

# Competência clínica de alunos de Medicina em estágio clínico: comparação entre métodos de avaliação

## *Clinical competence of medical students during clinical clerkship: a comparison of evaluation methods*

Rosângela Curvo Leite Domingues<sup>1</sup>

Eliana Amaral<sup>1</sup>

Angélica Maria Bicudo Zeferino<sup>1</sup>

Maria Ângela G. Monteiro Antonio<sup>1</sup>

Wilson Nadruz<sup>1</sup>

### RESUMO

**Objetivo:** Comparar quatro métodos de avaliação com base nas notas obtidas por alunos do quarto ano de Medicina em estágio clínico. **Método:** Ao final do estágio clínico em Ginecologia e Obstetrícia, Clínica Médica e Pediatria, em unidades de atenção primária, 106 alunos do quarto ano foram avaliados com base em quatro métodos: Prova Teórica (Pt), Portfólio (Pf), Avaliação Estruturada do Atendimento Clínico (AE) e Conceito Global Itemizado (CGI), (escala 0-10). A análise estatística utilizou coeficiente alfa de Cronbach, teste de Wilcoxon para dados pareados, coeficiente de correlação de Pearson, análise de componentes principais e distância euclidiana. **Resultados:** As medianas mais baixas foram de PT, e as mais altas, de AE. Houve diferença significativa nas distribuições das notas entre todos os pares de métodos ( $P < 0,001$ ). A correlação mais forte e a maior concordância ocorreram entre as notas de AE e CGI. A análise de componentes principais contrastou as notas de Pt e Pf com as de AE e CGI. **Conclusão:** Cada método parece enfocar diferentes aspectos da competência clínica. AE e CGI parecem avaliar competências semelhantes, em oposição a Pf e Pt. Portanto, nenhum método deve ser utilizado isoladamente para avaliar os alunos em estágio clínico.

### ABSTRACT

**Objective:** The objective was to compare four evaluation methods based on the grades received by fourth-year medical students in their clinical clerkship. **Method:** At the completion of the clinical clerkship in Obstetrics and Gynecology, Internal Medicine, and Pediatrics in primary care clinics, 106 fourth-year students were evaluated based on four methods, on a scale from 1 to 10: multiple-choice questions (MCQs), portfolio (P), real-case structured clinical assessment (RC-SCA), and global itemized rating (GIR). Statistical analysis used Cronbach alpha, Wilcoxon paired test, Pearson correlation coefficient, principal components analysis, and Euclidian distance. **Results:** The lowest median scores were for MCQs and the highest for RC-SCA. The distribution of scores for all pairs of methods showed significant differences ( $P < 0,001$ ). Agreement was strongest and highest between RC-SCA and GIR. Principal components analysis contrasted MCQs and P scores with those for RC-SCA and GIR. **Conclusion:** The various methods focused on different aspects of the expected clinical skills. RC-SCA and GIR appear to evaluate different skills as compared to P and MCQs. Thus, no method should be used alone to evaluate medical students in clinical clerkships.

### PALAVRAS-CHAVE

- Métodos de Avaliação.
- Competência Clínica
- Estágio Clínico
- Educação Médica

### KEY WORDS

- Evaluation Methods
- Clinical Competence
- Clinical Clerkship
- Medical Education

Recebido em: 23/01/2009

Reencaminhado em: 01/07/2009

Aprovado em: 23/07/2009

## INTRODUÇÃO

Com a crescente ênfase em preparar os estudantes para uma prática clínica responsável a serviço da comunidade, assegurar a qualidade da atenção e buscar padrões mínimos para certificação, a avaliação passou a ter um papel fundamental na graduação. Neste contexto, deixou de ser apenas uma questão de produzir medições confiáveis e válidas, expandindo-se para englobar fatores relacionados a desenho instrucional, impactos educacionais, aspectos de implementação e disponibilização de recursos<sup>1</sup>. Portanto, a seleção dos métodos mais adequados de avaliação numa situação específica deve resultar da combinação equilibrada de todos esses aspectos<sup>2</sup>.

