a sua formação.

Assim, os cursos que adotam currículos integrados e metodologias centradas no estudante favorecem o desenvolvimento cognitivo por valorizarem o processo de aprendizagem ao invés do produto, preparando o estudante para aprender a aprender. Neste contexto, o papel da avaliação seria verificar se esse progresso ocorreu ou não, principalmente no aspecto formativo. O processo avaliativo aponta as debilidades e potencialidades dos estudantes nas áreas avaliadas^{4,7}. Além disso, oferece a retroalimentação, possibilitando que o estudante estabeleça planos de estudo para buscar sua formação, em um processo de ação-reflexão-ação⁸. Isso traz um desafio para as escolas médicas, que devem implantar métodos e técnicas avaliativas que permitam acompanhar o desenvolvimento cognitivo, que Miller⁹ chama de avaliação para a aprendizagem, para além das finalidades somativa e certificativa. Portanto, é necessário buscar novos testes e métodos de análise dos resultados^{10,11}.

Verificam-se na literatura diversos métodos e técnicas, desde os testes objetivos escritos até os mais complexos, como o portfólio. Geralmente, recomenda-se a associação de técnicas para verificar se houve ou não o desenvolvimento cognitivo observado na resolução de problemas de saúde, que envolve a tomada de decisão, que requer várias competências complexas em nível de metacognição. Portanto, sugere-se que as técnicas empregadas alcancem altos níveis da taxonomia de Bloom e não somente as de memorização. Entre as diversas técnicas recomendadas, a mais empregada é o teste que contém questões objetivas, devido às vantagens de ser aplicado em larga escala, em processos de certificação dos sujeitos, acreditação de escolas e outras finalidades^{6,7,12,13}.

Educação Médica e Teste de Progresso

Atualmente, no cenário mundial da educação médica, o Teste de Progresso (TPMed) é utilizado para avaliar o desempenho cognitivo dos estudantes durante o curso e o próprio curso. É aplicado, principalmente, pelas escolas médicas que implantaram mudanças curriculares, introduzindo currículos basea-

dos/orientados na comunidade, aprendizagem baseada em problemas e currículos orientados por competências, entre outros, assim como por alguns programas de pós-graduação ou disciplinas isoladas¹⁴⁻¹⁷.

Verificam-se diferenças nos critérios dos testes aplicados pelas diversas escolas médicas, mas todos têm propósitos semelhantes: a avaliação do curso e a avaliação formativa do estudante^{14,15}. O TPMed avalia a relação entre conteúdo e estrutura curricular da graduação e o desenvolvimento cognitivo do estudante, e propicia a este a oportunidade de verificar o desenvolvimento de seu desempenho cognitivo nas diversas áreas do curso e do currículo, e a identificação de seus problemas potenciais. Permite implementar ações para o melhoramento contínuo tanto do estudante quanto do curso.

No Brasil, este teste vem sendo aplicado em algumas escolas médicas, isoladas ou em parceria, desde o final da década de 1990^{17,18}. Além disso, em 2004, o Ministério da Educação introduziu o novo modelo de Exame Nacional de Desempenho dos Estudantes (Enade), cujos objetivos são semelhantes aos do TPMed, mas é aplicado somente aos estudantes ingressantes e concluintes dos cursos, com periodicidade de quatro anos¹⁹.

O TPMed da universidade estadual em estudo é uma avaliação cognitiva longitudinal, construída com o conteúdo final do curso, aplicada a todos os estudantes, independentemente da série, e utiliza questões objetivas. Foi introduzido em 1998 como parte da mudança curricular do curso de Medicina. Nesta fase, os testes foram aplicados em diversos formatos: com questões do tipo múltipla escolha ou verdadeiro/falso; com as áreas de conhecimento contemplando ou não as de ciências básicas; e com periodicidade semestral ou anual. Assim, nos anos iniciais, de 1998 a 2002, os testes não foram padronizados. A partir de 2003, iniciou-se a definição de critérios para a elaboração, aplicação e análise dos resultados dos TPMed nos seguintes aspectos: áreas de conhecimento que integrariam o teste – ciências básicas, clínica médica, clínica cirúrgica, ginecologia e obstetrícia, pediatria e saúde coletiva; número total de questões; questões do tipo múltipla escolha; e periodicidade anual¹⁷.

Até 2006, todas as questões foram elaboradas na própria escola; após esse ano, a universidade passou a participar do Núcleo Interinstitucional de Estudos e Práticas de Avaliação em Educação Médica (Niepaem), constituído por mais oito escolas médicas brasileiras. Assim, atualmente, o TPMed é elaborado, aplicado e analisado em parceria com o Niepaem e composto por 120 questões de múltipla escolha com cinco opções. A maioria dos itens apresenta nível de maior complexidade pela taxonomia de Bloom e do tipo normorreferenciado.

