

ROSÂNGELA CURVO LEITE DOMINGUES

**CONCEITO GLOBAL PARA ALUNOS DE MEDICINA EM ESTÁGIO
CLÍNICO: COMPARAÇÃO ENTRE AUTO-AVALIAÇÃO,
AVALIAÇÃO DE DOCENTES E PARES**

Tese de Doutorado

**ORIENTADORA: Prof^ª. Dr^ª. ELIANA MARTORANO AMARAL
CO-ORIENTADORA: Prof^ª. Dr^ª. ANGÉLICA MARIA BICUDO ZEFERINO**

**Unicamp
2008**

ROSÂNGELA CURVO LEITE DOMINGUES

**CONCEITO GLOBAL PARA ALUNOS DE MEDICINA EM ESTÁGIO
CLÍNICO: COMPARAÇÃO ENTRE AUTO-AVALIAÇÃO,
AVALIAÇÃO DE DOCENTES E PARES**

Tese de Doutorado apresentada à Pós-Graduação da Faculdade de Ciências Médicas da Universidade Estadual de Campinas para obtenção do Título de Doutor em Tocoginecologia, área de Ciências Biomédicas

**ORIENTADORA: Prof^ª. Dr^ª. ELIANA MARTORANO AMARAL
CO-ORIENTADORA: Prof^ª. Dr^ª. ANGÉLICA MARIA BICUDO ZEFERINO**

**Unicamp
2008**

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA FACULDADE DE CIÊNCIAS MÉDICAS
UNICAMP**

Bibliotecário: Sandra Lúcia Pereira – CRB-8ª / 6044

D713c Domingues, Rosângela Curvo Leite
Conceito global para alunos de medicina em estágio clínico: comparação entre auto-avaliação, avaliação de docentes e pares / Rosângela Curvo Leite Domingues. Campinas, SP: [s.n.], 2008.

Orientadores: Eliana Martorano Amaral, Angélica Maria Bicudo Zeferino
Tese (Doutorado) Universidade Estadual de Campinas. Faculdade de Ciências Médicas.

1. Educação médica. 2. Estudantes universitários - avaliação. 3. Avaliação. I. Amaral, Eliana Martorano II. Zeferino, Angélica Maria Bicudo. III. Universidade Estadual de Campinas. Faculdade de Ciências Médicas. IV. Título.

Título em inglês: Global rating for medical students in clinical clerkship: comparison between self, faculty and peer assessment

Keywords:

- Medical education
- Students, medical - evaluation
- Evaluation

Titulação: Doutor em Tocoginecologia
Área de concentração: Ciências Biomédicas

Banca examinadora:
Profa. Dra. Eliana Martorano Amaral
Prof. Dr. Milton de Arruda Martins
Prof. Dr. Luiz Ernesto de Almeida Troncon
Prof. Dr. Gil Guerra Júnior
Prof. Dr. Aarão Mendes Pinto-Neto

Data da defesa: 25 - 11 - 2008

Diagramação e arte final: Assessoria Técnica do CAISM (ASTEC)

BANCA EXAMINADORA DA TESE DE DOUTORADO

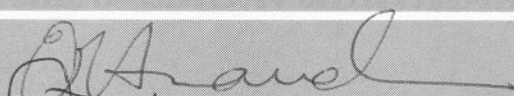
Aluna: ROSÂNGELA CURVO LEITE DOMINGUES

Orientadora: Prof^a. Dr^a. ELIANA MARTORANO AMARAL

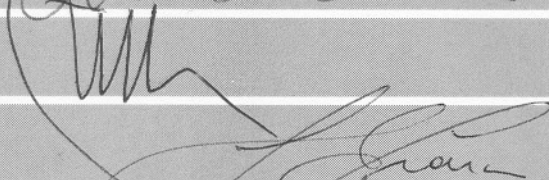
Co-Orientadora: Prof^a. Dr^a. ANGÉLICA MARIA BICUDO ZEFERINO

Membros:

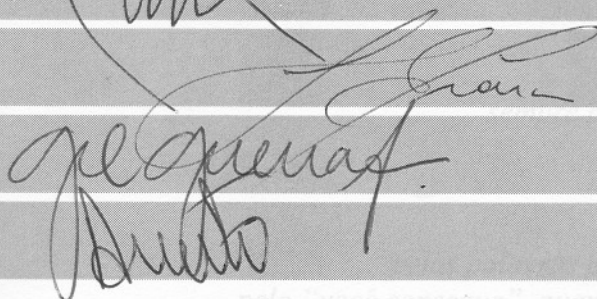
1.



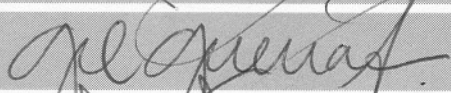
2.



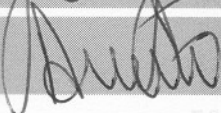
3.



4.



5.



Curso de Pós-Graduação em Tocoginecologia da Faculdade
de Ciências Médicas da Universidade Estadual de Campinas

Data: 25/11/2008

200904049

Dedico este trabalho...

*...aos meus pais Nicola e “Cida”,
cujo exemplo e incentivo
sempre impulsionaram a minha caminhada.*

*...à minha amada filha Paloma,
pelas palavras poderosas de otimismo e esperança,
pelo “você consegue”, quando as coisas pareciam impossíveis.*

*...ao meu querido marido Rodnei,
por me lembrar constantemente de que
por maior que fosse a montanha, ela nunca poderia tapar o sol.*

*ao meu irmão Cássio e à minha irmã Rita,
pelo carinho expresso nas ações mais sutis.*

Agradecimentos

“O valor das coisas não está no tempo em que elas duram, mas na intensidade com que acontecem. Por isso, existem momentos inesquecíveis, coisas inexplicáveis e pessoas incomparáveis.”

(Fernando Pessoa)

À Prof^a Dr^a Eliana Martorano Amaral, que com profissionalismo e competência, orientou minhas idéias durante a elaboração dessa pesquisa.

À Prof^a. Dr^a. Angélica Maria Bicudo Zeferino, que com coragem e determinação, apoiou essa linha de pesquisa e co-orientou o estudo.

Ao Prof^o Dr. Gil Guerra Jr e à Prof^a Dr^a Helaine Milanez pelas valiosas sugestões e comentários.

À amiga e Assessora Pedagógica do curso de Medicina da Faculdade de Ciências Médicas da Unicamp, Dr^a Silvia Maria Riceto Ronchim Passeri, pela colaboração, carinho a apoio em todos os momentos.

À amiga Vivian Cristina Rio, pelas cuidadosas revisões de texto, feitas com extrema competência.

Ao amigo e Prof. Dr. Armando Infante pelas preciosas orientações estatísticas e por me mostrar que é bom chegar quando se tem paciência e perseverança.

À estatística Sirlei Siani Morais pelas orientações, feitas com dedicação e carinho.

*Ao Prof. Dr. Wilson Nadruz Jr e à Profª Drª Maria Ângela Reis de Goes M Antonio,
pelas preciosas discussões sobre o escopo do projeto.*

*À Margarete Souza Donadon, secretária da pós-graduação, pelo carinho e disposição
de ajudar durante todo o curso.*

À Vanda F. F. Oliveira, bibliotecária do CAISM, pelas incansáveis buscas bibliográficas.

A todos os docentes e alunos participantes deste estudo.

Ao pessoal da ASTEC pela colaboração na formatação do trabalho.

Sumário

Símbolos, Siglas e Abreviaturas	xiii
Resumo	xv
Summary	xix
1. Introdução	23
2. Objetivos	35
2.1. Objetivo Geral.....	35
2.2. Objetivos Específicos	35
3. Sujeitos e Método	37
3.1. Tipo de estudo.....	37
3.2. Tamanho da amostra	37
3.3. Rotina das atividades do Módulo de Atenção Integral à Saúde	38
3.4. Seleção dos alunos	39
3.4.1. Critérios de inclusão.....	40
3.4.2. Critério de exclusão a posteriori.....	40
3.5. Seleção dos docentes	40
3.6. Variáveis.....	40
3.6.1. Dependentes	40
3.6.2. Independentes	42
3.7. Instrumentos para coleta de dados	43
3.7.1. Formulário de Conceito Geral	43
3.7.2. Formulário de Conceito Global Itemizado.....	43
3.7.3. Questionário de Opinião sobre a experiência de auto-avaliação e de avaliação dos pares	44
3.7.4. Tradução e Pré-teste dos Instrumentos.....	45
3.8. Coleta de dados	46
3.8.1. Docentes – Conceito Geral	46
3.8.2. Docentes – Conceito Global Itemizado.....	46
3.8.3. Alunos – Conceito Geral e Conceito Global Itemizado.....	47
3.8.4. Alunos – Opinião sobre a experiência de auto-avaliação e de avaliação dos pares.....	47
3.9. Aspectos gerais da análise estatística	47
3.10. Processamento e análise dos dados.....	50
3.11. Aspectos éticos.....	51
4. Publicações.....	53
4.1. Artigo 1 – Auto-avaliação e avaliação por pares - estratégias para o envolvimento do profissional médico	55
4.2. Artigo 2 – Feedback como estratégia de aprendizado no ensino médico	58
4.3. Artigo 3 – Avaliando competência clínica: o método de avaliação estruturada observacional	62
4.4. Artigo 4 – Conceito global: um método de avaliação de competência clínica	66
4.5. Artigo 5 – Competência clínica de alunos de medicina em estágio clínico: comparação entre métodos de avaliação	76
4.6. Artigo 6 – Global overall rating as a method of assessing students during clinical clerkship – what does it really show?	93

4.7. Artigo 7 – Global overall rating as method to assess students’ clinical competence on technical and humanistic skills: comparison between faculty, self and peer assessment.....	114
4.8. Artigo 8 – Peer assessment: how does it work in learning groups during clinical clerkship?.....	138
4.9. Artigo 9 – Os diferentes olhares na avaliação de alunos em estágio clínico supervisionado	162
5. Discussão	183
6. Conclusões.....	199
7. Referências Bibliográficas.....	201
8. Anexos	223
8.1. Anexo 1 – Objetivos do Módulo Atenção Integral à Saúde.....	223
8.2. Anexo 2 – Modelo de Portfólio	225
8.3. Anexo 3 – Avaliação Estruturada GO	227
8.4. Anexo 4 – Avaliação Estruturada CM	231
8.5. Anexo 5 – Avaliação Estruturada Ped.....	234
8.6. Anexo 6 – Formulário de Avaliação do Conceito Geral para Docentes.....	237
8.7. Anexo 7 – Formulário de Auto-avaliação do Conceito Geral	238
8.8. Anexo 8 – Formulário de Avaliação do Conceito Global Itemizado –Docentes.....	239
8.9. Anexo 9 – Formulário de Avaliação do Conceito Global Itemizado – Auto-avaliação ..	243
8.10. Anexo 10 – Formulário de Avaliação do Conceito Global Itemizado – Pares	247
8.11. Anexo 11 – Questionário de Opinião em relação à auto-avaliação e avaliação de pares.....	250
8.12. Anexo 12 – Parecer Comitê de Ética e Pesquisa.....	251

Símbolos, Siglas e Abreviaturas

- AA** – Auto-avaliação
- AAMC** – Association of American Medical Colleges
- ABEM** – Associação Brasileira de Educação Médica
- ABIM** – American Board of Internal Medicine
- ACGME** – Accreditation Council for Graduate Medical Education
- AE** – Avaliação Estruturada de Atendimento Clínico
- BEME** – Best Evidence in Medical Education
- CAISM** – Centro de Atenção Integral à Saúde da Mulher
- CM** – Clínica Médica
- CINAEM** – Comissão Interinstitucional Nacional de Avaliação do Ensino Médico
- CSA** – Clinical Structured Assessments
- D** – Docente
- DTG** – Departamento de Tocoginecologia
- FCM** – Faculdade de Ciências Médicas
- CG** – Conceito Geral
- CGI** – Conceito Global Itemizado
- CGIh** – Conceito Global Itemizado dos domínios humanísticos

- CGIt** – Conceito Global Itemizado dos domínios técnicos
- GR** – Global Rating
- GIR** – Global Itemized Rating
- GIRh** – Global Itemized Rating for humanistic domains
- GIRt** – Global Itemized Rating for technical domains
- GO** – Ginecologia e Obstetrícia
- GOR** – Global Overall Rating
- IM** – Internal Medicine
- MD756** – Módulo Atenção Integral à Saúde – 1º semestre
- MD856** – Módulo Atenção Integral à Saúde – 2º semestre
- MCQs** – Multiple-choice questions
- Mini-CEX** – Mini clinical evaluation exercise
- MSOP** – Medical School Objectives Project
- O&G** – Obstetrics and Gynecology
- OSCE** – Objective structured clinical examination
- P** – Par
- Ped** – Pediatria - Pediatrics
- Pf** – Portfólio
- Pt** – Prova Cognitiva
- RC-SCA** – Real-case - structured clinical assessment
- SUS** – Sistema Único de Saúde
- UBS** – Unidade Básica de Saúde

Resumo

Objetivo: Relacionar o Conceito Global Itemizado com o Conceito Geral dos alunos do 4º ano de Medicina, atribuídos por três tipos de avaliadores (docentes, colegas e próprio aluno) e compará-lo com os outros métodos de avaliação das competências esperadas no atendimento em unidade de atenção primária. **Sujeitos e métodos:** Foram estudadas as notas atribuídas a estudantes do 4º ano (n=106), por docentes (n=19), pelos próprios alunos e por colegas, ao final de um módulo de ensino que se desenvolve em 6 unidades básicas de saúde, sob supervisão de docentes e tutores de Ginecologia e Obstetrícia (GO), Clínica Médica (CM) e Pediatria (Ped). Os docentes forneceram uma nota (0 a 10) que resumia as competências do aluno no semestre, cuja média foi definida como Conceito Geral (CG). Depois de duas semanas, os mesmos docentes preencheram o formulário pré-testado do Conceito Global Itemizado (CGI) para os mesmos alunos, que continha 6 domínios de competências técnicas (CGIt) e 7 de competências humanísticas (CGIh). Os alunos preencheram formulários iguais (de CG e de CGI) para realizar sua auto-avaliação e apenas a parte das competências humanísticas (CGIh) para anonimamente avaliar 3-4 colegas que compunham o seu grupo. Por meio de um instrumento auto-respondido,

também expressaram sua opinião sobre o exercício de se auto-avaliar e de avaliar seus colegas. As notas de CGI foram comparadas com notas providas de outros métodos de avaliação: Provas Teóricas (Pt), Portfólio (Pf) e Avaliação Estruturada (AE). A análise estatística incluiu estatísticas descritivas, análise fatorial, coeficientes de Cronbach, testes de Friedman e de Wilcoxon para dados emparelhados, coeficientes de correlação de Pearson e de Spearman, modelos de análise de variância (ANOVA) e distâncias Euclidianas. Assumiu-se o nível de significância de 5%. Foi utilizado o programa SPSS Versão 10.0. O projeto foi aprovado pelo Comitê de Ética e Pesquisa da instituição. **Resultados:** Houve predominância de notas elevadas e discriminação reduzida entre os alunos. A consistência interna do questionário de CGI foi elevada para todos os avaliadores (alfa de Cronbach > 0,881). As notas de CG dos alunos foram significativamente inferiores às dos docentes (diferença de medianas = 0,41). As notas de CG foram significativamente inferiores às notas do CGI na avaliação docente (diferença de medianas = 0,13), mas não na auto avaliação e tiveram maior concordância e correlação com as notas de CGIt ($r = 0,873$ docente, $r = 0,715$ auto-avaliação) do que com as de CGIh ($r = 0,614$ docente, $r = 0,481$ auto-avaliação). As notas de CGIh foram significativamente mais elevadas do que as de CGIt, tanto na auto-avaliação (diferença de medianas = 0,17) quanto na avaliação docente (diferença de medianas = 0,19), com modestas correlações entre elas ($r = 0,62$ docentes, $r = 0,61$ auto-avaliação). Os alunos atribuíram notas significativamente mais baixas para os colegas do que para si mesmos (CGIh, diferença de medianas = 0,36). As correlações entre as notas dos três tipos de avaliadores para todos os domínios do CGI foram muito pobres e, em

geral, não significativas. As notas dos docentes de Ped foram significativamente inferiores às dos docentes de GO e CM (diferenças de medianas entre 0,50 e 0,67), com correlações baixas entre elas ($-0,02 < r < 0,48$). A maior discordância ocorreu entre as notas de GO e Ped, em todos os métodos. Os domínios que tiveram maior impacto na confiabilidade do CGI foram: *Exame físico*, *Solução de problemas*, *Capacidade de auto-reflexão* e *Julgamento clínico*. *Hábitos de trabalho* poderia ser excluído do instrumento itemizado. 90,6% dos alunos/avaliadores avaliaram todos os seus colegas de grupo. Houve indicações de reciprocidade e constância no padrão das notas entre pares e situações sugestivas de conflitos interpessoais. Os alunos se sentiram mais confortáveis e preparados para a auto-avaliação do que para a avaliação dos colegas (diferença de 34,4% e 17,7%, respectivamente) e julgaram a auto-avaliação mais justa e valiosa do que a avaliação dos colegas (diferença de 4,2% e 34,7%, respectivamente). Houve diferenças significativas entre as distribuições das notas de todos os métodos, com notas mais elevadas na AE (mediana = 9,27). Houve correlação forte e significativa ($r = 0,719$) e maior concordância entre as notas da AE e as do CGI. **Conclusão:** Os alunos foram mais rigorosos consigo mesmo do que os docentes no CG e nos atributos técnicos do CGI e com seus pares nos atributos humanísticos do CGI. As notas do CG foram inferiores às do CGI para os docentes, mas não para os próprios alunos e enfocaram principalmente as competências técnicas dos estudantes. O CGI mostrou-se uma estratégia válida e confiável para compor o conjunto de métodos de avaliação dos estudantes do 4º ano de medicina, em estágio clínico.

Summary

Purpose: To correlate the Global Itemized Rating with the Global Overall Rating attributed to Y4 medical students by three types of assessors (faculty, peers and self), and to compare it with other assessment methods used to evaluate learners in clinical clerkship, at the primary level of care. **Subjects and Methods:** The focus of the present study were ratings attributed to Y4 medical students (n= 106) by faculty members (n= 19), by own students and peers, at the end of a teaching module conducted at 6 primary care centers, and supervised by faculties and tutors from Obstetric and Gynecology (O&G), Internal Medicine (IM) and Pediatrics (Ped). Faculties provided a summative global numeric rating, ranging from 0 to 10, which summarized student's competencies during the clerkship. Faculty global scores across disciplines were averaged and defined as a Global Overall Rating (GOR). Two weeks after securing GOR ratings for all students, the same faculty members assessed the same students using in a pre-tested Global Itemized Rating, named GIR, which encompassed thirteen domains: six referred to technical skills (GIRt), and seven to humanistic skills (GIRh). Students completed similar GOR and GIR forms for self-assessment, and the humanistic scoring components of the GIR form to anonymously assess three to five working-group peers.

Students answered a questionnaire on their perceptions to the experience as self and peer raters. GIR ratings were compared with ratings from other assessment methods: Multiple-choice questions (MCQs), Portfolio (Pf) and Real-case structured clinical assessment (RC-SCA). Statistical analyses included descriptive statistics, factor analysis, Cronbach's alpha, Friedman test, Wilcoxon signed rank test, Pearson and Spearman correlation coefficients, ANOVA models and Euclidean distances. Statistical significance was set at $P < 0.05$. Statistical analyses were performed using SPSS program, version 10.0. Ethical approval was obtained.

Results: Distributions were negatively skewed and reduced discrimination among students was observed. GIR internal consistency was high, across all types of raters (Cronbach's alpha > 0.881). Students' GOR ratings were significantly lower than faculties' (median difference = 0.41). GOR ratings were significantly lower than GIR ratings for faculty (median difference = 0.13), but not for self, and showed stronger agreement and correlation with GIRt ($r = 0.873$ for faculty and $r = 0.715$ for self) than with GIRh ($r = 0.614$ for faculty and $r = 0.481$ for self). GIRh ratings were significantly higher than GIRt (median difference = 0.17 for self and 0.19 for faculty), with modest correlations between them ($r = 0.62$ for faculty and $r = 0.61$ for self). Learners attributed lower mean ratings to their fellow students than to themselves (CGIh median difference = 0.36). Correlations between faculty, self and peer scores were very poor in all domains, and generally non significant. Ped scores were significantly lower than O&G and IM (median differences between 0.50 and 0.67), with low correlations between them ($-0.02 < r < 0.48$). O&G and Ped scores showed greater disagreement, in all methods. The domains that had greater impact on the reliability of the scale were: *Physical examination, Problem-solving, Self-*

reflective skills and *Clinical Judgment*. *Work habits* could be excluded from the itemized form. 90.6% of the learners/raters assessed all their group-peers. Patterns of reciprocity and constancy in ratings, and some situations suggesting conflicts were noticed in working groups. Learners expressed to be more comfortable and prepared for self than for peer assessment (difference = 34.4% and 17.7%, respectively), and considered self-assessment more fair and valuable than peer assessment (difference = 4.2% and 34.7%, respectively). There were significant differences between score distributions in all assessment methods, with higher ratings in RC-SCA (median = 9.27). RC-SCA and GIR showed strong and significant correlations ($r= 0.719$) and greater agreement. **Conclusion:** Learners were stricter with themselves than faculty in GOR and in GIR technical domains, and with their colleagues in GIR humanistic domains. GOR ratings were lower than GIR ratings for faculty, but not for self, and focused mainly on students' technical skills. GIR has shown to be a valid and reliable strategy, which can make up the set of assessment methods used to evaluate fourth-year students in clinical clerkship.

1. Introdução

O ensino e a prática profissional são dimensões centrais do marco teórico-conceitual da educação médica e estão sujeitos a mudanças decorrentes de transformações da estrutura econômica, como acontece com os demais componentes da sociedade. O contexto atual, marcado pelo contínuo desenvolvimento científico-tecnológico e por complexos quadros demográfico-sanitários, em grande parte determina a importância, o lugar e a forma da medicina na estrutura social, trazendo novas expectativas em relação ao papel do médico no século XXI. Dessa forma, faz-se necessário que os profissionais ligados à educação médica reexaminem o conceito de profissionalismo e repensem os processos de como ele pode ser ensinado e avaliado (Almeida, 1999; Beauchamp, 2004).

Profissionalismo é uma questão central e complexa na educação e prática médicas, pois envolve ações, normas e valores presentes nas diversas interações com pacientes e seus familiares, colegas e outros profissionais (Swick, 2000). O termo está vinculado a um conjunto de diferentes elementos relacionados a atitudes e comportamentos dos profissionais na prática da medicina, ao comprometimento de

servir ao paciente e à sociedade e à obrigação moral de atuar com responsabilidade e ética. Quando um médico aceita a tarefa de cuidar de um paciente, subentende-se que ele apresente comportamentos indicativos de profissionalismo, incluindo, respeito pelo outro, comunicação eficaz, integridade, julgamento clínico, conhecimento médico, habilidade técnica e de trabalhar em equipe (ACGME, 1999; Arnold, 2002; Beauchamp, 2004; Duff, 2004).

Esse conjunto de atributos está contemplado nas Diretrizes Curriculares para os cursos de Medicina, promulgadas pelo Conselho Nacional de Educação em 2001. O perfil esperado do egresso/profissional é que ele tenha uma formação generalista, humanista, crítica e reflexiva e esteja capacitado a atuar com responsabilidade social e compromisso com a cidadania, como promotor da saúde integral do ser humano (Brasil, 2001; World Federation for Medical Education, 2001). Para que se possam atingir tais expectativas, é necessário um plano de curso abrangente e articulado, no qual a avaliação desempenha um papel relevante, orientando o processo ensino-aprendizagem (van der Vleuten, 2005).

Avaliação pode ser definida como um processo contínuo e sistemático de coleta, síntese e interpretação de informações qualitativas e quantitativas que facilitam tomadas de decisão, permitindo inferências sobre o desenvolvimento do aluno e a qualidade do ensino (Bordenave e Pereira, 1982; Haydt, 2002).

Na área médica, o processo de avaliação apresenta grandes desafios, começando pela definição de competência, passando pela descrição dos atributos que a constituem e encerrando com a seleção de medidas válidas e confiáveis,

que levem em consideração questões relacionadas a custo e à aceitação pelos avaliadores e avaliados (Souza e Rangel, 2003; AAMC, 2004; Silber, 2004).

Algumas entidades internacionais têm se debruçado sobre a discussão em torno do tema competência médica. A *Association of American Medical Colleges* (AAMC) iniciou um amplo projeto denominado *Medical School Objectives Project* (MSOP) com a finalidade de auxiliar as escolas médicas em seus esforços de responderem às demandas da sociedade quanto ao papel do médico no século XXI. Representantes dos 14 países participantes elaboraram um relatório de consenso que contempla os seguintes grupos de características esperadas do médico: (1) *altruísmo*: empatia, respeito à privacidade e à dignidade do paciente, atenção, cuidado, compreensão do contexto cultural e dos valores do paciente e de seus familiares; (2) *conhecimento*: compreensão das bases científicas da medicina e competência para colocá-las em prática; (3) *habilidade*: competência para fornecer cuidados médicos apropriados, obter história médica precisa, construir raciocínio clínico, realizar exames físicos, procedimentos diagnósticos e terapêuticos e para se comunicar eficazmente com pacientes e familiares; (4) *compromisso e colaboração*: compreensão dos fatores econômicos, psicológicos, ocupacionais, sociais e culturais envolvidos na atenção ao paciente; comprometimento para trabalhar em conjunto com outros profissionais da saúde; aplicação dos princípios da medicina baseada em evidência; capacidade para tomar decisões e lidar com recursos médicos limitados (AAMC, 1998; Horowitz, 2000).

Outro projeto foi iniciado em 1999, pela *ACGME (Accreditation Council for Graduate Medical Education)*, com o objetivo de imprimir uma abordagem

multidimensional à medida da competência clínica. Denominado *Outcome Project*, identificou seis competências gerais aplicáveis a qualquer especialidade ou área de atuação: (1) *cuidado com o paciente*, (2) *conhecimento médico*, (3) *aprendizagem e melhoria baseada na prática*, (4) *habilidades interpessoais e de comunicação*, (5) *profissionalismo* e (6) *prática baseada em sistemas* (Horowitz, 2000; Sliwa e Kowalske, 2000; AAMC, 2004).

Em 2000, o Projeto *CanMED* (*Canadian Medical Education Directions for Specialists*) descreveu sete papéis profissionais esperados, que deveriam ser incluídos em todos os níveis da educação médica, por representarem a estrutura de uma boa prática clínica: especialista, comunicador, colaborador, gerente, defensor da saúde, comprometido com educação continuada e profissional (Davis e Harden, 2003). Apesar dessa abordagem multidimensional de competência clínica, vários estudos sustentam que na prática ela é, em essência, bidimensional, pois os alunos são avaliados por duas dimensões substancialmente correlacionadas: (1) *competência humanística* e (2) *competência técnica* (DiMatteo e DiNicola, 1981; Ramsey, 1993; Silber, 2004).

Uma importante questão que tem permeado as discussões no campo da educação médica é a distinção entre os termos competência e desempenho. Embora ambos sejam utilizados de forma intercambiável, parece haver um consenso no entendimento de competência significando a mobilização de diversos saberes, relacionados a conhecimentos, habilidades e atitudes em situações controladas da prática profissional, enquanto desempenho está vinculado à ação propriamente dita, em situações profissionais reais (Depresbiteris, 2001; Epstein e

Hundert, 2002; Rethans et al., 2002; Wilkinson e Frampton, 2004). Dessa forma, o desempenho depende das competências a ele subjacentes e, portanto, permite inferi-las (Ramos, 2001).

Todas essas discussões teóricas repercutem na prática do ensino médico, uma vez que um aspecto essencial dessa abordagem baseada em competências relaciona-se à avaliação dos resultados atingidos. Assim, muitas universidades e órgãos de classe estão unindo esforços para repensar maneiras de sistematizar a avaliação, acrescentando-lhe a necessária perspectiva indutora de novas práticas pedagógicas na educação médica (DeLisa, 2000; Smith, Dollase, Boss, 2003).

A avaliação engloba funções que incluem assegurar que os alunos atinjam as qualificações pré-estabelecidas, estimar de forma precisa o nível dos alunos, oferecendo-lhes informações sobre seu progresso, disponibilizar ferramentas para os estudantes monitorarem seu aprendizado e fornecer a professores e gestores do ensino subsídios para a revisão dos métodos, processos pedagógicos e propostas de mudanças no currículo (Haydt, 2002; Carr, 2004; Ben-David, 2005). Portanto, a avaliação pode ser *formativa*, cujo objetivo central é orientar o processo de aprendizagem, ou *somativa*, que visa determinar se o estudante atingiu as competências esperadas (Ben-David, 2005). Para que essas funções sejam cumpridas adequadamente, o sistema de avaliação deve ser baseado em competências e conteúdos curriculares esperados, tendo como referência as Diretrizes Curriculares para os cursos de Medicina (Brasil, 2001).

Além disso, como a avaliação direciona o aprendizado (van der Vleuten, 2005), os instrumentos utilizados devem ser abrangentes, atrelando-se ao desenho instrucional, e capazes de validar os objetivos constantes no currículo (Wass, et al., 2001). Para assegurar a qualidade dos resultados, dois critérios devem ser cumpridos: (1) confiabilidade e (2) validade. A confiabilidade (*reliability*), de forma ampla refere-se à reprodutibilidade dos escores obtidos pelos diferentes tipos de avaliadores e à fidedignidade do instrumento (expressa pelo grau de homogeneidade dos itens da escala utilizada). A validade (*validity*) está relacionada a quão bem um instrumento mede aquilo a que se propõe. Vale ressaltar que os educadores devem levar em conta também o impacto dos resultados sobre a aprendizagem. Supõe-se que qualquer método possa ter uma influência (positiva ou negativa) sobre a aprendizagem, dependendo de como ele é usado e em que contexto. Assim, a avaliação deixa de ser apenas uma questão psicométrica e passa também a ser vista a partir de uma perspectiva de desenho educacional, englobando todo o currículo (AGCME, 2000; van der Vleuten, 2005).

É certo que, para produzir resultados válidos e confiáveis, a avaliação da competência médica deve utilizar diferentes métodos que enfoquem atributos distintos (Murray et al., 2000; Wilkinson e Frampton, 2004; Driessen et al., 2005; van der Vleuten, 2005). Dentre esses métodos incluem-se: (a) Avaliação Estruturada de Atendimento Clínico sob observação, que pode utilizar um instrumento do tipo “checklist” abrangendo as várias competências esperadas; (b) Portfólio, que visa mostrar as experiências clínicas, podendo ser associado à auto-avaliação e a um plano de aprendizagem; (c) Prova Teórica apropriada para avaliação cognitiva,

referente a conhecimentos técnicos específicos; (d) Avaliação 360°, voltada para a obtenção de informações do desempenho do aluno provenientes de múltiplas fontes (colegas, professores, outros profissionais e pacientes; (e) OSCE (*objective structured clinical examination*) ou CSA (*Clinical Structured Assessment*), destinado a avaliar o desempenho clínico durante atendimento simulado; (f) Mini-CEX (*mini-clinical evaluation exercise*), avaliação de desempenho clínico com uso de formulário específico; (g) Conceito Global, avaliação retrospectiva geral das competências clínicas (ACGME, 2000; Auewarakul, 2005; Farrell, 2005; Epstein, 2007; Norcini e McKinley, 2007).

Embora não haja um padrão-ouro para avaliar o desempenho do médico/estudante de medicina, o Conceito Global parece estar próximo de uma avaliação abrangente (van der Vleuten, 2005). No entanto, como ele não inclui nenhuma especificação dos atributos que o compõem, pode se tornar extremamente subjetivo e sujeito a vieses (Regehr et al., 1998), uma vez que esse julgamento geral é influenciado mais pela impressão que o avaliador forma do avaliado do que pelas habilidades deste (Norcini, 2003). Uma maneira de garantir que diferentes domínios da competência do aluno sejam contemplados pelo avaliador é usar um conceito com itens/atributos claramente especificados, ancorados por descritores que possam orientar a ação de avaliação (Bandiera et al., 2002). Mas, para produzir resultados confiáveis, o Conceito Global deve levar em consideração as impressões do avaliador a respeito das competências demonstradas pelo estudante em uma variedade de situações, por um determinado período de tempo, após inúmeras observações diretas (Norcini, 2003).

Independentemente do método utilizado, a avaliação tem sido uma responsabilidade essencialmente do corpo docente. No entanto, dentro da perspectiva de capacitar os alunos para seu desenvolvimento profissional continuado e para atuação em equipes, o projeto pedagógico do curso de medicina deve ter o professor como facilitador e mediador e o estudante como centro do processo de ensino-aprendizagem (CINAEM, 1997; Brasil, 2001). Há, portanto, uma crescente tendência de responsabilização e envolvimento dos estudantes, no sentido de torná-los agentes ativos de sua própria avaliação e da de seus pares (van der Vleuten, 2005).

A auto-avaliação representa uma importante estratégia de ensino, pois prepara o aluno para repensar os resultados das suas próprias ações profissionais, para refletir sobre o que aprendeu, para avaliar como tal aprendizado o preparou para realizar as tarefas esperadas e para perceber suas necessidades individuais de aprimoramento (Gordon, 1991; Fitzgerald, White, Gruppen, 2003; Edwards et al., 2003; Beyeler et al., 2004). Essa habilidade de identificar valores e atitudes pessoais, reconhecendo os próprios pontos fortes e fracos, pode ser desenvolvida, aperfeiçoada e modificada com *feedback* adequado e constante e é considerada essencial para manutenção e melhoria da proficiência médica e para o bom relacionamento com pacientes e colegas (Edwards et al., 2003).

São diversos os autores que defendem o uso da auto-avaliação em conjunto com a avaliação feita pelos pares e pelos professores, como parte integrante do desenvolvimento contínuo dos estudantes, auxiliando-os a refletir sobre questões como profissionalismo e habilidade de trabalhar em equipe

(Ramsey, 1993; Davis, 2002; Epstein e Hundert, 2002; Norman, Shannon, Marrin, 2004; Wilkinson e Frampton, 2004; van der Vleuten e Schuwirth, 2005). Entende-se que quanto mais cedo esses elementos de reflexão, julgamento crítico, elaboração e atualização de plano de ação forem acrescentados ao contexto de ensino-aprendizagem, maiores serão os benefícios para os estudantes (Henderson e Johnson, 2002; Mattheos, 2004).

A comparação da auto-avaliação com a avaliação feita pelos pares e professores tem como objetivo fornecer aos educadores informações úteis para o desenvolvimento e planejamento do currículo e para a revisão da metodologia adotada (Southgate et al., 2001; Davis, 2002). Como o docente não está observando o estudante todo o tempo, pode perder informações importantes a respeito de seu desempenho. Ouvir a perspectiva de colegas com quem convive por tempo prolongado e em diferentes situações pode levar a mudanças positivas de comportamento do aluno, especialmente no que se refere à competência humanista (van Rosendaal e Jennett, 1994; Sullivan, Hitchcock, Dunnington, 1999; Norcini, 2003; Evans, Elwyn, Edwards, 2004; Dannefer et al., 2005; Shue, Arnold, Stern, 2005; Ramsey e Owen, 2006). Portanto, a avaliação por pares também constitui um excelente material para *feedback* (Gordon et al., 2000; Rolfe e Sanson-Fisher, 2002; Henderson, Ferguson-Smith, Johnson, 2005).

A cultura da avaliação tem feito parte essencial do projeto pedagógico da Faculdade de Ciências Médicas da Universidade Estadual de Campinas desde 2001, quando se iniciou a implantação de um currículo integrado horizontal e verticalmente, visando à interdisciplinaridade. Foram criados módulos compostos

por diferentes áreas de conhecimento, em substituição às antigas disciplinas; houve um aumento significativo da integração ensino-serviço, uma exposição precoce às necessidades de saúde do indivíduo e das comunidades e uma ampliação dos cenários de prática médica. O novo currículo, voltado para a descompartmentalização do ser humano, facilita a integração entre teoria e prática (Campos, 2005). Especificamente, no quarto ano, inicia-se o estágio clínico nas Unidades Básicas de Saúde (UBS), com o objetivo de propiciar a visão clínica do paciente como um todo, permitindo ao estudante maior responsabilização pelo atendimento, maior vínculo com as equipes de saúde e melhor entendimento do Sistema Único de Saúde.

Nesse estágio, denominado *Módulo de Atenção Integral à Saúde*, o aluno é exposto, sob tutela, durante todo o ano letivo, a um modelo complexo de aprendizagem baseado no contato direto com o paciente, envolvendo as especialidades de Ginecologia e Obstetrícia, Clínica Médica e Pediatria. Diante desse novo cenário educacional, tornou-se imprescindível propor um programa abrangente de avaliação das competências dos alunos.

Quatro métodos de avaliação têm sido utilizados ao longo do ano: 1) Prova Teórica, com testes e respostas curtas; 2) Portfólio, dossiê elaborado pelo aluno, contendo registro dos casos atendidos, discussão de casos selecionados, baseada na literatura e auto-avaliação de aprendizado; 3) Avaliação Estruturada do Atendimento Clínico, que representa a observação direta de uma consulta com paciente real e 4) Conceito Geral, que reflete a impressão do docente acerca do desempenho do aluno no estágio clínico.

A natureza pouco específica do Conceito Geral não permite saber claramente quais competências estão sendo avaliadas. Assim, os docentes e estudantes não podem usufruir das vantagens formativas que um processo avaliativo pode proporcionar. Além disso, até este momento do curso, o estudante não é encorajado a refletir sobre a sua competência clínica e nem a de seus colegas. Finalmente, o sistema atual de avaliação não contempla informações acerca da sobreposição ou da complementaridade entre os resultados obtidos por meio de diferentes métodos de avaliação.