A avaliação deve ser parte integrante do processo de ensino-aprendizagem, já que pode ser muito útil para direcionar tal processo, oferecendo aos docentes orientações para eventuais melhorias no planejamento e aos alunos informações sobre seu progresso, motivando-os a redirecionar seu comportamento em relação ao que estão aprendendo<sup>3</sup>. Dessa forma, o objeto da avaliação depende dos objetivos educacionais propostos.

A avaliação envolve competências complexas que exigem informações quantitativas e qualitativas, provenientes de diversas fontes<sup>1</sup>. Porém, devido ao caráter polissêmico do termo competência, um primeiro desafio parece ser exatamente defini-lo e distingui-lo do termo desempenho<sup>4</sup>.

Para discutir a avaliação de competência clínica, Miller<sup>5</sup> concebeu um modelo conceitual hierárquico de quatro níveis, representado por uma pirâmide, cuja base envolve o conhecimento (“saber”); um segundo nível engloba a habilidade de aplicar o conhecimento em determinado contexto (“saber como”, competência); o próximo nível, “mostrar como” (desempenho), reflete a habilidade de agir corretamente numa situação simulada; e o último, “fazer” (ação), refere-se à prática em situações clínicas reais.

Rethans *et al.*<sup>6</sup> propuseram um refinamento do modelo da pirâmide de Miller, chamando o nível “mostrar como” de competência e o “fazer” de desempenho, considerando a competência como pré-requisito para o desempenho. Esses autores afirmam que outros fatores, além da competência, influenciam o desempenho clínico, incluindo disponibilidade de tempo, cansaço, humor do médico e do paciente, relacionamento do médico com o paciente, colegas e outros profissionais da saúde. Além disso, competência é contextual, pois reflete a relação entre as habilidades de um indivíduo e as tarefas que ele tem que realizar numa situação particular<sup>7</sup>. Afinal, nenhuma competência é exercida no vazio, mas em contextos com especificidades, peculiaridades e idiossincrasias<sup>8</sup>.

A competência também foi definida como um conjunto de elementos cognitivos, interpessoais, afetivos, emocionais e morais, passíveis de desenvolvimento e aprimoramento a partir da prática e reflexão<sup>9</sup>; ou como a capacidade de mobilização de diversos saberes: saber-fazer, saber-ser e saber-agir; ou ainda como a capacidade de utilizar conhecimentos e habilidades para o exercício de uma situação profissional<sup>8</sup>. Diante de tantas definições e considerações, não se pode afirmar que haja um claro consenso sobre a distinção e a relação entre competência e desempenho, o que faz com que esses dois termos sejam usados frequentemente como sinônimos<sup>10,11</sup>.

Entretanto, a literatura atual parece convergir para o entendimento de que, na área médica, tanto competência quanto desempenho representam uma síntese de inúmeras habilidades técnicas e humanísticas que os alunos/médicos devem possuir para poder melhorar sua prática diária em benefício dos pacientes, sendo que competência está ligada àquilo que o aluno/médico faz em situações controladas, e desempenho está vinculado às ações do aluno/médico durante a prática profissional real. A observação do desempenho permite identificar como o aluno mobiliza suas diferentes competências<sup>4, 7,12-14</sup>.

Durante muito tempo, os métodos de avaliação utilizados enfocaram exclusivamente o “saber” ou “saber como”, competências da esfera cognitiva. Mas, a partir da segunda metade do século 20, mudanças significativas foram incorporadas aos sistemas de avaliação. Concluiu-se que avaliar os aspectos cognitivos era necessário, mas insuficiente para prever a competência clínica dos estudantes<sup>14-16</sup>. A associação de métodos surge como uma alternativa para atingir resultados válidos e confiáveis na avaliação clínica<sup>5,12,13,16-19</sup>. Assim, diversas entidades passaram a implantar novas combinações de métodos de avaliação, acrescentando-lhes a necessária perspectiva formativa e indutora de novas práticas pedagógicas na educação médica<sup>10,17,20,21</sup>.