Além das áreas de conhecimento citadas acima, distribuídas equitativamente, o teste é composto por questões de ética médica. Todos os itens têm coerência com o currículo nuclear do curso e com as diretrizes curriculares nacionais para o curso de Medicina. Para os estudantes do curso, o teste é obrigatório, sendo realizado anualmente para todos os estudantes da primeira à sexta série. O teste é aplicado no mesmo dia e horário nas nove escolas que compõem o núcleo, e a retroalimentação aos estudantes é realizada por meio da entrega do caderno de questões e da divulgação do gabarito em edital.

Avaliação do Crescimento Cognitivo

Durante o curso, o estudante foi submetido a diferentes testes e obteve resultados distintos em cada um. Utilizando os escores médios obtidos pelos estudantes, poderia se afirmar que houve crescimento?

Considere-se que as medidas foram realizadas em tempos diferentes. Nesse intervalo de tempo, os estudantes receberam novas informações, e as provas foram diferentes, mesmo com a aplicação de critérios definidos para avaliar os mesmos conteúdos e habilidades. Assim, como medir o crescimento cognitivo dos estudantes por meio de testes diferentes?

A equiparação ou equivalência de escores é um processo estatístico que serve para comparar escores em formas diferentes de teste, ajustando níveis diferentes de dificuldade e de conteúdo entre os testes²⁰.

O propósito deste artigo foi avaliar o desenvolvimento cognitivo dos estudantes de Medicina de uma universidade estadual por meio da análise dos TPMed aplicados no período de 2004 a 2007, os quais foram equalizados para melhor estimar o progresso da aquisição do conhecimento pelos estudantes da primeira à sexta série.

METODOLOGIA

A trajetória metodológica deste estudo se baseou na busca das respostas aos questionamentos gerados pela implantação do currículo integrado, em 1998, e na análise das intervenções introduzidas nesse processo de mudança, em um movimento de ação-reflexão-ação. A pesquisadora e seus orientadores são sujeitos participantes ativos desse processo de mudança.

O desenho da pesquisa foi pesquisa-ação, e o estudo partiu da seguinte questão: qual foi o grau de crescimento cognitivo dos estudantes de um ano para outro no currículo integrado? Para isso, o estudo teve dois momentos, que contemplaram o planejamento e a implantação das ações, análise dos fatos observados, reflexão sobre as possíveis causas dos fatos observados; e replanejamento, implantação de novas ações e análise dos dados encontrados, com base no ciclo de Elliott²¹.

Os momentos da pesquisa foram caracterizados pelos distintos métodos utilizados para responder à pergunta, com base nos resultados dos TPMed obtidos pelos estudantes do curso de Medicina. Foram analisados todos os resultados dos estudantes da primeira à sexta série acadêmica no período de 2004 a 2007. A média de participação dos estudantes nos TPMed ficou acima de 90%. Observou-se menor participação dos internos (quinta e sexta séries), de 81,7% até 84,8%. Foram excluídos os escores dos estudantes que responderam ao teste de forma inadequada, como, por exemplo, os daqueles que assinalaram uma única opção para todos os itens.

Momento 1 – Análise dos Testes

A opção deste estudo foi se basear na psicometria, que procura explicar o sentido que têm as respostas dadas pelos estudantes a uma série de itens (questões). Utilizamos os dois modelos psicométricos: a Teoria Clássica dos Testes (TCT) e a Teoria de Resposta ao Item (TRI).

Teoria Clássica dos Testes (TCT)

A TCT busca explicar o resultado final total, ou seja, a soma das respostas dadas ao teste. Este modelo analisa apenas dois parâmetros do item (grau de dificuldade e de discriminação). Esta análise se baseia no escore total de acertos e erros, sem considerar as habilidades latentes dos estudantes e o acerto casual. A TCT analisa testes de qualidade²². Para esta etapa da análise, foi utilizado o programa Iteman 3.5. Foi considerado item de boa qualidade aquele que tinha correlação bisserial (correlação entre o acerto em determinado item e o número total de acertos na prova) acima de 0,40.

Teoria de Resposta ao Item (TRI)

A TRI não está interessada no escore total de um teste, mas, sim, no resultado obtido por cada um dos itens que compõem o teste e quer saber qual é a probabilidade e quais são os fatores que afetam a probabilidade de o estudante acertar ou errar um item de determinado teste, considerando os três parâmetros do item (grau de dificuldade, capacidade discriminativa e probabilidade de acerto casual) e as habilidades latentes do indivíduo. A TRI se preocupa em analisar itens de qualidade²².

O modelo da TRI testada foi o logístico unidimensional de três parâmetros para itens de múltipla escolha, dicotomizados em certo ou errado. Os parâmetros são: (a) capacidade discriminativa do item (discrimina os estudantes com maior ou menor habilidade cognitiva); (b) grau de dificuldade do item (mede se o item é fácil ou difícil); (c) acerto casual (capacidade de o estudante acertar o item pelo “chute”).

Para que os programas computacionais possam processar os cálculos, é necessário que cada respondente seja submetido a pelo menos 30 itens e cada item seja respondido por pelo menos 300 estudantes. Utilizamos o programa Bilog 3.0, que só consegue rodar na sua integralidade se todos os itens tiverem boa correlação bisserial. Para tanto, em primeiro lugar, realizou-se a análise pela TCT e, por meio desta, foram eliminadas as questões ruins.