Este estudo visou avaliar a experiência de utilização da auto-avaliação e avaliação de pares, comparando-as com as dos docentes. Pretendeu-se também analisar a utilidade de um conceito itemizado que fornece pontuação para distintos atributos relacionados a conhecimentos, habilidades e atitudes dos estudantes e que pode servir de base para um *feedback* posterior. Finalmente, objetivou-se compreender a inter-relação entre os diferentes métodos de avaliação. Os resultados encontrados poderão contribuir para orientar educadores e gestores a repensarem o processo de avaliação, em um sistema baseado nas competências dos profissionais que se pretende formar.

2. Objetivos

2.1. Objetivo Geral

Relacionar o conceito global itemizado com o conceito geral dos alunos do 4º ano de Medicina, atribuídos por três tipos de avaliadores (docentes, colegas e próprio aluno) e compará-lo com os outros métodos de avaliação dos alunos nas unidades de atenção primária.

2.2. Objetivos Específicos

- Comparar o conceito geral de cada aluno com o conceito global itemizado, atribuído pelos tipos de avaliadores: docente e próprio aluno.
- Comparar os conceitos globais itemizados que o aluno recebe de cada um dos três tipos de avaliadores (ele mesmo, docente, par).
- Comparar o conceito global itemizado do aluno dado pelos docentes dentre três especialidades médicas: Ginecologia e Obstetrícia, Clínica Médica e Pediatria.

- Verificar quais itens do CGI têm mais peso para os diferentes tipos de avaliadores.
- Identificar as características do processo de avaliação por pares e a opinião dos alunos sobre a auto-avaliação e a avaliação por pares.
- Comparar o conceito global itemizado dado pelo docente com as notas provindas de outros métodos de avaliação: provas teóricas, portfólio e avaliação estruturada do atendimento clínico.

3. Sujeitos e Método

3.1. Tipo de estudo

Trata-se de um estudo observacional cujos dados foram gerados por diversos métodos de avaliação, aplicados a alunos do 4º ano médico em estágio clínico, na atenção primária à saúde.

3.2. Tamanho da amostra

Não foi necessário calcular o tamanho amostral, uma vez que o estudo foi aplicado ao universo total de alunos ($n=106$) do *Módulo de Atenção Integral à Saúde* (MD 856) em 2005. Esse número é suficiente para aplicar a análise fatorial, utilizada para verificar a inter-relação entre os itens do conceito global itemizado, para a qual se recomenda um tamanho de amostra correspondente a (a) 5 vezes o número de questões do instrumento do conceito itemizado ($n=65$), ou (b) um mínimo de 100 sujeitos (Johnson e Wichern, 1992; Hair et al., 1998).

3.3. Rotina das atividades do Módulo de Atenção Integral à Saúde

As atividades de estágio clínico supervisionado do *Módulo de Atenção Integral à Saúde*, denominado MD 756 no primeiro semestre e MD 856 no segundo semestre, são realizadas em 6 Unidades Básicas de Saúde (UBS), selecionadas pela Universidade Estadual de Campinas, em parceria com a Secretaria Municipal de Saúde de Campinas. Esse estágio, com duração aproximada de 200 horas, distribuídas ao longo de 9 meses, tem por objetivo fundamental contribuir para a formação geral do estudante, inserindo-o na rede básica do Sistema de Saúde (a descrição completa dos objetivos do módulo é apresentada no **Anexo 1**). A supervisão dos alunos é realizada por docentes das três especialidades clínicas envolvidas no estágio – Ginecologia e Obstetrícia, Clínica Médica e Pediatria – com a colaboração de tutores, profissionais que trabalham na Rede Municipal de Saúde, capacitados para a docência e que foram esclarecidos sobre os objetivos do módulo e sobre o papel compartilhado entre a rede de saúde e a Universidade na formação de recursos humanos.

O estágio clínico na UBS acontece três tardes por semana. Os alunos são divididos em 24 grupos de até 5 alunos cada, sendo que 20 alunos (4 grupos) ficam fixos em uma das seis UBS durante os dois semestres e são supervisionados pelos mesmos tutores e docentes. Os grupos dirigem-se à unidade em semanas alternadas, mantendo, assim, o acompanhamento do paciente. Toda semana, dez alunos (dois grupos) se revezam, atendendo as três especialidades clínicas (4^a feiras, cinco alunos em Pediatria e cinco em Ginecologia e Obstetrícia; 5^a

feiras, o mesmo com Pediatria e Clínica Médica e 6ª feiras, Clínica Médica e Ginecologia e Obstetrícia).

Nos atendimentos individuais ao paciente, cada aluno deve praticar a anamnese e o exame físico completos, registrar as informações de modo claro e ordenado e integrar os achados clínicos com os aspectos relevantes de história de vida do paciente. Deve, ainda, indicar condutas diagnósticas e terapêuticas para as situações mais comuns, realizar prescrição sob supervisão, reconhecer a adequação da dieta e das doses de medicamentos. As consultas são complementadas por discussão de casos com docentes, tutores e outros membros da equipe de saúde, visitas domiciliares e participação em reuniões de projeto terapêutico multiprofissional.

A estrutura desse módulo de ensino incentiva a valorização do vínculo do aluno com o paciente, uma vez que ele pode fazer o processo integrativo, atendendo, por exemplo, a criança na puericultura, a mãe na revisão puerperal e a avó hipertensa e diabética na clínica. Essa inserção mais precoce do estudante na prática clínica na atenção básica permite considerar que o internato ou o treinamento em serviço teve início real no quarto ano do curso.

3.4. Seleção dos alunos

- Foram avaliadas as notas atribuídas a todos os alunos do 4º ano de Medicina da FCM/UNICAMP no MD 856 no ano de 2005, totalizando 106.

3.4.1. Critérios de inclusão

- Estar regularmente matriculado no MD 856.

3.4.2. Critério de exclusão a posteriori

- Foram excluídos os alunos que não tiveram as avaliações preenchidas por completo.

3.5. Seleção dos docentes

- Participaram do estudo todos os docentes que supervisionaram e avaliaram os alunos durante o MD 856, totalizando 19.

3.6. Variáveis

3.6.1. Dependentes

a) **Conceito Global Itemizado:** média atribuída ao estudante a partir do julgamento expresso pelos docentes, colegas e pelo próprio aluno em relação a 13 atributos indicativos da competência do estudante, classificados em:

- **Competência técnica:** conjunto de atributos que definem o desempenho técnico do estudante:
 - **Qualidade da história:** obtenção de informações clínicas críticas; qualidade das anotações: de 0 a 10.
 - **Exame clínico:** realização das manobras adequadas do exame clínico e/ou dos procedimentos considerados essenciais: de 0 a 10.

- **Conhecimento médico:** qualidade do conhecimento teórico: de 0 a 10.
 - **Julgamento clínico:** integração das informações da história com dados clínicos; qualidade do diagnóstico; desenvolvimento de plano terapêutico: de 0 a 10.
 - **Solução de problemas:** avaliação de informações, riscos e questões essenciais em relação à queixa do paciente: de 0 a 10.
 - **Hábitos de trabalho:** pontualidade; completude dos registros, adesão a princípios éticos: de 0 a 10.
- **Competência humanística:** conjunto de atributos que definem o desempenho humanístico do estudante:
 - **Comunicação e relacionamento com paciente e familiares:** interação estabelecida com o paciente e seus familiares; clareza das informações; habilidade de ouvir o paciente e seus familiares; envolvimento com paciente: de 0 a 10.
 - **Respeito ao paciente:** atenção e atitude acolhedora para com o paciente e seus familiares: de 0 a 10.
 - **Capacidade de auto-reflexão:** responsabilidade assumida em relação às próprias ações e decisões; atitude em relação às críticas recebidas: de 0 a 10.
 - **Percepção do contexto:** consideração e sensibilidade em relação às necessidades de conforto e ajuda aos pacientes e seus familiares: de 0 a 10.
 - **Interação com colegas:** relacionamento com os colegas; atitude de cooperação e ajuda: de 0 a 10.

- **Interação com docentes:** solicitação de orientação; verificação do próprio entendimento com o professor: de 0 a 10.
 - **Interação com demais profissionais:** cooperação com outros profissionais; habilidade para trabalhar em equipe: de 0 a 10.
- b) **Conceito geral:** nota atribuída ao estudante com base na impressão geral de seu desempenho, expressa pelo docente e pelo próprio aluno: de 0 a 10.
- c) **Prova teórica:** média obtida a partir das provas cognitivas: de 0 a 10.
- d) **Portfólio:** nota obtida a partir da avaliação do portfólio, feita pelos docentes: de 0 a 10. Um exemplo do formulário do portfólio está no **Anexo 2**.
- e) **Avaliação Estruturada de Atendimento Clínico:** média obtida a partir das notas de avaliação estruturada, preenchida pelos docentes das três especialidades (Ginecologia e Obstetrícia, Clínica Médica e Pediatria): de 0 a 10. Os formulários de Avaliação Estruturada usados pelas diferentes especialidades são apresentados nos **Anexo 3** (GO), **Anexo 4** (CM) e **Anexo 5** (Ped).

3.6.2. Independentes

- a) **Tipo de avaliador:** atuação do indivíduo no processo de avaliação, classificado em: D (docente), P (par), AA (auto-avaliação).
- b) **Especialidade:** área médica na qual o estudante será avaliado pelo docente, classificada em: GO (Ginecologia e Obstetrícia), CM (Clínica Médica) e Ped (Pediatria).

3.7. Instrumentos para coleta de dados

Neste estudo foram utilizados três instrumentos para coleta de dados, assim denominados: Formulário de Conceito Geral (**Anexos 6 e 7**), Formulário de Conceito Global Itemizado (**Anexos 8, 9 e 10**) e Questionário de Opinião dos Alunos sobre a experiência de auto-avaliação e de avaliação dos pares (**Anexo 11**).

3.7.1. Formulário de Conceito Geral

Esse formulário foi preenchido por dois tipos de avaliadores: docente e próprio aluno. Constava de apenas uma página e era dividido em 2 seções: (1) a primeira seção incluía explicações sobre o preenchimento do instrumento e dados de identificação do avaliador e do avaliado e (2) a segunda seção solicitava ao avaliador que atribísse um conceito geral ao estudante, marcando uma nota entre 0 e 10 que refletisse a impressão dele sobre o avaliado, ao longo do estágio. Abaixo dessa questão, havia um espaço reservado para comentários adicionais.

3.7.2. Formulário de Conceito Global Itemizado

Esse formulário foi elaborado com base no instrumento desenvolvido por Ramsey et al. (1993), modificado por Davis (2002), levando em consideração as competências estabelecidas pela ACGME (Chapman et al., 2004) e os objetivos do módulo de ensino do 4º ano. Após uma proposta inicial, os itens do formulário foram discutidos e validados pelo grupo gestor do *Módulo de Atenção Integral à Saúde*. Esse formulário foi preenchido integralmente pelos docentes e pelo próprio

aluno e parcialmente pelos colegas (apenas a parte do formulário referente às habilidades humanísticas).

O instrumento era composto de 2 seções: (1) a primeira seção incluía explicações sobre o preenchimento do instrumento e dados de identificação do avaliador e do avaliado, e (2) segunda seção era composta de 13 atributos, divididos em 2 partes: parte 1 - competência técnica (*qualidade da história, exame clínico conhecimento médico, julgamento clínico, solução de problemas e hábitos de trabalho*), parte 2 - competência humanística (*comunicação e relacionamento com o paciente, respeito, capacidade de auto-reflexão, percepção do contexto, interação com colegas, interação com os docentes, interação com os demais profissionais*).

Abaixo de cada atributo havia uma breve explicação para o escore de limite inferior e para o de limite superior na escala. O avaliador marcou uma nota entre 0 e 10, que representava a opinião dele sobre o aluno naquele domínio específico. Após cada item, havia um espaço reservado para comentários adicionais que o avaliador desejasse fazer.

3.7.3. Questionário de Opinião sobre a experiência de auto-avaliação e de avaliação dos pares

Esse questionário foi respondido pelos alunos e teve por objetivo conhecer a opinião deles sobre a experiência de auto-avaliação e de avaliação dos pares. Era composto de oito perguntas respondidas, para as quais cada estudante deveria marcar uma nota de 0 a 10 que melhor expressasse sua opinião sobre aquela

determinada questão. Abaixo de cada pergunta havia um espaço reservado para comentários adicionais que o aluno desejasse fazer.

3.7.4. Tradução e Pré-teste dos Instrumentos

Após ser traduzido, conforme os critérios explicitados a seguir, o instrumento do Conceito Global Itemizado foi pré-testado, no 2º semestre de 2005, em doze alunos do 5º ano de Medicina e um docente, ao final do Módulo Atenção Integral à Saúde da Criança e do Adolescente I, módulo que têm experiência clínica na atenção primária à saúde, semelhante a dos alunos do 4º ano.

O texto original foi primeiramente traduzido pela pesquisadora responsável, que é proficiente na língua inglesa e familiarizada com o conteúdo do material na linguagem-fonte. Isto, segundo Brislin (1970), é condição indispensável para uma tradução adequada. Uma segunda tradução foi realizada pela orientadora do projeto, que também é proficiente na língua inglesa. Os dois trabalhos foram revisados por outro pesquisador bilíngüe (Garyfallos et al., 1991), e que, ao examinar item por item, auxiliou na escolha das expressões que melhor reproduziam o sentido original. Assim, chegou-se a uma tradução de consenso, na qual cada afirmação trouxe o significado mais aproximado do questionário em inglês.

Um cuidado considerável foi necessário para assegurar que a tradução atingisse um nível de uso habitual de linguagem que fosse, ao mesmo tempo, equivalente à fonte inicial e compatível com a linguagem usada pelos sujeitos que se têm em perspectiva. De modo geral, as questões foram consideradas

compreensíveis e adequadas. Como resultado do pré-teste, foi feita uma alteração nesse instrumento: no item 9, a palavra *compaixão* foi substituída por *percepção do contexto*.

3.8. Coleta de dados

3.8.1. Docentes – Conceito Geral

Os docentes foram apresentados aos instrumentos de avaliação que seriam utilizados, em uma reunião específica, em horário normal de atividade acadêmica, para que pudessem conhecer os instrumentos e os procedimentos de preenchimento, assim como esclarecer dúvidas. Após esse momento, os docentes receberam o formulário de avaliação do Conceito Geral em suas respectivas UBS e preencheram um formulário para cada aluno, devolvendo-os em um prazo pré-estabelecido.

3.8.2. Docentes – Conceito Global Itemizado

Aproximadamente duas semanas após o preenchimento do formulário de Conceito Geral, os docentes receberam o formulário de avaliação do Conceito Global Itemizado em suas respectivas UBS. O preenchimento e a devolução seguiram as mesmas orientações descritas no formulário de Conceito Geral.

3.8.3. Alunos – Conceito Geral e Conceito Global Itemizado

Os alunos receberam instruções específicas de preenchimentos dos formulários de Conceito Geral e de Conceito Global Itemizado em uma sessão especial, marcada antecipadamente, em horário normal de atividade acadêmica. Após esclarecerem suas dúvidas, todos os alunos preencheram os formulários simultaneamente. Em seguida, na mesma sessão, cada aluno foi orientado a avaliar todos os seus colegas de grupo (entre 3 e 4), preenchendo apenas a parte 2 do formulário de Conceito Global Itemizado (competência humanística).

3.8.4. Alunos – Opinião sobre a experiência de auto-avaliação e de avaliação dos pares

Ao término do preenchimento dos instrumentos acima mencionados, os alunos expressaram sua opinião a respeito da experiência de auto-avaliação e avaliação do pares, preenchendo o formulário apropriado.

3.9. Aspectos gerais da análise estatística

Apresentamos a seguir algumas definições e considerações sobre a análise estatística:

1. A operação de comparação de notas inclui diversas tarefas. Nesse estudo tratou-se de:
 - Investigar diferenças na posição, dispersão e forma das distribuições das notas, considerando que o mesmo aluno fornece diversas notas. Esse é o motivo pelo qual a comparação das distribuições foi feita por

meio dos testes de Friedman e de Wilcoxon dos postos sinalizados. Não é suficiente apenas investigar a significância estatística das diferenças, pois elas podem ou não ser educacionalmente importantes e/ou relevantes na prática (Altman, 1990).

- Estudar as correlações entre as diferentes fontes e tipos de notas e entre os diferentes domínios dos questionários itemizados. A partir dessas correlações, pode-se procurar modelos que revelem sua estrutura interna. Isso foi feito via análise fatorial (Dunn, 1982).
 - Investigar a estrutura das notas itemizadas, que sofrem influências dos itens e dos alunos separadamente. Essa investigação foi feita mediante análise de confiabilidade (*reliability analysis*) (Devellis, 1991).
 - Estudar a concordância relativa das notas. As medidas de concordância são casos especiais de medidas de associação que avaliam as semelhanças entre as notas (*closeness*). Em nosso caso, utilizamos distâncias Euclidianas como instrumento de avaliação da concordância (Dunn, 1982).
2. Os formulários de CG e CGI utilizaram uma escala numérica de 0 a 10. De forma geral, uma escala é a regra que permite representar os níveis de um atributo mediante valores numéricos (DeVellis, 1991). A escala usada para medir competências no presente estudo é considerada *de intervalo*, pois: (a) as diferenças entre os números refletem diferenças entre níveis de competência, (b) os valores da escala podem ser alterados linearmente sem modificar sua mensagem e (c) a origem e a unidade da escala são arbitrárias. As escalas de intervalo são muito usadas na investigação educacional, pois permitem discutir e interpretar as distâncias que separam dois sujeitos em relação a um determinado atributo.
3. Em nosso estudo, a confiabilidade (*reliability*) de um instrumento foi investigada mediante a análise da concordância entre avaliadores e da

consistência interna do instrumento aplicado à população escolhida de alunos (cálculo do coeficiente alfa de Cronbach e análise de correlações entre notas dos domínios) (DeVellis, 1991). A validade foi procurada (a) recorrendo-se a um processo preexistente de elaboração de escala (Ramsey et al., 1993) (*face validity*), (b) investigando-se a concordância do instrumento com os objetivos e conteúdos do *Módulo de Atenção Básica à Saúde*, (c) apresentando-se instruções para aplicação do questionário, (d) pré-testando o instrumento (*content validity*), (e) comparando-se os resultados obtidos para os diferentes métodos (*construct validity*) e (f) identificando-se informações novas, não contidas em métodos pré-existentes de avaliação (*discriminant validity*) (DeVellis, 1991).

4. A avaliação dos pares apresentou características especiais de análise, devido ao caráter interativo, (produzido pelo requisito “A avalia a B e B avalia a A” das notas estudadas), que dependeu dos 24 grupos de colegas-alunos. As notas dos alunos de cada grupo adquiriram um caráter de interdependência que permitiu formar e investigar redes de avaliações mútuas: quantidade de avaliações no grupo, aderência do grupo ao requisito de avaliações mútuas, grau de leniência, padrão de reciprocidade e de constância das notas.
5. Para atender à exigência de uniformidade dos conjuntos de dados, ao realizarmos as comparações estatísticas, procedemos à exclusão dos alunos com dados omissos em qualquer um dos métodos comparados. Dessa forma, utilizamos avaliações que incluíam apenas os dados completos gerados pelos 106 alunos. Esse procedimento produziu uma variação aparente nos tamanhos amostrais N das diversas comparações.
6. Para fins de análise, os dados quantitativos obtidos com o instrumento que buscava conhecer a opinião dos alunos sobre o exercício de auto-avaliação e de avaliação dos colegas, originalmente coletados por meio de uma escala de intensidade de 0 a 10 foram transformados em três tipos de respostas: “não”, “eu não sei” e “sim”. As intensidades de

opinião de 0 a 3 no questionário original foram codificadas como “não”, as de 4 a 6 como “eu não sei” e as de 7 a 10 como “sim”.

3.10. Processamento e análise dos dados

Os questionários preenchidos pelos docentes e pelos alunos, a saber, o Formulário de Conceito Geral e o Formulário de Conceito Itemizado, foram agrupados por UBS e revisados pela pesquisadora responsável, a fim de se analisar a legibilidade e a qualidade das informações. Os Formulários de Opinião dos Alunos sobre a experiência de auto-avaliação e avaliação de pares foram organizados em ordem crescente e também revisados pela pesquisadora responsável. As respostas foram pré-codificadas, os dados coletados foram digitados duplamente e inseridos em meio eletrônico com o auxílio do programa Excel. A planilha de dados, assim obtida, foi exportada para o programa SPSS Versão 10, para tabulação, consistência e interpretação estatística.

Neste estudo, houve diversas variáveis que poderiam medir a competência clínica dos estudantes. Como se desejava medir a variável latente nos dados do questionário, que exibem escalas de valor, mediante as quais os avaliadores manifestam avaliações positivas, negativas ou mesmo neutras sobre os diversos atributos, utilizou-se a técnica multivariada de análise fatorial exploratória, que é uma ferramenta estatística para busca e identificação de fatores em um conjunto de medidas realizadas. A vantagem da utilização da análise fatorial é a habilidade de relacionar múltiplas variáveis em um único conceito para entender a relação complexa entre elas (Hair et al., 1998). Os fatores foram identificados

mediante o método de máxima verossimilhança, que fornece cargas e comunalidades para as variáveis estudadas. A consistência interna dos questionários nas diversas populações foi avaliada quantitativamente pelos coeficientes de Cronbach e qualitativamente pelas análises de confiabilidade.

Para efetuar comparações estatísticas das distribuições dos dados, foram utilizados testes de Friedman, seguidos por testes de Wilcoxon de postos sinalizados para comparações múltiplas, utilizando-se correções de Bonferroni. Os coeficientes de correlação de Pearson e de Spearman foram usados para verificar as associações entre os itens dos questionários e entre os métodos de avaliação. O cálculo das distâncias Euclidianas foi utilizado para medir a concordância relativa entre os diferentes tipos de notas atribuídas aos alunos. Os modelos de análise de variância (ANOVA) foram utilizados para investigar a fonte de variação das notas atribuídas pelos pares. Assumiu-se o nível de significância de 5%.

3.11. Aspectos éticos

O estudo foi aprovado pelos Comitês de Ética em Pesquisa do DTG/CAISM e FCM (nº 581/2005), com dispensa da assinatura de consentimento informado (**Anexo 12**).

4. Publicações

- Artigo 1** – Auto-avaliação e avaliação por pares – estratégias para o desenvolvimento profissional do médico. Domingues RCL, Amaral E, Zeferino AMB. Rev Bras Educ Med 2007; 31(2):173-5.
- Artigo 2** – Feedback como estratégia de aprendizado no ensino médico. Zeferino AMB, Domingues RCL, Amaral E. Rev Bras Educ Med 2007; 31(2):176-9.
- Artigo 3** – Avaliando competência clínica – o método de avaliação estruturada observacional. Amaral E, Domingues RCL, Zeferino AMBZ. Rev Bras Educ Med 2007, 31(3):287-90.
- Artigo 4** – Conceito global: um método de avaliação de competência clínica. Domingues RCL, Amaral E, Zeferino AMB. Aprovado pela Revista Brasileira de Educação Médica, em maio de 2008 (em lista para publicação).
- Artigo 5** – Competência clínica de alunos de medicina em estágio clínico: comparação entre métodos de avaliação. Domingues RCL, Amaral E, Zeferino AMB, Antonio MAGM, Nadruz W. Enviado para Revista Brasileira de Educação Médica - RBEM em 23/10/2008
- Artigo 6** – Global overall rating as a method of assessing students during clinical clerkship – what does it really show? Domingues RCL, Amaral E, Zeferino AMB. A ser enviado para Revista Medical Education.

Artigo 7 – Global rating as a method to assess students’ clinical competence on technical and humanistic skills: comparison between faculty, self and peer assessment. Domingues RCL, Amaral E, Zeferino AMB. A ser enviado para Revista Academic Medicine.

Artigo 8 – Peer assessment: how does it work in learning groups during clinical clerkship? Domingues RCL, Amaral E, Zeferino AMB. A ser enviado para Revista Advances in Health Sciences Education.

Artigo 9 – Os diferentes olhares na avaliação de alunos em estágio clínico supervisionado. Domingues RCL, Amaral E, Zeferino AMB Enviado para Revista da Associação Médica Brasileira (RBEM) em 31 de agosto de 2008.

4.1. Artigo 1 – Auto-avaliação e avaliação por pares - estratégias para o desenvolvimento do profissional médico

Auto-Avaliação e Avaliação por Pares – Estratégias para o Desenvolvimento Profissional do Médico

Self and Peer Assessment – Strategies for the Professional Development of the Physician

Rosângela Curvo Leite Domingues¹
Eliana Amaral²
Angélica Maria Bicudo Zeferino³

PALAVRAS-CHAVE:

- Auto-Avaliação;
- Aprendizagem;
- Comissão de Avaliação de Exercício Profissional.

KEY-WORDS:

- Self-assessment;
- Learnig;
- Peer Reviews.

Recebido em: 18/07/2006

Aprovado em: 14/03/2007

RESUMO

Muitos estudos recentes incluem a auto-avaliação e a avaliação feita por pares na nota final do aluno em estágio clínico. Por muito tempo as avaliações dos estudantes de Medicina têm sido feitas exclusivamente pelos professores de forma subjetiva. Entretanto, a avaliação dos pares tem sido apontada como um bom indicador do futuro desempenho profissional, sendo considerada consistente e confiável, fornecendo informações que não poderiam ser medidas pelos métodos tradicionais. De forma complementar, a auto-avaliação também contribui para o processo de aprendizagem contínua, pois auxilia os alunos a identificarem seus pontos fortes e fracos. Embora nenhum desses dois métodos de avaliação seja rotineiramente ensinado durante o curso, esse exercício é uma valiosa intervenção para o desenvolvimento e habilitação profissional do médico. Dessa forma, as escolas médicas deveriam ajudar os alunos a compreender os princípios básicos dessas estratégias educacionais, oferecendo treinamento adequado e apoio institucional, incluindo a confidencialidade das informações.

ABSTRACT

Many recent studies include self- and peer-assessment of medical students as a component in the final clerkship grades. For a long time the evaluation of the medical student depended exclusively on subjective faculty assessment. However, peer ratings have been pointed out as a good predictor of future professional performance. These ratings are consistent, reliable and provide information that is not measured by other traditional methods. Besides, self-assessment contributes to continued education and lifelong learning by helping students identify their own strengths and weaknesses. Although self- and peer-assessment are not routinely taught during the medical course this exercise represents a valuable intervention towards the professional development of the physician. Medical schools should therefore help students understand the basic principles of these educational strategies by offering appropriate training and institutional support including confidentiality of information.

¹ Departamento de Tocoginecologia, Faculdade de Ciências Médicas, Universidade de Campinas, Campinas, Brasil.

² Departamento de Tocoginecologia, Faculdade de Ciências Médicas, Universidade de Campinas, Campinas, Brasil.

³ Departamento de Pediatria, Faculdade de Ciências Médicas, Universidade de Campinas, Campinas, Brasil.

Os currículos mais atualizados na área de saúde visam preparar não apenas médicos qualificados, mas também alunos autônomos, capazes de atuar em ambientes complexos e dinâmicos e de lidar com a crescente quantidade de informações e situações profissionais a que estão expostos^{1,2}. A escola médica deve engajar os alunos em um processo contínuo de aprendizagem, no qual habilidades reflexivas, analíticas e de automonitoramento são fundamentais³.

Ao longo dos anos, a avaliação em educação médica tem sido uma responsabilidade essencialmente do corpo docente. No entanto, na perspectiva de capacitar médicos para desenvolvimento contínuo e atuação em equipes, há uma crescente tendência de responsabilização dos estudantes, incluindo o papel de agentes de sua própria avaliação e da avaliação de seus pares⁴. Os educadores têm buscado dados colhidos de múltiplas fontes, criando um “círculo completo” de avaliadores, que inclui professores, colegas, enfermeiros, outros profissionais e pacientes, além do próprio aluno⁵⁻⁷.

A auto-avaliação representa uma estratégia de ensino importante que prepara o aluno para repensar os resultados de suas próprias ações profissionais; refletir sobre o que aprendeu; avaliar como tal aprendizado o preparou para realizar as tarefas esperadas; perceber suas necessidades individuais de aprendizagem; elaborar um plano coerente para lidar com suas dificuldades; comparar os novos resultados com os anteriores e revisar e atualizar seu plano de aprendizado^{3,8}. Essa habilidade de identificar valores e atitudes pessoais, reconhecendo os próprios pontos fortes e fracos, pode ser desenvolvida, aperfeiçoada e modificada pela educação, sendo considerada essencial para a manutenção e melhoria da proficiência médica, para o bom relacionamento com pacientes e colegas, e para o desenvolvimento da identidade profissional^{3,5,9,10}.

No entanto, apesar de sua reconhecida importância no processo de aprendizagem contínua dos profissionais da área de saúde, a auto-avaliação tem sido negligenciada¹¹. Em parte, isso ocorre porque nem os docentes e nem os alunos parecem estar adequadamente preparados para a transição entre a abordagem de avaliação centrada no professor e a centrada no aluno. Parece que o momento ideal para essa transição é quando os alunos iniciam o estágio clínico do curso médico, pois eles começam a adquirir experiência em cuidar dos pacientes e podem usar esses encontros como guia de avaliação de seus próprios conhecimentos e habilidades. Uma vez conscientes de suas dificuldades, eles poderão decidir como irão usar o tempo limitado que possuem para atividades de auto-aprendizagem¹². Atividades de preparação quando os alunos

vão iniciar o estágio clínico, facilitando o desenvolvimento de seu autoconhecimento e encorajando-os a aprender a partir de seus erros, são estratégias pedagógicas que devem ser incorporadas aos currículos médicos³.

Diversos autores defendem o uso da auto-avaliação em conjunto com a avaliação feita por pares e professores^{4,6,13-16}. A comparação entre esses dois métodos de avaliação fornece aos educadores informações úteis para o desenvolvimento e planejamento do currículo e para a revisão da metodologia adotada, além de estimular a discussão e o *feedback*^{5,17}.

A avaliação pelos colegas talvez seja o método informal mais usado pelos médicos¹³. A Abim (American Board of Internal Medicine) incorporou a avaliação feita por pares no processo de recertificação de profissionais e no treinamento de futuros profissionais^{5,18}. Como os docentes não estão presentes 100% do tempo, eles podem perder informações importantes a respeito do desempenho do aluno. Por isso, ouvir a perspectiva de colegas com quem o aluno convive por um longo período de tempo e em diferentes situações pode levar a mudanças positivas de comportamento, especialmente no que se refere à competência humanista^{6,19-22}. Entretanto, o conceito dado pelos colegas pode sofrer o chamado “efeito halo”, isto é, pode ser influenciado mais pelas impressões gerais que o avaliador tem do avaliado do que pelo desempenho do mesmo. Sabe-se, porém, que esse efeito pode ser reduzido quando se tem uma amostra adequada de avaliadores, dados confidenciais e um instrumento estruturado^{21,23}.

CONCLUSÃO

A capacitação do aluno para ser agente atuante em seu desenvolvimento profissional é uma questão central da educação médica. Neste contexto, tanto a auto-avaliação quanto a avaliação feita por pares parecem ser métodos promissores. No entanto, é preciso compreender a relutância dos alunos em avaliarem seus pares e a si mesmos e desenvolver programas que preparem alunos e professores para realizarem avaliações de forma consciente e construtiva, criando paralelamente um ambiente propício e seguro de avaliação. Deve-se acompanhar a auto-avaliação do aluno ao longo do curso, comparando-a com as avaliações que ele recebe dos colegas e docentes. Cabe, então, às escolas médicas se mobilizarem e criarem condições adequadas para que estudantes e docentes consigam criar uma nova cultura de avaliação, como instrumento de aprendizado e crescimento profissional. Para tanto, a auto-avaliação e a avaliação pelos pares precisam ser integradas no conjunto de instrumentos necessários ao desenvolvimento do profissional médico.

REFERÊNCIAS

1. Mattheos N, Nattestad A, Falk-Nilsson E, Attstrom R. The interactive examination: assessing students' self-assessment ability. *Med Educ.* 2004; 38(4):378-89.
2. Armstrong E, Parsa-Parsi R. How can physicians' learning styles drive educational planning? *Acad Med.* 2005; 80(7):680-4.
3. Henderson P, Johnson MH. An innovative approach to developing the reflective skills of medical students. *BMC Med Educ.* [online]. 2002 [capturado 12 maio 2006]; Disponível em: <http://www.biomedcentral.com/1472-6920/2/4>.
4. Van der Vleuten CPM, Schuwirth LWT. Assessing professional competence: from methods to programmes. *Med Educ.* 2005; 39(3): 309-17.
5. Rudy DW, Fejfar MC, Griffith CH, Wilson JF. Self- and peer assessment in a first-year communication and interviewing course. *Eval Health Prof.* 2001; 24(4): 436-45.
6. Davis JD. Comparison of faculty, peer, self and nurse assessment of obstetrics and gynecology residents. *Obstet Gynecol.* 2002; 99(4): 657-51.
7. Rodgers KG, Manifold C. 360-degree feedback: possibilities for assessment of the ACGME core competencies for emergency medicine residents. *Acad Emerg Med.* 2002; 9(11):1300-4.
8. 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.
9. Woolliscroft JO, Tenhaken J, Smith J, Calhoun JG. Medical students' clinical self-assessments: comparisons with external measures of performance and the students' self-assessments of overall performance and effort. *Acad Med.* 1993; 68(4): 285-94.
10. Fitzgerald JT, White CB, Gruppen LD. A longitudinal study of self-assessment accuracy. *Med Educ.* 2003; 37(7): 645-9.
11. Belar C, Brown RA, Hersch LE, Hornyak LM, Rozensky RH, Sheridan EP, et al. Self-assessment in clinical health psychology: a model for ethical expansion of practice. *Prof Psychol Res Pr.* 2001; 32(2):135-41.
12. Gruppen LD, White C, Fitzgerald JT, Grum CM, Woolliscroft JO. Medical students' self-assessments and their allocations of learning time. *Acad Med.* 2000; 75(4): 374-9.
13. Ramsey PG, Wenrich MD, Carline JD, Inui TS, Larson EB, LoGerfo JP. Use of peer ratings to evaluate physician performance. *JAMA.* 1993; 269(13):1655-60.
14. Epstein RM, Hundert, EM. Defining and assessing professional competence. *JAMA.* 2002; 287(2): 226-35.
15. Norman GR, Shannon SI, Marrin ML. The need for needs assessment in continuing medical education. *BMJ.* 2004; 328(7446): 999-1101.
16. Wilkinson TJ, Frampton CM. Comprehensive undergraduate medical assessments improve prediction of clinical performance. *Med Educ.* 2004; 38(10):1111-6.
17. Southgate L, Hays RB, Norcini J, Mulholland H, Ayres B, Woolliscroft J, et al. Setting performance standards for medical practice: a theoretical framework. *Med Educ.* 2001; 35(5): 474-81.
18. Thomas PA, Geb KA, Hellmann DB. A pilot study of peer review in residency training. *J Gen Inter Med.* 1999;14(9): 551-4.
19. Van Rosendaal GMA, Jennett PA. Comparing peer and faculty evaluations in an internal medicine residency. *Acad Med.* 1994; 69(4): 299-303.
20. Sullivan ME, Hitchcock MA, Dunnington GL. Peer and self-assessment during problem-based tutorials. *Am J Surg.* 1999; 177(3): 266-9.
21. Norcini JJ. Peer assessment of competence. *Med Educ.* 2003; 37(6): 539-43.
22. Evans R, Elwyn G, Edwards A. Review of instruments for peer assessment of physicians. *BMJ.* 2004; 328(7450): 1240-3.
23. Arnold L. Assessing professional behavior: yesterday, today and tomorrow. *Acad Med.* 2002; 77(6): 502-15.

Conflito de Interesse.

Declarou não haver.

Endereço para correspondência

Rosângela Curvo Leite Domingues
Rua Luciano Venere Decourt, 858 – Cidade Universitária
13083-740 – Campinas – SP
e-mail: rosangela@cpdec.com.br

4.2. Artigo 2 – Feedback como estratégia de aprendizado no ensino médico

Feedback como Estratégia de Aprendizado no Ensino Médico

Feedback as a Teaching/Learning Strategy in Medical Education

Angélica Maria Bicudo Zeferino¹
Rosângela Curvo Leite Domingues¹
Eliana Amaral¹

PALAVRAS-CHAVE:

- Feedback Psicológico;
- Ensino;
- Aprendizagem;
- Aptidão.

KEY-WORDS

- Feedback;
- Teaching
- Learning;
- Aptitude.

Recebido em: 18/07/2006

Aprovado em: 14/03/2007

RESUMO

Fornecer feedback aos alunos é um importante aspecto da aprendizagem e um papel essencial dos docentes. No contexto da educação médica clínica, feedback se refere às informações que descrevem o desempenho dos alunos em determinada situação ou atividade. A habilidade de dar e receber feedback melhora os resultados da aprendizagem, uma vez que fornece a base para a aprendizagem autodirecionada e para a reflexão crítica, auxilia os alunos a corrigirem seus erros, reforça comportamentos desejáveis e mostra como o aluno pode melhorar. Apesar da evidente falta de feedback durante o curso médico, os alunos desejam e valorizam essa ferramenta construtiva, considerando-a um aspecto importante do ensino de qualidade. O feedback eficaz deve ser assertivo, respeitoso, descritivo, oportuno e específico. Docentes e alunos deveriam ser preparados para dar e receber feedback. Coordenadores e diretores deveriam reconhecer o papel do feedback como uma importante estratégia de ensino-aprendizagem na graduação.