Entre os métodos disponíveis incluem-se: provas teóricas; portfólio; avaliação estruturada por observação do atendimento; mini-CEX (*mini clinical evaluation exercise*); Osce (*objective structured clinical examination*); conceito global itemizado; simulação; autoavaliação; avaliação por pares; avaliação por usuários; avaliação por outros profissionais da equipe<sup>7</sup>.

Na Faculdade de Ciências Médicas da Universidade Estadual de Campinas (Unicamp), onde o estudo foi realizado, os alunos do quarto ano médico são iniciados na prática clínica durante o Módulo de Atenção Integral à Saúde, que visa propiciar a experiência clínica fora do ambiente hospitalar e facilitar o desenvolvimento de responsabilidade e vínculo com o paciente e com outros profissionais da equipe de saúde. Os estudantes são distribuídos em seis Unidades Básicas de Saúde (UBS) do SUS, realizando atendimen-

tos nas áreas de Ginecologia e Obstetrícia, Clínica Médica e Pediatria por nove meses. Durante as consultas, sob supervisão de docentes e tutores médicos da própria rede de saúde, os alunos devem realizar anamnese e exame físico completos, registrar as informações de modo claro e ordenado, e valorizar os achados relevantes de inserção social do indivíduo. Adicionalmente, devem indicar condutas diagnósticas e terapêuticas para as situações mais comuns. As atividades de atendimento individual são complementadas pela discussão dos casos clínicos com outros membros da equipe de saúde e participação em reuniões de projeto terapêutico. Para responder à necessidade de avaliar os diversos aspectos das competências mobilizadas nesta experiência de aprendizado, compatíveis com os objetivos constantes no currículo, o sistema de avaliação foi planejado de forma a englobar quatro métodos distintos, com notas que variam de 0 a 10: Prova Teórica (Pt), Portfólio (Pf), Avaliação Estruturada do Atendimento Clínico (AE) e Conceito Global Itemizado (CGI).

## MÉTODO

Ao final do Módulo Atenção Integral à Saúde (MD856), em outubro de 2005, foram avaliados 106 alunos do quarto ano de Medicina, com base nos seguintes métodos:

- Prova Teórica (Pt): consistiu na média de duas provas com questões dissertativas e de múltipla escolha;
- Portfólio (Pf): foi a nota atribuída ao dossiê do aluno, que conteve uma planilha de todos os casos atendidos, apresentação e discussão de três casos selecionados de Ginecologia e Obstetrícia, Clínica Médica e Pediatria, e autoavaliação da aprendizagem;
- Avaliação Estruturada (AE): consistiu na média das notas provenientes das observações docentes, com lista de verificação, de três consultas com paciente real, de Ginecologia e Obstetrícia, Clínica Médica e Pediatria. O formulário continha questões relacionadas à abordagem inicial da consulta, anamnese, exame físico, preenchimento dos formulários e orientações ao paciente;
- Conceito Global Itemizado (CGI): representou a média das notas atribuídas ao aluno por três docentes, utilizando instrumento com 13 itens que representam diferentes domínios da competência clínica (seis de habilidades técnicas: qualidade da história; exame clínico; conhecimento médico; julgamento clínico; solução de problemas; hábitos de trabalho; e sete de habilidades humanísticas: comunicação e relacionamento com pacientes e familiares; respeito; capacidade de autorreflexão; percepção do contexto; interação com colegas; interação com docentes; interação com demais profissionais).

A análise estatística incluiu a descrição das notas dos alunos nos quatro métodos estudados, mediante médias, desvios-padrão, medianas e intervalos de confiança para médias populacionais.