A análise pela TRI seguiu estes passos: as respostas dos estudantes eram modificadas para itens dicotômicos, onde 0 = errado e 1 = certo; roda-se o programa e escolhem-se os itens com melhores parâmetros A, B e C; escolhem-se os itens-âncora: a questão foi considerada como âncora quando respondida corretamente por uma grande população de estudantes (pelo menos 65%) com o nível de habilidade esperado e com pouco acerto por indivíduos com habilidade imediatamente inferior²³.

Ao se utilizar este programa computacional para analisar os dados dos estudantes de Medicina da UEL pela TRI, não foi possível ir até a etapa final, pois o número de itens com correlação bisserial acima de 0,40 ficou abaixo do recomendado e não tínhamos um número adequado de respondentes por grupo. Mas foi possível definir as questões-âncora que seriam utilizadas no momento seguinte.

Assim, o crescimento do conhecimento cognitivo dos estudantes de Medicina da universidade em estudo foi analisado pela TCT. A análise dos testes foi realizada com uso do alfa de Cronbach, e os itens foram avaliados utilizando-se os índices de dificuldade (porcentagem de acertos) e discriminação (que verifica a capacidade do item para diferenciar os 27% dos estudantes de maior habilidade em responderem corretamente as questões dos 27% dos estudantes de menor habilidade e, portanto, com os menores acertos), baseados na TCT^{12,22,24}.

De acordo com o grau de dificuldade, foi considerada como questão fácil aquela com taxa de acerto igual ou superior a 80%; como questão intermediária, aquela com taxa de acerto entre 20% e 80%; e como questão difícil, aquela com taxa de acerto igual ou inferior a 20%²⁴.

De acordo com a capacidade discriminativa, foram consideradas como questões de pobre capacidade aquelas cuja diferença de porcentagem de acertos entre os melhores e piores estudantes foi menor que 19%; como questões que deveriam ser melhoradas, aquelas cuja diferença estava entre 20% e 29%; como questões de boa capacidade discriminativa, aquelas cuja diferença estava entre 30% e 39%; e como questões de ótima capacidade, aquelas cuja diferença entre os melhores e piores estudantes era igual ou superior a 40%²⁴.

Momento 2 – equalização dos testes

Após selecionarem o método de análise do crescimento cognitivo dos estudantes, os responsáveis pela aplicação do TPMed passaram a estudar os métodos existentes para equiparação dos testes aplicados. O escolhido foi o método linear por meio do teste de ancoragem para grupos não equivalentes, que obedeceu às seguintes etapas: (a) escolha dos itens-âncora: escolhidos no teste aplicado em 2004 por meio da análise pela TRI; (b) elaboração e aplicação dos TPMed em 2005 e 2006, compostos de 120 questões, sendo 10% delas de itens-âncora, distribuídos equitativamente em cada área de conhecimento²³; assim, nos testes de 2005 e 2006, existiam 12 questões idênticas e comuns que permitiram aplicar o teste de ancoragem; as outras questões eram escolhidas pela equipe elaboradora do teste; (c) análise dos resultados: procedeu-se à equalização dos TPMed aplicando-se o teste de ancoragem segundo Angoff²⁵; (d) construção das curvas de crescimento.

Utilizamos a estatística descritiva para apresentar os dados em forma de figuras.

O projeto foi aprovado pelo Colegiado do Curso de Medicina em 23 de novembro de 2005 e pelo Comitê de Ética e Pesquisa em Seres Humanos da mesma instituição, conforme o parecer Cepe 188/06.

RESULTADOS

Os escores apresentados são as médias da porcentagem de acertos dos estudantes antes (Figura 1) e depois (Figura 2) da aplicação do teste de ancoragem (equalização), nos TPMed de 2004 a 2007. Neste artigo não apresentaremos os resultados segmentados por área de conhecimento.

A média de acerto dos estudantes, em todos os TPMed, 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 TPMed com equalização variaram de 31,0% até 73%, para a primeira e sexta séries, respectivamente (Figuras 1 e 2). A média de acertos de cada série ao longo dos quatro testes, sem equalização, foi de 36,9%, 42,8%, 47,7%, 52,2%, 57,4% e 64,4% para a primeira, segunda, terceira, quarta, quinta e sexta séries, respectivamente. Com equalização, foi de 38,9%, 45,0%, 50,0%, 54,6%, 59,9% e 66,9% para as seis séries, respectivamente. O escore obtido pelos estudantes da primeira série do TPMed 2007 chamou a atenção pelo valor, que foi de 41,4% e 48,2%, nos testes sem e com equalização, respectivamente.