ABSTRACT

Give feedback to the students is an important aspect in the learning process and an essential role of the teachers. In the context of medical education, feedback refers to information describing the students' performance in a given situation or activity. The ability on giving and receiving feedback improves learning outcomes since it provides a basis for self-directed learning and critical thinking, helps the student correct errors, reinforces desirable attitudes and shows the student how he can improve. Despite the patent lack of feedback throughout the medical course, students greatly desire and value this constructive tool and consider it as an important aspect of quality teaching. Effective feedback should be assertive, delivered with respect, descriptive, timely and linked to specific observed behaviors. Both faculty and students should be offered proper training in order to enhance their ability to give and receive feedback. Program directors should value the role of feedback as an essential teaching/learning strategy for the undergraduate curriculum.

O conceito de *feedback*, informações usadas pelo sistema para seu ajuste, foi definido por engenheiros¹. Na área educacional, *feedback* refere-se à informação dada ao aluno que descreve e discute seu desempenho em determinada situação ou atividade. O *feedback* gera uma conscientização valiosa para a aprendizagem, pois ressalta as dissonâncias entre o resultado pretendido e o real, incentivando a mudança; também aponta os comportamentos adequados, motivando o indivíduo a repetir o acerto². Se a informação for capaz de causar mudança no padrão de desempenho observado, teremos um processo de aprendizagem¹.

Entre os princípios ligados à aprendizagem dos adultos, inclui-se o desejo de receber *feedback*, juntamente com a necessidade de participação ativa no processo de aprender, a autodiretividade e a curiosidade de explorar soluções^{2,3}. Uma vez que reforça a prática reflexiva, o *feedback* está fortemente vinculado ao conceito de desenvolvimento profissional contínuo, habilidade considerada essencial no médico que se deseja formar⁴. A falta de *feedback* gera incertezas, amplifica o sentimento de inadequação e distancia o aluno dos objetivos propostos, podendo levá-lo a interpretar seus comportamentos de maneira totalmente inapropriada e a desenvolver uma "falsa confiança" ou medo exagerado do erro^{1,5}. Criar oportunidades de receber *feedback* no início da prática clínica permite que o aluno desenvolva e aprimore sua autopercepção e contribui para que ele aprenda a se auto-avaliar de maneira coerente e precisa¹.

No contexto da educação médica, em particular na área clínica, a habilidade de dar e receber *feedback* com eficácia tem uma importância singular e se mostrou efetiva em melhorar o desempenho^{2,5,6}. O *feedback* contribui com a prática reflexiva, ou capacidade do profissional de rever suas próprias conclusões, raciocínio e decisões, uma habilidade essencial para desenvolver profissionalismo e aprimorar as habilidades cognitivas, essenciais na prática clínica^{5,7,8}.

É um ingrediente essencial para a aprendizagem clínica, em especial quando fornecido imediatamente após o desempenho observado, como, por exemplo, na avaliação formativa ou somativa³. Estudos demonstram que o *feedback* está associado à melhoria do desempenho do aluno no exame físico e na prática clínica como um todo^{9,10}.

Os alunos consideram o *feedback* uma importante ferramenta de aprendizagem e valorizam os professores que sabem fornecê-lo de maneira eficaz¹¹. Num estudo com 3 mil alunos de último ano de Medicina e residentes, 95% destes consideraram que dar e receber *feedback* de forma eficaz é a segunda entre 36 habilidades de ensino, sendo ultrapassada apenas pela capacidade de estar aberto para perguntas¹².

COMO DAR FEEDBACK DE FORMA EFICAZ

Dar *feedback* exige habilidade, compreensão do processo, criação de um ambiente propício e de uma relação de confiança¹³. Não há como informar ao aluno que a sua hipótese diagnóstica estava errada ou que ele não colheu todos os dados necessários durante a história clínica sem causar uma sensação de desapontamento ou frustração. Por outro lado, essa informação é essencial e não pode ser omitida.

A validade do processo de *feedback* inicia-se na qualidade das observações feitas, e o observador deve estar comprometido com o processo, com o aluno e com a instituição. Além disso, deve utilizar um instrumento confiável de registro e seguir padrões bem estabelecidos de competência clínica¹. A preparação dos professores para utilizar a ferramenta é essencial, incluindo a maneira como as informações devem ser transmitidas, oralmente ou por escrito^{14,15}.

A maior eficácia do *feedback* só se consegue quando este enfoca, além de competências, as atitudes e outras dimensões do profissionalismo, repete-se em diferentes momentos do curso médico e quando o professor se preocupa em observar também as reações emocionais do aprendiz, estabelece os objetivos do aprendizado e formas de interação em consenso com o aluno e permite que este desenvolva seu raciocínio^{5,16}.

As múltiplas tarefas devidas ao professor no aprendizado da prática clínica parecem contribuir para limitar o tempo dedicado a esse recurso didático. Salerno *et al.*¹⁴ encontraram que 75% do tempo foram utilizados na apresentação e discussão do caso e que o *feedback*, com muitas informações negativas ou pouco específicas, ocupou apenas 30 segundos. A falta de treinamento, o desejo de não ofender e a vontade de manter a auto-estima dos alunos são outras causas citadas para justificar o pouco uso da estratégia¹⁵.

Aprender a partir do *feedback* requer que este seja fornecido de forma construtiva e positiva, colaborando para que o aluno reflita criticamente e elabore um plano de melhoria em prática^{2,5}. Uma boa estratégia é a utilização da "técnica do sanduíche", na qual o *feedback* sobre comportamentos inadequados é fornecido entre duas ações ou comportamentos positivos; essa estratégia é particularmente útil para os alunos considerados mais resistentes¹⁷. A efetividade do *feedback* é maior quando é:

- (1) Assertivo. A comunicação deve ser clara, objetiva e direta. Por temer o impacto das palavras, o professor pode não ser direto, falando de forma vaga, com afirmações ambíguas que ofuscam a mensagem principal. O aluno, temendo uma avaliação negativa, não procura esclarecimentos, reforçando a falta de clareza do professor. Como resultado, apesar das intenções educati-

vas, pouco é transmitido⁵. Assim, recomenda-se descrever os impactos e conseqüências de determinado comportamento, positivos ou negativos, assim como sugerir comportamentos alternativos¹⁸.

- (2) Respeitoso. Este é um elemento fundamental para o sucesso do *feedback*, independentemente das diferenças de conhecimento, experiência, hierarquia ou características pessoais entre os interlocutores. Como é um processo compartilhado, docente e aluno devem encontrar pontos de concordância sobre os comportamentos que devem ser trabalhados; entender e respeitar a opinião do outro geram o ambiente de respeito para um *feedback* construtivo⁵.
- (3) Descritivo. Embora o estudante, em geral, esteja ávido por ouvir a opinião dos professores, sua reação é menos resistente quando as palavras descrevem determinado comportamento ou ação, ao invés de julgá-lo^{3,5}.
- (4) Oportuno. O momento e o local para dar *feedback* ao aluno devem ser adequados, preferencialmente logo após a observação do comportamento e em ambiente reservado².
- (5) Específico. É fundamental que o docente indique claramente os comportamentos nos quais o aluno está tendo bom desempenho e aqueles nos quais o aluno pode melhorar. Exemplos e revisão dos fatos ocorridos contribuem para que o aluno reflita honestamente sobre seu desempenho¹³.

IMPLICAÇÕES PARA O ENSINO MÉDICO

O uso do processo de *feedback* emerge como um componente essencial no desenvolvimento da prática reflexiva e da construção do profissionalismo. Apesar de sabidamente eficaz, não tem sido uma estratégia pedagógica utilizada rotineiramente no curso médico, por vezes sendo visto como um "ato de coragem" do professor ou como um "ato de tortura" para o aluno. Para sua implementação bem-sucedida e eficaz, todos os envolvidos devem compreender o processo de *feedback* e aceitá-lo como uma oportunidade de crescimento pessoal e profissional, uma ferramenta de desenvolvimento de carreira.

Os docentes devem ser estimulados a praticar o *feedback* como estratégia regular de ensino-aprendizagem, inserindo-o precocemente na rotina de avaliação. Os alunos vão gradativamente se conscientizando do seu nível de competência e buscando aprimoramento, processos essenciais no desenvolvimento da prática reflexiva. Assim, o *feedback* pode ser incorporado de forma natural, tomando o ambiente de ensino seguro e os alunos menos hesitantes em aceitarem as críticas e

sugestões. E, ao aceitá-las, mais facilmente se tornarão agentes ativos no processo de aprender, buscando se capacitar em suas deficiências.

Para contribuir neste processo, os alunos devem ser encorajados e treinados para receber, mas também realizar *feedback*, em relação a si mesmos e aos colegas. Um exemplo de estratégia nesta direção é solicitar que o aluno analise o vídeo de uma consulta feita pelo colega, guiado por um instrumento estruturado. Em seguida, deve apresentar o *feedback* por escrito ao colega, reforçando os comportamentos que não precisam ser modificados e identificando os que devem ser mudados. Também deve preparar uma auto-avaliação, que seria, posteriormente, comparada com o *feedback* que recebeu de seu par e de seu tutor. Essa atividade seria acompanhada e orientada por um docente e utilizada tanto para comentar e corrigir os comportamentos observados na consulta, como para avaliar o *feedback* fornecido. Dessa forma, os alunos estariam gradativamente construindo e praticando suas habilidades de observar, ouvir atentamente, perguntar e se expressar com clareza, elementos essenciais e indispensáveis também na relação com os pacientes e familiares.

Mas, para essa mudança ocorrer, é preciso disponibilizar aos professores um conjunto de práticas e informações que aprimorem sua habilidade de fornecer *feedback* e de utilizá-lo como estratégia de ensino. Os programas de desenvolvimento docente para o ensino médico devem incluir tal componente. Há evidências de que programas de desenvolvimento são amplamente aceitos pelos docentes e apresentam efeitos positivos e que a habilidade de *feedback* está entre os temas sugeridos em sua composição³⁹. Demonstrou-se que os professores apresentaram melhoria significativa, embora modesta, em sua habilidade de dar *feedback* após terem participado de três seminários com duração de 90 minutos e que este tema deve ser constantemente revisto na capacitação docente¹⁴. Um exercício prático de avaliação das habilidades do professor como facilitador do aprendizado, seguida de *feedback*, pode ser uma vivência interessante que exemplifica o potencial dessa estratégia de aprendizagem.

A arte de ensinar e de aprender, visando à qualidade do desempenho clínico, pode indubitavelmente ser aprimorada com a prática do *feedback* regular. Neste artigo, buscou-se auxiliar os educadores médicos a pensarem criticamente sobre como incorporar ou aprimorar a habilidade de fornecer *feedback*. Reproduzindo o processo de incorporação de outras tecnologias, procuramos fornecer elementos teóricos para o entendimento da proposta, apresentando algumas evidências de sua efetividade.

REFERÊNCIAS

1. Ende J. Feedback in clinical medical education. *JAMA*. 1983; 250(6): 777-81.
2. Collins J. Education techniques for lifelong learning: principles of adult learning. *Radiographics* 2004; 24(5): 1483-9.
3. Rolfe IE, Sanson-Fisher RW. Translating learning principles into practice: a new strategy for learning clinical skills. *Med Edu*. 2002; 36(4): 345-52.
4. Rodgers KG, Manifold C. 360-degree feedback: possibilities for assessment of the ACGME core competences. *Acad Emerg Med*. 2002; 9(11): 1300-4.
5. Henderson P, Ferguson-Smith AC, Johnson MH. Developing essential professional skills: a framework for teaching and learning about feedback. *BMC Med Educ*. [online]. 2005; 5 [capturado maio 2006]. Disponível em: <http://www.biomedcentral.com/1472-6920/5/11>.
6. Veloski J, Boex JR, Grasberger MJ, Evans A, Wolfson DB. Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME guide n. 7. *Med Teach*. 2006; 28(2): 117-28.
7. Lachman N, Pawlina W. Integrating professionalism in early medical education: the theory and application of reflective practice in the anatomy curriculum. *Clin Anat*. 2006;19(5): 456-60.
8. Mamede S, Schmidt HG. Correlates of reflective practice in medicine. *Adv Health Sci Educ Theory Pract*. 2005; 10(4): 327-37.
9. Newble DI, Jaeger K. The effect of assessment and examinations on the learning of medical students. *Med Edu*. 1983; 17(3): 165-71.
10. Papa FJ, Aldrich D, Schumacker RE. The effects of immediate online feedback upon diagnostic performance. *Acad Med* 1999; 74(10Supl.): 16-8.
11. Rotenberg BW, Woodhouse RA, Gilbert M, Hutchison CR. A needs of assessment of surgical residents as teachers. *Can J Surg*. 2000; 43(4): 295-300.
12. Schultz KW, Kirby J, Delva D, Godwin M, Verma S, Birtwhistle R, et al. Medical students' and residents' preferred site characteristics and preceptor behaviours for learning in the ambulatory setting: a cross-sectional survey. *BMC Med Educ*. [online]. 2004; 4 [capturado abr. 2006] abril]. Disponível em: <http://www.biomedcentral.com/1472-6920/4/12>.
13. Gordon J. ABC of learning and teaching in medicine: one to one teaching and feedback. *BMJ*. 2003; 326(7388): 543-5.
14. Salerno SM, O'Malley PG, Pangaro LN, Wheeler GA, Moores LK, Jackson JL. Faculty development seminars based on the one-minute preceptor improve feedback in the ambulatory setting. *J Gen Intern Med*. 2002; 17(10): 779-87.
15. Dobbie A, Tysinger JW. Evidence-based strategies that help office-based teachers give effective feedback. *Fam Med*. 2005; 37(9): 617-9.
16. Menachery EP, Knight AM, Kolodner K, Wright SM. Physician characteristics associated with proficiency in feedback skills. *J Gen Intern Med*. 2006; 21(5): 440-6.
17. Lucas JH, Stallworth JR. Providing difficult feedback: TIPS for the problem learner. *Fam Med*. 2003; 35(8): 544-6.
18. Jerome PJ. Coaching through effective feedback: a practical guide to successful communication. California: R. Chang Associates; c1994.
19. Kilminster SM, Jolly BC. Effective supervision in clinical practice settings: a literature review. *Med Edu*. 2000; 34(10): 827-40.

Conflito de interesse

Declarou não haver.

Endereço para correspondência

Angélica Maria Bicudo-Zeferino
Rua Madre Paulina, 66 – Condomínio Residencial Parque
Rio das Pedras – 13085-150 – Campinas – SP
e-mail: ambz@fcm.unicamp.br

4.3. Artigo 3 – Avaliando competência clínica: o método de avaliação estruturada observacional

Avaliando Competência Clínica: o Método de Avaliação Estruturada Observacional

Assessing Clinical Competence: Structured Observation

Eliana Amara¹
Rosângela Curvo Leite Domingues¹
Angélica Maria Bicudo-Zeferino¹

PALAVRAS-CHAVE:

- Avaliação;
- Competência Clínica;
- Estudantes de Medicina;
- Educação Médica.

KEY-WORDS:

- Evaluation;
- Clinical Competence;
- Students, Medical;
- Education, Medical.

Recebido em: 18/07/2006

Reencaminhado em: 03/04/2007

Aprovado em: 19/04/2007

RESUMO

Apesar dos avanços tecnológicos, a história médica, o exame físico e a relação médico-paciente continuam as mais importantes ferramentas diagnósticas e terapêuticas. No entanto, deficiências em habilidades clínicas entre alunos têm sido relatadas. Competência clínica deve ser considerada não apenas a demonstração de comportamentos isolados, mas, sim, a habilidade integrada de pensar, sentir e agir na prática real, um determinado contexto. Por isso, ferramentas de observação direta são bastante úteis para avaliar o que os estudantes fazem ou deixam de fazer durante as consultas. O instrumento de avaliação estruturada parece ser adequado tanto para os avaliadores quanto para os avaliados, uma vez que pode ser usado como um guia de observação e também como uma ferramenta de feedback, ampliando as oportunidades de aprendizagem no cenário clínico. Para que esse instrumento seja aplicado de forma eficaz, justa e confiável, os avaliadores precisam ser adequadamente preparados.

ABSTRACT

Despite the technological advances, medical history, physical examination and patient-doctor relationship continue being the most important diagnostic and therapeutic tools. However, deficiencies in clinical skills among students have been related. Clinical competence is context-dependent; it is more than a demonstration of isolated behaviors, but an integrative ability to think, feel and act in real practice. Tools for direct observation are very helpful in evaluating what students are and are not able to do during clinical encounters. So, structured assessment seems to be a valuable instrument both for evaluators and students. It can be used as an observational guide and as a tool for feedback, and enhances the learning opportunities in clinical settings. To effectively apply this method, faculty should be provided with effective training to increase fairness and reliability of the assessment strategy.

INTRODUÇÃO

Apesar dos avanços tecnológicos, as habilidades para realizar história, exame físico e para se comunicar com o paciente continuam as mais importantes e eficazes ferramentas diagnósticas e terapêuticas diante de um caso clínico. No entanto, muitos alunos terminam o curso médico com deficiências nessas habilidades essenciais¹. Tal constatação reforça a necessidade de que os educadores voltem sua atenção para a avaliação da competência clínica, caracterizada como um conjunto de conhecimentos, habilidades técnicas e de comunicação, empatia, propedêutica e raciocínio clínico, durante a graduação médica².

Num ambiente clínico, muitos alunos se sentem frustrados pela dificuldade de recordarem conhecimentos adquiridos previamente e de serem incapazes de aplicá-los no caso clínico em questão. Os docentes, por sua vez, ficam surpresos com os alunos que parecem ter retido tão pouco. O problema talvez esteja na separação entre o aprendizado clínico e o ambiente e situação em que o mesmo será aplicado. Por isso, há uma forte tendência de se colocar o aluno em ação precocemente, tendo-se o cuidado de criar um ambiente propício à aprendizagem, que envolva os componentes cognitivos (o que aprender), afetivos (motivação para aprender) e metacognitivos (como aprender)³.

Se o aprender clínico tem sido modificado, com antecipação e diversificação de cenários, os métodos de avaliação das competências clínicas precisam ser adaptados a essa nova realidade. Em 1990, George Miller propôs um modelo hierárquico de avaliação da competência clínica que diferenciava o nível da "ação" (fazer) dos outros níveis considerados inferiores (saber, saber como e mostrar como)⁴. Avaliar a "ação" significa avaliar o que ocorre na prática. Embora nos últimos 30 anos inúmeros métodos tenham sido desenvolvidos e utilizados na avaliação dos alunos de Medicina, a maioria deles ainda tende a enfatizar o "saber" ou "saber como", habilidades da esfera cognitiva, essenciais mas insuficientes para compor a competência clínica. Esses níveis podem ser avaliados por testes de múltipla escolha, exame oral, dissertações ou trabalhos escritos, estudos de caso, cenários clínicos simulados⁵.

Um método bastante utilizado nos últimos anos é o Osce (*objective structured clinical examination*), que avalia o desempenho do aprendiz em situações delimitadas, baseadas em um roteiro predefinido, em que há interação com paciente simulado ou recursos didáticos por meio de estações de avaliação em rodízio⁶. As notas são dadas pelos observadores ou pelos pacientes, por meio de *checklists* isolados ou combinados com conceito. Além do grande desafio operacional

e custo que representa, três aspectos dificultam o Osce de se aproximar de situações de prática clínica: (1) o tempo limitado em cada estação exige que o aluno demonstre habilidades isoladas de uma consulta, o que desconstrói a relação médico-paciente; (2) ao se basear num *checklist* de tarefas, passa a mensagem de que a interação médico-paciente pode ser uma lista de ações; (3) como nem tudo pode ser simulado, há uma limitação dos problemas que podem ser abordados pelo ator⁷. O Osce não avalia o estudante em condições reais e se presta à avaliação do "saber como".

Para avaliar o "mostrar como", um degrau acima na pirâmide de Miller, utiliza-se em todo o mundo o "caso longo", no qual o aluno examina um paciente e depois apresenta o caso completo para o examinador de forma não estruturada. No formato tradicional, é um método que permite avaliar a integração do conhecimento pelo relato da interação do aluno com o paciente, mas é questionado por suas propriedades psicométricas, particularmente de validade. A inclusão de observação direta e o aumento do número de casos longos avaliados por estudante são estratégias para contornar estes problemas e recuperar a credibilidade do método na avaliação clínica⁸.

Mas o maior desafio dos educadores médicos está em avaliar o desempenho, o "fazer" real, com sua imprevisibilidade e com os aspectos emocionais envolvidos. Portfólios com documentação das experiências clínicas num período determinado de tempo e auto-reflexão têm sido usados como instrumentos de avaliação do que se fez⁹. Uma opção complementar de avaliação é o Mini CEX (*mini-clinical evaluation exercise*), que consiste numa observação estruturada da prática com guia de verificação ou *checklist*, na qual se atribuem notas de avaliação global para diversos itens de atitudes e habilidades, seguida de *feedback*, com duração total de 20-30 minutos, que pode ser repetida várias vezes para o mesmo aluno, aumentando a validade do método⁹.

Os autores concordam em que, para se avaliar a competência na verdadeira prática clínica, o nível "fazer", é preciso agregar a observação do desempenho no cuidado do aluno com pacientes reais^{1,2,5}. No entanto, raras são as oportunidades em que os docentes observam os atendimentos realizados por seus alunos durante o ciclo clínico. Frequentemente, os professores aceitam a veracidade da história e do exame físico apresentados pelos alunos, sem nunca terem realmente observado seu desempenho. De fato, muitos alunos referem que durante todo o seu curso foram observados em poucas oportunidades, muitas vezes por residentes e não pelos docentes. Indiscutivelmente, a observação direta é um método rico de avaliação, pois fornece uma visão mais realista e integrada das habilidades clínicas

dos alunos, facilitando a identificação das áreas de aprendizado a serem reforçadas. Por isso, deveriam receber alta prioridade por parte dos educadores, coordenadores e diretores de escolas médicas¹.

O uso de um instrumento estruturado, tipo *checklist*, para avaliação durante a observação de uma consulta auxilia o professor a focar sua atenção em habilidades específicas, aumentando a acurácia na detecção de falhas. Além de servirem como instrumento de avaliação, essas listas servem como ferramenta de aprendizagem, oferecendo elementos objetivos para dar *feedback* aos alunos, podendo ajudá-los a reforçar seus pontos fortes e a corrigir suas deficiências¹⁰.

Para diferenciar as nuances e gradações dos comportamentos esperados, não apenas sua presença ou ausência, o instrumento para avaliação estruturada deve conter uma escala de Likert¹¹. Os *checklists* do tipo sim/não, embora práticos, deixam de contemplar aspectos qualitativos importantes, que são essenciais como *feedback* aos alunos e para a comparação de seu desempenho. De fato, sabe-se que mudanças de comportamento são medidas de forma mais apurada como um contínuo progressivo do que como uma variável dicotômica^{11,12}.

A primeira dificuldade encontrada para o sucesso desse método está na qualidade das observações. Estudos demonstram falta de homogeneidade nos critérios usados pelos docentes para atribuir notas e falhas na detecção de erros cometidos pelos alunos^{13,14}. Certamente, a presença contínua do avaliador durante a consulta médica constitui uma dificuldade adicional, pois pode influenciar positiva ou negativamente o desempenho do aluno. Mas o uso apropriado do *feedback* e o esclarecimento de que essa é uma oportunidade para aprendizagem tornam a avaliação estruturada um método significativo para o aluno, que tende a agir mais naturalmente durante a consulta¹⁵.

Tais problemas podem ser minimizados se os docentes estiverem bem preparados como clínicos e como avaliadores, disponíveis, com tempo suficiente para a realização da observação e se o instrumento usado for válido e confiável^{10,16}.

A confiabilidade refere-se à repetibilidade do teste; quanto mais confiável, maior a probabilidade de se obter um resultado semelhante se o teste for novamente administrado. Inúmeros itens afetam a confiabilidade de um instrumento: tamanho, discriminação dos itens, heterogeneidade dos alunos, comportamento dos pacientes, atitudes e habilidades dos examinadores, número de observações por aluno e condições da avaliação. A validade mede o grau no qual o teste realmente avalia aquilo que se propõe e demonstra a ligação entre o conteúdo do teste e os objetivos de aprendizado. O sucesso

da avaliação estruturada está em se encontrar o equilíbrio entre validade e confiabilidade, e entre o ideal e o prático^{5,7,11}.

Embora a avaliação por observação estruturada possa gerar uma nota, com fins formativos ou somativos, é preciso decidir o padrão considerado como o mínimo aceitável. Avaliações somativas devem ser critério-referenciadas, isto é, cada aluno deve ser avaliado de acordo com o padrão estabelecido e não em relação aos colegas^{5,7,16}. Uma sugestão é a classificação da American Board of Internal Medicine (Abim), com os seguintes padrões e critérios: superior (o aluno excede as expectativas); satisfatório (o aluno sempre atinge as expectativas e ocasionalmente as excede); marginal (em geral, o aluno atinge as expectativas, mas ocasionalmente está abaixo delas); insatisfatório (consistentemente, o aluno está abaixo das expectativas). Restará definir, idealmente por um processo de consenso entre os avaliadores, que desempenhos específicos devem ser identificados para classificar o aluno segundo cada item da avaliação proposta¹⁷.

CONCLUSÃO E TENDÊNCIAS

A utilização de ferramentas que possuem uma forte característica observacional é uma tendência no ensino médico, pois permite avaliar o fazer, preferencialmente em condições reais. Mas é necessário atribuir pesos diferentes aos diferentes aspectos (habilidades ou comportamentos) observados, ponderando de acordo com que é esperado do aluno na etapa de aprendizagem em que se encontra¹⁸. Tal definição depende de consensos entre os educadores médicos e não pode ser arbitrada isoladamente ou definida no momento em que se vai atribuir nota, oferecendo-se uma abordagem mais homogênea e transparente dos critérios de avaliação clínica.

Os educadores devem ter em mente que toda avaliação é uma forma de aprendizagem que deve servir como guia e suporte para solucionar as necessidades de aprendizagem do aluno e aumentar sua autoconfiança e sua habilidade de pensar sobre sua forma de aprender. Por isso, ao selecionar o método de avaliação e ao utilizá-lo, o educador deve refletir sobre que conteúdo e quais habilidades serão avaliados, com que objetivo (formativo, somativo), em que contexto e como essa avaliação pode fomentar aprendizagens futuras². As escolas médicas devem se comprometer a experimentar as diversas opções descritas na literatura, aprimorando-as e adaptando-as às características do ensino médico que desejam construir. A avaliação deve ser adaptada ao contexto no qual será usada e voltada para os objetivos previamente traçados². Uma abordagem holística, que integre métodos, vai permitir a

avaliação de conhecimento, compreensão, solução de problemas, habilidades técnicas, atitude e ética¹⁹.

Neste contexto, a avaliação por observação estruturada do desempenho clínico na forma de mini-CEX, Osler (*Objective Structured Long Examination Record*), Osce, casos curtos e longos tem um lugar destacado, por abranger aspectos técnicos, éticos e relacionais. As dificuldades operacionais e a habilitação do corpo docente para trabalhar com os diferentes instrumentos devem ser discutidas no processo de decisão sobre os métodos de avaliação das habilidades clínicas a serem empregados, com destaque para a etapa de aprendizado em serviço, o Internato.

REFERÊNCIAS

- Holmboe ES. Faculty and the observation of trainees' clinical skills: problems and opportunities. *Acad Med.* 2004; 79(1):16-22.
- Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA.* 2002; 287(2): 226-35.
- Gordon J, Hazlett C, Ten Cate O, Mann K, Kilminster S, Prince K, et al. Strategic planning in medical education: enhancing the learning environment for students in clinical settings. *Med Educ.* 2000; 34(10): 841-50.
- Miller GE. The assessment of clinical skills/competence/performance. *Acad Med.* 1990; 65 Suppl 9: 63-7.
- Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet.* 2001; 357(9260): 945-9.
- Marks M, Humprey-Murto S. Performance assessment. In: Dent JA, Harden R, eds. *A practical guide for medical teachers.* 2nd ed. London: Elsevier; 2005. p.323-35.
- Smee S. Skill based assessment. *BMJ.* 2003;326(7391): 703-6.
- Wass V, Van der Vleuten C. The long case. *Med Educ.* 2004; 38(11): 1176-80.
- Norcini JJ, Blank LL, Duffy FD, Fortna GS. The mini-CEX: a method for assessing clinical skills. *Ann Intern Med.* 2003; 138(6):476-81.
- Norgaard K, Ringsted C, Dolmans D. Validation of a checklist to assess ward round performance of internal medicine. *Med Educ.* 2004; 38(7):700-7.
- Hayes BE. Medindo a satisfação do cliente: desenvolvimento e uso de questionários. Rio de Janeiro: Qualitymark; 2001.
- Lehmann F, Côté L, Bourque A, Fontaine D. Physician-patient interaction: a reliable and valid check-list of quality. *Can Fam Physician.* 1990; 36: 1711-6.
- Elliot DL, Hickam DH. Evaluation of physical examination skills. Reliability of faculty observers and patient instructors. *JAMA.* 1987; 258(23): 3405-8.
- Noel GL, Herbers JE, Caplow MP, Cooper GS, Pangaro LN, Harvey J. How well do internal medicine faculty members evaluate the clinical skills of residents?. *Ann Intern Med.* 1992; 117(9): 757-65.
- Katz NT, McCarty-Gillespie L, Magrane DM. Direct observation as a tool for needs assessment of resident teaching skills in the ambulatory setting. *Am J Obstet Gynecol.* 2003; 189(3): 684-7.
- Lynch DC, Swing SR. Key considerations for selecting instruments and implementing assessment systems. [Online]. USA. [Capturado abr. 2006]. Disponível em: www.acgme.org/outcome/assess/keyconsider.asp.
- Norcini JJ. Current perspectives in assessment: the assessment of performance at work. *Med Edu.* 2005; 39(9): 880-9.
- DeLisa JA. Evaluation of clinical competence. *Am J Phys Med Rehabil.* 2000; 79(5): 474-7.
- McMullan M, Endacott R, Gray MA, Jasper M, Miller CML, Scholes J, et al. Portfolios and assessment of competence: a review of the literature. *J Adv Nurs.* 2003; 41(3): 283-94.

CONFLITO DE INTERESSE

Declarou não haver.

ENDEREÇO PARA CORRESPONDÊNCIA

Eliana Amaral
Divisão de Obstetrícia / CAISM / Unicamp
Rua Alexander Fleming, 101
13081-883 – Campinas – São Paulo
e-mail: eliana@unicamp.br

4.4. Artigo 4 – Conceito global: um método de avaliação de competência clínica

(Artigo aceito pela revista Brasileira de Educação Médica - RBEM, em 08/05/2008 - em lista para publicação)

CONCEITO GLOBAL: UM MÉTODO DE AVALIAÇÃO DE COMPETÊNCIA CLÍNICA

GLOBAL RATING: AN ASSESSMENT METHOD FOR CLINICAL COMPETENCE

Rosângela Curvo Leite Domingues ⁽¹⁾, Eliana Amaral ⁽²⁾, Angélica Maria Bicudo-Zeferino⁽³⁾

- (1) Doutoranda, Departamento de Tocoginecologia, Faculdade de Ciências Médicas, UNICAMP, Brasil.
- (2) Professora-Assistente Doutora, Departamento de Tocoginecologia, Faculdade de Ciências Médicas, UNICAMP, Brasil
- (3) Professora-Assistente Doutora, Departamento de Pediatria, Faculdade de Ciências Médicas, UNICAMP, Brasil

Endereço para correspondência:

Rosângela Curvo Leite Domingues
Rua Luciano Venere Decourt, 858
Cidade Universitária
CEP 13083-740 - Campinas – SP
Fone: (19) 3289-2192
e-mail: rosangela@cpdec.com.br

CONCEITO GLOBAL: UM MÉTODO DE AVALIAÇÃO DE COMPETÊNCIA CLÍNICA

GLOBAL RATING: AN ASSESSMENT METHOD FOR CLINICAL COMPETENCE

Resumo

Muito se tem feito para desenvolver métodos que avaliem de forma confiável a proficiência em habilidade clínica dos alunos de graduação médica. Há fortes evidências de que a avaliação global pode ser uma alternativa viável e confiável na avaliação de estudantes de graduação no ambiente clínico. Rotineiramente, a competência clínica tem sido avaliada pelos docentes por meio de um conceito subjetivo. Essa nota, dada por um especialista, não é objetiva e está sujeita a vieses. O uso de um instrumento bem delineado, contendo diversos itens, pontuados numa escala, oferece a vantagem de especificar o que deve ser avaliado, auxiliando os avaliadores a distinguir os diferentes níveis do desempenho. Usando este método, os docentes podem expressar sua percepção global dos alunos, de forma mais objetiva, em relação a duas esferas distintas da competência: técnica (contendo itens como história clínica, exame físico, conhecimento médico, julgamento clínico, solução de problema e hábitos no trabalho) e humanista (incluindo comunicação com pacientes e familiares, respeito, habilidades reflexivas, sensibilidade ao contexto e trabalho em equipe).

Palavras-chave: conceito global, avaliação clínica, medida de avaliação, estudante de Medicina

Abstract

Considerable effort has been aimed at developing methods to be used to support reliable decisions regarding the proficiencies of graduate medical students on clinical skills. There is strong evidence that global assessment may be a viable and reliable alternative approach to evaluate undergraduates in a clinical setting. Clinical competence is routinely assessed by a preceptor through a global subjective mark. Although provided by an expert, this mark is usually subjective and prone to bias. The use of a well-delineated instrument, containing several items rated in a multipoint scale, offers the advantage of specifying what should be assessed, helping scorers to make distinctions among various levels of performance. Using this method, faculty can provide a clear global impression of students relating to two different spheres of competence: technical (containing items as clinical history, physical exam, medical knowledge, clinical judgment, problem solving, habits at work) and humanistic skills (including communication with patients, respect, reflective skills, sensitivity to context, team work).

Key-words: global rating, clinical assessment, educational measurement, medical student

CONCEITO GLOBAL: UM MÉTODO DE AVALIAÇÃO DE COMPETÊNCIA CLÍNICA

GLOBAL RATING: AN ASSESSMENT METHOD FOR CLINICAL COMPETENCE

Nos últimos anos, diversas entidades internacionais têm se debruçado sobre a discussão em torno do tema competência médica e o papel do médico no século XXI. Neste contexto, a *Association of American Medical Colleges (AAMC)* iniciou um projeto denominado *Medical School Objectives Project (MSOP)*. O relatório de consenso entre representantes dos 14 países participantes contempla os seguintes grupos de características esperadas do médico: (1) *altruísmo*: empatia, respeito à privacidade e à dignidade do paciente, atenção, cuidado, compreensão do contexto cultural e dos valores do paciente e de seus familiares; (2) *conhecimento*: compreensão das bases científicas da medicina e competência para colocá-las em prática; (3) *habilidade*: competência para fornecer cuidados médicos apropriados, obter história médica precisa, construir raciocínio clínico, realizar exames físicos, procedimentos diagnósticos e terapêuticos e se comunicar eficazmente com pacientes e familiares; (4) *compromisso e colaboração*: compreensão dos fatores econômicos, psicológicos, ocupacionais, sociais e culturais envolvidos na atenção ao paciente; comprometimento para trabalhar em conjunto com outros profissionais da saúde; aplicação dos princípios da medicina baseada em evidência; capacidade para tomar decisões e lidar com recursos médicos limitados^{1,2}.

Um outro projeto iniciado em 1999 pela *ACGME (Accreditation Council for Graduate Medical Education)* e denominado *Outcome Project*, identificou seis competências gerais para médicos de todas as especialidades: (1) *cuidado com o paciente*, (2) *conhecimento médico*, (3) *aprendizagem e melhoria baseadas na prática*,

(4) *habilidades interpessoais e de comunicação*, (5) *profissionalismo* e (6) *prática baseada em sistemas*.^{2,3,4} Em 2000, o Projeto *CanMED (Canadian Medical Education Directions for Specialists)* descreveu sete papéis que deveriam ser incluídos em todos os níveis da educação médica: especialista, comunicador, colaborador, gerente, defensor da saúde, comprometido com educação continuada e profissional.⁵

Apesar dessa abordagem multidimensional de competência clínica, vários estudos sustentam que na prática ela é apenas bidimensional, pois verdadeiramente os alunos são avaliados por duas dimensões substancialmente correlacionadas que, seguindo a taxonomia da ACGME, são denominadas de: (1) *competência humanista*, englobando as habilidades de comunicação interpessoal e profissionalismo e (2) *competência técnica*, constituída de conhecimento médico, cuidado com o paciente e atenção baseada no sistema.⁶⁻⁸

Todas essas discussões repercutem na prática do ensino médico. Muitas universidades e órgãos de classe estão unindo esforços para repensar maneiras de sistematizar a avaliação destas competências, acrescentando-lhe a necessária perspectiva formativa e indutora de novas práticas pedagógicas. A avaliação, como parte integrante do processo de ensino-aprendizagem, tem como funções: estabelecer a gama de conhecimentos e habilidades que um indivíduo precisa ter atingido em certo estágio de sua carreira, estimar de forma precisa o nível dos alunos, oferecer aos alunos informações sobre seu progresso, alterar o comportamento em relação à aprendizagem, motivar os alunos, conscientizando-os sobre a importância do que estão aprendendo e fornecer informações sobre a adequação do currículo aos professores e diretores.⁹⁻¹¹

É certo que a avaliação da competência médica deve partir de uma abordagem holística, que pressupõe a utilização de diferentes métodos que enfoquem atributos distintos, entendendo-se que apenas a combinação desses métodos, a triangulação, é capaz de produzir resultados válidos e confiáveis¹²⁻¹⁵.

