

O TESTE DE PROGRESSO COMO INDICADOR PARA MELHORIAS EM CURSO DE GRADUAÇÃO EM MEDICINA

Maria Inês da Rosa¹, Camila Carminati Isoppo², Kristian Madeira³, Olavo Franco Ferreira Filho⁴.

¹Ph.D., Laboratório de Epidemiologia da UNESC.

²Acadêmica de medicina da UNESC.

³MSc, professor do curso de Medicina da UNESC.

⁴Ph.D., professor da UEL, PR.

Autor Correspondente:

Maria Inês da Rosa

Rua Cruz e Souza, 510. Bairro Pio Correa. CEP 88811-550

Criciúma-SC - Brasil

mir@unesc.net

INTRODUÇÃO

A avaliação do conhecimento permanece um grande problema na educação médica. O Teste do Progresso (TP), muito popular na Holanda e em outros países há vários anos, somente agora esta sendo valorizado como ferramenta de avaliação nas instituições de ensino brasileiras. O TP tem sido utilizado por escolas médicas que implementaram mudanças curriculares, além de alguns programas de pós-graduação ou disciplinas isoladas (Aarts et al 2010, Bennett et al 2010, Freeman et al 2010, Nouns et al 2010, Schuwirth et al 2010, Swanson et al 2010).

Esta técnica foi aplicada pela primeira vez no início dos anos 1970, na Universidade de Missouri - *Kansas City School of Medicine e na Holanda – Universidade de Limburg* (Blake et al 1996).

O TP é uma forma de análise longitudinal, no qual participam os estudantes de todos os anos de graduação, elaborado com questões a partir do conhecimento considerado fundamental para o último ano de curso. O teste consiste em perguntas sobre um vasto campo do conhecimento médico, cujo conteúdo é baseado nas diretrizes curriculares nacionais para o curso de Medicina, instituídas pelo Ministério da Educação e Cultura. A literatura aponta diferentes maneiras de fazer o TP (Van der Vleuten et al 1996, Aarts et al 2010, Bennett et al 2010, Freeman et al 2010, Nouns et al 2010, Schuwirth et al 2010, Swanson et al 2010).

No Brasil, o primeiro Núcleo Interinstitucional em Educação Médica, formado por sete escolas paulistas, uma do Paraná e uma de Santa Catarina, está na décima edição do Teste de Progresso. O segundo grupo, Núcleo de Apoio Pedagógico Interinstitucional Sul II - NAPISUL II, formado por seis escolas catarinenses e três paranaenses, está em sua quarta edição. Com o apoio a Associação Brasileira de Educação Médica, outros núcleos têm sido formados em outros estados brasileiros. Em ambos os grupos citados, o TP é elaborado com 120 questões de múltipla escolha, com quatro alternativas, englobando as principais áreas de Medicina (Clínica Médica, Cirurgia, Pediatria, Ginecologia e Obstetrícia, Saúde Coletiva e Ciências Básicas, além de questões de Bioética). Cada instituição participante tem um coordenador geral do Teste de Progresso, que prepara, junto com seus pares, perguntas em formato específico, obedecendo aos rigores pedagógicos e as matrizes de referência, apresentando para cada questão a sua justificativa e bibliografia, que serão disponibilizados aos estudantes após a aplicação da prova. Os coordenadores e professores especialistas em cada área de conhecimento, constituintes do teste, se reúnem em data específica para a elaboração da prova final. Esta prova é aplicada, no Napisul-II, na primeira semana de outubro, simultaneamente em todas as instituições.

A Faculdade de Medicina da Universidade do Extremo Sul Catarinense (UNESC), uma das componentes do Napisul-II, participou de todas as edições do TP aplicados de 2011 a 2014. O objetivo deste estudo foi verificar se o resultado do Teste de Progresso, aplicados nas edições de 2011, 2012 e 2013, pode ser utilizado como indicador para melhorar o curso de graduação em Medicina, baseando-se na análise das três provas aplicadas na UNESC.