Os desempenhos dos estudantes nos TPMed de 2004 a 2007 mostraram uma mudança na média dos percentuais de acerto e no formato da curva dos escores nos resultados com a equalização em relação àqueles sem equalização (Figuras 1 e 2). A curva de desempenho de 2005 permanece a mesma para

FIGURA 1
Desempenho cognitivo dos estudantes nos TPMed de 2004 a 2007, sem equalização

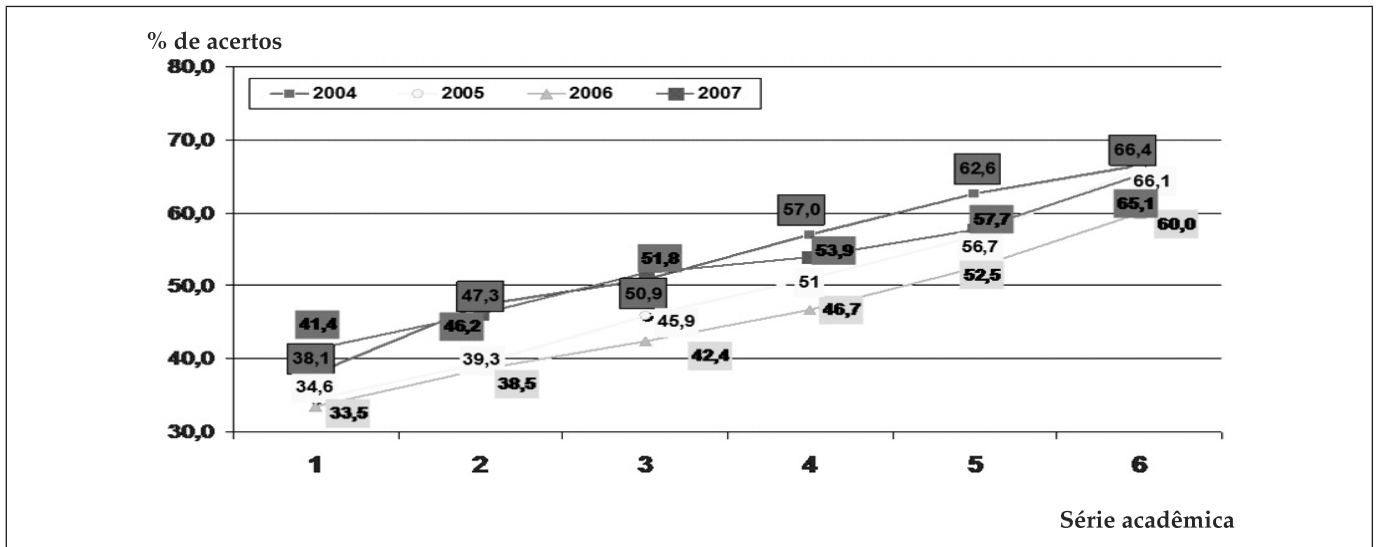
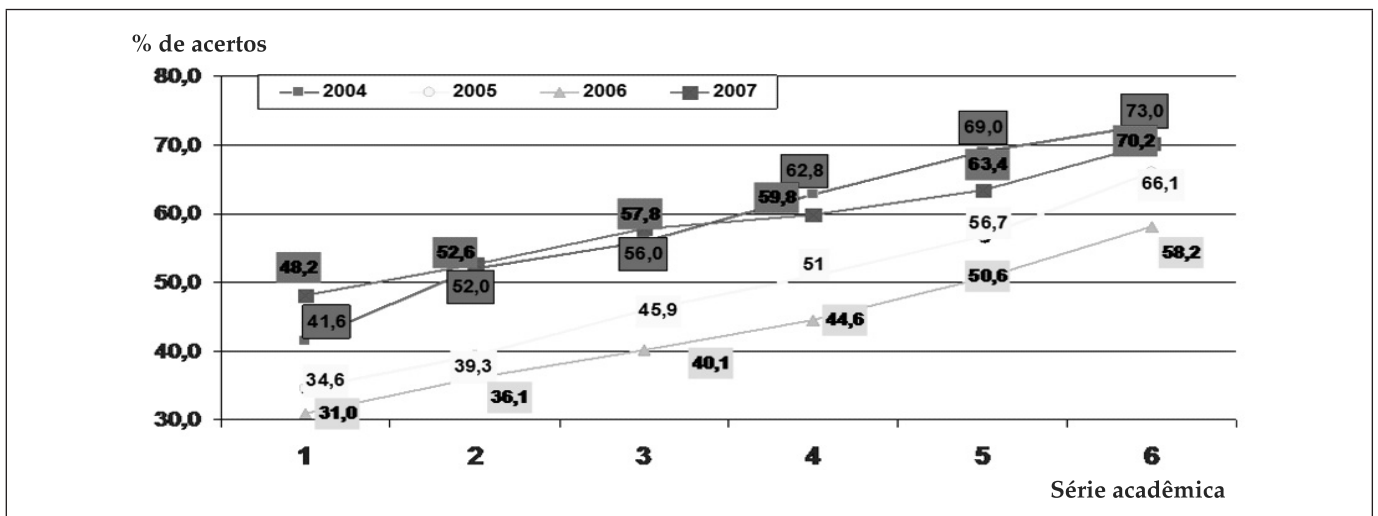


FIGURA 2
Desempenho cognitivo dos estudantes nos TPMed de 2004 a 2007, com equalização



os testes sem e com equalização, pois este teste foi escolhido como base para a equiparação.