Alguns dos métodos usados incluem: (a) avaliação estruturada, com instrumento do tipo “checklist”, útil para verificar diferentes subitens que compõem competências avaliadas por meio de comportamentos clínicos observáveis; (b) portfólio, utilizado tanto como instrumento de avaliação formativa quanto somativa, contempla a capacidade de se auto-avaliar e de executar um plano de aprendizagem; (c) prova cognitiva, apropriada para avaliar conhecimentos técnicos específicos; (d) avaliação 360⁰, voltada para obter informações das múltiplas fontes que observam o desempenho do aluno (docentes, colegas, outros profissionais e pacientes; (e) OSCE (*objective structured clinical examination*), observação de componentes de um atendimento clínico simulado ou CSA (*Clinical Structured Assessment*), destinado a avaliar o desempenho clínico durante atendimento simulado completo; (f) Mini CEX (*mini-clinical evaluation exercise*), que consiste na observação de parte de uma consulta, com avaliação global por checklist, seguida de uma sessão de *feedback* (g) conceito global, utilizado quando se desejam avaliar, de maneira retrospectiva, categorias gerais em vez de comportamentos específicos.^{3,4,8,14,16-18}

Para que se possam avaliar as competências identificadas pela ACGME, o conceito global não pode ser uma nota geral única, altamente subjetiva. Esta abordagem unidimensional é reducionista, apesar de prática. Entre suas limitações, a mais séria delas é que ele pode mascarar deficiências do aluno, uma vez que,

ao avaliar o todo, o docente perde a visão de distintos comportamentos ou habilidades que o compõe. Uma segunda limitação é que um índice geral tem baixo valor formativo, pois fornece muito pouca ou quase nenhuma informação que possa ser usada construtivamente como *feedback*.⁸

Uma alternativa é construir um conceito que contenha itens claramente especificados, refletindo a combinação de atributos (conhecimentos, atitudes, valores e habilidades) necessários para o bom desempenho profissional¹⁹. O conceito global deve contemplar itens como qualidade da história, exame clínico, conhecimento médico, julgamento clínico, solução de problemas, hábitos e organização do trabalho, comunicação e relacionamento com paciente e seus familiares, respeito, capacidade de auto-reflexão, percepção do contexto, interação com colegas, com docentes e com demais profissionais²⁰. O aluno deve ser avaliado por um ou mais docentes, utilizando um instrumento que contenha uma escala do tipo Likert, com descritores explícitos nas extremidades inferior e superior. A nota final seria a média das notas recebidas em cada um dos itens do instrumento.²¹

Elaborado dessa forma, o conceito global apresenta características psicométricas válidas, pois reduz a chance de um estudante conseguir compensar suas deficiências em uma área, obtendo um bom desempenho em outra. Além disso, quando combinado com outros métodos de avaliação, provê informações valiosas que podem ser usadas como *feedback* para os alunos, para os avaliadores, para os coordenadores de módulos e educadores em geral.^{8,21}

Embora ainda não se tenha conseguido definir um padrão-ouro para avaliar a competência do médico ou do estudante de medicina, a aplicação de um conceito global pode ser uma avaliação abrangente.^{14,22} Esse método de avaliação, que

leva em consideração as impressões formadas a partir de uma variedade de situações às quais o estudante é exposto, por um determinado período de tempo, é considerado confiável, sendo utilizado em várias partes do mundo.¹⁵

Como a maior parte dos avaliadores identifica apenas duas grandes competências, clínica e humanista, torna-se necessário oferecer treinamentos específicos para que eles possam reconhecer e validar os diferentes itens identificados no instrumento usado. Embora esses treinamentos signifiquem tempo adicional e custo, eles são vitais para a melhoria da qualidade das avaliações e para o desenvolvimento dos docentes nas instituições acadêmicas.

Referências Bibliográficas

1. Association of American Medical Colleges – AAMC. Learning objectives for medical student education. Guidelines for medical education. Report I, 1998.
2. Horowitz SD. Evaluation of clinical competencies Basic certification, subspecialty certification, and recertification. Am J Phys Med Rehabil 2000; 79:478-80.
3. Sliwa JA, Kowalske KJ. Assessing resident clinical competence. Am J Phys Med Rehabil 2000; 79:468-73.
4. Association of American Medical Colleges – AAMC. The SOAP approach to conquering the competencies challenges, 2004. Disponível em: <http://www.aame.org/members/gea>. Acesso em abril 2006.
5. Prideaux D, Alexander H, Bower A, Dacre J, Haisr S, Jolly B, et al. Clinical teaching: maintaining an educational role for doctors in the new health care environment. Med Educ 2000, 34:820-6.

6. DiMatteo MR, DiNicola DD. Sources of assessment of physician performance: a study of comparative reliability and patterns of intercorrelation. *Med Care* 1981; 19:829-39.
7. Ramsey PG, Wenrich MD, Carline JD, Inui TS, Larson EB, LoGerfo, JP. Use of peer ratings to evaluate physician performance. *JAMA* 1993; 269:1655-60.
8. Silber GC, Nasca TJ, Paskin DL, Eiger G, Robeson M, Veloski J. Do global rating forms enable program directors to assess the ACGME competencies? *Acad Med* 2004, 79:549-56.
9. Haydt RC. *Avaliação do Processo Ensino-Aprendizagem*. São Paulo: Editora Ática, 2002.
10. Carr SJ. Assessing clinical competency in medical senior house officers: how and why should we do it? *Postgrad Med J* 2004; 80:63-6.
11. Hulsman RL, Mollema ED, Hoos AM, De Haes JCJM, Donnison-Speijer JDD. Assessment of medical communications skills by computer: assessment method and student experiences. *Med Educ* 2004; 38:813-24.
12. DeLisa JA. Evaluation of clinical competency. *Am J Phys Med Rehabil* 2000; 79:474-7.
13. St George I, KAIGAS T, Mcavoy P. Assessing the competence of practicing physicians in New Zealand, Canada and the United Kingdom: progress and problems. *Fam Pract* 2004; 36:172-7.
14. Wilkinson TJ, Frampton CM. Comprehensive undergraduate medical assessments improve prediction of clinical performance. *Med Educ* 2004; 38:1111-6.

15. van der Vleuten CPM, Schuwirth LWT. Assessing professional competence: from methods to programmes. *Med Educ* 2005; 39:309-17.
16. Depresbiteris, L. Certificação de competências: a necessidade de avançar numa perspectiva formativa. *Formação* 2001; 2:27-38.
17. Chapman DM, Hayden S, Sanders, AB et al. Integrating the accreditation council for graduate medical education core competencies into the model of clinical practice of emergency medicine. *Acad Emerg Med* 2004;11:674-85.
18. Wang E, Sanders AB Global assessment tool for emergency medicine – specific core competency evaluation. *Acad Emerg Med* 2004; 11:1370-1.
19. McMullan M, Endacott R, Gray MA, Jasper M, Miller CML, Scholes J, Webb C. Portfolios and assessment of competence: a review of the literature. *J Adv Nurs* 2003; 41:283-94.
20. Davis JD. Comparison of faculty, peer, self and nurse assessment of obstetrics and gynecology residents. *Obstet Gynecol* 2002; 99:647-51.
21. Regehr G, MacRae, 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; 72:993-7.
22. Norcini JJ. Peer assessment of competence. *Med Educ* 2003; 37:539-43.

4.5. Artigo 5 – Competência clínica de alunos de medicina em estágio clínico: comparação entre métodos de avaliação

Artigo enviado para Revista Brasileira de Educação Médica - RBEM em 23/10/2008

Artigo e Submissão

**0186/2008 - COMPETÊNCIA CLÍNICA DE ALUNOS DE MEDICINA
EM ESTÁGIO CLÍNICO: COMPARAÇÃO ENTRE MÉTODOS DE
AVALIAÇÃO**

***CLINICAL COMPETENCE OF MEDICAL STUDENTS DURING CLINICAL
CLERKSHIP: COMPARISON BETWEEN ASSESSMENT METHODS***

COMPETÊNCIA CLÍNICA DE ALUNOS DE MEDICINA EM ESTÁGIO CLÍNICO: COMPARAÇÃO ENTRE MÉTODOS DE AVALIAÇÃO

CLINICAL COMPETENCE OF MEDICAL STUDENTS DURING CLINICAL CLERKSHIP: COMPARISON BETWEEN ASSESSMENT METHODS

Título Resumido: Diferentes métodos de avaliação de alunos em estágio clínico

Autores: Rosângela Curvo Leite Domingues¹, Eliana Amaral², Angélica Maria Bicudo Zeferino³, Maria Ângela G. Monteiro Antonio⁴, Wilson Nadruz⁵ – NAPEM/Faculdade de Ciências Médicas – Universidade Estadual de Campinas – UNICAMP – Endereço: Rua Alexander Fleming, 101 – Cidade Universitária – Campinas - São Paulo – SP - 13083-881

1. Doutoranda, Departamento de Tocoginecologia, Faculdade de Ciências Médicas, UNICAMP, Brasil.
2. Professora-Assistente Doutora, Departamento de Tocoginecologia, Faculdade de Ciências Médicas, UNICAMP, Brasil
3. Professora-Assistente Doutora, Departamento de Pediatria, Faculdade de Ciências Médicas, UNICAMP, Brasil
4. Professora-Assistente, Doutora, Departamento de Pediatria, Faculdade de Ciências Médicas, UNICAMP, Brasil
5. Professor-Assistente, Doutor, Departamento de Clínica Médica, Faculdade de Ciências Médicas, UNICAMP, Brasil

Endereço para correspondência:

Rosângela Curvo Leite Domingues
Rua Luciano Venere Decourt, 858
Cidade Universitária
CEP 13083-740 - Campinas – SP
Fone: (19) 3289-2192
e-mail: rosangela@cpdec.com.br

Conflito de interesse: nenhum

Resumo

Objetivo: Comparar as notas obtidas por alunos do 4º ano de Medicina, em estágio clínico, utilizando quatro métodos de avaliação. **Método:** Ao final do estágio clínico em Ginecologia e Obstetrícia, Clínica Médica e Pediatria, em unidades de atenção primária, 106 alunos do 4º ano foram avaliados utilizando 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 emparelhados, coeficiente de correlação de Pearson, análise fatorial e distâncias Euclidianas. **Resultados:** As medianas mais baixas foram de Pt e as mais altas da 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 fatorial contrastou as notas de Pt e Pf com as de AE e CGI. **Conclusão:** Cada método parece focar 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.

Palavras-chave: métodos de avaliação, competência clínica, estágio clínico, educação médica

Abstract

Background: This study aimed to compare Y4 medical students' scores obtained from four different assessment methods. **Method:** At the end of their clinical clerkship in Obstetrics&Gynecology, Internal Medicine and Pediatrics, 106 Y4 students were assessed through four distinct methods (a 0-10 rating scale): 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 tests, Pearson correlation coefficient, factor analysis and Euclidean distances. **Results:** The lowest median scores were related to MCQs and the highest to RC-SCA. The distribution of scores for all method pairs showed significant differences ($P < 0.001$). There were stronger and greater agreement between RC-SCA and GIR. Factor analysis contrasted MCQs and P scores with those from RC-SCA and GIR. **Conclusion:** Each method focused on different aspects of the expected competencies. RC-SCA and GIR seemed to assess similar domains, in opposition to MCQs and P scores. No isolated method should be used to assess learners in clinical settings.

Key-words: evaluation methods, clinical competence, clinical clerkship, medical education

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 durante a 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.¹ Portanto, a seleção dos métodos mais adequados de avaliação em uma situação específica deve ser resultante da combinação equilibrada de todos esses aspectos.²

A avaliação deve ser parte integrante do processo de ensino-aprendizagem, já que ela 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 redirecionarem seu comportamento em relação àquilo que estão aprendendo.³ 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.¹ Porém, devido ao caráter polissêmico do termo competência, um primeiro desafio parece ser exatamente sua definição e distinção com o termo desempenho.⁴

Para discutir a avaliação de competência clínica, George Miller⁵ 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 um determinado contexto (*saber como*, competência); o próximo nível, *mostrar como* (desempenho), reflete a habilidade de agir corretamente em uma situação simulada e o último, *fazer* (ação), refere-se à prática em situações clínicas reais.

Rethans et al.⁶ 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 em uma situação particular.⁷ Afinal, nenhuma competência é exercida no vazio, mas em contextos com especificidades, peculiaridades e idiosincrasias.⁸

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 reflexão⁹, 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.⁸ 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 freqüentemente como sinônimos.^{10,11}

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 poderem 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.^{4, 7,12-14}

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 XX, mudanças significativas foram incorporadas aos sistemas de avaliação. Concluiu-se que avaliar os aspectos cognitivos era necessário, porém não suficiente para prever a competência clínica dos estudantes.¹⁴⁻¹⁶ A associação de métodos surge como uma alternativa para se atingir resultados válidos e confiáveis na avaliação clínica.^{5, 12,13,16-19} Assim, diversas entidades passaram a implantar novas combinações de métodos de avaliação, acrescentando-lhe a necessária perspectiva formativa e indutora de novas práticas pedagógicas na educação médica.^{10,17,20,21}

Entre os métodos disponíveis incluem-se: (1) provas teóricas; (2) portfólio; (3) avaliação estruturada por observação do atendimento; (4) mini-CEX (mini clinical evaluation exercise); (5) OSCE (objective structured clinical examination); (6) conceito global itemizado; (7) simulação; (8) auto-avaliação; (9) avaliação por pares; (10) avaliação por usuários; (11) avaliação por outros profissionais da equipe.⁷

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 atendimentos 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 por 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: (1) Prova Teórica (Pt), (2) Portfólio (Pf), (3) Avaliação Estruturada do Atendimento Clínico (AE) e (4) Conceito Global Itemizado (CGI).

Método

O presente estudo teve como objetivo comparar as notas provenientes de quatro métodos, atribuídas aos 106 alunos do 4º ano de Medicina, ao final do *Módulo Atenção Integral à Saúde* (MD856) em outubro de 2005. Os métodos de avaliação aplicados aos alunos foram:

1. Prova Teórica (Pt): consistiu na média de duas provas com questões dissertativas e de múltipla escolha;
2. Portfólio (Pf): foi a nota atribuída ao dossiê do aluno, que conteve planilha de todos os casos atendidos, a apresentação e a discussão de três casos selecionados de Ginecologia e Obstetrícia, Clínica Médica e Pediatria e auto-avaliação da aprendizagem;
3. 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*;
4. Conceito global itemizado (CGI): representou a média das notas atribuídas ao aluno por 3 docentes, utilizando instrumento com 13 itens que representam diferentes domínios da competência clínica (6 de habilidades técnicas: *qualidade da história, exame clínico, conhecimento médico, julgamento clínico, solução de problemas, hábitos de trabalho* e 7 de habilidades humanísticas: *comunicação e relacionamento com pacientes* e

familiares, respeito, capacidade de auto-reflexão, 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 emparelhados. As correlações entre as notas obtidas foram estudadas por meio dos coeficientes de correlação de Pearson. A análise fatorial pelo método dos componentes principais foi aplicada para examinar as inter-correlações das notas nos quatro métodos. O coeficiente alfa de Cronbach foi calculado para investigar a confiabilidade das quatro notas. A concordância relativa entre os quatro métodos foi calculada recorrendo às distâncias Euclidianas entre notas. O nível de significância escolhido foi de 5%. O estudo foi aprovado pelo Comitê de Ética em Pesquisa, sob 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, detectou-se diferença significativa entre todas elas ($P < 0,001$). A **Tabela 1** mostra as medianas, médias, desvios-padrão e intervalos de confiança das notas provenientes dos quatro métodos.

Os coeficientes de correlação entre as notas dos quatro métodos estão apresentados na **Tabela 2**. 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 as notas dos demais 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 fatorial mostrou que dois fatores explicaram 74,5% da variação total dos dados; 45,4% da variabilidade foi explicada pelo primeiro fator, que se mostrou associado

a todos os métodos e 29,1% pelo segundo fator, que mostrou contraste entre as notas da Prova teórica e do Portfólio, de um lado, e as notas da Avaliação Estruturada e do Conceito Global Itemizado, do outro.

Observou-se um reduzido grau de confiabilidade das notas nos 4 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.

Discussão

Os resultados deste estudo mostraram que as medianas da Pt foram as mais baixas entre todos os métodos utilizados, enquanto as notas da AE obtiveram as maiores medianas. Esses resultados são de certa forma, contrários aos apresentados por Auewarakul et al.¹⁹, 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 uma maior ênfase dada pelos avaliadores a aspectos relacionados à competência cognitiva, isto é, ao *saber* obtido nos livros.²² 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 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 em uma fase inicial de atendimento clínico dentro do curso médico, podendo caracterizar o efeito de leniência dos docentes²³ ou a falta de um padrão pré-estabelecido²⁴, ou ainda a carência de preparo e treinamento prévio dos avaliadores.^{10,16} 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

média e diluído aqueles relacionados às habilidades clínicas, que começam a ser praticados no módulo que foi alvo de nosso estudo.

Os resultados referentes às correlações entre as notas dos métodos foram condensados pela análise fatorial, cujo segundo fator contrastou as notas da AE e do CGI com as do Pf e da Pt, evidenciando uma distinção entre métodos mais centrados no desempenho, na ação (*fazer*) e os mais focados na cognição (*saber*)⁵. Tais achados, aliados à forte correlação e à maior concordância relativa entre as notas da 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.²⁵

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.^{19,26} 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 componentes, dentre eles a auto-avaliação 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 avaliarem os alunos de forma semelhante a Pt, valorizando o acúmulo e o uso do conhecimento teórico.

Não foram observadas correlações estatisticamente significativas entre os demais métodos (Pt vs AE, Pt vs CGI, Pf vs AE e 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.²⁷ 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.^{10,19}

O CGI pareceu ser o mais indispensável 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.^{26, 28, 29}

Entretanto, nem sempre são utilizados instrumentos de avaliação contendo 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.³⁰

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, em uma única universidade, reduzindo sua generalização. A segunda limitação consiste no fato de que 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.²¹

Apesar das limitações apontadas, neste 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 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 de que nenhum método isoladamente deve ser usado para avaliar as distintas e complexas competências que a boa prática da medicina demanda.¹⁸ 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.^{17,31} O conteúdo, o formato e a frequência da avaliação devem estar alinhados com os objetivos propostos para módulo e com o 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.⁷

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, no sentido de se compreender como as informações obtidas podem contribuir para a qualificação dos estudantes que pretendemos formar.

Referências Bibliográficas

1. van der Vleuten CPM, 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. *Teaching and Teacher Education* 2007; 23:239-50
3. McAleer S. Choosing assessment instruments. In: Dent JA & Harden RM. *A practical guide for medical teachers*. Philadelphia: 2nd edition. Elsevier Limited, 2005.
4. Ramos MN. Qualificação, competência e certificação. *Formação* 2001; 2:17-26.
5. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990; 65 Suppl 9:S63-7.
6. Rethans JJ, Norcini JJ, Baron-Maldonado M, Blackmore D, Jolly BC, La Duca T, Lew S, Page GG, Southgate LH. The relationship between competence and performance: implications for assessing practice performance. *Med Educ* 2002; 36 (10):901-9.
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* 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 de Veulten CPM. 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, Vleuten VC, Shatzer J, Jones R. Assessment of clinical competence. *The 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):871-879.
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, Praditsuwan 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. ACGME [http:// www.acgme.org/outcome/assess/toolbox/asp](http://www.acgme.org/outcome/assess/toolbox/asp). Chicago. IL: Accreditation Council for Graduate Medical Education 2004. Accessed 30 June 2006
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):889-5

24. Boulet, JR, 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):131-40.
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 RJL, Scherpbier AJJA, Stehouwer CDA, Van der Vleuten, CPM. 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.

Tabela 1. Estatística descritiva das notas obtidas por meio de quatro métodos de avaliação (n= 103)

Método	Mediana (min-max)	Média	DP	IC 95%
Prova teórica * (Pt)	7,59 (5,67-9,16)	7,49	0,69	7,35-7,62
Portfólio * (Pf)	8,50 (6,33-9,83)	8,31	0,84	8,14-8,47
Avaliação estruturada *(AE)	9,27 (7,43-10,00)	9,20	0,47	9,11- 9,29
Conceito global itemizado * (CGI)	9,03 (7,62-9,95)	9,01	0,49	8,91-9,10

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

Tabela 2. 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 = 0,009

** P<0,001

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 *	R² (%) **	Alfa #
Prova Teórica (Pt)	0,24	7,2	0,479
Portfólio (Pf)	0,26	8,9	0,498
Avaliação Estruturada (AE)	0,35	51,8	0,403
Conceito Global Itemizado (CGI)	0,42	53,0	0,348

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

** R² = porcentagem da variação da nota explicada por regressão linear sobre as notas restantes.

Alfa obtido após a eliminação das notas do método.

4.6. Artigo 6 – Global overall rating as a method of assessing students during clinical clerkship – what does it really show?

Artigo a ser encaminhado para Revista Medical Education.
Fator de impacto = 2,467

Authors: Domingues RCL¹, Amaral E², Zeferino AMB³.

Institution: State University of Campinas – Brazil

1. PhD Researcher in Medical Education. O&G Department. State University of Campinas – Brazil
2. Associate professor, O&G Department - State University of Campinas – Brazil
3. Associate professor, Pediatrics Department. State University of Campinas – Brazil

Overview

What is already known?

There is no gold standard for assessing clinical competence. Global overall ratings are a common method being used, but they can be subjective and often measure constructs too broadly.

What does this study add?

Results indicated that global overall ratings of students obtained during initial clinical experience lacked specificity, focused mainly on students' technical skills, and undervalued humanistic aspects of their competence.

Suggestions for further research

Further research is needed to investigate whether global itemized ratings provide a more functional and relevant assessment of learner attributes and competencies.

Contributors:

RCLD participated in the development of the project, including the instrument and its implementation, prepared the draft manuscript and incorporated the contributions of the co-authors.

EA participated in the development and implementation of the project, and also made significant contributions to the conceptualization, presentation and editing of this manuscript.

AMBZ contributed in the development and implementation of the project and general editing of the manuscript.

Acknowledgments: the authors appreciate the statistical support and advice of Sirlei Siani and Armando Infante, and previous version inputs from Vanessa Burch and John Boulet.

Funding: none

Conflicts of interest: none

Ethical approval: CONEP 581/2005

Global overall rating as a method of assessing students during clinical clerkship – what does it really show?

Abstract

Context: Global overall ratings are often used as a method of assessing learners' clinical competence at the end of clerkship. The purpose of this study was to identify which particular aspects of clinical competence are being assessed through global overall ratings. **Methods:** At the end a clinical clerkship in primary health units, 106 Y4 students are routinely assessed by three discipline faculties (Obstetrics & Gynecology, Internal Medicine, Pediatrics), using a global numeric rating (scale 0-10). Faculty scores across disciplines for each learner are averaged to produce a Global Overall Rating (GOR). For this study, two weeks later, the same students were assessed by the same faculties, using a more detailed form composed of 13 domains, six related to technical skills and seven to humanistic skills, each scored 0-10. Scores for each domain across disciplines were averaged as Global Itemized Ratings (GIR1 to GIR 13). Mean GIR, mean GIR technical (GIRt) and humanistic (GIRh) were calculated. Statistical analysis included sign tests, Student test for dependent correlations, Pearson correlation coefficients, factor analysis and Euclidean distances. Significance level was set at $P < 0.05$. **Results:** GOR scores were significantly lower than GIR and GIRh scores, but higher than GIRt. There was a higher significant correlation between GOR and GIRt than between GOR and GIRh. GOR scores were closer to GIRt scores. **Conclusion:** When faculty attributed a global overall rating to a student's clinical competence, they tended to focus more on technical skills. Extra attention should be given to learners' humanistic skills assessment at this phase of the course.

Key words: global rating; clinical assessment, assessment in clinical clerkship

Global overall rating as a method of assessing students clinical competence – what does it really show?

INTRODUCTION

Assessment of medical students during clinical training is a complex undertaking. Through assessment, educators can gauge the progress learners make while becoming health care professionals and students can improve their knowledge, skills and attitudes while becoming aware of their weaknesses and strengths.^{1,2}

The effectiveness of the clinical education enterprise includes the achievement of clinical competence and qualified professional performance. But the literature does not offer a clear single consensual definition of clinical competence. Many authors state that clinical competence is composed of different skills, including the habitual and judicious use of medical knowledge, clinical reasoning skills, communication skills, reflection, and professional values.³ Others would argue that this is true performance rather than pure competence.⁴ Miller's framework suggests that competence predicts performance, but in the real environment, many other factors besides "knowing how" and "showing how", may influence clinical performance, including time availability, and tiredness of the health provider, but also mood of the patient.⁵⁻⁷ For other authors, competence is contextual, reflecting the relationship between a person's abilities and the tasks required in a particular situation, and developmental, once higher levels are achieved with deliberate practice.⁸

Although different methods to assess clinical competence have been described, a single global overall rating at the end of a clinical clerkship is common.^{4, 9-12} The more general the score, the greater the probability it is influenced by the evaluator's overall impression of the learner, subject to the halo effect.¹³ Besides,

general scores offer little guidance as to its application in relation to the core objectives of the clerkship.¹⁴ Also, as clinical competence requires consideration of a variety of skills, it is important to understand which of them have been captured, in order to assist learners in identifying which aspects of their clinical practice they might review, and improve, or reinforce.^{12,15} In an attempt to understand which competencies faculty were really focusing on when assessing through single global ratings, and to provide a most useful feedback to the learner, we have proposed a comparison of the overall global rating with a global itemized rating score.

METHOD

In the institution where the study was conducted, the medical curriculum lasts 6 years. The clinical phase starts in year 4, with a nine-month introductory clinical clerkship in which students undertake approximately 200 hours of supervised clinical practice at the primary level of care, into three disciplines: Obstetrics and Gynecology, Internal Medicine and Pediatrics. The students' main tasks are history taking and physical examination, medical record writing, and practicing skills in clinical reasoning and differential diagnosis, composed with humanistic aspects of clinical care, under the perspective of family medicine.

At the end of this period, faculties provide a summative global numeric rating, ranging from 0 to 10, which summarizes student's performance during the clerkship. Faculty global scores across disciplines are then averaged to form a single aggregated score for each student, defined as a Global Overall Rating, which hereafter will be referred to as GOR. GOR has routinely been one of the components of the learner's final assessment in the clerkship.

In the end of October, 2005, two weeks after securing GOR ratings for all students, the same faculty members were required to assess the same students using an instrument for a Global Itemized Rating, named GIR. The GIR ratings were based on thirteen domains, six related to technical skills (*quality of history, physical examination, medical knowledge, clinical judgment, problem-solving, work habits*), and seven to humanistic skills (*interpersonal and communication skills, respect to patients, self-reflective skills, compassion, relationship with peer, relationship with faculty, relationship with staff/ other health professionals*), selected on the basis of current literature,¹⁶⁻²⁰ and learning objectives. The rating scale was reviewed and validated by a core panel of faculty members from the clinical disciplines involved in the clerkship. The GIR instrument contained a brief descriptor for the extreme low and high scores, from 0 to 10; space was provided for narrative comments and observations to justify the grading. Raters were given brief oral instructions before completing the questionnaire.

Faculty scores for each domain across disciplines were averaged for each of the 13 domains, hereafter referred to as GIR1 to GIR13. A mean GIR was the result of averaging GIR1 to GIR13 for all three discipline faculties. Also, mean GIR for the technical subscale, named GIRt (mean for GIR1 to GIR6), and for the humanistic subscale, named GIRh (mean for GIR 7 to GIR 13) were provided. By comparing and correlating the GOR rating with GIRh, GIRt, and each individual GIR domain ratings, we aimed to identify which aspects of learners' clinical competence have been most highly valued by faculty assessors.

Descriptive statistics were calculated. GOR and GIR scores were compared using scatter plots and sign tests. Pearson correlation coefficients were used to

evaluate the statistical associations between the 13 domains of GIR and between GOR and GIR1 to GIR13. To determine correlation patterns between domains an exploratory factor analysis was conducted. The relative agreement between GOR scores and GIR domain scores was evaluated by Euclidean distances. Significance level was set at $P < 0.05$

RESULTS

A total of 106 students and 19 faculty members participated in the study. The mean age of students was 22.8 (SE 0.2) and 50% were female (SE 4.9). Faculty were mostly female (67%) and 58% were over the age of 40 years. Three students had missing data in one or more GIR items, so this study will present the results of 103 complete assessments. Mean ratings were high in GOR and in all domains of GIR (Table 1).

Representing GOR scores versus mean GIR scores through a scatterplot, we found that 74 out of the 103 points lay below the line $GOR = GIR$, and 27 lay over it, showing that GOR scores inferior to mean GIR scores predominated significantly in 73.3% of the cases (sign test, $z = 4.58$, $P < .0001$) (**Figure 1**).

An analogous result was found for GIRh scores ($z = 6.17$ for the two-tailed sign test), but the opposite was valid for GIRt scores ($z = 4.22$), with $P < .0001$ in both situations. Comparing GIRh against GIRt scores we found that GIRh scores were significantly higher than GIRt scores (sign test, $z = 7.88$, $P < .0001$) (data not shown). Thus, faculty consistently assigned higher ratings on the humanistic subscale than on the technical subscale. Using the two-tailed sign test with the Bonferroni correction, we observed that GOR scores were (a) significantly higher than all GIR

technical domains scores except *Quality of history* and *Work habits* ($z > 2.99$), (b) significantly lower than all GIR humanistic domains scores excluding the *Self-reflective skills* domain ($z > 5.11$) (data not shown).

Positive and significant correlations between the 13 GIR domains were found. The highest correlations were observed among the technical domains ($0.75 < r < 0.84$), with the exception of *Work habits*. Among the humanistic domains the correlations ranged from 0.51 to 0.76. *Respect* showed to be the least correlated with the technical domains ($.25 < r < .43$). The complete correlation matrix for the GIR domains is shown in Table 2.

In our data *Work habits*, among technical domains, and *Respect to patients*, among humanistic domains, were the least correlated with GOR, whereas *Clinical judgment* showed the highest correlation (Table 3). The correlation 0.873 between GOR and GIRt was significantly greater than the correlation 0.614 between GOR and GIRh (Student test to compare both dependent correlations sharing the variable GOR, $t = 6.45$, $df = 100$, $P < .0001$).

An exploratory factor analysis was performed using the maximum likelihood method. Two common factors explained 68.4% of the total variance, of which 55.4% was explained by the general Factor 1. Factor 2 made a distinction between technical and humanistic qualities, except *Work habits*. A varimax rotation did not produce any interpretive improvement in the results. The lower communalities of *Work habits*, *Relationship with peers* and *Relationship with staff / other health professionals* suggested that the variance of these domain scores was the least explained by the common factors (Table 3).

The relative agreement between GOR and GIR ratings was measured by computing the Euclidean distance between the columns of scores. It was observed that GOR scores were closer to the technical skills scores GIR1 to GIR 5, but not to GIR6. In fact, the agreement of GOR and GIR scores was higher with domains *Quality of history*, *Problem-solving* and *Physical examination*, and lower with *Respect*, *Relationship with faculty* and *Relationship with peers*, respectively (Table 4).

DISCUSSION

This study showed that GOR and GIR 1 to GIR 13 scores were skewed to the right end of the scale. GOR scores were significantly higher than GIR scores for technical domains except *Quality of history* and *Work habits*, and significantly lower than all GIR scores for humanistic domains, excluding the *Self-reflective skills* domain.

Faculty consistently assigned higher ratings on the humanistic subscale than on the technical subscale. This finding is consistent with the results of other studies in the literature, in which the mean ratings of students of non-cognitive aspects of clinical competence were significantly higher than the cognitive aspects.^{3,10} There are some possible explanations for these results. First, there seem to be a tendency among faculty to simplify humanistic skills and overemphasize the importance of technical skills, particularly at this phase of their clinical training.^{21,22} Second, faculty may be somehow lenient while assessing non-cognitive skills, because it seems more difficult to justify low scores for personal qualities than to point out and explain to a student that he or she needs to improve his or her technical skills.¹² Third, assessors in our study might have had too high a level of expectation of learners'

technical skills, which led them to be stricter while evaluating these domains. Finally, since learners were novice clinicians they may have demonstrated a true lack of technical skills appropriate to clinical practice.

This current study showed significant correlations between GOR and mean GIR scores. The fact that faculty who assigned GOR scores also assigned GIR scores may partially explain this correlation. However, as GIR forms were filled in two weeks after GOR forms, the effects of such lack of independence between these two assessment methods may have been reduced.¹⁰ In our data, stronger correlations were found between GOR and GIR for domains related to technical skills (except for *Work habits*) than for domains related to humanistic skills. As to the domains included in GIR questionnaire, there were positive and significant correlations between all domain scores, but stronger correlations were found among technical-skill domain scores than among humanistic-skill domain scores. In particular, the correlation between GOR and GIRt was significantly greater than the correlation between GOR and GIRh. Additionally, the higher agreement between GOR and GIR on *Quality of history, Problem-solving, and Physical examination* provided some evidence that when faculty attributed a GOR score, in fact, they were mostly focusing on the cognitive abilities.

There seemed to be a tendency among faculties to differentiate between technical and humanistic components of competence.² These results may be explained by the fact that in Brazilian educational context, clinical supervisors are used to examining clinical competence as a whole. In fact, a unidimensional approach has been commonly employed to define competence, even when the scores derive from the assessment of multiple skills.^{9,23}

The isolated status of *Work Habits*, which refers to learner's note-taking, self organization and thoroughness, suggests that it could not be adequately assessed by faculty raters. A possible explanation for this result may be that faculties conceive punctuality and self-organization neither as a technical nor a humanistic domain, but rather as a personal characteristic of the learners. Besides, for our population, *Work Habits* can be so intrinsically related to learners' activities in the health care centers that this domain may not have seemed necessary to be submitted to assessment.

Although GOR may have practical advantages in capturing general aspects of student competence, it has some shortcomings. The most important is that it is a very subjective approach, based on situations that are not specific in time, i.e., it covers an extended period, which may lead to some bias.²³ With such a single overall rating, faculties may be assigning grades based on their own value systems, so raters become interpreters of behaviors. If the quality of judgment is poor, the scores and derived feedback may not be very informative to the learners.^{24,25}

On the contrary, the use of the proposed multidimensional GIR, a more descriptive assessment method, if longitudinally applied like in the present study, may add quality to the assessment process for at least three good reasons. On this hypothesis, first, faculty may detect and describe significant changes in student performance thereby reducing the likelihood that marginal learners will be able to compensate for serious deficiencies in one area with strengths in another. Second, the GIR method may provide better, more meaningful and domain-specific feedback to learners and program directors, as previously highlighted, improving students' compliance with the evaluation system.²⁶ Third, there seems to be a consensus

among different medical professional associations and regulatory bodies that information gathering for the assessment of complex competencies will increasingly be based on more descriptive, multidimensional approaches, rather than a simple numerical score.^{14, 23}

The present study has some limitations. First, the results presented referred to a single class at a single time point. Second, there may have been some kind of rating bias, once GIR and GOR assessment methods were used by the same teachers to assess the same learners within a short interval. Third, lack of previous extensive rater training may have affected the rating process. Global ratings accuracy may be heavily influenced by raters' training and experience in assessment issues, once they require precise knowledge of learning objectives and expected competence of students from raters. Being a good clinician and teacher may not, therefore be a sufficient condition to achieve rating accuracy. Training may reduce uncertainty in judgment and familiarize raters with the assessment process and content, enabling them to recognize characteristics that might otherwise have been unnoticed and to interpret and integrate students' behaviors on specific domains.^{21, 23, 27} Finally, as there are inherent difficulties associated with defining and operationalizing the various domains, the instrument should use rich appropriate vocabulary not only describe the extreme scores, but also for the intermediate levels of competence.²⁶

It is generally agreed that professional competence includes not only medical knowledge and technical skills, but also interpersonal and humanistic skills, but even so, medical educators have generally paid more attention to assessing the cognitive abilities.²⁸ It still remains unclear whether raters reward some attributes

more than others, but it is possible that humanistic skills may be less rewarded than medical knowledge and efficiency.²⁹

Our results led us to conclude that, while assessing learners through GOR, faculties were inclined to associate it mainly to technical-skill domains⁹; consequently humanistic domains were being unappreciated. Nevertheless, the use of a global itemized rating method, that integrates multiple domains of competence seems feasible and useful to students and to faculty, since it fosters reflection and change of habits. If we wish to cultivate humanistic qualities in our graduates as early as possible in medical training, it would be helpful to encourage and prepare faculty to look beyond the cognitive components of medical professionalism, to notice and evaluate the nuances and complexities that constitute true humanistic behaviors in a wide variety of contexts. The implementation of a standardized protocol would be a valuable strategy in order to guide faculty raters to consider humanistic attributes.

REFERENCES

1. Wass V, van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *The Lancet* 2001; **357**:945-9.
2. Arnold L. Assessing professional behavior: yesterday, today and tomorrow. *Acad Med* 2002; **77**: 502-15.
3. Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA* 2002; **287**: 226-35
4. Wilkinson TJ, Frampton CM. Comprehensive undergraduate medical assessments improve prediction of clinical performance? *Med Educ* 2004, **38**:1111-6.

5. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990; **9**:S62-7.
6. Rethans JJ, Norcini JJ, Baron-Maldonado M, Blackmore D, Jolly BC, La Duca T, Lew S, Page GG, Southgate LH. The relationship between competence and performance: implications for assessing practice performance. *Med Educ* 2002; **36**:901-9.
7. Carr SJ. Assessing clinical competency in medical senior house officers: how and why should we do it? *Postgrad Med J*; 2004; **80**:63-6.
8. Epstein RM. Assessment in medical education. *N Engl J Med* 2007; **356**:387-96.
9. 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 Educ Theory Pract* 2000; **5**: 131-140.
10. Morgan PJ, Cleave-Hogg D, Guest CB. A comparison of global ratings and checklist scores from an undergraduate assessment using an anesthesia simulator. *Acad Med* 2001; **76**: 1053-5.
11. Daelmans HEM, van der Hem-Stokroos HH, Hoogenboom RJL, Scherpbier AJJA, Stehouwer CDA, van der Vleuten CPM. Global clinical performance rating, reliability and validity in undergraduate clerkship. *Neth J Med* 2005; **63**: 279-84.
12. Pulito AR, Donnelly MB, Plymale M. Factors in faculty evaluation of medical students' performance. *Med Educ* 2007; **41**:667-75.
13. Norcini JJ. Peer assessment of competence. *Med Educ* 2003; **37**: 539-43.
14. Battistone MJ, Pendleton B, Milne C, Battistone ML, Sande MA, Hemmer PA, Shomaker TS. Global descriptive evaluations are more responsive than global

- numeric ratings in detecting students' progress during the inpatient portion of an internal medicine clerkship. *Acad Med* 2001; **76**: S105-7.
15. Troncon LEA. Significance of experts' overall ratings for medical student competence in relation to history-taking. *São Paulo Med J* 2006; **124**: 101-4.
 16. Ramsey PG, Wenrich MD, Carline JD, Inui TS, Larson EB, LoGerfo JP. Use of peer ratings to evaluate physician performance. *JAMA* 1993; **269**: 1655-60.
 17. Chapman DM, Hayden S, Sanders AB, Binder LS et al. Integrating the Accreditation Council for graduate medical education core competencies into the model of the clinical practice of emergency medicine. *Acad Emerg Med* 2004; **11**:674-85.
 18. American Board of internal Medicine (ABIM). Evaluating clinical competence in internal medicine 2003. <http://www.abim.org/residence/tools>. Accessed 6 May 2008.
 19. Dannefer E, Henson LC, Bierer SB, et al. Peer assessment of professional competence. *Med Educ* 2005; **39**:713-22.
 20. Elwyn G, Lewis M, Evans R, Hutchings H. Using a "peer assessment questionnaire" in primary medical care. *Br J Gen Pract* 2005; **55**:690-5.
 21. Misch DA. Evaluating physicians' professionalism and humanism: the case for humanism "connoisseurs". *Acad Med* 2002; **77**:489-95.
 22. Mattheos N, Nattestad A, Falk-Nilsson E, Attstrom R. The interactive assessment: assessing students' self-assessment ability. *Med Educ* 2004; **38**:378-89.
 23. Silber GC, Nasca TJ, Paskin DL, Eiger G, Robeson M, Veloski J. Do global rating forms enable program directors to assess the ACGME competencies? *Acad Med* 2004; **79**:549-56.

24. Norcini J, Boulet J. Methodological issues in the use of standardized patients for assessment. *Teach Learn Med* 2003; **15**:293-7
25. van der Vleuten C. Validity of final examinations in undergraduate medical training. *BMJ* 2000; **321**:1217-9.
26. Farrell SE. Evaluation of student performance: clinical and professional performance. *Acad Emerg Med* 2005; **12**:302e6-e10.
27. Nendaz MR, Perrier A, Simonet ML, Huber P, Junod A, Vu NV. Appraisal of clinical competence during clerkship: how knowledgeable in curriculum and assessment development should a physician-examiner be? *Acad Med* 2001; **76**:S99-101.
28. Sullivan ME, Hitchcock MA, Dunnington MD, Dunnington GL. Peer and self-assessment during problem-based tutorials. *Am J Surgery* 1999, **177**: 266-9.
29. Lurie SJ, Lambert DR, Nofziger AC, Epstein, RM, Grady-Weliky TA. Relationship between peer assessment during medical school, dean's letter rankings, and ratings by internship directors. *Soc Gen Intern Med* 2007; **22**:13-6.

Table 1. Mean, standard deviation, median and extreme quartiles of GOR and GIR faculty assessments (n=103)

	Mean	SD	Median	Extreme quartiles
GOR	8.85	.580	8.90	8.50 – 9.33
Mean GIR	9.02	.490	9.02	8.72 – 9.38
GIRt*	8.67	.634	8.67	8.28 – 9.17
GIRh**	9.33	.456	9.33	9.00 – 9.62
Individual Items				
1. Quality of history	8.73	.770	8.67	8.33 – 9.33
2. Physical examination	8.67	.732	8.67	8.00 – 9.17
3. Medical knowledge	8.43	.722	8.33	8.00 – 9.00
4. Clinical judgment	8.43	.752	8.33	8.00 – 9.00
5. Problem-solving	8.60	.764	8.67	8.00 – 9.00
6. Work habits	9.28	.579	9.33	9.00 – 9.67
7. Interpersonal and communication skills	9.26	.610	9.33	9.00 – 9.67
8. Respect to patients	9.57	.482	9.67	9.33 – 10.00
9. Self-reflective skills	9.03	.676	9.00	8.67 – 9.67
10. Compassion	9.12	.635	9.33	8.67 – 9.67
11. Relationship with peers	9.34	.532	9.33	9.00 – 9.67
12. Relationship with faculty	9.48	.437	9.67	9.33 – 9.67
13. Relationship with staff/other health professionals	9.13	.547	9.00	8.67 – 9.67

* mean of GIR1 to GIR6

**mean of GIR7 to GIR13

Table 2. Correlation matrix for the GIR domains (GIR1 to GIR13) * (n=103)

Domain	GIR1	GIR2	GIR3	GIR4	GIR5	GIR6	GIR7	GIR8	GIR9	GIR10	GIR11	GIR12
GIR2	0.81											
GIR3	0.80	0.80										
GIR4	0.84	0.75	0.81									
GIR5	0.80	0.79	0.79	0.80								
GIR6	0.54	0.55	0.61	0.50	0.55							
GIR7	0.59	0.48	0.45	0.64	0.59	0.48						
GIR8	0.25	0.37	0.26	0.29	0.35	0.43	0.61					
GIR9	0.58	0.54	0.64	0.55	0.67	0.54	0.54	0.55				
GIR10	0.57	0.57	0.58	0.59	0.62	0.56	0.63	0.68	0.76			
GIR11	0.39	0.33	0.28	0.44	0.43	0.31	0.64	0.56	0.57	0.51		
GIR12	0.34	0.39	0.36	0.43	0.44	0.43	0.65	0.74	0.56	0.58	0.69	
GIR13	0.39	0.33	0.39	0.42	0.54	0.40	0.54	0.52	0.56	0.76	0.57	0.51

* All Pearson correlations were significant.

Table 3. Factor analysis results of faculty assessment scores for the 13 items of the questionnaire and Pearson correlations (r) between GOR and GIR ratings (n= 103)

	Factor Analysis			r*
	First Factor	Second Factor	Communalities	
GIRt (<i>Technical skills</i>)				.873
1. Quality of history	.865	-.307	.841	.804
2. Physical examination	.835	-.243	.756	.722
3. Medical knowledge	.846	-.307	.810	.805
4. Clinical judgment	.866	-.230	.804	.835
5. Problem-solving	.881	-.160	.802	.808
6. Work habits	.658	.052	.436	.594
GIRh (<i>Humanistic skills</i>)				.614
7. Interpersonal and communication skills	.727	.308	.624	.583
8. Respect for patients	.544	.665	.738	.314
9. Self-reflective skills	.763	.224	.633	.618
10. Compassion	.784	.331	.724	.600
11. Relationship with peers	.571	.460	.538	.398
12. Relationship with faculty	.611	.560	.687	.424
13. Relationship with staff/ other health professionals	.607	.386	.517	.470

* All values are significant (p<0.01) under two-tailed normal tests of independence.

Table 4. Euclidean distances between GOR and GIR scores (n= 103)

Domains	Distance to GOR
Technical skills	
1. Quality of history	4.80
2. Physical examination	5.46
3. Medical knowledge	6.04
4. Clinical judgment	6.03
5. Problem-solving	5.21
6. Work habits	6.85
Humanistic skills	
7. Interpersonal and communication skills	6.86
8. Respect	9.63
9. Self-reflective skills	5.88
10. Compassion	6.14
11. Relationship with peers	7.94
12. Relationship with faculty	8.52
13. Relationship with staff/ other health professionals	6.53

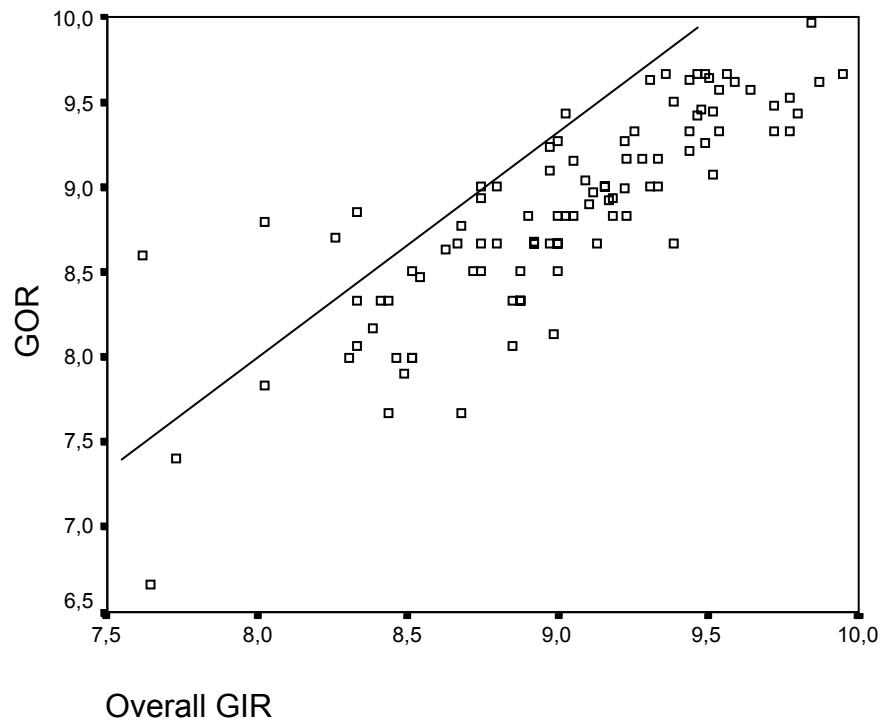


Fig 1. Scatterplot of GOR versus mean GIR scores (n=103).

4.7. Artigo 7 – Global overall rating as method to assess students’ clinical competence on technical and humanistic skills: comparison between faculty, self and peer assessment

Artigo a ser encaminhado para Revista Academic Medicine
Fator de impacto = 2,607

This paper has not been previously published and is not under consideration in the same or substantially similar form in any other journal.

This paper represents part of the results of a more comprehensive study on undergraduate medical students’ assessment. Other manuscripts are currently under consideration at other journals.

Contributors:

RCLD participated in the development of the project, including the instrument and its implementation, prepared the draft manuscript and incorporated the contributions of the co-authors.

EA participated in the development and implementation of the project, and also made significant contributions to the conceptualization, presentation and editing of this manuscript.

AMBZ contributed in the development and implementation of the project and general editing of the manuscript.

Conflicts of interest: none

Funding: none

Corresponding author:

Rosângela Curvo Leite Domingues

Rua Luciano Venere Decourt, 858- Cidade Universitária

Campinas SP- Brasil – 13.083-740

Fone para contato (19) 9207-6404 / (19) 3289-2182

e-mail: rosangela@cpdec.com.br

Global rating as a method to assess students' clinical competence on technical and humanistic skills: comparison between faculty, self and peer assessment

Authors: Domingues, Rosângela Curvo Leite¹; Amaral Eliana²;

Zeferino, Angélica Maria Bicudo³

1. PhD Researcher in Medical Education. O&G Department. State University of Campinas – Brazil
2. Associate professor, O&G Department - State University of Campinas – Brazil
3. Associate professor, Pediatrics Department. State University of Campinas – Brazil

Corresponding author:

Rosângela Curvo Leite Domingues

Rua Luciano Venere Decourt, 858- Cidade Universitária

Campinas SP- Brasil – 13.083-740

Fone para contato (19) 9207-6404 / (19) 3289-2182

e-mail: rosangela@cpdec.com.br

Abstract

Purpose: To compare global itemized ratings generated through faculty, self and peer assessment, at the end of Y4 medical students' clinical clerkship. **Method:** In 2005, at State University of Campinas – Unicamp, Brazil, faculty were required to score students using a global itemized rating form (GIR), scale 0-10, consisting of thirteen domains; six referred to technical skills (GIRt), and seven to humanistic skills (GIRh). Students were instructed to complete a similar GIR form for self-assessment, and the humanistic domains of GIR to anonymously assess their peers (3-5). **Results:** 737 assessment events were generated: 42.7% (faculty-alpha 0.935), 13.6% (self-alpha 0.881), 43.7% (peers-alpha 0.904). Self ratings

were higher than faculty ratings in mean GIR, lower in GIRT, with no significant difference in GIRh. GIRh was significantly higher than GIRT (faculty and self). There were significant correlations between GIRT and GIRh (self $r=0.61$), (faculty $r=0.62$). Very weak correlations were found between self, peer and faculty ratings for each domain. There was not a uniform pattern among raters in relation to the impact of each domain on the reliability of the questionnaire. For the merged data, five factors were obtained, explaining 60.2% of total data variation. There was a higher disagreement between self and faculty scores for technical than for humanistic skills. Peer scores were closer to faculty than to self scores. **Conclusion:** Students and faculty seemed to apply different assessment criteria when evaluating learners' clinical competence. Raters were more lenient when assessing the non-cognitive aspects of care. Learners attributed lower ratings to their fellow students than to themselves.

Key-words: self, peer and faculty assessment, global rating, clinical competence

Introduction

Medical education has faced several challenges in the last few decades, but the most important of them is to ensure that physicians who leave medical school are capable of delivering competent professional services.¹ It is critical therefore, that adequate and reliable assessment methods be developed and implemented, in which evaluation of clinical competence would be a relevant outcome measurement.^{2,3,4} But as clinical competence is not an absolute attribute, its definition must be clarified. Clinical competence can be considered a set of technical and humanistic skills needed to effectively manage specific clinical situations. It is qualified by the context in which these skills are demonstrated, since it conjures up the relationship the student/physician establishes with a health-related environment.^{2,4}

Evaluation of clinical competence has primarily been the responsibility of supervisors and faculty members.^{5,6} But medical professional associations and regulatory bodies have suggested that self and peer assessment may provide information complementary to faculty ratings that is useful in the assessment of undergraduate medical students as well as relevant in the development of self-directed learning and medical professionalism.^{5,7,8}

The ability to assess one's skills, weaknesses and strengths deepens students' learning experiences, and helps them select continuing education activities that will address any specific gaps.⁹⁻¹¹ Consequently, as an attempt to prepare students for the demands of professional life, undergraduate medical education is expected to incorporate early and frequent opportunities for student self-assessment in order to develop self-regulatory behavior, thereby contributing to the improvement of self-confidence.¹²⁻¹⁴

Similarly, peer assessment has been considered a reliable and valid method for evaluating difficult-to-assess skills, particularly humanistic non-cognitive skills, since peers might be more favorably positioned to observe one another regularly over a wide range of circumstances where faculty may not be present.^{12,15-17} Peer assessment may promote professionalism, teamwork and interpersonal communication, and been found to be highly predictive of later medical students' performance evaluations.^{18,19}

The clinical clerkship environment in medical curricula, with its emphasis on extensive training in primary care, collaborative learning and team-working provides a unique context in which to explore alternative forms of assessment. So, the potential exists for faculty assessment to be combined with or compared to peer and self-assessment to assess a range of skills and to improve the comprehensiveness of the evaluations generated.

Method

The State University of Campinas, Unicamp, Brazil offers a six-year vertically and horizontally integrated medical curriculum based on multidisciplinary teaching modules. The clinical phase starts in year 4, with a nine-month introductory clerkship at six primary health care centers in which students, in groups of three to five, undertake approximately 200 hours of supervised clinical practice in three clinical disciplines (Obstetrics & Gynecology, Internal Medicine and Pediatrics). The students' main tasks are history taking and physical examination, medical record writing and practicing skills in clinical reasoning and differential diagnosis.

In the end of October 2005, teachers were required to score students they had supervised using a global rating form, which consisted of thirteen domains assessing clinical competence. Six domains referred to technical skills: *quality of history, physical examination, medical knowledge, clinical judgment, problem-solving, work habits*, and seven to humanistic skills: *interpersonal and communication skills, respect, self-reflective skills, compassion, relationship with peer, relationship with faculty, relationship with staff/ other health professionals*. The domains, selected on the basis of current literature and the program learning objectives, were reviewed by a core panel of faculty members from the three clinical disciplines involved in the clerkship.^{4,15,20,21} This assessment tool used an eleven-point rating scale anchored by behavior descriptors for the extreme low and high scores, i.e. worst performance (0) to best performance (10). The form also provided space for narrative comments. Ratings across domains and disciplines were averaged and will be hereafter referred to as overall global itemized rating (overall GIR). GIR for technical and for humanistic skills were also calculated, and will be named mean GIRt and mean GIRh, respectively, and used to make further comparisons. Additionally, for the first time, students were instructed to complete a similar GIR form for self-assessment, and the humanistic scoring components of the GIR form to anonymously assess their peers, with whom they shared similar scheduled activities (three to five colleagues).

This paper presents the comparative results between GIR scores generated through faculty, self and peer assessment, at the end of Y4 medical students' clinical clerkship. Ethical approval was obtained from the State University of Campinas' Ethical and Research Committee (number 581/2005).

Reliability, as determined by internal consistency across the domains of the questionnaire was measured by Cronbach's alpha coefficient. Friedman's test and Wilcoxon signed rank tests were used to compare score distributions. Correlations between (1) overall GIR, mean GIRt and mean GIRh for faculty and self scores, and (2) mean GIRh for self, faculty and peer scores were assessed by means of two-sided binormal tests of independence. Factor analysis was used to examine the correlation patterns of scores. The relative agreement between scores was evaluated by Euclidean distances. Statistical significance was set at $P \leq 0.05$.

Results

A total of 106 students and 19 faculty members participated in the study. The mean age of students was 22.8 (SE 0.2) and 50% were female (SE 4.9). Faculty were mostly female (67%) and 58% were over the age of 40 years. A total of 737 assessment events were considered: 42.7% by faculty, 13.6% by self and 43.7% by peers. Only students with no missing values were considered in the analyses, so this study will present the results obtained for the 96 students with complete forms. Alpha coefficients for the 13-domain questionnaire were: 0.881 for self-assessment and 0.935 for faculty assessment. For the 7-domain questionnaire filled in by peers it was 0.904. The distributions of scores were negatively skewed. Table 1 shows the GIR medians and ranges of values for faculty, self and peer ratings.

At the domain level, medians of self ratings were generally slightly higher than medians of faculty ratings, but *Medical knowledge* was the only domain exhibiting a significant difference between both distributions (Wilcoxon $W=5.21$, $P<.001$). Self and faculty ratings were significantly higher for humanistic than for

technical competencies ($W=8.23$ for self and $W=7.80$ for faculty, both with $P<0.001$). Faculty mean GIRt was significantly higher than mean GIRt for self evaluation scores ($W=2.54$, $P=0.011$) and nonsignificantly lower than mean GIRh for self ($W=1.32$, $P=0.186$). The distributions of self, faculty and peer mean GIRh differ significantly (Friedman test, $\chi^2=8.25$, $df=2$, $P=0.016$), the difference being due to self mean scores distribution being higher than the distribution of peer scores ($W=2.70$, $P=0.007$). Mean GIRh for peers was not significantly lower than mean GIRh faculty ($W=1.86$, $P=0.063$). At the domain level, only *Respect*, *Relationship with peers* and *Relationship with faculty* exhibited significantly different ratings ($W=3.70$, $P<0.001$, $W=2.84$, $P=0.005$ and $W=2.80$, $P=0.005$, respectively).

There were significant correlations between mean scores of technical and humanistic skills for self ($r=0.61$) and faculty assessment ($r=0.62$). As shown in Table 2, very weak and some negative correlations were found between self, peer and faculty ratings for each domain. A significant, but modest correlation was observed between peer and self scores only in the domain *Relationship with peers*.

To explore the statistical association between scores for all domains for the three types of raters (self, faculty and peer), we computed Pearson correlation coefficients. The correlations between faculty scores for the 13 domains were all significant and ranged from 0.19 to 0.84. Despite reaching statistical significance for the majority of scores, the correlations between self scores for the 13 domains were low to modest (0.13 to 0.67). Peer scores correlation coefficients were all significant and ranged from 0.43 to 0.76 (data not shown).

The reliability analysis of faculty, self and peer assessment scores demonstrated that *Physical examination* (for self), *Problem-solving* (for faculty), and *Self-reflective skills* (for peers) were the most associated with the sum of the other scores. *Respect* (for self), *Clinical Judgment*, and *Medical Knowledge* (for faculty), and *Self-reflective skills* (for peers) were the most explained by the remaining domains. There was a very small reduction of Cronbach's alpha after elimination for all domains for faculty, self and peer scores; however *Physical Examination* (for self), *Problem Solving* (for faculty) and *Self-reflective skills* (for peers) showed to have more impact in the reliability of the scale than the other domains (Table 3).

An exploratory factor analysis of the 13 GIR faculty, the 13 GIR self and the 7 GIR peer domain scores was conducted on their 33×33 correlation matrix. After examination of the scree plot, the parameters for a five-factor solution were estimated by the maximum likelihood method. The factors explained 60.2% of total variation. As the solution was not easy to interpret, the factors were subjected to a varimax procedure. After rotation we obtained factors with mainly positive loadings, showing that all domains were aligned with the factors. Factor 1 represented *faculty technical domain* scores; Factor 2 can be called a *peer factor*, since it represented the seven peer humanistic domain scores; Factor 3 was related to *faculty humanistic domain scores*; Factor 4 encompassed all self domain scores and was called a *self factor*, and Factor 5 represented *self technical domain scores* (data not shown).

The relative agreement between technical and humanistic ratings was investigated by constructing the Euclidean distance between columns of scores. Table 4 shows that the distance between self and faculty scores for technical skills

was greater than for humanistic skills, thus exhibiting higher disagreement. Peer scores were closer to faculty than to self scores, for humanistic skills.

To explore levels of relative agreement between all self, faculty and peer domain scores we used a two-dimensional map, constructed using multidimensional scaling (Figure 1). Four clusters were formed: 1. self scores for technical domains (S1 to S5), except *Work Habits* (S6), 2. self scores for humanistic domains (S7 to S13), including *Work Habits* (S6); 3. faculty scores for technical domains (F7 to F13, except *Work Habits* (F6); 4. faculty and peer scores for humanistic domains (F7 to F13 and P7 to P13), with *Relationship with peers* (for peer scores, P11) being more distant than the other humanistic domains.

Discussion

The overall results of the present study provide evidence that students and faculty apply different assessment criteria when evaluating learners' clinical competence.

Ratings given by faculty, self and peers were skewed to the right end of the scale, with similar rating ranges, limiting the discrimination among students. Concerning faculty ratings, our data are consistent with previous studies that report high faculty ratings of learners' often exhibiting a "ceiling effect".²² Supervisors may not be sufficiently objective when asked to assess the competence of their own students. In our study, either the "halo effect", in which a good or bad performance in one area affects the assessor's judgment in other areas, or "leniency", in which assessors are reluctant to attribute low rates may have played a role. In addition, a variety of factors may influence faculty decisions when rating student performance, including a

fear of impairing student-teacher relationships, invoking negative emotional reactions from students, or concern that poor student performance may be attributed to limited expertise on the part of the teacher.²³

Some studies have shown that raters are generally more lenient when assessing the non-cognitive aspects of care.^{12,24} This tendency to simplify humanistic skills and overemphasize the importance of technical skills, observed among both teachers and students²⁵, may be one explanation for our results which showed that humanistic skills were rated higher than technical skills.

When we compared the distributions of scores attributed by the three types of raters, we observed that there were no significant differences between most of them; *Medical Knowledge* was the only domain rated significantly higher by self. Our findings are somewhat different from the results of other studies, which showed that learners at early phases of clinical practice tend to overestimate their competence, probably because they lack both the skills needed to perform well and the necessary meta-cognitive skills to accurately perceive and rate their knowledge and behaviors.^{11, 25-27}

Learners attributed lower mean ratings to their fellow students. These findings are different from the ones shown in other studies, where students were ranked higher by their peers than by themselves.^{7,9} However, these results should be interpreted with caution, since mean ratings were very close and, in accordance with other studies, they may be a reflection of learners' lack of experience, fear of being misinterpreted, and reluctance to provide low ratings resulting in social isolation.^{4-,7,14,16,28,29}

Although the questionnaire used in this study showed to have high internal consistency, as expressed by Cronbach's alpha coefficients, and attempted to take into

account various relevant dimensions of clinical competence,³⁰ this was not an easy task, since medical competence cannot be simply reduced to a set of items.⁶ The results of the reliability analysis of the domains per se included in the questionnaire were not very enlightening. However, when we examined the different population of raters, we observed that there seemed not to be much in common between the way faculty and learners interpreted each of the thirteen test domains.

These findings were confirmed by factor analysis, whose results also suggested that faculty ratings seemed not to be explained by the same factors that explained self and peer ratings. The factors related to self ratings showed some overlapping of information, suggesting that there was a higher level of arbitrariness in self rating process than in faculty and peer rating processes. Additionally, self ratings showed to have a smaller agreement with peer ratings than with faculty ratings for the humanistic domains. Our findings are in accord with the general consensus in medical education that students are most likely to provide untrustful self ratings, which makes self assessment an inexact measure.^{8,27} The factors found in faculty analysis were somehow reflected in the agreement analysis, since in both statistical methods we could observe that raters made a distinction between technical and humanistic domains.

Correlations between the scores obtained from faculty, self and peer assessment were low. Although correlation is not necessarily a reflection of validity, the low degree of correlation between faculty evaluations and peer and self evaluations shows that self and peer ratings should be interpreted with caution. Our results coincide with the ones found in other studies, in which self-ratings failed to correlate significantly with either faculty and peer ratings, showing that learners are not fully prepared to

accurately assess their own performance.^{8,31-35} This may be explained by the fact that self and peer assessment are indeed difficult tasks, which include the development of a clear understanding of the self and of others as individuals and professionals. Besides, neither self nor peer evaluation skills are taught in our medical curriculum. According to the current literature, the discrepancy between self, peer and faculty assessment may also be a matter of time and experience. Some studies have evaluated the temporal stability of medical students' self-and peer assessment accuracy, and their results indicated that time, continued practice and effective feedback helped learners "calibrate" their self-and peer assessments to meet those of faculty.^{8,31,36}

Our results could be used to argue against the use of self and peer assessment by Y4 medical students. Many self and peer assessment studies have pointed out some methodological problems, including the lack of gold standard to assess clinical competence, the inconsistent use of the rating scales by individual raters, issues of group rating behaviors (poor quality assessors may dilute the effect of good quality assessors), and direct comparison of median scores attributed by different types of raters, each of which with his own principles, values, knowledge, experience and expectations.^{17,31,37}

On the other hand, there is a significant body of literature that supports the idea that self and peer assessment are the basis for every physician to maintain clinical competence, being the fundamental components upon which the whole process of independent learning rests.^{5,8,19,31,38}

Our results should be interpreted within the context of potential limitations. Our data represent medical students from within a single institution. It remains to be seen if these results can be generalized to other settings.^{39,40} Another weakness is

the lack of advance rater preparation and training. The behavior of raters might have been influenced significantly if more detailed information and instructions had been given about the aim and nature of peer and self assessment.^{20,41} Norcini²⁸ suggests that the purpose of the assessment be stated in writing, and that assessment criteria be developed and communicated to all participants (how many and which peers will participate, what they will assess, when they will assess it, what constitutes the acceptable range of quality, and, if necessary, what is considered a suitable performance).

A further limitation is that all of the students who provided scores in this study worked together in the same groups, in the same primary care units, contributing to a potential bias, the “cluster” effect”, meaning that learners tend to be closely together, making it difficult for them to have an accurate view of each individual. Nevertheless, only colleagues who work together in close proximity could be reasonably expected to have sufficient information to undertake peer ratings.²⁰ Finally, more explicit anchors for the evaluation criteria should be provided, in order to address the problems with differential use of the scale among faculty and students.³⁵

This study tried to make some contribution to knowledge in the area of peer-and self assessment, the most significant one being the introduction of both strategies at the beginning of the clinical phase of an undergraduate medical program. This practice needs to be viewed from a self-improvement perspective, because it represents a shift from “How good I am?” to “How can I get better?” By judging the work and behavior of their colleagues early in medical school, learners gain insight into their own performance.^{7,8} So, we understand that self and peer evaluation would enhance our current method of assessing Y4 medical students in clinical clerkship,

since both processes may encourage learners to reflect and elaborate their own development plans and should be addressed throughout the medical curriculum.^{9,25} A student with a well-developed sense of self and a capacity for self-acceptance and constructive peer criticism is more likely to relate to patients and colleagues with enthusiasm, clarity and empathy and is more prone to undergo the stresses of the medical roles.⁴² There are considerable data in the literature that support the idea that students' self perception and perception of others can increase with increasing clinical maturation and with specific, intense and continuous feedback.^{10,31}

The use of faculty, self and peer ratings in early clinical practice can stimulate discussion and provide useful information for curriculum development and planning, as well as help identifying clinical areas where students need more attention and where educators should re-evaluate the teaching methodology.

REFERENCES

- 1- Belar CD, Hersh LE, Rozensky RH, Brown RT et al. Self-assessment in clinical psychology: a model for ethical expansion of practice. *Prof Psychol Res Pr.* 2001; 2:135-41
- 2- Klass D. Reevaluation of clinical competency. *Am J Phys Med Rehabil.* 2000; 79:481-6.
- 3- Silber GC, Nasca TJ, Paskin DL, Eiger G, Robeson M, Veloski J. Do global rating forms enable program directors to assess the ACGME competencies? *Acad Med.* 2004; 79: 549-56.
- 4- Dannefer E, Henson LC, Bierer SB, et al. Peer assessment of professional competence. *Med Educ.* 2005; 39:713-22.

- 5- Davis JD. Comparison of faculty, peer, self and nurse assessment of obstetrics and gynecology residents. *Obstet Gynecol.* 2002; 99:647-51.
- 6- van der Vleuten, CPM., Schuwirth, LWT. Assessing professional competence: from methods to programmes. *Med Educ.* 2005; 39:309-17.
- 7- Bryan, RE, Carmichael SW, Viggiano TR, Pawlina W. Assessing professionalism in early medical education: experience with peer evaluation and self-evaluation in the Gross Anatomy Course. *Ann Acad Med Singapore.* 2005; 34:486-91.
- 8- Papinczak T, Young L, Groves M, Haynes M. An analysis of peer, self and tutor assessment in a problem-based learning tutorials. *Med Teach.* 2007; 29:e122-e132.
- 9- Rudy DW, Fejfar MC, Griffith CH, Wilson JFIII. Self-and peer assessment in a first year communication and interviewing course. *Eval Health Prof.* 2001; 24:436-45
- 10- Edwards RK, Kellner KR, Siström CL, Magyari EJ. Medical student self-assessment of performance on an obstetrics and gynecology clerkship. *Am J Obst Gynecol.* 2003; 188:1078-82
- 11- Rees C, Shepherd M. Students' and assessors' attitudes towards students' self-assessment of their personal and professional behaviours. *Med Educ.* 2005; 39:30-9.
- 12- Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA.* 2002; 287:226-35.
- 13- Ben-David MF. Principles of assessment. In: DENT, J.A., HARDEN, R.M. A practical guide for medical teachers. Second edition. London: Elsevier, 2005. p.282-92.

- 14- Langedyk V. Not knowing that they do not know: self-assessment accuracy if third-year medical students. *Med Educ.* 2006; 40: 173-9.
- 15- Ramsey PG, Wenrich MD, Carline JD, Inui TS, Lasrson EB, LoGerfo, JP. Use of peer ratings to evaluate physician performance. *JAMA.* 1993; 269:1655-60.
- 16- Arnold L. Assessing professional behavior: yesterday, today and tomorrow. *Acad Med.* 2002; 77:502-15.
- 17- Evans R, Elwyn G, Edwards A. Review of instruments for peer assessment of physicians. *BMJ.* 2004; 328:1240-3.
- 18- Epstein RM. Assessment in medical education. *N Engl J Med.* 2007. 356:387-96.
- 19- Lurie SJ, Lambert DR, Nofziger AC, Epstein RM, Grady-Weliky TA. Relationship between peer assessment during medical school, dean's letter rankings, and ratings by internship directors. *Soc Gen Inter Med.* 2007; 22:13-6.
- 20- Elwyn G, Lewis M, Evans R, Hutchings H. Using a "peer assessment questionnaire" in primary medical care. *Br J Gen Pract.* 2005; 55: 690-5.
- 21- ABIM - American Board of Internal Medicine – Attending Physicians Competency Card, available at www.abim.org.
- 22- 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 Educ Theory Pract.* 2000; 5:131-40.
- 23- McKinsty BH, Cameron HS, Elton RA, Riley SC. Leniency and halo effects in marking undergraduate short research projects. *BMC Med Educ* 2004; 4:28 doi:10.1186/1472-6920-4-28. This article is available from <http://www.biomedcentral.com/1472-6920/4/28>.

- 24- Morgan PJ, Cleave-Hogg D, Guest CB. A comparison of global ratings and checklist scores from an undergraduate assessment using an anesthesia simulator. *Acad Med.* 2001; 76:1053-5.
- 25- Mattheos N, Nattestad A, Falk-Nilsson E, Attstrom R. The interactive assessment: assessing students' self-assessment ability. *Med Educ.* 2004; 38: 378-89.
- 26- Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J Pers Soc Psychol.* 1999. 77:1121-34.
- 27- Sullivan ME, Hitchcock MA, Dunnington MD, Dunnington GL. Peer and self-assessment during problem-based tutorials. *Am J Surg.* 1999; 117:266-9.
- 28- Norcini JJ. Peer assessment of competence. *Med Educ.* 2003; 37:539-43.
- 29- Shue CK, Arnold L, Stern DT. Maximizing participation in peer assessment of professionalism: the students speak. *Acad Med.* 2005; 80:51-5.
- 30- ACGME- Accreditation Council for Graduate Medical Education - Toolbox of assessment methods. Version 1.1, 2000
- 31- Woolliscroft, JO, TenHaken J., Smith J, Calhoun JG. Medical students' clinical self-assessments: comparison with external measures of performance and the students' self assessments of overall performance and effort. *Acad Med.* 1993; 68:285-94.
- 32- Morton JB, Macbeth WAAG. Correlations between staff, peer and self assessments of fourth-year students in surgery. *Med Educ.* 1977; 11:167-70.
- 33- Calhoun JG, Woolliscroft JO, Hockman, EM et al. Evaluating medical student clinical skill performance: relationships among self, peer and expert ratings *Res Med Educ.* 1984; 23:205-10.

- 34- Risucci DA, Tortolani AJ, Ward RJ. Ratings of surgical residents by self, supervisors and peers. *Surg Gynecol Obstet.* 1989; 169: 519-26.
- 35- Ward M, Gruppen L, Regehr G. Measuring self-assessment: current state of art. *Adv Health Sci Educ Theory Pract.* 2002; 7:63-80.
- 36- Fitzgerald JT, White CB, Gruppen LD. A longitudinal study of self-assessment accuracy. *Med Educ.* 2003; 37:645-9
- 37- Eva KW, Cunningham JPW, Reiter HI, Keane DR, Norman GR. How can I know what I don't know? Poor self assessment in a well-defined domain. *Adv Health Sci Educ Theory Pract* 2004; 9:211-24.
- 38- Moss AC, Dugal T, Silke B. Attitudes to peer review as a competence assurance structure – results of a survey of Irish physicians. *Ir J Med Sci.* 2005; 174:43-6.
- 39- Violato C, Lockyer J. Self and peer assessment of pediatricians, psychiatrists and medicine specialists: implications for self-directed learning. *Adv Health Sci Educ Theory Pract* 2006; 11:235-44.
- 40- Lurie SJ, Nofziger AC, Meldrum S, Mooney C, Epstein TM. Temporal and group-related trends in peer assessment amongst medical students. *Med Educ.* 2006; 40: 840-7.
- 41- Norcini J. Boulet J. Methodological issues in the use of standardized patients for assessment. *Teach Learn Med.* 2003; 15:293-7.
- 42- Henderson P, Johnson MH. An innovative approach to developing the reflective skills of medical students. *BMC Med Educ.* 2, 2002. [on-line] [acesso em mai. 2006]. Disponível em <http://www.biomedcentral.com/> 1472-6920/2/4.

Table 1. Median and ranges of values of GIR – faculty, self and peer ratings (n=96)

	Faculty Median (range)	Self Median (range)	Peer Median (range)
Mean GIR *	9.04 (7.64-9.95)	9.11 (7.15-10)	
Mean GIRt **	8.67 (6.50-9.94)	8.50 (6.67-10)	
Mean GIRh***	9.38 (7.71-10)	9.57(7.43-10)	9.21 (7.50-9.95)
Individual items			
1.Quality of history	8.75 (5.33-10)	9.00 (6-10)	
2.Physical examination	8.67 (7-10)	9.00 (7-10)	
3.Medical knowledge	8.33 (7-9.83)	8.00 (4-10)	
4.Clinical judgment	8.33 (5-9.83)	8.00 (5-10)	
5.Problem-solving	8.67 (5.67-10)	9.00 (6-10)	
6.Work habits	9.33 (7.33-10)	10 (7-10)	
7.Interpersonal and communication skills	9.33 (6.67-10)	10 (6-10)	9.25 (7-10)
8.Respect	9.67 (8-10)	10 (8-10)	9.50 (7.8-10)
9.Self-reflective skills	9.08 (6.67-10)	9.00 (7-10)	9.00 (6.5-10)
10.Compassion	9.33 (7.33-10)	10 (8-10)	9.33 (7.5-10)
11.Relationship with peers	9.33 (8-10)	10 (2-10)	9.25 (5.3-10)
12.Relationship with faculty	9.67 (8-10)	10 (7-10)	9.33 (7.5-10)
13.Relationship with staff/ other health professionals	9.08 (7.67-10)	9.00 (7-10)	9.00 (6.7-10)

* Mean of items 1 to 13

**Mean of items 1 to 6

*** Mean of items 7 to 13

Table 2. Pearson correlations between self, faculty and peer scores for GIR (n=96)

	Self vs. faculty	Peer vs. faculty	Peer vs. self
Mean GIR *	.12		
Mean GIRt **	.12		
Mean GIRh***	.11	.04	.13
Individual items			
1.Quality of history	.18		
2.Physical examination	.17		
3.Medical knowledge	.08		
4.Clinical judgment	.19		
5.Problem-solving	-.03		
6.Work habits	-.05		
7.Interpersonal and communication skills	-.03	.04	.05
8.Respect	-.02	-.02	.00
9.Self-reflective skills	.06	.19	.03
10.Compassion	.05	.07	.04
11.Relationship with peers	.16	.02	.48#
12.Relationship with faculty	.07	.06	.11
13.Relationship with staff/ other health professionals	.09	.14	.08

* Mean of items 1 to 13

**Mean of items 1 to 6

*** Mean of items 7 to 13

#Significant value under the two-sided binormal test of independence.