Para investigar as diferenças nas quatro distribuições, foram feitas comparações múltiplas mediante o teste de Wilcoxon para dados pareados. As correlações entre as notas obtidas foram estudadas por meio dos coeficientes de correlação de Pearson. O método das componentes principais foi utilizado para buscar as componentes que explicassem a maior parte da variabilidade dos dados. O coeficiente alfa de Cronbach foi calculado para investigar a confiabilidade das quatro notas. A concordância relativa entre os quatro métodos foi calculada recorrendo-se às distâncias euclidianas entre notas. O nível de significância escolhido foi de 5%. O estudo foi aprovado pelo Comitê de Ética em Pesquisa, sob o número 581/2005.

## RESULTADOS

Foram excluídos três alunos que não tinham suas notas de CGI registradas. Portanto, os resultados se referem aos 103 alunos restantes. As medianas e médias mais baixas corresponderam às notas da Prova teórica (Pt), e as mais altas, às notas da Avaliação Estruturada (AE). Ao compararmos as distribuições das notas obtidas nos quatro métodos, detectamos diferença significativa entre todas elas ( $P < 0,001$ ) (Figura 1).

Os coeficientes de correlação entre as notas dos quatro métodos estão apresentados na Tabela 1. As notas da Avaliação Estruturada e do Conceito Global Itemizado apresentaram correlação forte e significativa, mas a correlação foi baixa e significativa entre as notas da Prova Teórica e do Portfólio. Não houve correlação significativa entre os demais pares de métodos.

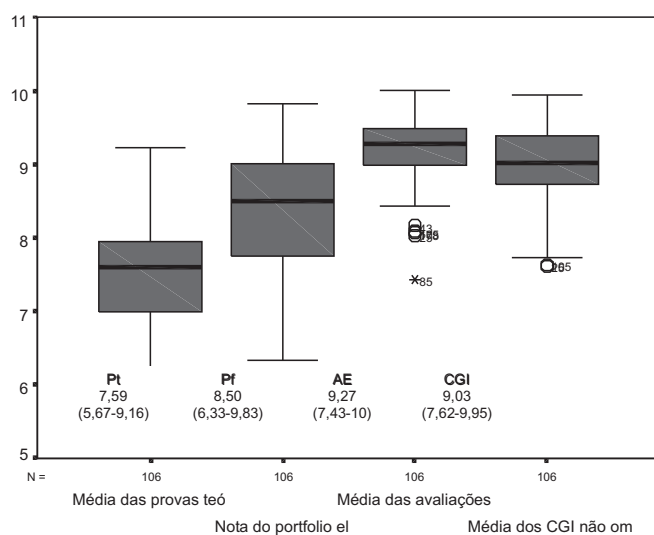
O resultado da análise de concordância relativa mostrou que as notas da AE e as do CGI exibiram a máxima concordância (menor distância), seguidas das notas de Pf e CGI, Pf e Pt, Pf e AE, Pf e AE, Pt e CGI e, por último, as menos concordantes foram as notas de AE e Pt.

A análise das componentes principais mostrou que duas delas (autovalor maior do que 1) explicaram 74,5% da variação total dos dados; 45,4% da variabilidade foi explicada pela primeira componente, que mediu o desempenho comum nos métodos; e 29,1% pela segunda componente, cujos sinais indicaram que existe uma diferença entre o par de notas da Prova Teórica e do Portfólio, de um lado, e o par de notas da Avaliação Estruturada e do Conceito Global Itemizado, do outro (Tabela 2).

Observou-se um reduzido grau de confiabilidade das notas nos quatro métodos (Cronbach = 0,498). A Tabela 3 mostra que a nota da Prova Teórica é a menos associada à soma das notas dos demais métodos ( $r$ ) e também a menos explicada pelas notas dos outros métodos ( $R^2$ ). A eliminação de qualquer um dos métodos gera uma queda discreta na confiabilidade geral, sendo a maior queda produzida com a eliminação da nota do Conceito Geral Itemizado.