Os testes apresentaram um alfa de Cronbach (coeficiente de fidedignidade) de 0,825, 0,906, 0,872 e 0,834, respectivamente, para os TPMed 2004, 2005, 2006 e 2007. Este foi o motivo da escolha do teste de 2005 como base para a equiparação, pois apresentava o melhor coeficiente de fidedignidade.

O índice de dificuldade dos itens mostrou que a maioria das questões dos TPMed foi considerada de nível intermediário e a dificuldade foi aumentando a cada ano. No TPMed

2006, houve redução no percentual de questões fáceis em relação aos anos anteriores (Figura 3).

Os valores do índice de discriminação foram melhorando a cada teste, pois se verificou um aumento na proporção de questões ótimas e boas, que variaram de 23,7% a 47,6% na composição dos testes, e redução no percentual de itens pobres, de 54,6% para 36,2%. O TPMed 2005 foi o que teve a menor proporção de questões pobres e apresentou a maior porcentagem de itens de ótima capacidade discriminativa (Figura 4).

FIGURA 3
Distribuição dos itens dos TPMed 2004, 2005, 2006 e 2007 de acordo com o grau de dificuldade

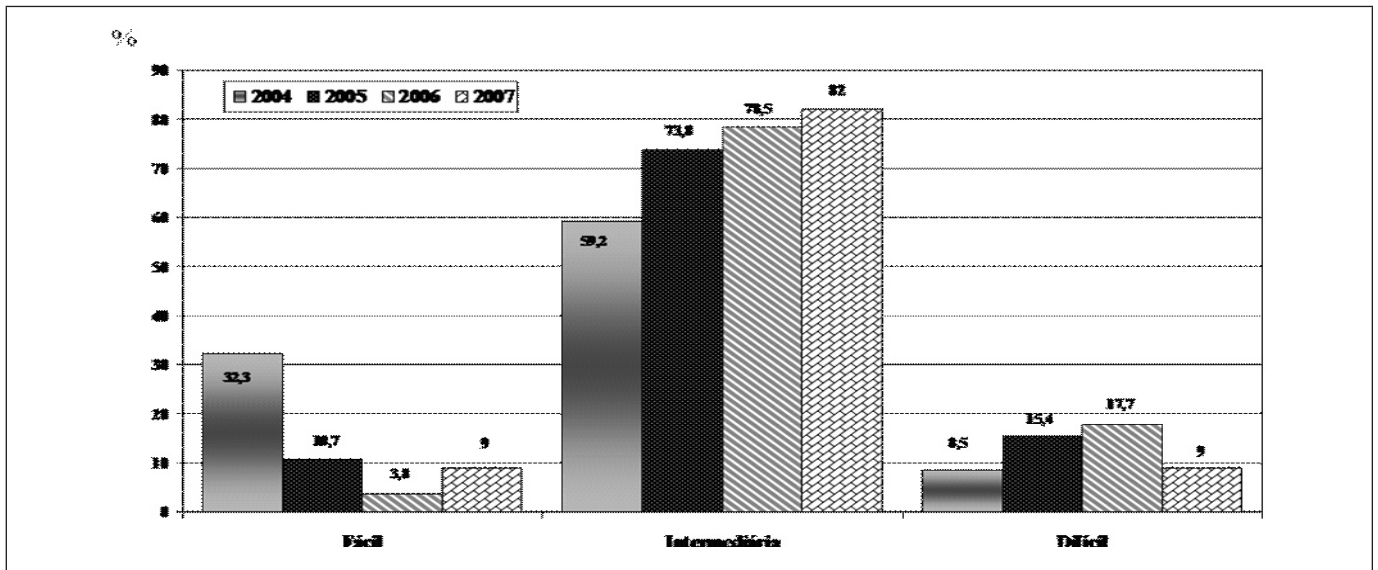
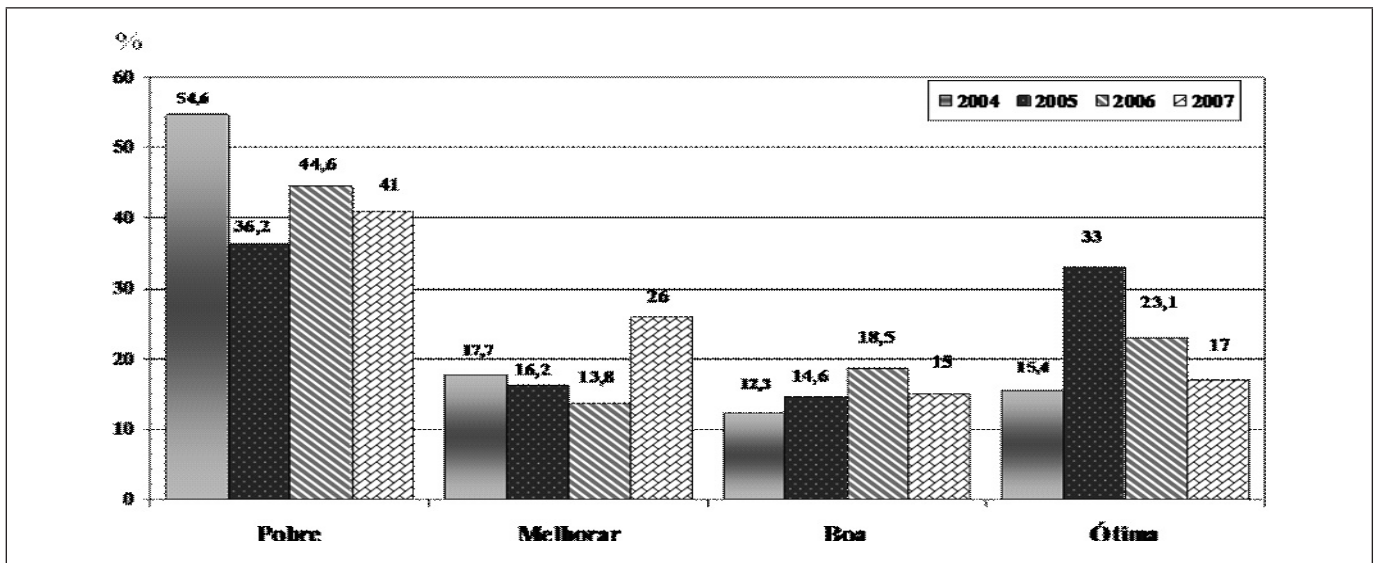