Table 3. Results of the Reliability analysis of the 13 domains for self (Cronbach's alpha .881) and faculty ratings (Cronbach's alpha .935), and of the 7 domains for peer ratings (Cronbach's alpha .904) (n=96)

Domain	Self			Faculty			Peer		
	r *	R2** (%)	AE #	r *	R2** (%)	AE #	r *	R2** (%)	AE #
1. Quality of history	.478	35.2	.877	.777	82.2	.928			
2. Physical examination	.662	54.1	.867	.752	80.6	.928			
3. Medical knowledge	.565	48.9	.873	.762	83.3	.928			
4. Clinical judgment	.634	53.8	.869	.794	83.3	.927			
5. Problem-solving	.651	57.5	.868	.829	81.2	.926			
6. Work habits	.484	31.0	.876	.602	43.5	.933			
7. Interpersonal and communication skills	.611	59.3	.870	.704	68.3	.930	.641	45.0	.898
8. Respect	.659	61.5	.872	.534	72.3	.935	.778	66.2	.887
9. Self-reflective skills	.636	45.6	.868	.739	73.0	.929	.790	69.3	.882
10. Compassion	.593	48.5	.871	.783	82.9	.927	.723	61.1	.890
11. Relationship with peers	.402	35.7	.884	.562	65.3	.934	.776	66.2	.888
12. Relationship with faculty	.551	48.6	.873	.610	67.7	.933	.683	56.4	.896
13. Relationship with staff/ other health professionals	.578	42.1	.872	.658	73.5	.932	.739	57.4	.888

*r= Pearson correlation between each domain and the sum of the remaining domains

** R2 = percentage of domain's score variation explained by the linear regression on the other domains' scores

Alpha obtained after domain elimination

Table 4. Euclidean distances between the GIRt and GIRh scores for different raters (n= 96)

Rater	Technical Skills		Rater	Humanistic skills		
	Self	Faculty		Self	Faculty	Peer
Self	0		Self	0		
Faculty	8.87	0	Faculty	6.95	0	
			Peer	7.82	7.05	0

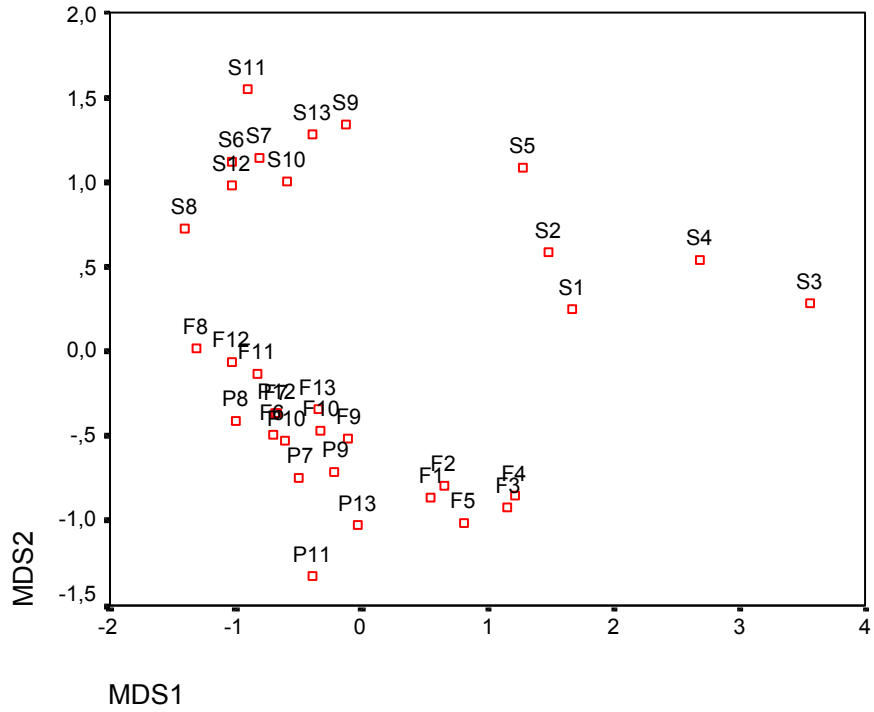


Figure 1. Map of the distances of the 33 domains, 13 related to self scores (S), 13 to faculty scores (F) and 7 to peer scores (P).

4.8. Artigo 8 – Peer assessment: how does it work in learning groups during clinical clerkship?

Artigo a ser encaminhado para Revista Advances in Health Sciences Education.
Fator de impacto = 1,065

Authors: Domingues RCL¹, Amaral E², Zeferino AMB³

1. PhD Researcher in Medical Education. O&G Department. State University of Campinas – Brazil
2. Associate professor, O&G Department - State University of Campinas – Brazil
3. Associate professor, Pediatrics Department. State University of Campinas – Brazil

Corresponding author:

Rosângela Curvo Leite Domingues

Rua Luciano Venere Decourt, 858- Cidade Universitária

Campinas SP- Brasil – 13.083-740

Fone para contato (19) 9207-6404 / (19) 3289-2182

e-mail: rosangela@cpdec.com.br

Peer assessment: how does it work in learning groups during clinical clerkship?

Abstract

Purpose: This study summarized how medical students assessed humanistic skills of their peers at the end of clinical clerkship. **Method:** In 2005, 106 Y4 students performed peer assessment in 6 health care units, by anonymously rating all his/her group peers (3-5, scale 0-10). Four groups were formed in each unit. Seven humanistic domains were assessed; opinions about the process were obtained. Statistical analysis used Student paired tests, ANOVA models, Cronbach's alpha and factorial analysis. Statistical significance was set at $P < 0.05$. **Results:** In average, each learner was assessed by 3 colleagues; 322 sets of domain scores were generated. There were 139 complete pairs of raters (in which A rated B and B rated A) - 75.5% of the total expected number. There were some reciprocal (11.5%) and constant ratings (19.8%). Median and mean domain scores were high and significantly different. Individual raters were the main component of data variance (64.5%). Cronbach's alpha was 0.946. There were significant moderate to strong Pearson correlations between group domain scores. One factor explained 74.5% of total variance of the data. 54% of learners felt *comfortable*, and 50% felt *prepared* to assess their colleagues; 46% considered peer assessment *valuable*. **Conclusion:** The observed number of reciprocal and constant ratings suggested a tendency of leniency and courtesy bias. Raters were the dominant component of data variation. Peer assessment was not seen as a valuable assessment strategy by most of the learners. The quality of assessment may be improved by additional raters' training.

Key words: assessment of humanistic domains, clinical assessment, learning groups, peer assessment.

Peer assessment: how does it work using learning groups during clinical clerkship?

Introduction

Medical education programs have been searching for reliable methods to assess multiple dimensions of competence, which encompass cognitive, technical, reflective, humanistic and affective domains (Danneffer et al., 2005; Epstein and Hundert, 2007). With the emphasis on promoting more active participation of students in the learning process, peer assessment has been suggested and implemented as an important strategy (Ramsey and Wenrich, 1999).

But, judgment of colleagues in health professions is not new; in fact, it has been used in the US since the 1980s and in Canada since the 1990s (Elwyn et al., 2005). Peer assessment has been increasingly explored in the literature, particularly during the clinical years, when students generally work in groups (Ramsey and Wenrich, 1999) and are able to observe one another regularly over a wide range of circumstances, where faculty may not be present (Ramsey et al., 1993; Ramsey et al., 1996; Arnold, 2002; Danneffer et al., 2005; Shue et al., 2005; Lurie et al., 2006; Epstein and Hundert, 2007). As members of learning groups, peers might be in a more favorable position to evaluate difficult-to-assess skills, and consequently add unique and valuable information to the assessment process, helping fellow students identify specific mutable behaviors and develop learning agendas (Danneffer et al., 2005; Epstein and Hundert, 2007).

Peer assessment is also a core component of group-based learning (Levine et al., 2007), since it highlights the value of respecting others' opinions, allows students to gain insight validating the judgments of colleagues, and provides faculty

with relevant information about individual personal and professional behavior and group interaction (Ramsey and Owen, 2006). When performed early in medical course, peer evaluation has great predictive value of later ratings, as well as positive correlations with other assessment methods, particularly when based on observed clinical work (Levine et al., 2007; Lurie et al., 2007).

However, peer evaluation is a complex accomplishment. There may be some kind of resistance to the process, because students may misunderstand the objectives of the exercise or believe they would not benefit from it, or still fear it would interfere in their relationship with fellow students (Levine et al., 2007).

The institute where the study was conducted offers a six-year vertically and horizontally integrated medical curriculum based on multidisciplinary teaching modules. During a Y4-module, medical students attend an introductory 200-hour clinical clerkship, under close faculty supervision. Their tasks are history taking, physical examination, medical record writing, and practicing skills in clinical reasoning and differential diagnosis. For educational purposes, learners work in self-selected groups of 4-5, in six different primary health care centers. Twenty-four learning groups are formed, four in each health care center.

This study is part of a larger comprehensive project on assessment of undergraduates and has the objective of describing results of student behaviors and perceptions in a novel peer evaluation process at a primary care clinical setting.

Method

As a new initiative in our medical course, peer assessment was performed in October 2005, at the end of their primary care clerkship. Each learner was asked to

anonymously assess all the peers in his/her learning group by completing a questionnaire composed of seven domains related to humanistic skills, using a 0-10 scale. The domains were selected considering current literature (Ramsey et al., 1993; Abim, 2001; Danneffer et al., 2005; Elwyn et al., 2005), after validation by a core panel of faculty members: *interpersonal and communication skills, respect* (for patients and family members), *self-reflective skills, compassion, relationship with peer, relationship with faculty, and relationship with staff/ other health professionals*. Each domain contained a brief descriptor for the extreme scores. There was also space for narrative comments. Before peer assessment process began, students received an explanation of the objectives, and were assured that their responses and comments would be kept confidential. Ethical approval for the study was provided by the School of Medical Sciences Institutional Review Board (581/2005).

Immediately after completing the peer evaluation forms, students also answered a questionnaire on their perceptions to the experience as raters. They were asked four questions: “*Did you feel comfortable assessing your working-group colleagues?*”; “*Do you think Y4 medical students are prepared to evaluate their colleagues in a responsible way?*”; “*Were you fair when you evaluated your colleagues?*”; “*Do you think this peer-assessment exercise is valuable?*”. For analytic purposes, the 0-10 intensity scale originally used was transformed into a *yes-no-I don't know* type of answer. The intensities of opinion from 0-3 in the original questionnaire were coded as *no*, from 4-6 as *I don't know*, and from 7 to 10 as *yes*.

Statistics of the domain scores (median, range, mean, standard deviation and 95% confidence interval) were obtained. A two-way ANOVA model and paired Student tests were used to compare means. Pearson correlation coefficients evaluated the

statistical associations and factor analysis was used to examine correlations (Kim and Muller, 1978). The reliability of the scale was assessed using Cronbach's alpha (Dunn, 1982). A three-factor nested random ANOVA model (Scheffé, 1959) was fitted to describe the variation sources of means of seven domain scores, incorporating *health care units*, *groups within units* and *raters within groups* as factors. Statistical significance was set at $P < 0.05$.

To analyze the data, we considered: (a) the score attributed by each rater to each of the seven domains of each ratee as *domain score* (b) the mean of the set of seven domain scores for each ratee-rater pair, hereafter referred to as an *event score* and (c) the mean of the domain scores, for each domain, across all the ratee-rater pairs in a group, hereafter referred to as a *group domain score*.

Extremely low scores were defined as values that lay more than 3 box lengths below the lower hinges of the seven boxplots of 322 sets of domain scores (Tukey, 1970).

We could aggregate the scores over one of the three possible factors: *health care units*, *groups*, or *raters*. Summing up over *health care units* would condense the data too strongly; adding over *raters* would ignore the fact that raters were also ratees. Thus, the best alternative was to sum up the domain scores over *groups*, obtaining 24 sets of seven group domain scores.

Results

A total of 106 students were expected to participate as raters, but 10 students did not rate any peer. So the present study will present the results of 96 effective raters. There was an average number of raters/ratees per group of 3.04. 322 event

scores, 322 sets of domain scores, 2254 domain scores (322 x 7 domains) and 24 sets of group domain scores were generated. Table I shows the characteristics of the scoring provided by raters when assessing their colleagues.

Table II shows the statistics for event scores and group domain scores. As group domain scores were calculated as averages, their dispersion measures (ranges and SD) were reduced.

The seven means of group domain scores exhibited significant differences (ANOVA, $p < .0001$). The highest mean was attributed to *Respect*, significantly different from all other mean scores, except for *Relationship with faculty* (paired Student tests, $p = 0.001$). *Relationship with faculty* mean score was significantly higher than the means of *Self-reflective skills* and *Relationship with staff* scores ($p = 0.001$ and 0.002 , respectively).

Direct reliability measures for event scores could not be obtained for learning group level because the *within ratee variance* of event scores was higher than their *among ratee variance* (Dunn, 1982) in 20 out of 24 groups, suggesting an expressive variation due to raters in most groups. The dominating component of variance of the 322 event scores was the *raters within groups* variance (64.5% of total variance, $df = 71$). Both the *health care units* variance ($df = 5$) and the *groups within units* variance ($df = 18$), showed reduced contributions to the total variance (1.8% and 2.5%), leaving the remaining 31.2% for the residual variance, composed by variation sources like ratees and rater-ratee interactions.

From now on group domain scores will be used to present our results. The reliability analysis of group domain scores (Table III) demonstrated that the *Relationship with faculty* domain score was the least associated with the sum of the

others, and also one of the least explained by the remaining group domain scores. There were small reductions of Cronbach's alpha after elimination of each domain.

Table IV shows that there were significant moderate to strong Pearson correlations between group domain scores. The highest correlation was found between *Relationship with peers* and *Respect* ($r=0.85$) and the lowest between *Relationship with faculty* and *Respect* ($r=0.52$).

To describe the correlation patterns of Table IV we used a unifactorial model, validated by the maximum likelihood method ($\chi^2=14.08$, $df=14$, $P=0.44$), which explained 74.5% of total variance (Data not shown)

The lowest loading and communality of the *Relationship with faculty* domain in Table V show its small correlation with the factor and its weak status, also detected by the reliability analysis in Table III.

Student perception of peer assessment process

Questionnaires about learners' opinions on the peer assessment process were completed by 96 students, out of 106. About half of the learners expressed they felt *comfortable and prepared* assessing their colleagues; approximately 94% answered they were *fair* while evaluating their peers, and 46% agreed that the exercise of peer assessment was *valuable* (Table VI).

Fifty three students justified their answers or wrote any comments (50%). The students who expressed their opinion reported feelings such as: "I didn't feel prepared", "I don't like to assess other people", "I feel that friendship may influence assessment"; "I tried to be fair, but I might have been influenced by the relationship I have with my colleagues"; "I don't understand the meaning of peer assessment".

Discussion

The goal of this study was to gather pertinent data concerning how students performed peer assessment in a clinical setting where students have a one-year lasting primary care experience. Our results suggest that peer assessment can be a useful evaluation tool because it allowed us to understand and quantify the group dynamics. The productivity and the characteristics of the learning groups, such as lenience and very low scores, as well as the reliability and correlations between domains helped us study the behaviors of the groups.

Rater behavior as shown in Table 1 had an important effect on assessment outcomes. Rater training and experience in assessment, as well as rater-ratee contact and contextual factors may increase reliability.

The number of peer ratings recommended to be reliable (between 6 to 15) was not achieved (Ramsey et al., 1993; Thomas et al., 1999; Epstein and Hundert, 2002; Daelmans et al., 2005; Lurie et al., 2006). Nevertheless, Elwyn et al. (2005) showed that although median number of ratings per ratee was 12, on average, only three of them were “true peers” (fellow doctors working in the same local setting, for about one year), and the remaining nine were considered secondary care colleagues. In our setting, group selection and size were determined by logistic and educational issues, as the number of consultation rooms available per health care unit, number of patients per day, and number of faculty supervisors per unit, but all peers can be considered “true peers”.

Domain and event scores were skewed to the right end of the scale. Our findings coincide with other studies, in which peers provided uniformly high ratings

(Ramsey et al., 1996; Govaerts et al., 2007). There are some possible explanations for this finding. First, peer evaluation may have been mainly based on observation of best behaviors. Second, there may also have been some *leniency* effect, probably due to the fact that, while forming the groups students were more likely to assign themselves to classmates whom they liked. So they might have felt reluctant to attribute low ratings to those they relate on a daily basis (Davis, 2002; Lurie et al., 2006). *Courtesy bias*, a code of behavior people try to maintain in order to keep the atmosphere pleasant, agreeable, free from conflict or contradiction (Gagliete and Barbosa, 2006) could have influenced raters' behaviors.

According to Norcini (2003), there seems to be no way to completely avoid this issue of high ratings, but it may help to ensure previous training and the anonymity of evaluators when possible. In the present study, all evaluations were anonymous, in an attempt to protect both the rater and the ratee, and to encourage genuine student participation. However, in a study conducted by (Arnold et al., 2005), some students predicted that peers would not take anonymous evaluations as seriously as those that were signed.

Conversely, there were some very low scores. Although learners were not asked to compare themselves with their peers, comparative judgment may have occurred and the self was probably used as a basis for trying to understand and rate others (Krueger and Dunning, 1999). According to Moore and Small (2007), people attribute low grades to their colleagues because they routinely overestimate themselves relative to others, particularly in those domains in which they feel very capable. Very low peer assessment scores may also have been a consequence of the *hallo effect*, in which early impressions of a subject may influence subsequent

evaluations by a rater (Arnold, 2002; Davis, 2002; Evans et al., 2004; Mckinstry et al., 2004). An individual who is known to have performed poorly is judged to be worse than others, whereas an individual who is known to have performed well is judged better than others. The effects of over-or underestimating the performances of others can be attenuated if people obtain accurate information about the person being judged (Moore and Small, 2007).

In the studied sample, there were 24.5% of non observed complete pairs of raters, 11.5% of reciprocity and 19.8% of constancy in rating pattern. These results may be due to some characteristics of the peer assessment system itself and the environment in which it occurred (Arnold et al., 2005). Learners may not have felt motivated to assess peers; they may have been concerned about who would receive this information, its use and consequences. Also, as they are beginning their clinical practice and are not used to this kind of exercise, they may not have had sufficient grounds on which to make judgments about the humanistic domains of competence (Elwyn et al., 2005). Finally, as the evaluation process was held at the end of the day and took approximately one hour, learners may have felt it was a long and tiring process. This paper leaves for further analysis the difficulties encountered in the application and interpretation of a peer assessment system.

Among the variation sources, i.e., *health care units, groups within units and raters within groups*, individual raters accounted for the dominant variation component. Raters play a central role in assessment; they are active information processors and, their behaviors, motivations, judgment and decision making processes may compromise assessment results (Govaerts et al., 2007). So, improving the quality of the raters' preparation is likely to produce a great impact on the reliability of ratings

(Norcini, 2003). Accuracy of rating can be achieved when raters are prepared to apply both analytical (related to the recall of factual details) and non-analytical strategies (related to global, holistic impressions) while rating (Govaerts et al., 2007).

Although contextual factors such as *health care units* and *groups within unit* showed reduced contributions to the total variance of the scores in the present study, there is increasing recognition of the role of context and educational environment in the quality of ratings (Arnold et al., 2005). The interaction between the learning context and the peer evaluation activity is an essential element toward reinforcing the core goals of module (Levine et al., 2007). Similarly, rater-ratee interactions and ratees' actual competencies are regarded as important sources of score variation, since they may influence raters' perceptions of the relative importance of the domains being assessed (Govaerts et al., 2007).

The domains included in the peer assessment form were coherent, as shown by the high reliability (internal consistency) and by the unifactorial structure of the instrument. *Respect* scores were the highest and its mean was significantly different from the majority of the other domain peer scores means. Additionally, *Respect* group scores showed to be more associated with the other group domain scores. This may be partly explained by the fact that the two initial teaching modules (Y1 and Y2) of our medical curriculum focus on the early integration of alumni to the community, encouraging students to perceive the basic needs of the patients, family members and accompanying persons.

Conversely, *Relationship with faculty* domain was the least associated with the other domains, increasing Cronbach's alpha by its elimination. These findings are in line with the factor analytic results, which showed an outlying behavior for this

domain. These data suggest that *Relationship with faculty* could be eliminated from the peer rating questionnaire with no substantial loss.

Most learners did not regard peer assessment as a valuable exercise. These results are somehow different to the findings of other studies, in which peer assessment was reported as an important evaluation method both by educators, who have been using it as one component of clinical clerkship grades, and by the students themselves, who have reported acceptance and perceived value of the process (Rudy et al., 2001; Danneffer et al., 2005). But, our findings coincide with the results presented by Levine et al. (2007), whose study showed that students disliked peer evaluation process. Some contextual factors may have led our students to underestimate the value of peer assessment. Assessment purposes and the use of assessment results may not have been clear enough to learners. The “colleague/friend factor” may have caused some resistance to the process and influenced assessors’ perception of the nature of peer ratings (Ramsey and Wenrich, 1999). Finally, since learners were not used to assessing fellow students, they may not have trusted the appraisal process or believed that it would not bring any positive consequences (Govaerts et al., 2007). Trust in and acceptance of the assessment system is a crucial factor, accounting for more than 30% of the variance in ratings, with raters who feel a high degree of trust providing less lenient ratings (van Bernardin et al., 1981)

Although almost all learners considered themselves fair while assessing their peers, only half of the students felt comfortable when doing so. Students may either felt somehow unease to rate their colleagues, or feared that the process would undermine the group function (Arnold, 2002; Norcini, 2003). A clear

explanation of the underlying assessment theory would be useful to help students feel more comfortable and prepared to perform peer assessment.

This study has other limitations. First, the results presented referred to a single class at a single time point, so possible effects of group culture could be present. Second, because groups were not formed randomly, it is likely that the nature of group selection has contributed to the observed ceiling effect (Arnold, 2002; Evans et al., 2004; Mckinstry et al., 2004; Lurie et al., 2006). A final drawback to the study was the lack of an intense preparation of assessors for the experience, including a more detailed instruction on how to interpret each of the seven domains of the questionnaire (Thomas et al., 1999). Norcini (2003) states that learners should be offered some kind of training, which could range from simple oral descriptions of what is expected to intense videotape-based sessions. Additionally, he suggests that the purpose of peer assessment be stated in writing and assessment criteria be clearly communicated to participants. However, we understand that peer assessment should be introduced gradually in our institution, so the first step has already been taken.

In summary, peer assessment showed to be a promising and reliable assessment strategy when true peers are involved. It can be a valuable complement to other assessment methods. However, caution must be taken in terms of clarity of the process purpose, interpretation of the domains included in the instrument, their implication as competencies of professionalism, and the ability of the students to assess. Rater motivation and the impact of context on raters' judgment and decision making processes must be taken into consideration by medical educators when planning peer assessment tools. We suggest that educators recognize the

central active role of raters in the evaluation process and the need of additional rater training in performing peer assessment and in offering supportive feedback. All these aspects should be included in program planning; after all, institutional support for peer assessment is crucial to its success. As the use of peer ratings becomes more common in undergraduate medical education, it is likely that learners will become more accepting of them and will be able to profit from them along their professional career. Further research on peer evaluation in team-based learning is warranted.

References

ABIM - American Board of Internal Medicine. Attending Physicians Competency Card. Available at [www.abim.org]. Accessed April 15, 2008

Arnold, L, Shue, C.K., Kritt, B., Ginsburg, S., Stern D.T. (2005). Medical students' views on peer assessment of professionalism. *Journal of General Internal Medicine* **20**: 819-824

Arnold, L. (2002) Assessing professional behavior: yesterday, today and tomorrow. *Academic Medicine* **77**: 502-515.

Daelmans, H.E.M., van der Hem-Stokroos, H.H., Hoogenboom, R.J.I., Scherpbier, A.J.J.A, Stehouwer, C.D.A, van der Vleuten, C.P.M. (2005). Global clinical performance ratings, reliability and validity in and undergraduate clerkship. *The Netherlands Journal of Medicine* **43**: 279-284.

Danneffer, E.F., Henson, L.C., Bierer, S.B., Grady-Weliky, T.A., Meldrum, S, Nofziger, S.C., Barclay, C., Epstein, R.M. (2005). Peer assessment of professional competence. *Medical Education* **39**: 713-722.

- Davis, J.D. (2002). Comparison of faculty, peer, self and nurse assessment of obstetrics and gynecology residents. *Obstetrics & Gynecology* **99**: 746-751
- Dunn, G. (1982). Design and Analysis of reliability studies. Edward Arnold: London, Chapter 3.
- Elwyn, G., Lewis, M., Evans, R., Hutchings, H. (2005). Using a “peer assessment questionnaire” in primary medical care. *British Journal of General Practice* **55**: 690-695.
- Epstein, R.M. & Hundert, E.M. (2007). Defining and assessing professional competence. *JAMA*; **356**: 387-396.
- Evans, R., Elwyn, G., Edwards, A. (2004). Review of instruments for peer assessment of physicians. *British Medical Journal*; 328, 1240. doi:10.1136 /bmj.328.7450.1240
- Gagliete, M., Barbosa, M.L.S. (2006). A Brasilidade no entre-lugar: leituras de Dyonélio Machado e Sérgio Buarque de Holanda. *Letras de Hoje* **41**:29-36
- Govaerts, M.J.B., van der Veluten, C.P.M., Schuwirth, L.W.T., Muijtjens, A.M.M. (2007). Broadening perspectives on clinical performance assessment: rethinking the nature of in-training assessment. *Advances in Health Sciences Education* **12**: 239-60.
- Kim, J.O., Mueller, C.W. (1978). Factor Analysis. Statistical Methods and Practical Issues, Sage: Beverly Hills.
- Kruger, J., Dunning, D.(1999). Unskilled and unaware of it: how difficulties in recognizing one’s own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology* **77**: 1211-1234.
- Levine, R.E., Kelly, P.A., karakoc, T., Haidet, P. (2007). Peer evaluation in a clinical clerkship: students’ attitudes, experiences, and correlations with traditional assessments. *Academic Psychiatry* **31**: 19-24.

Lurie, S.J., Nofziger, A.C., Meldrum, S., Mooney, C., Epstein, T.M. (2006). Temporal and group-related trends in peer assessment amongst medical students. *Medical Education* **40**: 840-847.

Lurie, S.J., Lambert, D.R., Nofziger, A.C., Epstein, R.M. Grady-Weliky T.A. (2007). Relationship between peer assessment during medical school, dean's letter rankings, and ratings by internship directors. *Society of General Internal Medicine*; **22**: 13-16.

McKinstry, B.H., Cameron, H.S., Elton, R.A., Riley, S.C. (2004). Leniency and halo effects in marking undergraduate short research projects. *BMC Medical Education*; 4:28 doi:10.1186/1472-6920-4-28. Available from [<http://www.biomedcentral.com/1472-6920/4/28>]. Accessed April 15, 2008

Moore, D.A., Small, D.A. (2007) Error and bias in comparative judgment: on being both better and worse than we think we are. *Journal of Personality and Social Psychology* **92**: 972-989

Norcini, J.J. (2003). Peer assessment of competence. *Medical Education* **37**: 539-543.

Ramsey, P.G., Wenrich, M. D., Carline, J.D., Inui, T.S., Larson, E.B., LoGerfo, J.P. (1993). Use of peer ratings to evaluate physician performance. *JAMA* **269**: 1655-1660.

Ramsey, P.G., Carline, J.D., Blank, L.L., Wenrich, M.D. (1996). Feasibility of hospital-based use of peer ratings to evaluate the performances of practicing physicians. *Academic Medicine* **71**: 364-370.

Ramsey, P.G., Wenrich, M.D. (1999). Peer Ratings: an assessment tool whose time has come. *Journal of General Internal Medicine* **14**: 581-582

Ramsey, W., Owen, C. (2006). Is there a role for peer review in performance appraisal of medical students? *Medical Education* **40**: 95-96.

Rudy, D.W., Fejfar, M.C., Griffith, C.H., Wilson, J.F. (2001). Self- and peer assessment in a first-year communication and interviewing course. *Evaluations & the Health Professions* **4**: 436-443.

Scheffé, H. *The Analysis of Variance*, Wiley: New York (1959). Chapters 5 and 7.

Shue, C.K., Arnold, L., Stern, D.T. (2005). Maximizing participation in peer assessment of professionalism: the students speak. *Academic Medicine* **80**: 51-55.

Thomas, P.A., Gebo, A.K., Hellmann, D.B. (1999). A pilot study of peer review in residency training. *Journal of General internal Medicine* **4**: 551-554

Tukey, J.W. (1970). *Exploratory Data Analysis*, Addison-Wesley: Reading MA.

van Bernardin, C., Orban, J.A., Carlyle, J.J. (1981) Performance ratings as a function of trust in appraisal and rater individual differences. *Academy of Management Proceeding* 311-315.

Table I– Results of peer scoring

Characteristic	Number	%
1. Raters who assessed all their group-peers*	87	90.6
2. Complete pairs of raters (in which A rated B and B rated A) **	139	75.5
3. Reciprocal sets of domain scores (complete pairs in which A attributed to B exactly the same 7 ratings B attributed to A)	16	11.5
4. Reciprocal sets of domain scores composed only by the maximum score 10	14	10.1
5. Sets of domain scores equal 10 (maximum score)	89	27.6
6. Extremely low domain scores	42	1.9
7. Raters with constant rating pattern on all domains*	19	19.8
8. Event scores across all raters equal 10 (maximum score)	0	0

* Number of effective raters = 96

**The maximum expected number of complete pairs was 184.

Table II. Statistics for the event scores (N=322) and group domain scores (N=24)

Scores	Median	Min-Max	Mean	SD*	CI **	
					Inf	Sup
Event scores	9.571	5.29 –10	9.174	0.912	9.074	9.273
Group domain scores						
Interpersonal and communication skills	9.333	7.89 –10	9.186	0.482	8.982	9.389
Respect	9.500	8.33 –10	9.406	0.422	9.228	9.585
Self-reflective skills	9.211	8.12 –9.88	9.095	0.481	8.891	9.298
Compassion	9.139	8.11 –9.92	9.204	0.462	9.009	9.399
Relationship with peers	9.189	7.74 –10	9.047	0.641	8.776	9.317
Relationship with faculty	9.355	8.60 –9.92	9.326	0.341	9.182	9.470
Relationship with staff/ other health professionals	9.175	7.67 –9.78	8.981	0.609	8.724	9.238

* SD= standard deviation.

** Inf and Sup are the extremes of the 95% confidence interval CI for the population mean.

Table III. Reliability analysis of group domain scores (N = 24, Cronbach's alpha=0.946)

Domain	r *	R² (%) **	AE #
1. Interpersonal and communication skills	0.853	75.3	0.935
2. Respect	0.886	83.8	0.934
3. Self-reflective skills	0.866	82.6	0.934
4. Compassion	0.866	80.4	0.935
5. Relationship with peers	0.870	82.3	0.937
6. Relationship with faculty	0.731	68.7	0.948
7. Relationship with staff/ other health professionals	0.798	68.5	0.943

*Pearson correlation between each domain score and the sum of the remaining domain scores.

** Percentage R2 of the domain score variation explained by the linear regression on the other domain scores.

Alpha obtained after domain elimination.

Table IV. Correlation coefficients between group-domain scores (N=24) (*)

Domain	1	2	3	4	5	6
1. Interpersonal and communication skills						
2. Respect	.80					
3. Self-reflective skills	.78	.77				
4. Compassion	.81	.84	.77			
5. Relationship with peers	.80	.85	.83	.74		
6. Relationship with faculty	.64	.58	.78	.71	.62	
7. Relationship with staff/ other health professionals	.69	.78	.67	.75	.75	.63

(*) All correlations are significant under the two-sided binormal test of independence.

Table V. Loadings and communalities for the one-factor model for the group domain scores (N = 24)

Domain	Loading	Communality
1. Interpersonal and communication skills	0.883	0.779
2. Respect	0.912	0.832
3. Self-reflective skills	0.884	0.782
4. Compassion	0.890	0.792
5. Relationship with peers	0.900	0.810
6. Relationship with faculty	0.741	0.549
7. Relationship with staff/ other health professionals	0.820	0.673

Table VI – Learners' opinions on peer assessment

Questions	n	% of affirmative answers	CI*
1. Did you feel comfortable assessing your working-group colleagues?	96	54.2	46.8 to 61.5
2. Do you think Y4 medical students are prepared to evaluate their colleagues in a responsible way?	96	50.0	42.9 to 57.1
3. Were you fair when you evaluated your colleagues?	96	93.8	89.1 to 98.4
4. Do you think this peer-assessment exercise is valuable?	95	46.3	39.5 to 53.1

* Extremes of the 95% confidence interval CI for the population percentage.

4.9. Artigo 9 – Os diferentes olhares na avaliação de alunos em estágio clínico supervisionado



Revista da Associação Médica Brasileira

ASSOCIAÇÃO MÉDICA BRASILEIRA
Rua São Carlos do Pinhal, 324 CEP: 01333-903 - Caixa Postal:
8904 - São Paulo SP - Brasil
Tel.: (11) 3178-6800 - Email: ramb@amb.org.br

São Paulo, domingo, 31 de agosto de 2008

Ilmo(a) Sr.(a)
Prof.(a), Dr(a) Rosângela Curvo Leite Domingues

Referente ao código de fluxo: 726

Classificação: Artigos Originais

Informamos que recebemos o manuscrito Os diferentes olhares na avaliação de alunos em estágio clínico supervisionado será enviado para apreciação dos revisores para possível publicação/participação na(o) Revista da Associação Médica Brasileira. Por favor, para qualquer comunicação futura sobre o referido manuscrito cite o número de referência apresentado acima.

Obrigado por submeter seu trabalho a(o) Revista da Associação Médica Brasileira.

Atenciosamente,

Bruno Caramelli
Editor

Os diferentes olhares na avaliação de alunos em estágio clínico supervisionado

Autores: Rosângela Curvo Leite Domingues, Eliana Amaral, Angélica Maria Bicudo Zeferino - NAPEM/Faculdade de Ciências Médicas – Universidade Estadual de Campinas – UNICAMP – Endereço: Rua Alexander Fleming, 101 – Cidade Universitária – Campinas - São Paulo – SP - 13083-881

Carta de Apresentação:

Rosângela Curvo Leite Domingues, Angélica Maria Bicudo Zeferino e Eliana Amaral participaram da elaboração deste trabalho, tanto na concepção e desenho deste estudo, na análise e interpretação dos dados, assim como na redação deste texto.

Todos os autores fizeram a revisão final deste artigo e o aprovaram para ser encaminhado à publicação. O presente trabalho nem outro com conteúdo substancialmente semelhante de minha autoria foram publicados ou submetidos à aprovação do Conselho Editorial de outra revista. Não há nenhum tipo de conflito de interesse de nenhum dos autores. Esse trabalho foi aprovado pelo Comitê de Ética e Pesquisa da Faculdade de Medicina da Unicamp, sob número 581/2005

Autor para correspondência:

Rosângela Curvo Leite Domingues

Rua Luciano Venere Decourt, 858- Cidade Universitária

Campinas SP- Brasil – 13.083-740

Fone para contato (19) 9207-6404 / (19) 3289-2182

e-mail: rosangela@cpdec.com.br

Os diferentes olhares na avaliação de alunos em estágio clínico supervisionado

Resumo:

Objetivo: comparar notas do conceito global itemizado e da avaliação estruturada, atribuídas por docentes de três especialidades (ginecologia-GO, clínica médica-CM e pediatria-Ped) a alunos em final de estágio clínico. **Metodologia:** 106 alunos do 4º ano foram avaliados por um docente de cada especialidade, por meio do Conceito Global Itemizado (CGI), que compreendia 6 habilidades técnicas (média CGIt) e 7 habilidades humanísticas (média CGIh), escala 0-10, e da Avaliação Estruturada – AE. A análise estatística usou coeficientes de Cronbach, testes de Friedman e pareados de Wilcoxon, coeficientes de correlação de Spearman e Pearson e distâncias Euclidianas. Nível de significância=5%. **Resultados:** Consistência interna >0,92, nos três CGI. As distribuições das notas concentraram-se no extremo superior da escala. As notas de Ped foram significativamente inferiores às de GO e CM. As correlações entre as notas das três especialidades foram baixas. Na AE, houve diferenças significativas entre todas as especialidades. No CGI, houve diferenças significativas em 6 dos 13 domínios: em 5 domínios entre GO e Ped, em 1 deles entre GO e CM, em todos eles entre CM e Ped. Não houve consenso entre as especialidades sobre os impactos relativos dos itens do CGI. As notas de GO e Ped mostraram a menor concordância; as de GIRT de Ped foram as que mais discordaram das outras notas. **Conclusão:** Cada especialidade tem uma visão distinta sobre como avaliar alunos; Ped possui uma visão particular; GO e CM possuem percepções semelhantes. Essas diferenças estão vinculadas à “cultura” de cada especialidade. Permanece o desafio de minimizá-las, em um módulo de ensino integrado.

Palavras-chave: avaliação de alunos, especialidades médicas, estágio clínico.