MÉTODOS

Foram realizados três estudos observacionais transversais de agregados institucionais. Foram analisados todos os resultados dos estudantes, que são convidados a participar da prova de maneira voluntária, da 1ª à 12ª fases acadêmicas, no período de 2011 a 2013. Foram excluídos os escores dos estudantes que responderam ao teste de forma inadequada, como por exemplo, assinalaram uma única alternativa para todos os itens.

O grau de confiabilidade da prova foi medido pelo cálculo do α de Cronbach.

Os escores apresentados são as médias da porcentagem de acertos dos estudantes na prova total e em cada área do conhecimento, tendo como variável preditora a fase acadêmica. Não havia diferenças significativas de acordo com o sexo. Para comparação das médias dos escores foi utilizado o teste de *Mann-Whitney-Wilcoxon*, com nível de significância de 0,05.

A pesquisa foi aprovada pelo Comitê de Ética e Pesquisa em Seres Humanos da Universidade do Extremo Sul Catarinense, sob o parecer nº: 625.470/2014.

RESULTADOS

A média de participação dos estudantes nos TP foi acima de 90%. A confiabilidade das provas, consistência interna (α de Cronbach), foi acima de 80% em todos os anos.

No primeiro teste, em 2011, o curso de medicina da UNESC tinha 419 alunos matriculados e 98,3% participaram do Teste de Progresso, sendo 60,2% do sexo feminino.

Comparada à média geral das escolas do NAPISUL II, a UNESC teve desempenho médio significativamente menor nas fases nona ($p = 0,027$), décima ($p = 0,046$) e décima segunda ($p = 0,022$) e não apresentou diferença estatisticamente significativa nas demais fases ($p > 0,05$) (Tabela 1).

Em relação ao desempenho por grande área em 2011, a oitava fase da UNESC apresentou número de acertos significativamente maior ao NAPISUL II em Bioética ($p = 0,011$), enquanto demais fases não apresentaram diferença estatística significativa (Tabela 2).

Em Ciências Básicas não houve diferença significativa entre os dois grupos (Tabela 3). Na Cirurgia os estudantes apresentaram desempenho significativamente maior nas fases quarta ($p = 0,001$) e quinta ($p = 0,038$) e menor na nona ($p = 0,002$) e décima segunda ($p = 0,001$) fases (Tabela 4). Já em Clínica Médica, a oitava fase da UNESC teve um maior desempenho quando comparado às oitavas fases componentes do NAPISUL II ($p < 0,001$) (Tabela 5).

A UNESC apresentou resultado significativamente maior na quinta ($p = 0,001$) e sexta ($p < 0,001$) fase na área de Ginecologia e Obstetrícia, e significativamente menor ($p = 0,014$) na quarta fase (Tabela 6). Em Pediatria não foi demonstrado diferenças estatisticamente significativas em comparação ao restante do NAPISUL II (Tabela 7).

Na grande área de Saúde Coletiva, apresentou desempenho significativamente menor nas fases quarta ($p = 0,042$), quinta ($p = 0,024$), sexta ($p = 0,005$), nona ($p = 0,012$), décima ($p = 0,001$) e décima segunda ($p = 0,013$); nas demais fases não houve diferença significativa (Tabela 8).

No segundo teste, em 2012, o curso de medicina da UNESC tinha 434 alunos matriculados e 97,4% participaram do Teste de Progresso, sendo 60% do sexo feminino.

Comparada à média geral das oito escolas que compõe o NAPISUL II, a UNESC teve desempenho médio significativamente maior nas fases oitava ($p = 0,003$) e décima ($p = 0,022$) e menor nas fases décima primeira ($p = 0,002$) e décima segunda ($p = 0,002$), e não apresentou diferença estatisticamente significativa nas demais fases ($p > 0,05$) (Tabela 1).

Comparando o desempenho por grande área entre UNESC e o NAPISUL II, em Bioética (Tabela 2) a UNESC apresentou um número de acertos significativamente maior na primeira ($p < 0,001$) e oitava ($p = 0,030$) fases, enquanto na terceira ($p = 0,007$), quarta ($p = 0,002$) e quinta ($p = 0,002$) fases obteve um desempenho inferior ao NAPISUL II.