Figura 1

Comparação das distribuições das notas dos diferentes métodos de avaliação – Prova teórica (Pt\*), Portfólio (Pf\*), Avaliação Estruturada (AE\*) e Conceito Global Itemizado (CGI\*) n= 103



\* Wilcoxon para dados emparelhados,  $P < 0,001$  (todos os métodos)

Tabela 1  
Coeficientes de correlação de Pearson das quatro notas (n = 103)\*

Nota	PT	Pf	AE
Portfólio (Pf)	0,258*		
Avaliação Estruturada (AE)	0,081	0,093	
Conceito Global Itemizado (CGI)	0,119	0,174	0,719*

\*  $P < 0,001$

Tabela 2

Resultados da análise de componentes principais das notas obtidas com os quatro métodos (n = 103)

	Componente 1	Componente 2
Autovalor	1,818	1,164
Variância explicada (%)	45,45	29,10
Conceito Global Itemizado (CGI)	0,894	-0,247
Avaliação Estruturada (AE)	0,864	-0,341
Portfólio (Pf)	0,396	0,681
Prova Teórica (Pt)	0,340	0,723

Tabela 3

Resultado da análise de confiabilidade com as notas obtidas por meio dos quatro métodos de avaliação (n = 103, alfa de Cronbach = 0,498)

Método	r *	R <sup>2</sup> (%) **	Alfa #
Pt	0,24	7,2	0,479
Pf	0,26	8,9	0,498
AE	0,35	51,8	0,403
CGI	0,42	53,0	0,348

\*r = correlação de Pearson entre a nota de cada método e a soma das outras três notas.

\*\* R<sup>2</sup> = percentagem da variação da nota explicada por regressão linear sobre as notas restantes.

# Alfa obtido após a eliminação das notas do método.

## DISCUSSÃO

Os resultados deste estudo mostraram que as medianas da Prova Teórica (Pt) foram as mais baixas entre todos os métodos utilizados, enquanto as notas da Avaliação Estruturada (AE) obtiveram as maiores medianas. Esses resultados, de certa forma, são contrários aos apresentados por Auewarakul *et al.*<sup>19</sup>, para os quais os métodos voltados à prática clínica, como as observações de atendimento, apresentaram maior grau de dificuldade para os alunos do que as provas teóricas.

Os resultados encontrados também podem refletir maior ênfase dos avaliadores a aspectos relacionados à competência cognitiva, isto é, ao “saber” obtido nos livros<sup>22</sup>. Pode ter ocorrido mais rigidez na elaboração e correção das provas teóricas, já que há uma elevada expectativa em relação à quantidade de informação que o aluno deve possuir nessa etapa do curso médico. Também a abrangência e a complexidade de cada um dos temas podem ter gerado uma dificuldade adicional para os alunos no direcionamento de seus estudos.

Contrariamente, durante a avaliação estruturada, o que se avalia é o conhecimento aplicado, o “saber fazer”, que retrata as habilidades clínicas em situações reais. Surpreendentemente, em nosso estudo, os alunos foram avaliados com notas elevadas numa fase inicial de atendimento clínico dentro do curso médico, o que pode caracterizar o efeito de leniência dos docentes<sup>23</sup>, a falta de um padrão preestabelecido<sup>24</sup> ou a carência de preparo e treinamento prévio dos avaliadores<sup>10,16</sup>. Pode ter ocorrido pouca sensibilidade do instrumento utilizado na AE para captar as competências esperadas nesta fase. Finalmente, os avaliadores podem ter valorizado aspectos de comunicação e relação interpessoal, que são trabalhados com os alunos em módulos iniciais da escola médica, e diluído aqueles relacionados às habilidades

clínicas, que começam a ser praticados no módulo que foi alvo de nosso estudo.