The different perspectives in the assessment of learners during supervised clinical clerkship

Abstract:

Objective: To compare medical students' global itemized ratings and real-case structured clinical assessment, generated by faculty from three different clinical specialties (Gynecology-O&G, Internal Medicine-IM, Pediatrics-Ped). **Method:** 106 Y4 learners were assessed by one faculty member from each specialty, who filled in a Global Itemized Rating form (GIR), consisting of 6 technical domains (mean score GIRt) and 7 humanistic domains (mean score GIRh) and a real-case structured clinical assessment (RC-SCA), 0-10 scale. Statistical analyses used Cronbach coefficient, Friedman and Wilcoxon paired tests, Pearson and Spearman correlation coefficients, and Euclidean distances. Significance level=5%. **Results:** Internal consistency > 0.92, in the three GIR. Rates given by teachers were negatively skewed. Ped scores were significantly lower than O&G and IM. There were low correlations among the three specialties' scores. In RC-SCA there were significant differences between the 3 specialties. In GIR, there were significant differences in 6 out of the 13 domains: in 5 domains between O&G and Ped, in 1 domain between O&G and IM, in the six domains between IM and Ped. No consensus was reached among specialties in relation to the relative impact of the 13 GIR domains. In all methods, O&G and Ped scores showed the smallest relative agreement. GIRt Ped scores showed the greatest disagreement with all the other scores. **Conclusion:** Each specialty may have different views on how to value students' skills. Ped has a unique perspective; O&G and IM have similar views. These differences may be due to each specialty "culture". The challenge remains to minimize them in an integrated clinical module.

Key-words: learners' assessment, medical specialties, clinical clerkship

Introdução

Desenvolver competências técnicas e humanísticas para o exercício da profissão tem se tornado um aspecto central da educação médica, para a qual o estágio clínico supervisionado assume um papel relevante.¹ A necessidade de preparar profissionais críticos e reflexivos, capazes de atuar em equipes multidisciplinares, em diversos cenários da prática e aptos a responderem às demandas da sociedade pressupõe grandes transformações educacionais. Dentre elas, destaca-se a capacitação dos docentes/supervisores para garantir maior qualidade no estabelecimento de critérios de desempenho, nas observações da atuação dos alunos, na prática de *feedback* e no processo de avaliação das competências esperadas.^{2,3}

Segundo a Associação Européia de Educação Médica (AMEE), o termo supervisão é definido como o fornecimento de orientações e *feedback*, relacionados ao desenvolvimento profissional e educacional do aluno, para ajudá-lo a prover um atendimento médico adequado.⁴ A avaliação é um poderoso componente da supervisão eficaz, pois direciona a aprendizagem, serve como guia para os estudantes em relação às competências que devem ser aprimoradas e auxilia os supervisores a identificarem oportunidades de melhoria da qualidade do ensino.^{5,6}

Para que a avaliação do aluno em estágio clínico tenha um forte impacto educacional, é importante que haja entendimento dos objetivos a serem atingidos, consenso sobre conteúdo a ser avaliado e conhecimento do nível de competência dos alunos.⁴

No entanto, a definição de quais domínios da competência os alunos devem atingir, nas várias etapas de sua formação, pode não ser clara ou ainda não ser

igualmente compreendida por supervisores de diferentes estágios clínicos. Essas diferenças tendem a se acentuar quando se contrasta a supervisão oferecida por especialistas, que geralmente possuem diferentes percepções de seu papel docente e da importância relativa de sua especialidade na formação do médico.^{4, 7}

O objetivo do presente estudo foi verificar como os docentes/supervisores de diferentes especialidades avaliam os alunos em estágio clínico, em um módulo de ensino integrado.

Método

Na instituição onde o estudo foi realizado, os alunos do quarto ano iniciam a prática clínica durante o Módulo *Atenção Integral à Saúde*, que visa propiciar experiência fora do hospital de referência, contribuindo para o desenvolvimento de responsabilidade e o vínculo com o paciente e com outros profissionais da equipe de saúde. Os estudantes são divididos em grupos e alocados em seis Unidades Básicas de Saúde (UBS) do SUS, onde realizam atendimentos de Ginecologia e Obstetrícia (GO), Clínica Médica (CM) e Pediatria (Ped), sob supervisão de docentes e tutores da própria unidade de saúde, durante nove meses. Os estudantes realizam anamnese e exame físico, indicam condutas diagnósticas e terapêuticas para as situações mais comuns, registram o atendimento, discutem o caso clínico com docentes, tutores e membros da equipe de saúde e participam de reuniões de projeto terapêutico.

Os alunos são avaliados por meio de um Conceito Global Itemizado (CGI) composto de 13 domínios, 6 referentes à competência técnica (denominado a partir de agora como CGI_t) - *qualidade da história, exame clínico, conhecimento*

médico, julgamento clínico, solução de problemas, hábitos de trabalho) - e 7 à competência humanística (denominado a partir de agora como CGIh) - *comunicação e relacionamento com pacientes e familiares, respeito, capacidade de auto-reflexão, percepção do contexto, interação com colegas, interação com docentes, interação com demais profissionais*. As notas atribuídas a cada domínio variavam de 0 a 10, com descritores ancorando o menor e o maior valor.

Adicionalmente, os alunos são submetidos à Avaliação Estruturada (denominada a partir de agora como AE), que tem como base a observação docente de uma consulta completa, em cada uma das especialidades (GO, Ped, CM). O desempenho dos alunos é registrado em uma lista de verificação detalhada, que contém questões relacionadas à *abordagem inicial da consulta, anamnese, exame físico, preenchimento dos formulários e orientações ao paciente*, em número e conteúdo, adaptados para cada especialidade, sendo que 62,9% (Ped), 62,5% (GO) e 54,5% (CM) das questões se relacionam a domínios técnicos da competência.

O presente estudo comparou as notas provenientes de CGI (média ponderada de CGI_t e CGI_h), CGI_t (média dos domínios 1 a 6), CGI_h (média dos domínios 7 a 13) e AE (média das notas de todas as questões do formulário de desempenho), atribuídas por docentes das três especialidades clínicas aos 106 alunos do 4º ano de Medicina, ao final do Módulo *Atenção Integral à Saúde*, em outubro de 2005. A análise estatística incluiu medianas e distâncias interquartis para descrição das notas dos alunos nas quatro estratégias de avaliação. Para investigar as diferenças nas quatro distribuições foram aplicados os testes de Friedman e as comparações múltiplas com testes de Wilcoxon (W) para dados emparelhados, com correção de Bonferroni. As correlações entre as notas das três especialidades foram estudadas

por meio de coeficientes de correlação de Spearman. O coeficiente alfa de Cronbach foi calculado para investigar a confiabilidade dos questionários de CGI em cada uma das três especialidades. O coeficiente de correlação de Pearson foi usado para a análise de confiabilidade entre os 13 domínios do CGI para cada especialidade. A concordância relativa entre as notas foi calculada recorrendo a distâncias euclidianas. O nível de significância escolhido foi de 5%. O estudo foi aprovado pelo Comitê de Ética em Pesquisa, sob o número 581/2005

Resultados

Dos 106 alunos participantes, 3 tiveram alguma nota incompleta. Por isso, foram analisados resultados de 103 alunos, avaliados por 19 docentes, de forma independente. A confiabilidade do questionário foi elevada em todas as especialidades e está representada pelos coeficientes de Cronbach, 0,933 para GO, 0,939 para CM e 0,916 para Ped.

A **Tabela 1** mostra que as distribuições das quatro notas estão concentradas no extremo direito da escala, com pouca variação entre elas, nas três especialidades. O teste não paramétrico de Friedman detectou diferenças significativas entre as distribuições das notas das três especialidades, em todas as quatro notas.

Tabela 1 - Medianas, distâncias interquartis e comparação das distribuições das notas entre as especialidades* (N= 103)

Estratégias de avaliação	GO Mediana	GO Distância inter quartis	CM Mediana	CM Distância Inter quartis	Ped Mediana	Ped Distância Inter quartis	F*	P**
CGI	9,38	8,77-9,69	9,31	8,85-9,73	8,77	8,23-9,23	13,96	0,001
CGIt	9,00	8,50-9,50	8,83	8,33-9,50	8,33	7,75-9,00	10,14	0,006
CGIh	9,71	8,86-10	9,79	9,14-10	9,14	8,57-9,57	19,65	<0,001
AE	9,60	9,20-10	9,40	9,00-9,75	9,00	8,25-9,50	36,35	<0,001

* F=valor da estatística qui-quadrado do teste de Friedman, GL=2.

** Nível de significância alcançado no teste de Friedman.

Observou-se diferença significativa entre as distribuições das notas de Ped e CM (W=5,16 para CGI, W=4,04 para CGIt, W=5,34 para CGIh e W=4,42 para AE, todas com $P < 0,001$) e entre as notas de Ped e GO (W=4,08 para CGI, W=3,95 para CGIt, W=3,63, para CGIh e W=6,24 para AE, todas com $P < 0,001$). Portanto, as notas de Ped foram significativamente inferiores às de GO e CM em CGI, CGIt, CGIh e AE. As diferenças entre GO e CM não foram significativas nem para CGI (W=0,09, $P=0,927$) nem para CGIt (W=0,85, $P=0,395$), nem para CGIh (W=0,77, $P=0,443$), mas sim para AE (W=3,10, $P=0,002$).

As correlações entre as notas das três especialidades variaram de muito baixas a moderadas ($-0.02 < r < 0.48$), nas quatro estratégias de avaliação. Houve correlação significativa, mas negativa ($r = -0,23$) entre as notas da GO e Ped no GIRh (**Tabela 2**).

Tabela 2- Correlações de Spearman entre as notas dos docentes de GO, CM e Ped, para cada estratégia de avaliação (N= 103)

Especialidade	CGI		CGIt		CGIh		AE	
	GO	CM	GO	CM	GO	CM	GO	CM
CM	0,19		0,24 [*]		0,11		-0,03	
Ped	-0,02	0,48 ^{**}	0,21 [*]	0,48 ^{**}	-0,23 [*]	0,33 ^{**}	0,19	0,20 [*]

* Significativa para $P = 0,05$ - para o teste bilateral de correlação populacional de Spearman nula.

** Significativa para $P = 0,01$ - para o teste bilateral de correlação populacional de Spearman nula.

Houve diferenças significativas nas distribuições das notas em seis dos 13 domínios do CGI: *Solução de problemas*, *Capacidade de auto-reflexão*, *Percepção do contexto*, *Interação com colegas* e *Interação com demais profissionais*, após correção de Bonferroni. Essas diferenças estiveram presentes em todos os seis domínios, entre CM e Ped, em cinco deles entre GO e Ped (com exceção de *Interação com colegas*), mas em nenhum desses domínios entre GO e CM.

O impacto dos diferentes domínios na confiabilidade da avaliação para cada especialidade mostrou diferenças relevantes. A **Tabela 3** mostra que *Respeito* (GO), *Hábitos de trabalho* (CM) e *Interação com colegas* (Ped) foram os menos correlacionados com a soma dos outros domínios. *Exame físico* (GO) e *Hábitos de trabalho* (CM e Ped) foram os menos explicados pelos outros domínios. *Julgamento clínico* (GO), *Solução de problemas* (CM) e *Julgamento clínico e Capacidade de auto-reflexão* (Ped) produziram a maior queda no alfa, diminuindo a confiabilidade da escala, se eliminados.

Table 3: Análise de confiabilidade para GO, CM e Ped (N= 103)

Domínios	r*			R ² ** (%)			Alfa#		
	GO	CM	Ped	GO	CM	Ped	GO	CM	Ped
1. Qualidade da história	0,747	0,790	0,717	64,1	83,3	86,9	0,926	0,932	0,906
2. Exame físico	0,635	0,720	0,776	56,6	75,1	88,2	0,929	0,934	0,904
3. Conhecimento medico	0,660	0,669	0,712	68,1	78,7	77,9	0,929	0,935	0,907
4. Julgamento clínico	0,788	0,573	0,785	71,0	72,3	80,9	0,924	0,938	0,903
5. Solução de problemas	0,673	0,849	0,642	62,4	86,8	63,6	0,928	0,929	0,910
6. Hábitos de trabalho	0,645	0,526	0,603	60,4	56,4	47,3	0,929	0,939	0,911
7. Comunicação	0,703	0,809	0,585	64,4	79,6	72,8	0,927	0,930	0,912
8. Respeito	0,633	0,668	0,549	71,6	72,5	76,9	0,930	0,935	0,913
9. Capacidade de auto-reflexão	0,712	0,736	0,815	61,6	65,8	75,7	0,927	0,933	0,903
10. Percepção do contexto	0,676	0,798	0,693	75,1	75,4	72,9	0,928	0,931	0,907
11. Interação com colegas	0,737	0,683	0,449	82,6	77,4	54,3	0,926	0,935	0,916
12. Interação com docentes	0,781	0,770	0,455	79,2	75,8	73,9	0,926	0,933	0,916
13. Interação com demais profissionais	0,676	0,712	0,575	69,3	81,9	67,0	0,928	0,934	0,912

*r= correlação de Pearson de cada domínio com a soma dos outros

**R²= porcentagem da variação explicada pelas notas dos outros domínios

alfa = alfa após a eliminação do domínio.

A análise da concordância relativa entre as doze notas: CGI, CGIt, CGIh e AE, cada uma das três especialidades (GO, CM e Ped) mostrou que as notas mais concordantes foram encontradas dentro das especialidades e não entre elas. Nas três especialidades, as notas mais concordantes foram CGI e CGIh. Tomando-se por base os métodos, em CGI e em CGIh as menores distâncias foram entre CM e Ped; em CGIt e AE entre GO e CM. Analisando todos os métodos conjuntamente, as notas de GO e Ped exibiram as maiores discordâncias. Na análise de métodos e especialidades como um todo, as maiores distâncias foram encontradas entre CGIt de Ped e CGIh de GO, CGIt de Ped e AE de GO e CGIt de Ped e CGIh de CM. Observou-se que as três notas de Clínica Médica possuem maior concordância relativa entre si (distância média = 8,05) do que as três notas de GO (distância média 8,30) ou de Pediatria (distância média 8,47) (dados não apresentados).

Discussão

O estágio clínico representa um momento especial para a formação global do estudante, pois o ajuda a ampliar seu conhecimento teórico e a compreender que a medicina cuida de pessoas inseridas em contextos socioculturais diversos, com necessidades distintas, que esperam ser compreendidas, respeitadas e cuidadas.⁸ Mas isso acarreta aos docentes/supervisores um desafio adicional: avaliar os estudantes em um cenário ágil, mutável, com alto impacto educacional.

Em nosso estudo, observou-se que em CGI, CGIt, CGIh e AE as notas ficaram concentradas no extremo superior da escala, com medianas mais elevadas no CGIh e com pouca variação nas distribuições, nas três especialidades. Esse resultado pode ter diversas explicações: mostrar um efeito de leniência entre os supervisores/

docentes, especialmente nos domínios humanísticos, ou demonstrar que a maioria dos estudantes atingiu satisfatoriamente o nível de competência esperada.

Embora nossos dados não nos permitam descartar nenhuma das alternativas, a tendência de atribuir notas elevadas para os alunos está relatada em outros estudos.^{9,10,11} Portanto, nenhuma das estratégias de avaliação discriminou claramente os alunos com desempenho superior daqueles com desempenho inferior, em nenhuma das especialidades.

As notas de Ped foram significativamente inferiores às de GO e CM, nas quatro estratégias. As notas de GO e CM só foram estatisticamente diferentes na AE. Ao analisarmos a concordância relativa entre as notas das três especialidades, observamos que as notas de GO e Ped foram as mais discordantes. As notas de CM mostram maior concordância relativa com as de GO no CGIt e na AE, mas não no CGIh. Tais achados sugerem que os docentes de Ped são mais rígidos nas avaliações, apresentando uma forma particular de avaliar os alunos e que há certa similaridade entre a percepção dos docentes de GO e CM sobre o desempenho dos alunos.

Essas diferenças podem estar apoiadas em aspectos psicológicos que diferenciam a escolha da especialidade entre os profissionais médicos. Diversos estudos sugerem que há uma relação entre atitudes e características pessoais e escolha da especialidade.¹²⁻¹⁴

Essa relação foi confirmada por um estudo entre alunos chilenos, tendo como base os tipos psicológicos do Inventário de Indicadores de Tipos Psicológicos de Myers Briggs.^{15,16} A maior parte dos alunos que optou por CM é do tipo denominado INTp (pessoas que procuram e desenvolvem uma explicação lógica para tudo;

têm habilidade especial para solucionar problemas, são lógicas, céticas, analíticas e críticas, geralmente quietas, contidas, com maior interesse nas idéias do que na interação social, são flexíveis e facilmente adaptáveis). Já os alunos que optaram por Ped são caracterizados como ESFJ, cujas qualidades predominantes são o cuidado e a atenção com as pessoas e o desejo de criar um clima harmonioso no ambiente de trabalho. São pessoas, empáticas, leais, colaboradoras, organizadas, detalhistas, gostam de trabalhar em equipe, de cumprir prazos e de serem reconhecidos e valorizados. Nessa pesquisa, os autores não fazem referência específica à especialidade de GO. Como há características comuns entre GO e Cirurgia, uma vez que ambas realizam atividades clínicas-cirúrgicas, a partir de agora usaremos o perfil do profissional de Cirurgia para representar o de GO. Para os pesquisadores, os cirurgiões são predominantemente do tipo ESTJ: pessoas práticas, decididas, realistas e diretas, voltadas para a implementação das decisões. Geralmente, são focadas em resultados, cuidam dos detalhes da rotina, gostam de passar rapidamente para a próxima etapa e pensam nos fatos de forma pragmática¹⁷.

Ao apresentarem alguns exemplos de práticas de supervisão realizadas em diferentes especialidades, Kilminster et al.⁴ evidenciaram também as características particulares de cada uma delas, faltando dados específicos sobre GO. Na cirurgia, o aluno é avaliado por meio de um formulário que contém itens relacionados aos procedimentos essenciais que um aluno deve realizar. Em CM é dada ênfase a aspectos como diagnóstico, qualidade das anotações feitas e das prescrições. Em Ped, auto-controle emocional, relações interpessoais e habilidade de trabalhar em equipe são altamente valorizados.

De forma geral, os alunos que optam por especialidades incluídas na área de atenção primária (dentre elas, Pediatria, Clínica Médica e Ginecologia e Obstetrícia) valorizam a oportunidade de fornecer atendimento direto ao paciente e de se envolver com os aspectos psico-sociais do cuidado médico e apresentam atitudes mais centradas no paciente.^{13,14,18} Mas a área de Ped parece ter uma orientação humanística acentuada, enquanto tanto GO quanto CM parecem colocar mais foco em resultados.¹⁷ Essas características podem nos ajudar a entender a relativa proximidade entre as notas de GO e CM e a evidente particularidade das notas de Ped.

A avaliação estruturada foi o método no qual houve diferenças significativas entre todas as especialidades, mostrando que, durante a observação de uma consulta, as diferenças de cada especialidade se tornam ainda mais evidentes, apesar de os formulários utilizados apresentarem porcentagens relativamente semelhantes de aspectos técnicos (62,9%, 62,5%, 54,5% para Ped, GO e CM, respectivamente). Esse resultado pode ser parcialmente explicado pelo fato de que, sabidamente, os docentes/supervisores diferem em sua visão daquilo que constitui a gama de conhecimentos e habilidades essenciais para os estudantes em estágio clínico e, por isso, avaliam os alunos de forma distinta, incluindo um viés de observação direcionado pelas características de sua própria especialidade.⁷ Além disso, como os alunos estão iniciando sua prática clínica, é possível que haja uma supervalorização dos aspectos técnicos, dos procedimentos, do conhecimento das doenças, em detrimento dos aspectos de relacionamento com o paciente, como sugerem alguns autores.⁸

Em relação ao CGI, a consistência interna do questionário foi elevada, mostrando que os docentes das três especialidades percebem uma homogeneidade entre os domínios do questionário. Entretanto, ao compararmos as distribuições das treze notas, observamos que a maior parte das diferenças entre as três especialidades esteve concentrada nos domínios humanísticos, especificamente nos domínios *Capacidade de auto-reflexão*, *Percepção do contexto* e *Interação com demais profissionais*. Uma possível explicação para esses resultados é que, como as definições desses domínios são subjetivas e passíveis de ambigüidade, há uma intensificação das diferenças de “cultura” e de “personalidade” de cada especialidade.¹⁹

CM foi a especialidade que apresentou a maior concordância relativa entre as notas dos diferentes métodos, enquanto Ped apresentou a maior discordância, sugerindo um comportamento de avaliação mais uniforme entre os docentes de CM. Embora não seja possível ignorar as diferenças entre as especialidades, vale ressaltar a necessidade de se buscar uma uniformização no processo de avaliação, particularmente num módulo integrado.¹¹ O fato de diferentes docentes avaliarem o mesmo aluno com critérios muito distintos pode acarretar conseqüências educacionais negativas tanto para o estudante, levando-o a ter dificuldade de elaborar um plano de melhoria e desenvolvimento, quanto para o corpo docente, prejudicando a distinção clara entre desempenho aceitável e inaceitável.²⁰

Faz-se necessário, portanto, estabelecer critérios de avaliação do desempenho esperado dos estudantes ao longo do curso. Sem dúvida, essa tarefa se torna mais fácil quando se avaliam as competências cognitivas do que as clínicas, especialmente em um cenário em que múltiplas especialidades estão envolvidas. Nesse caso, o sistema de avaliação deveria ser composto de duas partes. Uma delas

englobando atributos comuns às diferentes áreas médicas, mensuráveis, relevantes para a prática clínica, relacionadas ao nível de competência atual do estudante e com alto impacto educacional. Uma segunda englobaria um núcleo de conteúdos e domínios específicos para cada especialidade. Esse conjunto permitiria a elaboração de uma matriz de desempenho do aluno que fosse mais facilmente comparável entre as disciplinas.²¹

Sabe-se que a percepção de qualidade do estágio clínico (pré-internato e internato) pelo estudante, para o qual a avaliação tem um papel relevante, é um fator crucial na escolha da carreira. Estágios bem delineados, combinando atividades práticas e teóricas e supervisores entusiasmados e envolvidos com o trabalho, são decisivos na opção que o estudante faz e na identificação da “personalidade da especialidade”¹⁹.

Pesquisas realizadas em diversas partes do mundo, incluindo os Estados Unidos e o Brasil, mostraram que mais de 40% dos alunos escolhem sua especialidade entre a fase pré-clínica e o início do internato.^{19,22} Pediatria foi apontada como a especialidade preferida pelos alunos (20%), seguida de cirurgia (18%) e de semelhante proporção de indecisos (19%). As opções dos 43% restantes estavam pulverizadas, sendo 8% para clínica médica e 4% para ginecologia e obstetrícia.²³

Entre os aspectos determinantes para a escolha da especialidade destacam-se: características da própria instituição, incluindo o fato de ela ser pública ou particular, grade curricular, influência de docentes, variáveis demográficas, características próprias da especialidade, tipos de problemas/doenças dos pacientes, estilo de vida do médico, prestígio e questões financeiras.²³

Uma limitação do nosso estudo é que ele foi realizado em uma turma de alunos, avaliados por um pequeno número de supervisores/docentes. Portanto, não sabemos se esses dados podem ser generalizados para outros contextos. No entanto, são raras as publicações sobre a relação entre comportamentos de avaliação características das especialidades clínicas. Assim sendo, esse estudo representa uma contribuição inicial para procurar compreender como a “cultura” intrínseca de cada especialidade se reflete no processo de avaliação realizado por docentes/supervisores.

Conclusões

Há uma tendência dos docentes em atribuírem notas elevadas para os alunos, dificultando a discriminação entre o melhor e o pior desempenho. Os docentes de Ped têm uma forma particular de avaliar os alunos, enquanto os docentes de GO e CM os avaliam de forma semelhante. Enquanto GO e CM valorizam os domínios técnicos, Ped valoriza os humanísticos, o que parece estar de acordo com as “culturas” das três especialidades e com a tipologia psicológica dos profissionais que atuam em cada uma delas. Dentre os desafios a serem superados em relação à decisão sobre as estratégias de avaliação do aluno nas diversas fases do curso, estão as particularidades das diferentes especialidades.

Disponibilizar métodos que contemplem diversas competências não garante necessariamente que todas elas estejam sendo avaliadas adequadamente. Uma das estratégias para minimizar o impacto dos diferentes olhares sobre a avaliação em um módulo clínico integrado é o oferecimento de um programa de desenvolvimento docente, no qual se busque encontrar uma uniformização na avaliação, sempre

tentando obter critérios confiáveis, válidos e justos. A qualificação do processo ensino-aprendizagem demanda o preparo dos docentes para a aplicação e interpretação de diferentes estratégias de avaliação.

Referências

1. Daelmans HE, Hoogenboom RJ, Donker AJ, Scherpbier AJ, Stehouwer CD, van der Vleuten CP. Effectiveness of clinical rotations as a learning environment for achieving competences. *Med Teach*. 2004; 26(4):305-12.
2. Kilminster S & Jolly B. Effective supervision in clinical practice: a literature review. *Med Educ*. 2000; 34:827-9.
3. Brasil, Ministério da Educação, Conselho Nacional de Educação. Câmara de Educação Superior. Resolução CNE/CES n.4. 7 de nov. 2001- Diretrizes Nacionais para Curso de Graduação em Medicina. Brasília, legislação oficial, 2001.
4. Kilminster S, Cottrell D, Grant J, Jolly B. AMEE Guide nº 27: Effective educational and clinical supervision. *Med Teach*. 2007; 29:2-19.
5. Daelmans HE, Overmeer RM, van der Hem-Stoktoos HH, Scherpbier AJ, Stehouwer CD, van der Vleuten CP. In-training assessment: qualitative study of effects on supervision and feedback in undergraduate clinical rotation. *Med Educ*. 2006; 40:51-8.
6. Swick S, Hall S, Beresin E. Assessing the ACGME competencies in psychiatry training programs. *Acad Psych*. 2006; 30:330-51.

7. Koens F, Rademakers JJDMJ, Cate OTJT. Validation of core medical knowledge by postgraduates and specialists. *Med Educ.* 2005; 39:911-7.
8. Ribeiro, MMF, Amaral, CFS. Medicina centrada no paciente e ensino médico: a importância do cuidado com a pessoa e o poder do médico. *Rev Bras Edu Med.* 2008; 32:90-7.
9. Bergovec M, Kuzman T, Rojnic M, Makovic A. is there grade inflation at medical schools? Case study of the Zagreb University School of Medicine. *Croat Med J.* 2003; 44:92-7.
10. Speer AJ, Solomon DJ, Fincher RM. Grade inflation in internal medicine clerkship: results of a national survey. *Teach Learn Med.* 2000; 12:112-6.
11. Nahum GG, Facs F. Evaluating medical student obstetrics and gynecology clerkship performance: which assessment tools are most reliable? *Am J Obst Gyn.* 2004; 191:1762-71.
12. Fincher RME, Lewis LA, Jackson TW. Why students choose primary care or nonprimary care career. *Am J Med.* 1994; 97:410-7.
13. Batenburg V, Smal JA, Lodder A, Melker RA. Are Professional attitudes related to gender and medical specialty? *Med Educ.* 1999; 33:489-93.
14. Haidet P, Dains JE, Paterniti DA, Hechet L. Chang T, Tseng E, et al. Medical student's attitudes toward the doctor-patient relationship. *Med Educ.* 2002; 36:568-74.
15. Myers IB, MacCaulley M, Queen N, Hammer A. *MBTI Manual: a guide in the development and use of the Myers-Briggs Type Indicator.* Palo Alto. California: Consulting Psychologist Press Inc. 1998.

16. Calegari ML, Gemignani OH. Temperamento e Carreira. São Paulo: Summus Editorial, 2006.
17. Bitran MC, Zuniga D, Lafuente MG, Viviani PG, Mena BC. Influencia da personalidade y el estilo de aprendizaje en la elección de especialidad médica. Rev Med Chile. 2005; 133:1191-9.
18. Kassler WJ, Wartman SA, Silliman RA. Why medical students choose primary care careers. Acad Med. 1991; 66: 41-43.
19. Bland KI, Isaacs G. Contemporary trends in student selection of medical specialties. Arch Surg. 2002; 137:259-67.
20. Bouriscot KAM, Roberts TE, Pell G. Standard setting for clinical competence at graduation from medical school: a comparison of passing scores across five medical schools. Adv Health Sci Edu Theory Pract. 2006; 11:173-83.
21. Southgate L, Hays RB, Norcini J, Mulholland H, Ayres B, Wooliscroft J et al. Setting performance standards for medical practice: a theoretical framework, Med Educ. 2001; 35:474-81
22. Bellodi P. Surgery or general medicine – a study of the reason underlying the choice of medical specialty. São Paulo Med J. 2004; 122:81-6.
23. Compton MT, Frank E, Elon L, Carrera J. Changes in US medical students' specialty interests over the course of medical school. J Gen Inter Med. 2008, 23: 1095-100 DOI: 10.1107/s11606-008-0579-z

5. Discussão

As Diretrizes Curriculares Nacionais do Curso de Graduação em Medicina de 2001 instituem que o aluno deve ter uma formação generalista, humanista, crítica e reflexiva, ser capacitado para atuar em equipe, com responsabilidade social e compromisso com a cidadania, agindo como promotor da saúde integral do ser humano (Brasil, 2001). O módulo de ensino *Atenção Integral à Saúde* cria um cenário favorável a essa formação profissional, porque propicia que o aluno inicie sua prática clínica como médico da família, próximo à comunidade, fora do hospital de referência, permanecendo em uma mesma Unidade Básica de Saúde, com a mesma equipe de profissionais e supervisores docentes por dois semestres. Esse perfil de médico generalista, que consegue reunir a prática e a teoria, que apresenta habilidades de comunicação e princípios éticos e que utiliza uma abordagem integrativa em relação ao cuidado com o paciente é vislumbrado internacionalmente (World Federation for Medical Education, 2001).

Para atender às necessidades e expectativas de um perfil diferenciado para o profissional médico, os métodos de avaliação das competências clínicas precisam

ser adaptados a essa nova realidade. A utilização de múltiplos métodos que enfoquem diferentes domínios das competências clínicas é fundamental para que o processo de avaliação do aluno seja válido, confiável e justo. Nesse sentido, o sistema de avaliação do *Módulo Atenção Integral à Saúde* contempla avaliações por Conceito Geral/Conceito Global Itemizado, Portfólio, Avaliação Estruturada do Atendimento Clínico e Provas Teóricas do tipo múltipla escolha e respostas curtas.

O Conceito Geral (CG), aqui entendido como um valor unidimensional, isto é, uma nota única, atribuída pelo docente pela impressão de desempenho do aluno, não calcada em observação de atendimento, é uma estratégia de avaliação muito utilizada na prática do ensino nas fases clínicas (Farrell, 2005). Do ponto de vista psicométrico, essa nota possui limitações importantes, como a ausência dos fatos ou atributos que devem ser avaliados, a falta de uma definição das notas da escala e a carência de uma base objetiva não tendenciosa (Thorndike, 1912). Essa nota compôs o conjunto de avaliações deste módulo de forma transitória, até que fosse possível compará-la com outros métodos e com um conceito multidimensional que inclui diversos domínios da competência clínica do estudante. A introdução do Conceito Global Itemizado (CGI) objetivou superar a tendenciosidade do instrumento de CG, chamando a atenção dos avaliadores para a necessidade de considerar individualmente distintos atributos técnicos e humanísticos.

Os resultados das análises comparativas entre os dois tipos de conceitos mostraram que, ao atribuírem um conceito geral, os docentes privilegiaram os domínios técnicos, em detrimento dos humanísticos, atributos essenciais da prática médica. Os atributos *Respeito* (a pacientes e familiares) e *Hábitos de Trabalho*

(pontualidade, assiduidade) do CGI demonstraram ser os menos correlacionados com o Conceito Geral. Embora inicialmente definido como técnico, *Hábitos de trabalho* comportou-se como atributo humanístico, mas ficou isolado de todos os outros domínios. Ainda, a eliminação de *Hábitos de Trabalho* do questionário não mostrou prejuízo na confiabilidade do instrumento, o que não se repetiu para nenhum outro domínio avaliado. Excetuando-se essa situação, a avaliação propiciada pelo instrumento de conceito itemizado mostrou-se adequada, confiável e válida.

Encontramos correlações positivas e significativas entre os treze domínios do CGI, especialmente entre os atributos da competência técnica e uma alta consistência interna do questionário para todos os avaliadores. Embora esses resultados reforcem a confiabilidade do CGI, eles podem refletir uma dificuldade dos avaliadores em distinguirem claramente as peculiaridades de cada atributo, não conseguindo julgá-los de forma independente, visto que não estão habituados a utilizar o CG ancorado a descritores dos respectivos domínios (Bandiera, Morrison, Regehr, 2002). É possível que tenha ocorrido o “efeito halo” isto é a nota atribuída a um determinado domínio pode ter influenciado positiva ou negativamente a nota atribuída ao domínio seguinte (Newble, 2000). Apesar dos descritores dos valores limítrofes estarem impressos no formulário, a ausência de um programa mais elaborado para implementação do instrumento, com discussão prévia do significado de cada item e de seus descritores pode ter causado variações de interpretação e compreensão entre os avaliadores, levando-os a continuarem usando critérios pessoais, semelhantes aos que vinham sendo utilizados ao atribuírem o Conceito Geral.

Observamos que as notas atribuídas aos alunos pelos docentes tanto no Conceito Geral quanto no Conceito Global Itemizado concentraram-se essencialmente no extremo direito da escala, sugerindo leniência dos supervisores, especialmente nos domínios humanísticos da competência. Esse “efeito teto” também foi encontrado nas notas da auto-avaliação e da avaliação dos pares, assim como em todos em outros métodos de avaliação analisados nesse estudo, evidenciando que os avaliadores parecem resistir ao uso do espectro total da escala.

As estratégias de auto-avaliação e de avaliação dos pares/colegas foram iniciativas pioneiras para os alunos do 4º ano de Medicina da Faculdade de Ciências Médicas da Unicamp. Pela primeira vez em um módulo de ensino clínico elas foram incorporadas de forma combinada, visando o desenvolvimento da metacognição, cuja importância no processo de aprendizagem é amplamente reconhecida (Ribeiro, 2003; Stedile e Freindlander, 2003). No âmbito educacional, metacognição pode ser entendida como *conhecimento sobre o conhecimento* (a tomada de consciência das competências necessárias para a realização de uma tarefa) e como *auto-regulação* (capacidade de avaliar a execução de uma tarefa e fazer as correções necessárias) (Ribeiro, 2003), dois aspectos essenciais durante o processo de avaliação.

Nossos resultados sugerem que a qualidade da auto-avaliação foi questionável, uma vez que houve um maior grau de arbitrariedade nas notas de auto-avaliação do que nas dos colegas e dos docentes. Apesar do amplo reconhecimento em relação à importância da auto-avaliação como elemento essencial para o crescimento profissional, esse método apresenta menor acurácia do que o método de avaliação feita por colegas (Ward, Gruppen, Regehr, 2002;

Papinczak, 2007). Como a prática da auto-avaliação não é enfatizada durante o curso médico, talvez tenha faltado aos alunos o claro reconhecimento de seus pontos fortes e suas limitações. Tais achados coincidem com os de outros estudos que apontam baixa correlação entre a auto-avaliação e a avaliação realizada por docentes (Sullivan, 1999; van Rosendaal e Jennet, 1994; Rudy et al., 2001; Papinczak, 2007), reforçando a idéia de que, de forma geral, as pessoas não são boas julgadoras de seu próprio desempenho (Ward, Gruppen, Regehr, 2002).

Ao analisarmos os resultados das auto-avaliações, das avaliações docentes e de colegas observamos que, embora cada tipo de avaliador tenha expressado uma visão particular sobre o processo de avaliação, houve alguma coincidência em relação aos domínios que mais impactaram na confiabilidade do Conceito Global Itemizado: *Exame físico e Solução de problemas* (na auto-avaliação), *Solução de problemas* (para os docentes em geral), *Capacidade de auto-reflexão e Julgamento clínico* (para os docentes de Pediatria), *Julgamento clínico* (para os docentes de Ginecologia e Obstetrícia), *Solução de problemas* (para os docentes de Clínica Médica) e *Capacidade de auto-reflexão e Respeito* (para os colegas) tiveram o maior impacto na confiabilidade do CGI.

A avaliação de colegas que pertencem aos mesmos grupos de trabalho tem sido considerada uma questão central na educação médica (Norcini, 2003). Além de contribuir para o desenvolvimento do pensamento crítico, permite ao aluno adquirir habilidades que podem ser transferidas para a tarefa de auto-avaliação (Papinczak, 2007). De forma geral, os alunos atribuíram notas mais baixas para os colegas do que para si mesmos. De fato, estudos mostram a tendência das pessoas

superestimarem suas próprias notas (Wooliscroft et al., 1993; Mattheos et al., 2004), julgando-se melhores do que os colegas, especialmente em atributos com alto grau de subjetividade, sujeitos à ambigüidade, como são os domínios humanísticos de nosso questionário (Ward, Gruppen, Regehr, 2002). Tal fenômeno de comparação social é denominado “efeito melhor que a média” e tem por objetivo manter uma imagem de superioridade em relação aos outros (Larrick, Burson, Soll, 2007).

A estratégia utilizada no 4º ano de Medicina de se criar pequenos grupos de trabalho nas UBS traz ganhos pedagógicos e profissionais importantes. A percepção do valor do trabalho em equipe, a aprendizagem por meio de observações e discussões, o desafio de lidar com as diferenças, a motivação para a aprendizagem e o comprometimento com os colegas são alguns dos benefícios encontrados. Entretanto, a avaliação de colegas que pertencem ao mesmo grupo de trabalho ou de aprendizagem pode ter vieses, especialmente em situações como a do presente estudo, nas quais os colegas mantêm uma relação duradoura entre si, com vínculos de amizade (Papinczak, 2007).