Em Ciências Básicas houve um maior desempenho da UNESC nas fases segunda ($p = 0,014$) e oitava ($p = 0,030$) e menor na quarta ($p = 0,048$) e décima primeira ($p = 0,034$) (Tabela 3).

Na Clínica Cirúrgica os estudantes tiveram desempenho significativamente menor nas fases segunda ($p = 0,012$), décima primeira ($p = 0,002$) e décima segunda ($p = 0,050$) da UNESC em relação ao NAPISUL II, mas significativamente maior na segunda ($p = 0,012$) e oitava fase ($p = 0,001$) (Tabela 4). Os acertos dos estudantes na Clínica Médica apresentaram diferença estatística na oitava ($p = 0,001$) e décima fase ($p = 0,038$) com desempenhos superiores e desempenho inferior na décima segunda ($p = 0,001$) em relação ao NAPISUL.

Na grande área de Ginecologia e Obstetrícia (Tabela 6), a UNESC apresentou resultado significativamente maior na quinta e sexta fase ($p = 0,009$ e $p = 0,002$ respectivamente), porém diferença estatística significativa menor nas fases quarta ($p = 0,027$) e décima segunda ($p = 0,003$). Em Pediatria a diferença foi significativamente menor apenas na décima segunda fase ($p = 0,03$) e maior na quarta, oitava e décima fase ($p = 0,034$, $p = 0,002$ e $p = 0,010$ respectivamente). (Tabela 7).

Em de Saúde Coletiva foi demonstrada diferença significativa inferior da UNESC em relação ao NAPISUL II na quarta ($p = 0,040$), quinta ($p = 0,022$), sexta ($p = 0,010$), décima primeira ($p = 0,013$) e décima segunda ($p = 0,001$) (Tabela 8).

Em 2013, no terceiro teste, o curso de medicina da UNESC tinha 474 alunos matriculados e 96,8% participaram do Teste de Progresso, sendo 59,4% do sexo feminino.

Comparando a média de desempenho geral da UNESC em relação ao das oito escolas que compõe o NAPISUL II (Tabela 1), neste teste a UNESC apresentou diferença significativamente maior na décima e décima segunda fase ($p = 0,005$ e $p = 0,018$

respectivamente) e menor na quarta fase ($p = 0,020$). As demais fases não apresentaram diferença estatística significativa ($p > 0,05$).

Em relação ao desempenho em cada grande área, vê-se que em Bioética (Tabela 2) a UNESC apresentou diferença estatística significativa menor na sexta fase ($p = 0,019$) em relação à média das escolas do NAPISUL II, porém maior que essa nas fases oitava, décima, décima primeira e décima segunda ($p = 0,045$; $p = 0,003$; $p = 0,005$; $p < 0,001$; respectivamente). Já em Ciências Básicas, a média foi significativamente menor na segunda ($p = 0,002$) e quarta fase ($p = 0,036$), não apresentando diferença significativa nas demais fases.

Em Clínica Cirúrgica (Tabela 4), a média da UNESC foi significativamente menor apenas na décima fase ($p = 0,012$), sem diferenças significativas nas fases restantes. Clínica Médica obteve diferença estatística significativa maior na décima segunda fase ($p = 0,005$), também sem diferenças significativas nas demais fases (Tabela 5).

Na grande área de Ginecologia e Obstetrícia, a média da UNESC em relação ao NAPISUL II demonstrou diferença significativamente menor na quarta fase ($p = 0,017$), enquanto nas fases sexta ($p < 0,001$), nona ($p = 0,033$) e décima ($p < 0,001$) essa diferença foi significativamente maior (Tabela 6). Em Pediatria houve diferença significativamente maior na média da décima fase ($p < 0,001$) da UNESC em relação ao NAPISUL, e as demais fases não apresentaram diferença estatística significativa (Tabela 7).