A dimensionalidade dos dados foi reduzida por meio da análise das componentes principais. A segunda componente contrastou as notas da Avaliação Estruturada (AE) e do Conceito Global Itemizado (CGI) com as do Portfólio (Pf) e da Prova Teórica (Pt), evidenciando uma distinção entre métodos mais centrados no desempenho, na ação (“fazer”), e os mais focados na cognição (“saber”)<sup>5</sup>. Tais achados, aliados à forte correlação e à maior concordância relativa entre as notas de AE e CGI, sugerem que, embora esses dois métodos não capturem exatamente os mesmos aspectos da competência e não sejam realizados no mesmo momento, ambos avaliam o aluno em função de seu desempenho em consultas clínicas, dentro do consultório<sup>25</sup>.

Em nosso estudo, os mesmos docentes avaliaram os mesmos alunos nos métodos AE e CGI, o que pode explicar parcialmente a correlação observada. De qualquer forma, esse resultado coincide com os de outros estudos, que demonstraram que métodos de observação direta, simulação e conceito global estão correlacionados, podendo, assim, ser usados em situações semelhantes para avaliar desempenho<sup>19,26</sup>. Esses resultados nos levam a refletir sobre a necessidade de revisão do instrumento de AE, do processo de avaliação e do preparo dos avaliadores.

Embora o portfólio incluísse diversos elementos, dentre eles a autoavaliação e a crítica comentada dos atendimentos prévios, há um forte conteúdo cognitivo na preparação, descrição e discussão dos casos clínicos, o que pode ter levado os docentes a avaliar os alunos de forma semelhante à da Pt, valorizando o acúmulo e o uso do conhecimento teórico.

Não foram observadas correlações estatisticamente significativas entre os demais pares de métodos (Pt vs. AE; Pt vs. CGI;



Pf vs. AE; Pf vs. CGI). Esses resultados coincidem com os de outros autores que também não observaram correlação entre métodos que avaliam conhecimento e os que avaliam habilidades<sup>27</sup>. A reduzida correlação entre os quatro pares de métodos de avaliação acima mencionados não surpreende, uma vez que eles são intrinsecamente distintos e apresentam propósitos diferentes, embora alguns estudos apresentem correlações moderadas entre provas teóricas e simulações e observações de consultas<sup>10,19</sup>.

O CGI pareceu ser o mais adequado no conjunto de métodos estudados, uma vez que sua eliminação produziu a maior queda no coeficiente de Cronbach, se comparado aos outros métodos. Uma possível explicação para esse resultado é o fato de que o CGI consegue capturar dimensões amplas da competência do aluno, o que justifica ser um dos métodos mais utilizados em educação médica. A validade do CGI aumenta quando ele consegue integrar impressões de diferentes avaliadores, provenientes de múltiplas situações clínicas, durante um longo período de contato com o avaliado, baseadas em domínios específicos, como aconteceu em nosso estudo<sup>26,28,29</sup>.

Entretanto, nem sempre são utilizados instrumentos de avaliação com diferentes itens e seus correspondentes descritores de bom e mau desempenho, como o CGI. Frequentemente, professores atribuem aos alunos, no final do estágio clínico, um conceito geral, expresso em nota única, subjetivo, não baseado em critérios previamente definidos. Estudos mostram que apenas avaliadores muito experientes conseguem formar essa visão integrada das competências de um aluno, atribuindo-lhe um conceito geral que se assemelha às avaliações baseadas em observação direta<sup>30</sup>.

Podem-se identificar algumas limitações neste estudo. A primeira delas é que foram utilizados dados referentes a uma turma de alunos, durante um módulo específico, numa única universidade, o que reduz sua generalização. Além disso, os professores utilizaram os métodos de AE e CGI após breve capacitação, sem terem tido a oportunidade de experimentá-los e discuti-los extensivamente em sessões de treinamento, o que é recomendado para o aprimoramento das habilidades de avaliação<sup>21</sup>.

