

Avaliação Formativa entre Alunos de Medicina do Quarto Ano: Relato de Experiência

Formative Assessment Among Fourth-year Medical Students: an Experience Report

Ricardo Francalacci Savaris¹

PALAVRAS-CHAVE:

- Estudantes de Medicina.
- Avaliação do programa.
- Educação Médica

KEYWORDS:

- Students, Medical.
- Program Evaluation.
- Medical Education

Recebido em: 05/04/2008

Reencaminhado em: 15/06/2008

Reencaminhado em: 21/06/2008

Aprovado em: 21/08/2008

RESUMO

Objetivo: Comparar grau de dificuldade, satisfação e escores finais entre os métodos de avaliação formativa e somativa em alunos de Medicina do quarto ano, na disciplina de Ginecologia e Obstetrícia. **Método:** Trinta e oito alunos foram submetidos a um exame escrito (alternativo) e, um mês depois, ao método tradicional de escolha múltipla. O exame alternativo consistia em casos clínicos reais, enfocando as hipóteses diagnósticas e o manejo médico. As questões eram graduadas em 0, 0,5 ou 1 ponto. Após o resultado inicial, os exames eram devolvidos e, em duas semanas, os alunos tinham que escrever uma dissertação estruturada para cada resposta errada. Pontos adicionais (+ 0,5 ou + 1) eram dados para cada resposta. Dissertações com incorreções descontavam pontos da nota original. O teste tradicional consistia em conceitos clínicos e prevalência de diferentes doenças e de agentes etiológicos. Nenhum retorno era dado, a não ser a nota final. O grau de dificuldade, a média final de ambos os exames e a resposta dos alunos a essa iniciativa foram comparados entre os dois métodos. **Resultados:** O exame escrito alternativo teve maior grau de dificuldade ($7,6 \pm 0,2$ vs. $5,2 \pm 0,4$; $p < 0,001$). Nenhuma diferença estatística foi encontrada entre as médias dos escores finais de ambos os métodos ($8,3 \pm 0,9$ vs. $8,5 \pm 0,9$; $p > 0,05$). Os alunos preferiram o método alternativo, quando comparado ao tradicional ($8,6 \pm 0,2$ vs. $5,4 \pm 0,4$; $p < 0,001$). **Conclusão:** A avaliação formativa foi mais difícil do que a tradicional, mas os escores finais da classe dos alunos não foram menores do que os escores dos testes tradicionais. Os alunos preferiram a avaliação formativa.

ABSTRACT

Objectives: To compare the degree of difficulty, satisfaction and the final scores between the formative and summative assessment method among fourth-year medical students in the discipline of obstetrics and gynecology. **Method:** Thirty-eight students took a written exam (alternative) and 1 month later they were tested using the traditional multiple choices method. The alternative exam consisted in real clinical cases focusing on diagnostic hypothesis and medical management. The questions were scored 0, 0.5 or 1 point. After the initial result, the exams were returned and, within 2 weeks, the students had to prepare a structured essay in relation to each wrong answer. Additional points (+ 0.5, or +1) were given for each correct answer. Points were deduced for incorrect essays. The traditional exam consisted in clinical concepts, the prevalence of different diseases and etiological agents. No feedback but the final score was given. The degree of difficulty, the final mean score of both exams, and the students' responses to this initiative were compared between both methods. **Results:** The alternative written exam had a higher degree of difficulty (7.6 ± 0.2 vs. 5.2 ± 0.4 ; $p < 0.001$). No statistical difference was found between the means of the final scores of both methods (8.3 ± 0.9 vs. 8.5 ± 0.9 ; $p > 0.05$). The students preferred the alternative method when compared to the traditional one (8.6 ± 0.2 vs. 5.4 ± 0.4 ; $p < 0.001$). **Conclusion:** The formative assessment was more difficult than the traditional one. Nevertheless, the final scores of the medical class for the formative test were not lower than the scores for the traditional test. All students declared to prefer the formative assessment.