Por fim, Saúde Coletiva (Tabela 8) demonstrou uma média com diferença significativamente maior na décima fase ($p = 0,004$) e menor na segunda, quarta e sexta fase ($p = 0,04$; $p = 0,044$; $p = 0,002$).

DISCUSSÃO

Os resultados do presente estudo demonstraram que o TP mostrou-se um importante indicador para a tomada de decisões e melhoria do curso de medicina da UNESC.

O Teste do Progresso é uma forma de avaliação em que a prova é administrada simultaneamente para alunos em diferentes fases da sua formação, para que o seu progresso possa ser monitorado ao longo do tempo. A prova é elaborada com um nível de dificuldade equivalente a estudantes em final de formação, e é nesse nível que todos os alunos, da primeira a décima segunda fase, são testados.

Algumas limitações desse estudo devem ser levadas em conta: trata-se de três estudos transversais em uma análise temporal e o grau de dificuldade das provas são diferentes em cada ano. Assim, as curvas de crescimento de conhecimento cognitivo dos estudantes comparadas ano a ano, devem ser analisadas com muita parcimônia. Para minimizar esses vieses procura-se fazer uma análise estatística apurada, inclusive utilizando-se de testes que medem a comparabilidade entre as provas dos diferentes anos, como, por exemplo, o teste da ancoragem (Pasquali, 2003).

Preocupado com a equiparação das provas, o NAPISUL-II está aplicando testes de equivalência de provas, cujos resultados serão apresentados em outro artigo.

Apesar de suas limitações, o TP além de fornecer subsídios para tomadas de decisões dos gestores dos cursos é também um valioso *feedback* individual para o aprendizado do aluno. Larsen et al (2009) demonstraram que repetidos testes com *feedback* parece resultar, no estudante, em uma maior retenção do conhecimento a longo prazo, e sugeriram que os TPs deveriam ser considerados importantes, não apenas como uma ferramenta de avaliação, mas, também, como um poderoso instrumento de aprendizagem (Larsen et al 2009).

No curso de medicina da Universidade de McMaster, no Canadá, um estudo demonstrou que o efeito imediato da introdução dos TPs foi uma redução, de 19% para 4,5%, no índice de reprovação, dos estudantes daquela instituição, no exame nacional de licenciamento (Norman et al 2010).

No Brasil, antes do NAPISUL-II, algumas instituições já aplicavam rotineiramente TP, e seus resultados eram divulgados nos Congressos Brasileiros de Educação Médica, enfatizando-se a importância do mesmo na avaliação da escola e dos estudantes. Assim, gerou-se uma grande expectativa, entre professores e estudantes, com aplicação do teste nas escolas do Paraná e Santa Catarina. As oito instituições que compõe núcleo se uniram na elaboração das provas, formando um grupo coeso, capacitando-se pedagogicamente na elaboração de itens, esforçando-se em proporcionar uma boa adesão dos alunos, tanto quantitativa como qualitativa. Assim, os coordenadores de cada instituição ficam ansiosos em receber os resultados do teste, que sempre é feito de maneira sigilosa e individual, não permitindo qualquer “ranqueamento” entre as escolas. A análise é realizada por serviço externo contratado pelo grupo.

Com o resultado em mãos, o coordenador do curso de medicina da UNESC reuniu primeiramente seu Núcleo Docente Estruturante (NDE) e equipe de coordenadores de fase para uma ampla discussão e definição de estratégias para melhorias, que posteriormente foram apresentadas no colegiado pleno com participação discente.

O curso de medicina da UNESC nesse primeiro TP (2011), comparada à média geral das escolas do NAPISUL II, teve um número de acertos significativamente menor no internato médico (nona, décima e décima segunda fases) e não apresentou diferença nas demais fases, sendo que a Saúde Coletiva teve um desempenho aquém do esperado na maioria das fases. Analisando cuidadosamente o plano pedagógico do curso, concluiu-se que nas fases em que se usa a metodologia da Aprendizagem Baseada em Problemas (ABP), da primeira a oitava fase, estávamos na média das outras escolas, mas que, no internato, estávamos abaixo da média. Quanto à Saúde Coletiva, embora os alunos tivessem inseridos na rede de atenção básica desde a primeira fase, percebemos que o conteúdo teórico-conceitual estava insuficiente.