Apesar das limitações deste estudo, que retratou a experiência clínica inicial de alunos de Medicina num cenário de atenção primária à saúde, no qual o aluno estava começando a aprender a se relacionar com os pacientes e a utilizar as habilidades clínicas, o sistema de avaliação utilizado pareceu englobar de forma adequada todos os níveis de competências esperados de um aluno.

## CONCLUSÃO

Há consenso entre pesquisadores em que nenhum método deve ser usado isoladamente para avaliar as distintas e complexas competências que a boa prática da medicina demanda<sup>18</sup>. A estratégia da triangulação, isto é, a combinação de informações provenientes de diferentes fontes, é essencial para que o sistema de avaliação possa ser abrangente o suficiente e ter credibilidade<sup>17,31</sup>. O conteúdo, o formato e a frequência da avaliação devem estar alinhados aos objetivos propostos para o módulo e ao planejamento curricular. Além disso, os vários domínios da competência clínica, com seus aspectos cognitivos e humanísticos, devem ser avaliados de forma integrada, coerente e longitudinal<sup>7</sup>.

Nesse sentido, nosso estudo apresentou a utilização de diferentes métodos que contemplaram informações provenientes de todos os níveis da pirâmide de Miller. Entretanto, a utilização de um método é apenas a materialização de um conceito pedagógico. Para que se possa aumentar o impacto educacional das avaliações, há que se promover reflexões mais abrangentes sobre os conceitos subjacentes aos métodos, para se compreender como as informações obtidas podem contribuir para a qualificação dos estudantes que pretendemos formar.

## AGRADECIMENTOS

Os autores agradecem a colaboração dos estatísticos Sirlei Siani Moraes e Armando Mário Infante, que contribuíram com a análise dos dados nas diferentes versões deste manuscrito.

## REFERÊNCIAS

1. Van der Vleuten CP, Schuwirth LWT. Assessing professional competence: from methods to programmes. *Med Educ*. 2005;39(3):309-17
2. Norcini JJ, McKinley DW. Assessment methods in medical education. *Teach Educ*. 2007;23(3):239-50
3. McAleer S. Choosing assessment instruments. In: Dent JA, Harden RM. *A practical guide for medical teachers*. 2<sup>nd</sup> ed. Philadelphia: Elsevier; 2005.
4. Ramos MN. Qualificação, competência e certificação: visão educacional. *Formação (Brasília)*. 2001;2:19-28.
5. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med*. 1990;65( Suppl 9):63-7.
6. Rethans JJ, Norcini JJ, Baron-Maldonado M, Blackmore D, Jolly BC, La Duca T, et al. The relationship between competence and performance: implications for assessing practice performance. *Med Educ*. 2002;36(10):901-09.
7. Epstein RM. Assessment in medical education. *N Engl J Med*. 2007;356(4):387-96.