INTRODUÇÃO

A educação médica dos alunos de graduação tem recebido atenção especial nesses últimos anos^{1,2}. O principal enfoque tem sido o aprendizado baseado em problemas (*problem-based learning*)³, e os métodos tradicionais de ensino estão sendo substituídos por essa nova tendência⁴. O método tradicional de ensino nas escolas médicas do Brasil tem sido a avaliação somatória. Esse método tenta resumir o aprendizado do estudante em determinado ponto no tempo, geralmente no final do curso. Esse sistema não foi projetado para oferecer um retorno útil ao aluno e ao professor, o que auxiliaria no processo de aprendizado.

Por outro lado, a avaliação formativa tem sido utilizada por alguns autores⁵⁻⁸. Os erros cometidos pelos estudantes de Medicina são utilizados como uma oportunidade para aprender por meio da participação, revisão e reflexão⁹. Através desse método, o aluno aprende sem a intervenção ativa do professor¹⁰. A avaliação formativa deve ser usada para auxiliar o aluno a identificar seus pontos fortes e fraquezas, estabelecendo, assim, metas educacionais. Alguns autores sugerem que seria necessário estabelecer se o retorno da avaliação formativa tem algum impacto sobre a avaliação somatória¹¹.

O objetivo geral deste estudo é descrever os diferentes aspectos do método da avaliação formativa entre alunos do quarto ano de Medicina na disciplina de Ginecologia e Obstetrícia. Os objetivos específicos são: (a) quantificar o grau de satisfação dos alunos com o método formativo de aprendizado, em comparação com a avaliação somatória; (b) comparar o grau de dificuldade dos testes nas avaliações formativa e somatória; (c) comparar o escore final da classe de Medicina submetida à avaliação formativa com os escores finais do método tradicional.

MÉTODOS

Os 38 alunos do quarto ano da Faculdade de Medicina da Universidade Federal do Rio Grande do Sul (UFRGS) inscritos na disciplina de Ginecologia e Obstetrícia, no ano de 2004, participaram deste estudo, de coorte observacional cruzado. No final do primeiro mês, os alunos foram submetidos a uma avaliação formativa; um mês depois, realizaram um exame tradicional. A disciplina tem caráter obrigatório, carga horária de 190 horas e 19 créditos. Como parte dos objetivos do ensino, a disciplina ministra aos alunos a prevenção de doenças da mulher, além de condições patológicas prevalentes relacionadas à mulher grávida e não grávida. Os alunos têm aulas teóricas e práticas. As aulas práticas são ministradas por meio de rodízio nas diferentes equipes – oncologia ginecológica, emergência, endocrinologia, estática pélvica, gestação de

alto risco (hipertensão, diabetes), urodinâmica, infertilidade, pré-natal normal, unidade básica de saúde. Nesses estágios, os alunos entram em contato com as habilidades cognitivas da disciplina. O processo de ensino-aprendizado tem por base a interação entre aspectos teóricos e práticos.

A avaliação formativa (alternativa) se iniciava com uma prova escrita baseada em casos clínicos vistos nos estágios práticos. As questões abordavam hipóteses diagnósticas e manejo médico inicial. Os exames tradicionais de escolha múltipla focalizavam conceitos clínicos, prevalência de doenças e identificação de agentes etiológicos. O processo da avaliação formativa está detalhado nos Quadros 1 e 2.

QUADRO 1

Processo do método alternativo de ensino (formativo)

1. As questões são feitas por um grupo de professores.
2. Os alunos respondem as questões.
3. Os professores recebem suas questões de volta e corrigem como 0 (zero), 0,5 (meio ponto) ou 1 (correto). Comentários são feitos para as respostas incorretas (0 e 0,5) e qual o aspecto que deve ser revisto.
4. Os alunos recebem as notas iniciais do teste.
5. Os alunos revisam as respostas incorretas e escrevem uma dissertação estruturada em duas semanas, de acordo com o Quadro 2.
6. As redações (arquivos eletrônicos em Word®) e os testes originais são devolvidos ao grupo de professores. Pontos adicionais são dados ao escore inicial, dependendo da qualidade da dissertação: se a resposta estiver incorreta, o aluno perde o ponto recebido na nota inicial ou mantém o 0 (zero); pontos adicionais (0,25-1) são dados ao escore adicional, até 1 ponto por questão. Comentários são adicionados para cada questão feita pelo professor.
7. As melhores respostas são compiladas e anexadas à redação de cada estudante, para comparação.
8. O escore final é calculado.
9. As redações são devolvidas aos alunos por e-mail com a nota final.