Resolveu-se então inserir atividades semanais sob a forma de aulas expositivas dialogadas, seminários, discussão e análise crítica de artigos e textos de conteúdos gerais além de uma hora de temas abordando Saúde Coletiva, fazendo parte da avaliação do internato

médico. Nos outros anos, incentivou-se a realização de seminários de Saúde Coletiva nas unidades básicas de Saúde pelos professores de Interação Comunitária.

No segundo TP a UNESC teve desempenho médio significativamente maior na oitava e décima fases e menor nas fases décima primeira e décima segunda, demonstrando já uma melhoria no aprendizado, pois em duas fases do internato ficamos significativamente melhor, e nas demais ficamos na média com exceção do último ano, que não havia passado ainda pela intervenção realizada no início do internato.

Por outro lado em 2013, a média de desempenho geral da UNESC em relação ao das oito escolas que compõe o NAPISUL II apresentou diferença significativamente maior na décima e décima segunda fase (último ano) enquanto que na nona e décima fases manteve-se na média das outras escolas.

Assim, para a UNESC, os resultados do teste de progresso aplicados de 2011 a 2013, analisados criteriosamente pela equipe gestora do curso de medicina, foi um importante instrumento de melhoria do curso, principalmente em nível de internato, onde o teste apontou as maiores falhas. Porém, a maior dificuldade para os gestores é comparar os resultados ano a ano, pois, por mais que as provas sejam elaboradas pela mesma equipe, elas não são idênticas, permanecendo o viés de que a melhora ou piora do desempenho dos estudantes seja devido à aplicação de uma prova com itens mais fáceis ou mais difíceis, respectivamente. Como perspectiva futura há necessidade de elaboração e análise das questões do teste de progresso de maneira a se permitir a equiparação dos mesmos, detectando-se que os escores encontrados sejam determinados exclusivamente pelo desempenho dos estudantes e não pelo grau de dificuldade da prova.

REFERÊNCIAS

1. Aarts R, Steidell K, Manuel BAF, Driessen EW. Progress testing in resource-poor countries: a case from Mozambique. *Med Teacher* 2010; 32:461–3.
2. Bennett J, Freeman A, Coombs L, Kay L, Ricketts C. Adaptation of medical progress testing to a dental setting. *Med Teacher* 2010; 32:500–2.
3. Freeman A, Van der Vleuten C, Nouns Z, Ricketts C. Progress testing internationally. *Med Teacher* 2010; 32:451–5.
4. Nouns Z, Georg W. Progress testing in German speaking countries. *Med Teacher* 2010; 32:467–70.
5. Schuwirth L, Bosman G, Henning R, Rinkel R, Wenink A. Collaboration on progress testing in medical schools in the Netherlands. *Med Teacher* 2010; 32:476–9.
6. Swanson D, Holtzman K, Butler A, et al. Collaboration across the pond: the multi-school progress testing project. *Med Teacher* 2010; 32:480–5.

7. Blake JM, Norman GR, Keane DR, Mueller B, Cunnington J, Didyk N. Introducing progress testing in McMaster University's problem-based medical curriculum: psychometric properties and effect on learning. *Acad Med* 1996; 71(9):1002-7.
8. Van der Vleuten CPM, Verwijnen GM, Wijnen HFW. Fifteen years of experience with progress testing in a problem-based learning curriculum. *Med Teacher* 1996; 18:103–10.
9. Pasquali L. *Psicometria- teoria dos testes na psicologia e na educação*, Petrópolis-RJ: Vozes, 2003; 261-278.
10. Larsen DP, Butler AC, Roediger HL. Repeated testing improves long-term retention relative to repeated study: a randomized controlled trial. *Med Educ* 2009; 43:1174–81.
11. Norman G, Neville A, Blake J, Mueller B. Assessment steers learning down the right road: impact of progress testing on licensing examination performance. *Med Teacher* 2010; 32:496–9.