8. Depresbiteris, L. Certificação de competências: a necessidade de avançar numa perspectiva formativa. *Formação (Brasília)*. 2001;2:27-38.
9. Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA*. 2002; 287(2):226-35.
10. Edelstein RA, Reid HM, Usatine R, Wilkes MS. A comparative study of measures to evaluate medical students' performances. *Acad Med*. 2000;75(8):825-33.
11. Sturmberg JP, Atkinson K, Farmer EA. Attainment and maintenance of professional capabilities. *Aust Fam Physician*. 2005;34(5):371-3.
12. Wilkinson TJ, Frampton CM. Comprehensive undergraduate medical assessments improve prediction of clinical performance. *Med Educ*. 2004;38(10):1111-6.
13. Schuwirth LWT, Van der Veulten CP. Changing Education, changing assessment, changing research? *Med Educ*. 2004;38(8):805-12.
14. Norcini JJ. Current perspectives in assessment: the assessment of performance at work. *Med Educ*. 2005;39(9):880-9
15. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet*. 2001; 357(9260):945-9.
16. Carr SJ. Assessing clinical competency in medical senior house officers: how and why should we do it? *Postgrad Med J*. 2004;80(940):63-6.
17. Murray E, Gruppen L, Catton P, Hays R, Woolliscroft JO. The accountability of clinical education: its definition and assessment. *Med Educ*. 2000;34(10):871-9.
18. St George IS, Kaigas T, McAvoy P. Assessing the competence of practicing physicians in New Zealand, Canada, and the United Kingdom: progress and problems. *Fam Med*. 2004; 36(3):172-7.
19. Auewarakul, C, Downing SM, Jaturatamrong U, Praditsuwon R. Sources of validity evidence for an internal medicine student evaluation system: an evaluative study of assessment methods. *Med Educ*. 2005;39(3):276-83.
20. IL: Accreditation Council for Graduate Medical Education 2004. [documento online]. [access 2006 jun. 30]. Available em: <http://www.acgme.org/outcome/assess/toolbox/asp>.
21. Holmboe ES, Hawkins RE, Huot SJ. Effects of training in direct observation of medical residents clinical competence. *Ann Intern Med*. 2004;140(11):874-81.
22. Misch DA. Evaluating physicians' professionalism and humanism: the case for humanism "connoisseurs". *Acad Med*. 2002;77(6):489-95.
23. Bandiera GW, Morrison LJ, Regehr G. Predictive validity of the global assessment form used in a final-year undergraduate rotation in emergency medicine. *Acad Emerg Med*. 2002;9 (9):889-95
24. Boulet JR, De Champlain AF, McKinley DW. Setting defensible performance standards on OSCEs and standardized patient examinations. *Med Teach*. 2003;15(3):245-9.
25. Solomon DJ, Szauter K, Rosebraugh CJ, Callaway MR. Global ratings of student performance in a standardized patient examination: is the whole more than the sum of the parts? *Adv Health Sci Edu Theory Pract*. 2000;5(2):31-40.
26. Morgan PF, Cleave-Hogg D, Guest CB. A comparison between global ratings and checklist scores from an undergraduate assessment using an anesthesia simulator. *Acad Med*. 2001; 76(10):1053-5.
27. Nuovo J, Bertakis KD, Azari R. Assessing resident's knowledge and communication skills using four different evaluation tools. *Med Educ*. 2006;40(7):630-6.
28. Daelmans HEM, Van der Hem-Stokroos HH, Hoogenboom RJJ, Scherpbier AJJA, Stehouwer CDA, Van der Vleuten CP. Global clinical performance rating, reliability and validity in undergraduate clerkship. *Neth J Med*. 2005;63(7):279-84.
29. Troncon LEA. Significance of experts' overall ratings for medical student competence in relation to history-taking. *São Paulo Med J*. 2006;124(2):101-4.
30. Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Acad Med*. 1998;73(9):993-7.
31. Driessen E, Van de Vleuten C, Schuwirth L, Tartwijk JV, Vermunt J. The use of qualitative research for portfolio assessment as an alternative to reliability evaluation: a case study. *Med Educ*. 2005;39(2):214-20.

## CONTRIBUIÇÃO DOS AUTORES

Rosângela Curvo Leite Domingues participou do desenvolvimento deste estudo. Da interpretação dos dados, da redação da primeira versão do artigo e incorporou as contribuições dos autores. Eliana Amaral, Angélica Maria Bicudo-Zeferino, Maria Ângela G. Monteiro Antonio e Ilson Narduz participaram de forma suficiente da concepção e desenvolvimento do estudo e da redação do texto e suas revisões. Todos os autores participaram da edição do artigo, fizeram a revisão final, aprovando-a para ser encaminhada para publicação

**CONFLITO DE INTERESSES**

Declarou não haver.

**ENDEREÇO PARA CORRESPONDÊNCIA**

Rosângela Curvo Leite Domingues

Rua Luciano Venere Decourt, 858

Cidade Universitária - Campinas

CEP. 13083-740 SP

E-mail: rosangela@cpdec.com.br