FIGURA 4
Distribuição dos itens dos TPMed 2004, 2005, 2006 e 2007 de acordo com o índice de discriminação



DISCUSSÃO

Optou-se, neste estudo, pela psicometria para a análise dos resultados, pois esse método leva em conta a aquisição do conhecimento como uma competência a ser desenvolvida e as inovações curriculares introduzidas pelo curso de Medicina em estudo, além de permitir comparar os diferentes testes aplicados a grupos também diferentes^{10,11,23}. Tal escolha diferiu daquela de outras escolas médicas que utilizaram métodos

como testes estatísticos simples, Anova e *mixed longitudinal design* também para verificar se houve ou não crescimento cognitivo dos estudantes^{1,18,26-29}.

Neste estudo, não foi possível aplicar a TRI, embora Embretson¹¹ e Andrade²³ enfatizem que se trata do modelo recomendado para ser aplicado nas avaliações dos testes objetivos, ao invés da TCT, por considerar as habilidades latentes dos estudantes e por 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 teste de progresso, este é um fator de grande importância, pois, por se tratar de um teste elaborado em nível de sexta série e 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 itens.

Ao se analisar a qualidade dos TPMed, observa-se que houve uma melhora contínua, pois o alfa de Cronbach de todos eles mostrou alta consistência interna dos testes, e os índices de dificuldade e de discriminação dos itens mostraram boa evolução do nível das questões²². Portanto, os testes podem ser considerados de ótima qualidade.

A tendência de as escolas médicas formarem parcerias para elaborar e aplicar os testes aponta uma melhoria da qualidade da prova, pois isto permite a troca de experiências entre elas, a constituição de um grupo de trabalho com as especialidades envolvidas e a montagem de um banco de itens com questões testadas e avaliadas. Isto não só facilitaria o processo de seleção das questões para compor a prova com base em critérios preestabelecidos, como otimizaria o uso dos recursos financeiros^{16,17,30}. 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 testes²³. Assim, as análises levariam em conta as habilidades latentes dos estudantes que seriam mais adequadas aos itens que avaliam o raciocínio clínico e a resolução de problemas de saúde, que geralmente compõem os testes de progresso.

Entretanto, por mais que se reconheçam e valorizem os traços latentes dos indivíduos, mesmo a TRI apresenta uma limitação de uso, pois, de acordo com os autores^{11,22,23}, os modelos de um, dois ou três parâmetros são unidimensionais, ou de um traço dominante, e não permitem a análise das questões de caráter interdisciplinar para a solução de problemas, cujo emprego tem se tornado frequente, o uso de itens interdisciplinares que propiciam a integração das sete áreas de conhecimento do TPMed e a elaboração de questões com alta taxonomia de Bloom.

Já quanto à seleção das questões-âncora, o procedimento utilizado neste trabalho mostrou-se adequado e otimizador. Fez-se uma primeira seleção utilizando a TCT, que eliminou as questões com baixo poder discriminatório, para em seguida aplicar a TRI e obter os itens-âncora. Além disso, a avaliação criteriosa de cada item pela TRI levaria à melhoria da qualidade da prova em geral e do banco de questões.