As redações da segunda fase eram corrigidas sistematicamente. Cada professor corrigia as mesmas questões, para manter a coerência e identificar plágio. Todas as redações eram redigidas em português, para evitar o “recorta-e-cola” dos sítios da internet. A qualidade e a confiabilidade das respostas eram verificadas pelo grau de evidência dos artigos citados na redação. No final do bimestre, os mesmos alunos foram avaliados por meio do sistema tradicional, tendo o restante das matérias da disciplina que englobava aspectos da Ginecologia e Obstetrícia. No final da disciplina, os alunos completaram um questionário anônimo que comparava suas experiências com as dos sistemas alternativo e tradicional. A análise objetiva de cada método era feita numa escala de 0 a

QUADRO 2

Instruções para uma redação estruturada

Diretrizes: a ênfase é na alta qualidade, não na quantidade. Tente usar essas redações como “notas de plantão” para auxiliá-lo em casos semelhantes.
Por que errei: o aluno explica a lógica de sua resposta.
Resposta correta: o aluno escreve a resposta correta baseada nas melhores evidências da literatura médica. É dada uma orientação (<i>rubric</i> – uma ferramenta para o escore que lista os critérios para a redação). A pesquisa no Pubmed é fortemente enfatizada. As imagens e os <i>links</i> da <i>web</i> são encorajados. Os livros-textos são desencorajados, a não ser que estejam atualizados. As afirmações importantes sem referências são inaceitáveis. Comentários sobre a qualidade das referências devem ser adicionados para cada resposta. Em caso de tratamento médico, o nível de recomendação deve ser citado na resposta: A (princípios geralmente aceitos para o manejo dos pacientes que refletem um alto grau de certeza (p.ex., baseado em “força de evidência Classe I” ou forte evidência de estudos “Classe II”, que abordam todas as questões); B (recomendações para o manejo de pacientes que identificam uma estratégia em particular ou uma gama de estratégias que refletem certeza clínica moderada (p.ex., baseado na força de “evidência Classe II”, estudos que abordam diretamente a questão ou consenso com estudo com força de “evidência Classe III”); C (outras estratégias para o manejo de pacientes baseadas em evidências preliminares, inconclusivas ou conflituosas, ou, na ausência de qualquer literatura publicada, baseadas num painel de consenso).
O que aprendi: o aluno escreve o que aprendeu com a atividade. O grau de aprendizado será avaliado. Por exemplo: “aprendi a prestar mais atenção à questão” é diferente de “aprendi que a ausência de atraso menstrual não é suficiente para descartar uma gravidez ectópica”.
Qualidade da questão: o aluno deve avaliar a relevância da questão para a vida profissional. Sugestões para melhorar a questão são bem-vindas.

10 (0 = muito ruim; 10 = excelente). Uma questão subjetiva era formulada no final do questionário: em sua opinião, o método alternativo de avaliação é (melhor) (igual) (pior) do que o sistema tradicional?

O grau de dificuldade dos diferentes tipos de testes foi avaliado entre cinco residentes de Ginecologia e Obstetrícia do terceiro ano. Eles graduaram os exames tradicionais e alternativos de 0 (muito fácil) a 10 (muito difícil).

O teste *t*-Student foi utilizado nas variáveis contínuas, enquanto as variáveis categóricas foram analisadas pelo teste exato de Fisher, considerando uma diferença significativa um $p < 0,05$.

RESULTADOS

A mesma classe de estudantes, composta de 38 alunos, foi submetida aos métodos de avaliação formativa (alternativa) e somativa (tradicional) durante o bimestre. Todos os estudan-

tes completaram os questionários finais de avaliação. O grau de dificuldade do teste alternativo foi maior do que o do tradicional, segundo a avaliação dos residentes do terceiro ano de Ginecologia e Obstetrícia ($7,6 \pm 0,2$ vs. $5,2 \pm 0,4$; média \pm DPM, $p < 0,001$). Não houve diferença significativa entre as médias finais dos alunos ao se compararem os exames tradicionais e alternativos ($8,3 \pm 0,9$ vs. $8,5 \pm 0,9$; média \pm DPM, $p > 0,05$), embora a média da nota inicial da prova alternativa tenha sido significativamente menor quando comparada com a nota após a recuperação da prova ($4,1 \pm 1,1$ vs. $8,3 \pm 0,9$; média \pm DPM, $p < 0,001$). Os estudantes apresentaram maior grau de satisfação com a avaliação formativa, quando comparada com a somativa ($8,6 \pm 0,2$ vs. $5,4 \pm 0,4$; $p < 0,001$). Todos os alunos informaram que a avaliação formativa era melhor do que a somativa (dados não apresentados).

DISCUSSÃO

O uso das teorias de aprendizado de Piaget não é novo¹², mas o presente método oferece um duplo retorno ao estudante. Para os residentes do terceiro ano, o exame escrito com casos clínicos foi considerado mais difícil do que o exame objetivo de escolha múltipla (tradicional). Esse grau de dificuldade pode ser verificado pela média inicial da primeira parte da avaliação alternativa. Uma possível causa para essa nota reduzida seria a natureza prática das questões. Ao invés de solicitar o conceito de “ovo cego” na ultra-sonografia do primeiro trimestre da gravidez e dar opções, o aluno recebia um laudo de ultra-sonografia incompatível com o tamanho do saco gestacional demonstrado na foto do laudo. O aluno tinha que desenvolver uma lógica baseada na experiência e no conhecimento dos conceitos, fato este que corresponde mais à realidade da prática médica. Além disso, a ausência de opções escritas e a necessidade de explicar sua conduta tornavam o exame mais difícil. Todavia, os escores finais do método alternativo, após a segunda correção, não foram estatisticamente diferentes da média dos escores finais do método tradicional.

A avaliação formativa exigia mais trabalho dos alunos. As questões eram corrigidas e comentadas pelo professor que as elaborou. Esse retorno era necessário para o aluno entender seu erro. As melhores respostas das redações feitas pelos alunos para recuperar suas notas eram compiladas e devolvidas aos alunos. Esse segundo retorno era ainda mais importante e parece ser uma nova abordagem de ensino. Apesar do trabalho, os estudantes preferiram a avaliação formativa, que lhes dava a oportunidade de concordar ou discordar da resposta dada como correta no gabarito. Se o aluno demonstrasse evidências científicas suficientes, pontos adicionais eram dados à resposta, permitindo que a nota fosse recuperada. Os estudan-

tes aprendiam com seus erros, e os professores poderiam atualizar a nova informação para aprimorar a questão. No exame tradicional, essas oportunidades não ocorrem. Apesar de os alunos terem duas semanas para escrever suas redações em casa, consultando diferentes bases de dados, nenhum conseguiu atingir 100% dos acertos. Contudo, a média final do exame formativo foi 8,3, que não foi significativamente diferente da obtida no exame tradicional.

O método formativo exigiu grande esforço por parte tanto dos alunos como dos professores. A queixa mais comum dos alunos era o tempo gasto em procurar as melhores evidências para responder as perguntas. Da mesma forma, o professor precisava de um tempo considerável para corrigir cada questão. O fato de o mesmo professor corrigir a mesma questão, além de manter a coerência, facilitou a identificação de plágio e fraude entre os alunos, valores que nem sempre são avaliados nos exames. Além disso, a submissão eletrônica das respostas em editores de textos permitia identificar a autoria do texto por meio das propriedades do arquivo, recurso normalmente presente nos editores de texto.