O crescimento cognitivo dos alunos foi evidente em todos os testes aplicados, sendo, em média, de 36,9% na primeira

série e 64,4% na sexta série, ou seja, um ganho de conhecimento em torno de 30% ao longo do curso de Medicina. Desde o início da aplicação do teste de progresso na UEL, encontramos a taxa de acerto dos estudantes da primeira série muito elevada, como também estranhamos o “pouco” ganho após seis anos de curso. Estas inquietações foram amenizadas ao encontramos trabalhos nacionais que obtiveram o mesmo resultado, como o estudo de Tomic *et al.*¹⁸, que mostraram sua experiência com a aplicação do teste durante quatro anos na Universidade de São Paulo, uma tradicional escola de Medicina do País. A alta taxa de acertos dos estudantes da primeira série no teste de 2007 pode ser explicada pela melhor qualidade dos itens aplicados, conforme mostram as Figuras 3 e 4; ou pode ser atribuída ao currículo integrado, que desde a primeira série tem conteúdos baseados no perfil epidemiológico; ou, então, foi mero acaso.

A equiparação dos testes com aquele aplicado em 2005 mostrou que o teste de 2006 foi mais difícil, tendo em vista que os escores foram inferiores; os de 2004 e 2007 foram mais fáceis, o que também pode reforçar a justificativa de os estudantes da primeira série terem altas taxas de acerto em 2007.

Esses resultados mostraram que os estudantes apresentaram crescimento cognitivo progressivo de uma série acadêmica para as outras e, também, que o currículo integrado não comprometeu a aquisição de conhecimento do estudante, preocupação inicial da comunidade docente do curso de Medicina.

Não encontramos na literatura qualquer trabalho que utilizasse o método estatístico de equiparação em testes de progresso aplicados em momentos diferentes, somente alguns autores que salientaram a importância do uso da TRI^{27,31,32}. Ressaltamos que o método utilizado neste trabalho (teste da ancoragem) é preconizado para testes homogêneos aplicados a estudantes da mesma série, como, por exemplo, uma prova de biologia aplicada a alunos do segundo ano do ensino médio.

Os valores encontrados nas curvas de crescimento nos propiciaram uma delimitação de valores mínimos e máximos de escores esperados para cada série acadêmica. Isto poderá ser considerado um dos indicadores quantitativos de avaliação do curso, bem como poderá fornecer um *feedback* ao estudante se o desempenho observado estiver dentro do esperado para a série. O estabelecimento desses valores poderá ser mais bem evidenciado com os resultados de uma série histórica de seis anos. Permitiria, ainda, acompanhar uma coorte de estudantes que fizerem todos os testes analisados. Porém, à medida que se estratificam mais os resultados, a análise é dificultada pela redução do ‘n’.

Além disso, as análises por série e área de conhecimento, que não foram apresentadas neste trabalho, oferecem uma re-

troalimentação ao estudante, que poderá estabelecer um plano de estudo em busca de seu próprio melhoramento contínuo e, também, do curso.

Um aspecto positivo foi a adesão dos estudantes à participação no TPMed, que não tem finalidade somativa e de progressão. Isto foi possível por dois motivos: o caráter obrigatório desde 2003 e o trabalho intenso de divulgação da importância e das vantagens do teste no processo de aprendizagem dos estudantes, pois o *feedback* é imediato.

CONCLUSÃO

Ocorreu crescimento cognitivo entre estudantes de Medicina da universidade estadual da primeira à sexta série. A média de acerto dos estudantes, em todos os TPMed, variou de um mínimo de 33,5%, observado na primeira série, até um máximo de 66,4%, na sexta série, nos resultados sem equalização. Já os escores médios nos TPMed com equalização variaram de 32,1% até 73%, para a primeira e sexta séries, respectivamente, com um ganho aproximado de 30%.

CONSIDERAÇÕES FINAIS

A institucionalização do TPMed foi importante e fundamental para a consolidação do currículo integrado de Medicina, pois confirmou o ganho cognitivo dos estudantes.

Nossa pesquisa é pioneira e deve ser lapidada ao longo do tempo. No entanto, apontamos que uma parceria entre as escolas médicas favoreceria a realização e a análise dos resultados com a aplicação da TRI.

AGRADECIMENTOS

Aos membros atuais e anteriores da Comissão de Avaliação do Curso de Medicina; aos coordenadores, atual e predecessores, do Colegiado do Curso de Medicina, pela implantação e institucionalização do Teste de Progresso; e ao professor Dalton Francisco Andrade, pelas contribuições para a análise do TPMed.

Apoio: Promed-LD.