Alguns alunos colocaram referências bibliográficas inexistentes ou resultados diferentes dos que o artigo citado mencionava. Esse problema pode refletir a natureza dos testes tradicionais, em que a quantidade de trabalho é enfatizada em detrimento da qualidade. Ao invés de focar o processo de aprendizagem, principal característica do processo formativo, o aluno procura a resposta mais completa, sem identificar os aspectos práticos do problema, como, por exemplo, custo dos medicamentos e praticidade da posologia. Por isso, um roteiro (*rubric*) era dado para cada questão, a fim de balizar a avaliação. A principal ameaça da avaliação formativa é o não engajamento dos professores em sua realização, recaindo o método sob a responsabilidade de poucos.

Diferentemente do exame oral baseado em problemas (POE), defendido por alguns autores¹³, no qual o grau de dificuldade pode variar entre os professores e nenhum retorno é dado, o presente método trabalhou com aspectos novos, como a oportunidade de o aluno pontuar a qualidade das questões feitas. Além disso, a avaliação formativa pode prover mais do que o conhecimento em si, pois o aluno aprende o processo de procurar a melhor resposta possível. Isso acontecia por intermédio de trabalho em grupo, divisão de tarefas em busca de evidências, discussão em pequenos grupos. Em avaliações somativas análogas, pode-se prever o desempenho discente a partir da nota no exame formativo¹⁴.

Em conclusão, os alunos preferiram a avaliação formativa, embora esse método seja considerado mais difícil do que o tradicional, sem que isso leve a uma redução na média dos

escores finais, quando comparada com a média dos testes tradicionais.

REFERÊNCIAS

1. Drake RL. Anatomy education in a changing medical curriculum. *Anat Rec.* 1998;253:28-31.
2. Dolmans DH, Gijsselaers WH, Moust JH, de Grave WS, Wolfhagen IH, van der Vleuten CP. Trends in research on the tutor in problem-based learning: conclusions and implications for educational practice and research. *Med Teach.* 2002;24:173-80.
3. Donner RS, Bickley H. Problem-based learning in American medical education: an overview. *Bull Med Libr Assoc.* 1993;81:294-98.
4. Fowell SL, Bligh JG. Recent developments in assessing medical students. *Postgrad Med J.* 1998;74:18-24.
5. Philips BU, Mahan JM, Perry RR. A formative and summative evaluation model for special educational programs. *J Med Educ.* 1976;51:836-43.
6. Konje JC, Taylor DJ. Formative assessment within structured training in obstetrics and gynaecology. *Br J Obstet Gynaecol.* 1998;105:139-41.
7. Townsend AH, McLlvenny S, Miller CJ, Dunn EV. The use of an objective structured clinical examination (OSCE) for formative and summative assessment in a general practice clinical attachment and its relationship to final medical school examination performance. *Med Educ.* 2001;35:841-46.
8. Rahman SA. Promoting learning outcomes in paediatrics through formative assessment. *Med Teach.* 2001;23:467-70.
9. Hughes M, Ventura S, Dando M. On-line interprofessional learning: introducing constructivism through enquiry-based learning and peer review. *J Interprof Care.* 2004;18:263-68.
10. La Taille Y. O erro sob a perspectiva de Piaget. In: Aquino JG, editor. *Erro e fracasso na escola.* São Paulo: Summus; 1997. p.25-44.
11. Chur-Hansen A, Koopowitz LF. Formative feedback in teaching undergraduate psychiatry. *Acad Psychiatry.* 2005;29:66-8.
12. Qayumi S. Piaget and his role in problem based learning. *J Invest Surg.* 2001;14:63-5.
13. Walubo A, Burch V, Parmar P et al. A model for selecting assessment methods for evaluating medical students in African medical schools. *Acad Med.* 2003;78:899-906.
14. Krasne S, Wimmers PF, Relan A, Drake TA. Differential effects of two types of formative assessment in predicting performance of first-year medical students. *Adv Health Sci Educ Theory Pract.* 2006;11:155-71.

CONFLITOS DE INTERESSE

Declarou não haver.

ENDEREÇO PARA CORRESPONDÊNCIA

Ricardo Françalacci Savaris

Universidade Federal do Rio Grande do Sul

Rua Ramiro Barcelos, 2350 / 1125

Rio Branco

CEP: 90035-903 Porto Alegre – RS

E-mail: rsavaris@hcpa.ufrgs.br