REFERÊNCIAS

- Verhoeven BH, Verwijnen GM, Scherpbier AFFA, Vleuten CPMvd. Growth of medical knowledge. *Med Educ*. 2002;36:711-7.
- Miller GE. The assessment of clinical skills/competence/performance. *Acad Med*. 1990;65(9 Suppl):S63-7.
- Sternberg RJ. Intelligence as developing expertise. *Contemporary Educational Psychology*. 1999;24:359-75.
- Perrenoud P. Avaliação: da excelência à regulação das aprendizagens entre duas lógicas. Porto Alegre: Artmed; 1999.
- Luckesi CC. Avaliação da aprendizagem escolar. São Paulo: Cortez; 1995.
- Gronlund N. Assessment of student achievement. 6 ed. Boston: Allyn and Bacon; 1998.
- Anastasiou LGC, Aleves LP, Wachowicz L, org. Processo de ensinagem na universidade: pressupostos para as estratégias de trabalho em aula. 5 ed. Joinville: Ed. Univille; 2005.
- Schön DA. The reflective practitioner: how professionals think in action. London: Ashgate; 2002.
- Miller GE. Continuous assessment. *Med Educ*. 1976;10:81-6.
- Sternberg RJ. Ability testes, measurements, and markets. *Journal of Educational Psychology*. 1992;84(2):134-40.
- Embretson SE. The new rules of measurement. *Psychological Assessment*. 1996;8(4):341-9.
- Linn RL, Gronlund N. Measurement and assessment in teaching. 8 ed. Upper Saddle River, New Jersey: Prentice Hall; 2000.
- Coll C, Pozo JI, Sarabia B, Valls E. Os conteúdos na reforma: ensino e aprendizagem de conceitos, procedimentos e atitudes. Porto Alegre: Artmed; 2000.
- Van Der Vleuten CPM, Verwijnew GM, Wijnen WHFW. Fifteen years of experience with progress testing in a problem-based learning curriculum. *Medical Teacher*. 1996;18(2):103-9.
- Blake JM, Norman GR, Keane DR, Mueller CB, Cunningham J, Didyk N. Introducing progress testing in McMaster University's problem-based medical curriculum: psychometric properties and effect on learning. *Acad Med*. 1996;71(9):1002-7.
- Van Der Vleuten CPM, Schuwirth LWT, Muijtens AMM, Thorben AJNM, Cohen-Schotanus J, Van Boven CPA. Cross institutional collaboration in assessment: a case on progress testing. *Medical Teacher*. 2004;26(8):719-25.
- 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). Disponível em: <http://www.scielo.br/pdf/rbem/v32n2/a14v32n2.pdf>
- Tomic ER, Martins MA, Lotufo PA, Bonsenor IM. Progress testing: evaluation of four years of application in the medical school of medicine, University of São Paulo. *Clinics*. 2005;60(5):389-96.
- Limana A, Brito MRFd. O modelo de avaliação dinâmica e o desenvolvimento de competências: algumas

- considerações a respeito do ENADE. *Rev Avaliação*. 2005;10(2):9-32.
20. Pasquali L. Equiparação de escores. In: *Psicometria: teoria dos testes na psicologia e na educação*. Petrópolis: Vozes; 2003.
21. Robson C. *Real world research: a resource for social scientists and practitioner-researchers*. Oxford: Blackwell; 1999.
22. Pasquali L. Os modelos da psicometria: TCT e TRI. In: *Psicometria: teoria dos testes na psicologia e na educação*. Petrópolis: Vozes; 2003.
23. Andrade DF. Comparando desempenho de grupos de alunos por intermédio da teoria de resposta ao item. *Estudos em Avaliação Educacional*. 2001;23:31-69
24. Oermann MH, Gaberson KB. *Evaluation and testing in nursing education*. New York: Springer Publishing Company; 1998.
25. Angoff WH. Scales, norms, and equivalent scores. Princeton, NJ: Angoff, WH and Ford, SF. Item-race interaction on a test of scholastic aptitude. *Journal of Educational Measurement*. 1984;10:95-105.
26. Diest RV, Dalen JV, Bak M, Schruers K, Vleuten CVD, Muijtjens A, et al. Growth of knowledge in psychiatry and behavioral sciences in a problem-based learning curriculum. *Med Educ*. 2004;38:1295-301.
27. McHarg J, Bradley P, Chamberlain S, Ricketts C, Searle J, McLachlan JC. Assessment of progress tests. *Med Educ*. 2005;39:221-7.
28. Basu S, Roberts C, Newble DI, Snaith M. Competence in the musculoskeletal system: assessing the progression of knowledge through an undergraduate medical course. *Med Educ*. 2004;38:1253-60.
29. Gurpinar E, Musal B, Aksakoglu G, Ucku R. Comparison of knowledge scores of medical students in problem-based learning and traditional curriculum on public health topics. *BMC Medical Education* [periodico na internet]. 2005[acesso em 5 ago.2010];5(7):2-8. Disponível em: <http://www.biomedcentral.com/1472-6920/5/7>.
30. Verhoeven BH, Snellen-Balendong HAM, Hay IT, Boon JM, Van Der Linde MJ, Blitz-Lindeque JJ, et al. The versatility of progress testing assessed in an international context: a start for benchmarking global standardization? *Med Teach*. 2005;27(6):514-20.
31. Downing SM. Item response theory: applications of modern test theory in medical education. *Med Educ*. 2003;37:739-45.
32. Guilera G, Gómez J. Item response theory test equating in health sciences education. *Adv Health Sci Education*. 2008;13:3-10.

CONTRIBUIÇÃO DOS AUTORES

Todos os autores participaram na concepção e desenho deste estudo, da análise e interpretação dos dados, assim como da redação deste texto.

CONFLITO DE INTERESSES

Declarou não haver.

ENDEREÇO PARA CORRESPONDÊNCIA

Marcia Hiromi Sakai
Rua Belo Horizonte, 1126 – apto 1001
Centro – Londrina
CEP 86020-060 – PR
E-mail: sakai.marcia@gmail.com