Os 24 grupos participantes do nosso estudo apresentaram diferentes comportamentos durante o processo de avaliação dos pares. Em alguns grupos, nem todos os componentes foram avaliados por seus pares e notas extremamente baixas foram atribuídas. Observou-se algum grau de reciprocidade e de constância nas notas. Uma possível explicação para esses achados é que como a avaliação dos pares foi realizada pela primeira vez por alunos do 4º ano, é possível que alguns deles tenham ficado em dúvida sobre a utilidade dessa estratégia, adotando, assim, uma atitude de ceticismo ou incredulidade. Essa percepção foi corroborada

pelos achados do Questionário de Opinião sobre o processo de auto-avaliação e de avaliação dos colegas, inserido em nosso estudo. Sabe-se que a implantação desse tipo de avaliação causa um desconforto inicial e dúvidas sobre sua validade, qualquer que seja o contexto. Entretanto, essas percepções aparentemente negativas poderão ser revertidas se os alunos compreenderem claramente os objetivos a serem atingidos, conscientizarem-se sobre a relevância dessa estratégia, forem constantemente expostos a ela e receberem *feedback* construtivo e estruturado sobre seu desempenho (Papinczak, 2007).

A variação das notas que um mesmo avaliado recebeu foi maior do que entre os diferentes avaliados no grupo, ressaltando que o principal componente de variabilidade dos dados foi o próprio avaliador. Ao estudarem as características dos grupos e os papéis desempenhados pelos atores sociais que deles participam, Krueger e Stanke (2001) concluíram que cada avaliador realmente utiliza critérios particulares para atribuir notas aos colegas, ou generalizando para o grupo as características dos indivíduos mais conhecidos por ele, ou usando os estereótipos do grupo para inferir as características de cada membro. Nossos resultados sugerem que as percepções que cada avaliador formou dos outros membros de seu grupo dependeram, fortemente, do grau de conhecimento que ele tinha de cada colega e dos sentimentos que existiam entre eles.

Outra questão que não pode deixar de ser considerada na dinâmica grupal refere-se às projeções sociais e às percepções que os diferentes membros têm de si mesmos (Sande, Goethals, Radloff, 1988; Krueger, Stanke, 2001). Auto-imagem e auto-estima são referenciais utilizados no processo de avaliação intra-

grupais. Pessoas com elevada auto-estima apresentam um padrão mais consistente de auto-avaliação e de avaliação dos colegas. Elas tendem a se avaliar melhor do que a seus colegas em atributos considerados positivos, mas não tentam minimizar seus próprios atributos negativos. Já as pessoas com baixa auto-estima tendem a avaliar melhor os colegas em relação a atributos considerados positivos, mas ao mesmo tempo, tendem a acentuar suas próprias características positivas, sem conseguirem minimizar as negativas (Suls, Lemos, Stewart, 2002). Portanto, neste estudo e em qualquer processo que inclua avaliação auto-avaliação e avaliação por pares, não se deve ignorar os efeitos que estes aspectos psicológicos podem causar nos julgamentos que cada avaliador faz de seus avaliados.

Adicionalmente, estudamos os comportamentos dos docentes das três especialidades envolvidas no estágio e nossos resultados mostraram que cada uma delas enfocou aspectos distintos da competência, evidenciando a “personalidade” da especialidade. Não se sabe se os estudantes médicos se amoldam às características da especialidade que escolheram, ou se a especialidade atrai estudantes com características semelhantes à dela. De qualquer forma, vale ressaltar a importância do papel do docente/supervisor durante o estágio clínico, na avaliação, orientação e formação profissional dos alunos. Supervisores motivados, comprometidos e envolvidos com as atividades de ensino influenciam o desempenho do aluno durante o estágio, influenciando-o, inclusive no direcionamento de sua carreira (Bland, Isaacs, 2002; Cottrell et al., 2002).

A variedade de notas atribuídas pelos docentes a partir de um mesmo formulário de avaliação nos leva a refletir sobre o uso diferencial da escala.

Apesar de apresentarmos as escalas mediante valores discretos, muitos avaliadores atribuíram valores intermediários. Portanto, utilizaram as mesmas de forma contínua, tanto no formulário de CG quanto no de CGI. Essa característica de continuidade dos dados, embora não contemplada em nossa proposta original, foi considerada nos procedimentos estatísticos. Assim, apresentamos medidas de tendência central, com as quantidades apropriadas de casas decimais, incorporando o grau de acurácia usado pelos avaliadores.

É improvável que todos os avaliadores constituam um grupo monolítico, que interpretem igualmente as diversas dimensões da competência de forma consistente, que se baseiem em critérios semelhantes, ou ainda que sejam homogêneos em sua habilidade de avaliação. Além disso, o significado amplo do termo “competência” por si só pode criar uma dificuldade intrínseca para os avaliadores (Ward, Gruppen, Regehr, 2002). Assim sendo, diferenças entre avaliadores são esperadas, embora constituam, inevitavelmente, fontes potenciais de erros e vieses (Mazor et al., 2007). Portanto intercorrelações entre docentes de especialidades diferentes não deveriam ser usadas como medida exclusiva de validade (DiMatteo, 1981). Ao invés de haver um único avaliador atribuindo notas em diferentes situações, recomendam-se diferentes olhares sobre o desempenho dos estudantes nas atividades clínicas, no sentido de contribuir para qualificar a avaliação, aumentando sua confiabilidade (Crossley et al., 2002; Haes, Oort, Hulsman, 2005).

Como ocorreu em nosso estudo, as auto-avaliações são geralmente comparadas com as avaliações dos docentes e supervisores, que forneceriam “a

nota verdadeira”, aproximando-se de um “padrão-ouro”. No entanto, pode haver inconsistência mesmo entre docentes experientes. Além, disso, as avaliações sob forma de conceitos, que não se baseiam em observação direta do desempenho, são realizadas ao término de um longo estágio, limitando-se à lembrança do que aconteceu. Embora docentes experientes tenham maior probabilidade de concordar entre si, a idéia de um avaliador infalível é vista com ceticismo na literatura (Ward, Gruppen, Regehr, 2002). Estudos demonstram que a confiabilidade do processo avaliativo aumenta quando o mesmo contempla um círculo completo de avaliadores, denominado de avaliação 360 graus, incluindo o próprio aluno, docentes, colegas, membros da equipe multiprofissional e pacientes (Rodgers e Manifold, 2002; Jackson, Blaxter, Lewando-Hundt, 2003).

Todas essas questões evidenciam a necessidade de qualificação do corpo docente em relação às habilidades de avaliação (Newble, 2000). No entanto, essa capacitação não deve ocorrer como um evento único e pontual. Para que se possa observar redução de vieses e melhorias duradouras em confiabilidade e validade, o processo de treinamento em avaliação deve fazer parte do currículo de aprimoramento profissional e desenvolvimento docente, de forma continuada (Bandiera, Morrison, Regehr, 2002). É importante ressaltar que ações de aprimoramento das habilidades de avaliação deveriam ser estendidas também para os alunos, que assim iniciariam este processo de capacitação ainda durante o curso médico (Sullivan, Hitchcock, Dunnington, 1999; van Rosendaal e Jennet, 1994; Rudy et al., 2001; Papinczak, 2007). Diversos estudos foram publicados sobre o papel dos residentes como docentes e avaliadores. Os resultados demonstram

que os benefícios da prática de avaliação durante a graduação se refletem na residência médica, quando o desconforto em avaliar os colegas diminui e a eficácia de avaliação dos estudantes aumenta (Wendling e Hoekstra, 2002; Busari et al., 2003).

Reconhece-se que a avaliação da competência clínica deve ser realizada em diversos momentos, em contextos semelhantes aos da prática real, utilizando a combinação de múltiplos métodos e o julgamento de diversos avaliadores (Newble, 2000). Ao compararmos os diferentes métodos utilizados neste módulo, pudemos evidenciar as peculiaridades de cada um deles. Nossos resultados mostraram a maior concordância entre notas da Prova teórica e do Portfólio, demonstrando que esses métodos enfocam essencialmente o nível do “saber”, em oposição ao Conceito Global Itemizado e à Avaliação Estruturada, que também concordantes, valorizam o saber aplicado ao atendimento do paciente real.

De fato, o Portfólio deve conter informações sobre as evidências do processo de aprendizagem do aluno, incluindo as planilhas de atendimento, a revisão de casos clínicos apoiada na bibliografia, uma auto-análise relacionada aos seus pontos fortes e fracos e ao seu nível de competência e uma proposta para remediar as dificuldades. Entretanto, o cerne do Portfólio está na qualidade das análises que o aluno faz de si mesmo e dos casos atendidos, na clareza como ele expõe suas idéias e como esse conjunto pode sinalizar que os objetivos propostos foram atingidos (Driessen et al., 2005).

Embora a Avaliação Estruturada tenha como objetivo avaliar o aluno durante o contato com o paciente (Farrell, 2005), esse método é pouco prático,

pois demanda muito tempo do avaliador e tem como base a observação de uma única consulta. Esse fato pode explicar parcialmente as notas elevadas que foram atribuídas aos alunos. Por um lado, como o estudante estava sendo observado, ele pode ter desempenhado melhor do que em outras situações. Por outro, o docente pode ter sido leniente, por considerar que o eventual desempenho insatisfatório do aluno pudesse ser devido ao viés de sua presença.

De qualquer forma, nossos resultados apontam para a importância de se estabelecer padrões referenciais mínimos de desempenho que possam orientar as ações de avaliação dos docentes e servir de guia para o desenvolvimento profissional dos estudantes. Estudos futuros devem ser realizados no sentido de validar critérios que possam diferenciar alunos competentes dos não competentes (Boulet, Champlain, McKinley, 2003). Afinal, dentre as múltiplas funções da avaliação destacam-se o estabelecimento de parâmetros que garantam a qualidade dos serviços prestados à sociedade e o fornecimento de orientações necessárias para o aprimoramento contínuo do estudante (Swick, Hall, Beresin, 2006).

A partir dos resultados do nosso estudo, algumas ações poderão ser tomadas como parte do processo de aprimoramento do sistema de avaliação. Sugerimos uma revisão do instrumento de Conceito Global Itemizado que englobe: (a) a reanálise dos descritores pelos docentes, (b) revisão da linguagem utilizada para minimizar a sobreposição de informações), (b) a inclusão de descritores nos pontos intermediários da escala, (c) a exclusão do domínio *Hábitos de trabalho* e (d) a eliminação da divisão das competências em técnicas e humanísticas.

Além disso, os critérios de avaliação deveriam ser discutidos com os próprios alunos. Docentes, tutores e profissionais de outras áreas da saúde deveriam ser envolvidos na definição dos comportamentos e do nível de competência esperado dos estudantes (Ward, Gruppen, Regehr, 2002). Uma possibilidade seria que os avaliadores, conjuntamente, selecionassem os domínios a partir de uma amostragem de comportamentos com forte impacto educacional, observados durante a prática clínica. Esse instrumento “piloto” deveria ser testado; os comportamentos seriam discutidos entre os diferentes tipos de avaliadores, antes da decisão de incluí-los definitivamente no questionário. Paralelamente, a estratégia de trabalho e aprendizagem em grupo ainda deveria ser objeto de novos estudos para que se possa melhor compreender a dinâmica psicológica e social que circunda os grupos (Levine et al., 2004).

No sentido de legitimar o processo de auto-avaliação, os alunos deveriam calibrar suas notas, por meio de prática, observação e discussão com docentes (Ward, Gruppen, Regehr, 2002). Após uma consulta, por exemplo, os estudantes poderiam completar um formulário de auto-avaliação; posteriormente, eles avaliariam algumas consultas gravadas, realizadas com pacientes simulados, que apresentassem diferenças na qualidade do desempenho do médico. Em seguida, eles discutiriam suas notas com os professores, comparando-as com as das consultas assistidas. Por fim, os alunos fariam uma revisão de suas próprias notas. Essas estratégias auxiliariam os estudantes a melhorarem seu desempenho, a aprimorarem suas habilidades metacognitivas e a realizarem auto-avaliações de forma mais confiável (Mckinstry, 2007; Srinivasan et al., 2007).

Dessa forma, o foco da auto-avaliação seria alterado, de uma abordagem inter-individual, na qual há uma inevitável comparação com os colegas, para uma abordagem intra-individual, que enfatiza a habilidade do aluno identificar seus pontos fortes e fracos, reconhecer os domínios da competência que necessitam maior aprimoramento e os que necessitam de menos trabalho, para assim poder elaborar um plano de desenvolvimento profissional.

Quanto mais oportunidades os estudantes tiverem para demonstrar suas diversas competências, mais estável e eficaz o processo de avaliação se tornará. Entretanto, antes da definição e do detalhamento dos métodos a serem utilizados, é importante considerar os princípios e valores subjacentes, como por exemplo, o que se pretende com o processo de avaliação, como ele será viabilizado, quais parâmetros serão considerados, que profissionais estarão envolvidos, como os resultados serão apresentados e tratados e, finalmente, como será conduzido o processo de melhoria. Essas foram as motivações que permearam a realização desse estudo.

Não se pode desconsiderar o valor preditivo da avaliação nas escolas médicas. Um processo avaliativo bem delineado e adequadamente conduzido durante a graduação torna-se um importante indicador de qualidade do currículo e dos métodos educacionais, podendo fornecer uma base de predição do futuro desempenho dos alunos durante o período de residência médica. Futuros estudos devem ser conduzidos para se possa comparar a competência clínica dos alunos em diferentes momentos (Hamdy, 2008).

Parte integrante de um processo de avaliação, o fornecimento de *feedback* aos alunos destaca-se como uma estratégia de aperfeiçoamento pessoal e profissional e de melhoria efetiva de desempenho, pois oferece ao aluno a oportunidade de refletir sobre suas ações e elaborar um plano de melhoria (Ende, 1983; Menachery, et al., 2006). Um *feedback* estruturado, bem embasado e construtivo representa um diferencial da atuação formativa do docente e permite que os alunos valorizem as informações nele contidas, preocupando-se menos com a nota final e mais com as experiências de aprendizagem que esta sumariza (Buchel e Edwards, 2005). As “regras de Pendleton” (Pendleton et al., 1984) são uma estratégia simples, prática, útil e eficaz: 1) perguntar ao estudante o que foi bem feito; 2) perguntar o que pode ser melhorado; 3) expor ao estudante o que foi bem feito; 4) mostrar a ele o que ainda pode ser melhorado e 5) resumir os principais pontos levantados pelo estudante e pelo docente (Troncon, 2008).

No *Módulo Atenção Integral à Saúde*, o processo de *feedback* já está incorporado aos métodos de Portfólio e de Avaliação Estruturada. A introdução do método do Conceito Global Itemizado, substituindo o Conceito Geral, permitirá mais um momento valioso de *feedback*. Como próximo passo, os alunos devem ser preparados para fornecer *feedback* aos colegas e receber *feedback* deles também. Essas distintas oportunidades de troca de informações são fundamentais para o aprimoramento da metacognição dos estudantes, auxiliando-os no desenvolvimento gradual de sua capacidade de identificar, controlar e aprimorar seus processos cognitivos e utilizá-los na resolução de problemas.

Por último, a proposta de implantação de um conjunto de novos métodos, embora válida e necessária, esbarrou inevitavelmente em algumas questões importantes, como a compartimentalização de conhecimentos e habilidades, o viés de subjetividade, a sobreposição de idéias na compreensão dos diferentes domínios da competência clínica e a falta de critérios uniformes para o julgamento do desempenho. Além disso, ressaltamos o desafio inerente ao processo de avaliação que demanda balancear rigor (confiabilidade e validade) e praticidade (viabilidade e aceitabilidade). Nesse sentido, enfatizamos a relevância e necessidade do envolvimento, dedicação e comprometimento do corpo docente em todas as fases do processo de avaliação.

Em síntese, este estudo procurou realçar o valor educacional do processo avaliativo, em uma experiência de atendimento clínico, no cenário da atenção primária. Trouxe uma visão mais clara do sistema de avaliação do *Módulo Atenção Integral à Saúde* do 4º ano de Medicina, permitiu validar a proposta de integração de método e introduziu novas estratégias que poderão se expandir para outros módulos. Adicionalmente, nossos resultados poderão desencadear um processo de aprimoramento de todos os instrumentos de avaliação do *Módulo Atenção Integral à Saúde* e de outros módulos de ensino. Desta forma, a avaliação alcança um dos seus objetivos principais, contribuir para a melhoria da qualidade de um curso que visa formar profissionais aptos a cuidar dos pacientes com profissionalismo, a interagir com os outros membros da equipe, atendendo, assim às necessidades e expectativas da sociedade.

6. Conclusões

- As notas do Conceito Geral foram inferiores às do Conceito Global Itemizado para os docentes, mas não para os próprios alunos, estiveram mais fortemente correlacionadas e foram mais concordantes com as notas dos domínios técnicos do que com as dos humanísticos.
- No Conceito Global Itemizado, os alunos se avaliaram mais rigorosamente do que os docentes nos domínios técnicos, mas não nos domínios humanísticos. As notas dos pares foram mais baixas do que as dos próprios alunos, nos atributos humanísticos. As notas dos domínios humanísticos foram mais elevadas do que as dos técnicos para alunos e docentes.
- As notas dos docentes de Pediatria foram as mais baixas, em todos os métodos. A maior discordância ocorreu entre as notas de Ginecologia e Obstetrícia e Pediatria.
- Os domínios com maior peso na confiabilidade do conceito global itemizado foram: *Exame físico* (para os alunos), *Solução de problemas* (para os docentes) e

Capacidade de auto-reflexão (para os colegas). A exclusão do domínio *Hábitos de trabalho* não implicaria em uma queda da confiabilidade do instrumento de CGI.

- Os alunos/avaliadores dentro dos grupos, não os grupos em si, nem as UBS, foram o maior componente de variação das notas. Padrões de reciprocidade e constância das notas foram evidenciados, assim como efeitos de leniência e cortesia. Os alunos sentiram-se mais confortáveis e preparados para a auto-avaliação do que para a avaliação dos colegas e julgaram a auto-avaliação mais justa e valiosa do que a avaliação dos colegas.
- A maior concordância ocorreu entre as notas do Conceito Global Itemizado e da Avaliação Estruturada, que se agruparam de um outro lado, contrastando-se com as notas da Prova Teórica e do Portfólio, que se agruparam de outro, evidenciando uma distinção entre os métodos mais centrados no desempenho e os mais focados na cognição. A Avaliação Estruturada do Atendimento Clínico foi o método que apresentou notas mais elevadas, enquanto que as notas de Prova teórica foram as mais baixas.

7. Referências Bibliográficas

AAMC – Association of American Medical Colleges, Learning objectives for medical student education. Guidelines for medical education. Report I, 1998.

AAMC - Association of American Medical Colleges, The SOAP approach to conquering the competencies challenges [on-line] 2004. [acesso em fev 2008] Disponível em <http://www.aamc.org/members/gea>.

ABIM - American Board of Internal Medicine, Evaluation Tools, 2004. [on-line] [acesso em jun. 2008]. Disponível em <http://www.abim.org/residency/tools.aspx>.

ACGME - Accreditation Council for Graduate Medical Education. Outcome Project. Competencies. Versão 1.3, 1999. [on-line] [acesso em jun. 2008]. Disponível em <http://www.acgme.org/outcome>.

ACGME - Accreditation Council for Graduate Medical Education. Outcome Project. Toolbox of assessment methods. Versão 1.1, 2000. [on-line] [acesso em jun. 2008]. Disponível em <http://www.acgme.org/outcome>.

Almeida MJ Ensino médico e o perfil do profissional de saúde para o século XXI. Interface - Comunic, Saúde, Educ 1999; 4: 123-32.

Armstrong E, Parsa-Parsi R. How can physicians' learning styles drive educational planning? Acad Med 2005; 80: 680-4.

Arnold L, Willoughby TL, Calkins EV. Self-evaluation in undergraduate medical education: a longitudinal perspective. *J Med Edu* 1985; 60:21-8.

Arnold L. Assessing professional behavior: yesterday, today and tomorrow. *Acad Med* 2002; 77: 502-15.

Arnold L, Shue CK, Kritt B, Ginsburg S, Stern DT. Medical students' views on peer assessment of professionalism. *J Gen Inter Med* 2005; 20: 819-24.

Auewarakul C, Downing SM, Jaturatamrong U, Praditsuwan R. Sources of validity evidence for an internal medicine student evaluation system: an evaluative study of assessment methods. *Med Educ* 2005; 39:276-83.

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; 9:889-95.

Batenburg V, Smal JA, Lodder A, Melker RA. Are professional attitudes related to gender and medical specialty? *Med Educ*. 1999; 33:489-93.

Battistone MJ, Pendleton B, Milne C, Battistone ML, Sande MA, Hemmer PA, Shomaker TS. Global descriptive evaluations are more responsive than global numeric ratings in detecting students' progress during the inpatient portion of an internal medicine clerkship. *Acad Med* 2001; 76:S105-7.

Beauchamp G. The challenge of teaching professionalism. *Ann Acad Med* 2004; 33: 697-705.

Belar CD, Hersch LE, Rozensky RH, Brown RT, Brown RA, Hornyak LM, Sheridan EP, Reed GW. Self-assessment in clinical health psychology: a model for ethical expansion of practice. *Prof Psychol Res Pr* 2001; 32:135-41.

Bellodi P. Surgery or general medicine – a study of the reason underlying the choice of medical specialty. *São Paulo Med J.* 2004; 122:81-6.

Ben-David MF. Principles of assessment. In: DENT JA, HARDEN RM. A practical guide for medical teachers. 2nd edition. London: Elsevier, Chapter 31, p.282-92, 2005.

Bergovec M, Kuzman T, Rojnic M, Makovic A. is there grade inflation at medical schools? Case study of the Zagreb University School of Medicine. *Croat Med J.* 2003; 44:92-7.

Beyeler C, Westkamper R, Villiger PM, Aeschliman A. Self-assessment in continuous professional development: a valuable tool for individual physicians and scientific societies. *Ann Rheum Dis* 2004; 63:1684-6.

Bitran MC, Zuniga D, Lafuente MG, Viviani PG, Mena BC. Influencia da personalidade y el estilo de aprendizaje en la elección de especialidad médica. *Rev Med Chile.* 2005; 133:1191-9.

Bland KI, Isaacs G. Contemporary trends in student selection of medical specialties. *Arch Surg.* 2002; 137:259-67.

Bordenave JD, Pereira AM. Estratégias de ensino-aprendizagem. 4^a. ed. Petrópolis: Editora Vozes, 1982.

Boulet JR, Champlain AF, Mckinley DW. Setting defensible performance standards on OSCEs and standardized patient examinations. *Med Teach* 2003; 15: 245-9.

Bouriscot KAM, Roberts TE, Pell G. Standard setting for clinical competence at graduation from medical school: a comparison of passing scores across five medical schools. *Adv Health Sci Edu Theory Pract* 2006; 11:173-83.

Brasil, Ministério da Educação, Conselho Nacional de Educação. Câmara de Educação Superior. Resolução CNE/CES n.4. 7 de nov. 2001 - *Diretrizes Nacionais para Curso de Graduação em Medicina*. Brasília, legislação oficial, 2001.

Brislin RW. Back translation for cross cultural research. *J Cross Cultural Psychol* 1970; 1:185-216.

Bryan RE, Carmichael SW, Viggiano TR, Pawlina W. Assessing professionalism in early medical education: experience with peer evaluation and self-evaluation in the Gross Anatomy Course. *Ann Acad Med Singapore* 2005; 34:486-91.

Buchel TL, Edwards FD. Characteristics of effective clinical teachers. *Fam Med* 2005; 37:30-5.

Busari JO, Scherpbier AJJA, van de Vleuten CPM, Essed GGM. The perceptions of attending doctors of the role of residents as teachers of undergraduate clinical students. *Med Educ* 2003; 37:241-7.

Calegari ML, Gemignani OH. *Temperamento e Carreira*. São Paulo: Summus Editorial, 2006.

Calhoun JG, Woolliscroft JO, Hockman EM et al. Evaluating medical student clinical skill performance: relationships among self, peer and expert ratings. *Res Med Educ* 1984; 23:205-10.

Campos GWS. Papel da rede de atenção básica em saúde na formação médica – Diretrizes - Artigo elaborado sob encomenda da Associação Brasileira de Educação Médica (ABEM). Departamento de Medicina Preventiva e Social da Faculdade de Ciências Médicas da UNICAMP, abril 2005.

Carr SJ. Assessing clinical competency in medical senior house officers: how and why should we do it? *Postgrad Med J* 2004; 80:63-6.

Chapman DM, Hayden S, Sanders AB. et al. Integrating the accreditation council for graduate medical education core competencies into the model of clinical practice of emergency medicine. *Acad Emerg Med* 2004; 11:674-85.

CINAEM – Comissão Interinstitucional Nacional de Avaliação do Ensino Médico. Relatório geral 1991-1997: avaliação do ensino médico no Brasil. Brasília: CINAEM, 1997.

Collins J. Education techniques for life long learning. Principles of adult learning. *RadioGraphics* 2004; 24: 1483-9.

Compton MT, Frank E, Elon L, Carrera J. Changes in US medical students' specialty interests over the course of medical school. *J Gen Inter Med.* 2008, 23: 1095-100.

Cottrell D, Kilminster S, Jolly B, Grant J. What is effective supervision and how does it happen? A critical incident study. *Med Educ* 2002; 36:1042-9.

Crossley J, Humphris GM, Jolly B. Assessing health professionals. *Med Educ* 2002; 36:800-4.

Daelmans HE, Hoogenboom RJ, Donker AJ, Scherpbier AJ, Stehouwer CD, van der Vleuten CP. Effectiveness of clinical rotations as a learning environment for achieving competences. *Med Teach.* 2004; 26:305-12.

Daelmans HEM, van der Hem-Stokroos HH, Hoogenboom RJL, Scherpbier AJJA, Stehouwer CDA, van der Vleuten CPM. Global clinical performance rating, reliability and validity in undergraduate clerkship. *Neth J Med* 2005; 63:279-84.

Daelmans HE, Overmeer RM, van der Hem-Stokroos HH, Scherpbier AJ, Stehouwer CD, van der Vleuten CP. In-training assessment: qualitative study of effects on supervision and feedback in undergraduate clinical rotation. *Med Educ.* 2006; 40:51-8.

Danneffer EF, Henson LC, Bierer SB, Grady-Weliky TA, Meldrum S, Nofziger AC, Barclay C, Epstein RM. Peer assessment of professional competence. *Med Educ* 2005; 39:713-22.

Davis JD. Comparison of faculty, peer, self and nurse assessment of obstetrics and gynecology residents. *Obstet Gynecol* 2002; 99:647-51.

Davis MH, Harden RM. Competency-based assessment: making it a reality. *Med Teach* 2003; 25:565-8.

DeLisa JA. Evaluation of clinical competency. *Am J Phys Med Rehabil* 2000; 79:474-7.

Depresbiteris L. Certificação de competências: a necessidade de avançar numa perspectiva formativa. *Formação* 2001; 2: 27-38.

DeVellis RF. Scale development. Theory and applications. California: Sage, 1991.

DiMatteo MR, Dinicola DD. Sources of assessment of physician performance: a study of comparative reliability and patterns of intercorrelation. *Med Care* 1981; 19:829-39.

Dobbie A, Tysinger JW. Evidence-based strategies that help office-based teachers give effective feedback. *Fam Med* 2005; 37:617-9.

Driessen E, van Der Vleuten C, van Schuwirth L, Tartwijk J, Vermunt J. The use of qualitative research for portfolio assessment as an alternative to reliability evaluation: a case study. *Med Educ* 2005; 39: 214-20.

Duff P. Teaching and assessing professionalism in medicine. *Obstet Gynecol* 2004; 104: 1362-6.

Dunn G. Design and Analysis of reliability studies. London: Edward Arnold, Chapter 3, 1982.

Edelstein RA, Reid HM, Usatine R, Wilkes MS. A comparative study of measures to evaluate medical students' performances. *Acad Med* 2000; 75:825-33.

Edwards RK, Kellner KR, Sistrom CL, Magyari EJ. Medical student self-assessment of performance on an obstetrics and gynecology clerkship. *Am J Obst Gynecol* 2003; 188:1078-82.

Elliot DL, Hickam DH. Evaluation of physical examination skills. Reliability of faculty observers and patient instructors. *JAMA* 1987; 258:3405-8.

Elwyn G, Lewis M, Evans R, Hutchings H. Using a "peer assessment questionnaire" in primary medical care. *Br J Gen Pract* 2005; 55:690-5.

Ende J. Feedback in clinical medical education. *JAMA* 1983; 250:777-81.

Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA* 2002; 287:226-35.

Epstein RM. Assessment in medical education. *N Engl J Med* 2007; 356: 387-96.

Eva KW, Cunnington JPW, Reiter HI, Keane DR, Norman GR. How can I know what I don't know? Poor self assessment in a well-defined domain. *Adv Health Sci Edu* 2004; 9:211-24.

Evans R, Elwyn G, Edwards A. Review of instruments for peer assessment of physicians. *BMJ* 2004; 328:1240-3.

Farrell S. Evaluation of student performance: clinical and professional performance. *Acad Emerg Med* 2005; 12:302e6-10.

Fincher RME, Lewis LA, Jackson TW. Why students choose primary care or nonprimary care career. *Am J Med.* 1994; 97:410-7.

Fitzgerald JT, White AB, Gruppen LD. A longitudinal study of self-assessment accuracy. *Med Educ* 2003; 37:645-9.

Fraga CF. Ensino Médico: bases e diretrizes para sua reformulação. *Rev Bras Educ Med* 1986; 10:34-6.

Gagliete M, Barbosa MLS. A Brasilidade no entre-lugar: leituras de Dyonélio Machado e Sérgio Buarque de Holanda. *Letras de Hoje* 2006; 41:29-36.

Garyfallos GK, Karastergiou A, Adamopoulou A, Moutzoukis C, Alagiozi E, Mala D. et al. Greek version of the general health questionnaire: Accuracy of translation and validity. *Psych Scand* 1991; 84:371-8.

Gordon J, Hazlett C, Cate O, Mann K, Kilminster S, Prince K et al. Strategic planning in medical education: enhancing the learning environment for students in clinical settings. *Med Educ* 2000; 34:841-50.

Gordon MJ. A review of the validity and accuracy of self-assessments in health professions training. *Acad Med* 1991; 66:762-9.

Gordon MJ. One to one teaching and feedback. *BMJ* 2003; 326:543-5.

Govaerts MJB, van der Veluten CPM, Schuwirth LWT, Muijtjens AMM. Broadening perspectives on clinical performance assessment: rethinking the nature of in-training assessment. *Adv Health Sci Edu Theory and Pract* 2007;12:239-60.

Gruppen LD, White CJ, Fitzgerald T, Grum CM, Wooliscroft JO. Medical students' self-assessments and their allocations of learning time. *Acad Med* 2000; 75: 374-9.

Haes JCJM, Oort FJ, Hulman RL. Summative assessment of medical students' communication skills and professional attitudes through observation in clinical practice. *Med Teach* 2005; 27:583-9.

Haidet P, Dains JE, Paterniti DA, Hechet L, Chang T, Tseng E, et al. Medical students' attitudes toward the doctor-patient relationship. *Med Educ.* 2002; 36:568-74.

Hair JFJ, Anderson RE, Tatham RL, Black WC. *Multivariate Data Analysis.* 5 ed. New Jersey: 1998.

Hamdy H, Prasad K, Anderson MB, Scherpbier A, Williams R, Zwierstra R et al. BEME Systematic Review: Predictive values of measurement obtained in medical schools and future performance in medical practice BEME 5. *Med Teach.* 2006;28:103-16. [on line] [acesso em set. 2008]. Disponível em <http://www.bemecollaboration.org/beme/pages/published.html>.

Hanley E, Higgins A. Assessing of clinical practice in intensive care: a review of literature. *Intensive Crit Care Nurs* 2005; 21:268-75.

Haydt RC. *Avaliação do Processo Ensino-Aprendizagem.* 6ª ed. São Paulo: Editora Ática, 2002.

Hayes BE. *Medindo a satisfação do cliente. Desenvolvimento e uso dos questionários.* Tradução de Luiz Liske. Rio de Janeiro: Qualitymark Ed., 2001.

Henderson P, Johnson MH. An innovative approach to developing the reflective skills of medical students. *BMC Med Educ;* 2, 2002. [on-line] [acesso em mai. 2006]. Disponível em <http://www.biomedcentral.com/1472-6920/2/4>.

Henderson P, Ferguson-Smith AC, Johnson MH. Developing essential professional skills: a framework for teaching and learning about feedback. *BMC Med Educ,* 2005. [on-line] [acesso em mai. 2006]. Disponível em <http://www.biomedcentral.com/1472-6920/5/11>.

Holmboe ES. Faculty and the observation of trainees' clinical skills: problems and opportunities. *Acad Med* 2004; 79:16-22.

Holmboe ES, Hawkins RE, Huot SJ. Effects of training in direct observation of medical residents' clinical competence. *Ann Intern Med* 2004; 140:874-81.

Horowitz SD. Evaluation of clinical competencies Basic certification, subspecialty certification, and recertification. *Am J Phys Med Rehabil* 2000; 79:478-80.

Hulsman RL, Mollema ED, Hoos AM, De Haes, JCJM, Donnison-Speijer JDD. Assessment of medical communications skills by computer: assessment method and student experiences. *Med Educ* 2004; 38:813-24.

Jackson A, Blaxter L, Lewando-Hundt G. Participating in medical education: views of patients and carers living in deprived communities. *Med Educ* 2003; 37:532-8.

Jerome PJ. Coaching through effective feedback. California: Richard Chang Associates Inc., 1994.

Johnson RA, Wichern DW. Factor analysis and inference for structured covariance matrices In: Johnson, RA; Wichern, DW. Applied multivariate statistical analysis. 3rd edition. New Jersey: Prentice Hall International, p.396-457, 1992.

Kassler WJ, Wartman SA, Silliman RA. Why medical students choose primary care careers. *Acad Med*. 1991; 66:41-43.

Katz TN, Mccarty-Gillespie L, Magrane DM. Direct observation as a tool for needs assessment of resident teaching skills in the ambulatory setting. *Am J Obstet Gynecol* 2003; 189:684-7.

Kilminster SM, Jolly BC. Effective supervision in clinical practice settings: a literature review. *Med Educ* 2000; 34:27-40.

Kilminster S, Cottrell D, Grant J, Jolly B. AMEE Guide n° 27: Effective educational and clinical supervision. *Med Teach* 2007; 29:2-19.

Kim JO, Mueller CW. Factor Analysis. Statistical Methods and Practical Issues. Beverly Hills: Sage, 1978.

Klass D. Reevaluation of clinical competency. Am J Phys Med Rehabil 2000; 79:481-6.

Koens F, Rademakers JJDMJ, Cate OTJT. Validation of core medical knowledge by postgraduates and specialists. Med Educ. 2005; 39:911-7.

Krueger J, Stanke D. The role of self-reference and other-reference knowledge in perceptions of group characteristics. J Pers Soc Psychol 2001; 27:878-88.

Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. J Pers Soc Psychol 1999; 77:121-34.

Lachman N, Pawlina W. Integrating professionalism in early medical education: The theory and application of reflective practice in the anatomy curriculum. Clin Anat 2006; 19:56-60.

Langedyk V. Not knowing that they do not know: self-assessment accuracy if third-year medical students. Med Educ 2006; 40:173-9.

Larrick RP, Burson KA, Soll JB. Social comparison and competence: when thinking you're better than average predicts overconfidence (and when it does not). Organization Behavior and Human Decision Processes 2007; 102:76-94.

Lehmann F, Côté L, Bourque A, Fontaine D. Physician-patient interaction: a reliable and valid check-list of quality. Can Fam Physician 1990; 36:1711-6.

Levine RE, O'boyle M, Haidet P, Lynn DJ, Stone MM, Wolf DV, Paniagua FA. Transforming a clinical clerkship with team learning. Teach Learn Med 2004; 16:3:270-5.

Levine RE, Kelly PA, Karakoc T, Haidet P. Peer evaluation in a clinical clerkship: students' attitudes, experiences, and correlations with traditional assessments. *Acad Psychiatry* 2007; 31:19-24.

Lucas JH, Stallworth JR. Providing difficult feedback: TIPS for the problem learner. *Fam Med* 2003; 35:544-6.

Lurie SJ, Nofziger AC, Meldrum S, Mooney C, Epstein TM. Temporal and group-related trends in peer assessment amongst medical students. *Med Educ* 2006; 40:840-7.

Lurie SJ, Lambert DR, Nofziger AC, Epstein RM, Grady-Weliky TA. Relationship between peer assessment during medical school, dean's letter rankings, and ratings by internship directors. *Soc Gen Inter Med* 2007; 22:13-6.

Lynch DC, Swing SR. Key considerations for selecting instruments and implementing assessment systems. ACGME Research Department. [on-line] [acesso em abr. 2006]. Disponível em: [http:// www.acgme.org/outcome/assess/keyconsider.asp](http://www.acgme.org/outcome/assess/keyconsider.asp).

Mamede S, Schmidt HG. Correlates of reflective practice in medicine. *Adv Health Sci Educ Theory Pract* 2005; 10:327-37.

Marks M, Humprey-Murto S. Chapter 35. Performance assessment. In: Dent, JA, Harden, RM. *A practical guide for medical teachers*. 2nd edition. London: Elsevier, p.323-35, 2005.

Mattheos N, Nattestad A, Falk-Nilsson E, Attstrom R. The interactive assessment: assessing students' self-assessment ability. *Med Educ* 2004; 38:378-89.

Mazor KM, Zanetti ML, Alper EJ, Hatem D, Barrett S, Meterko V. et al. Assessing professionalism in the context of an objective-structured clinical examination: an in-depth study of the rating process. *Med Educ* 2007; 41:331-40.

Mcaleer S. Choosing assessment instruments. In: Dent JA & Harden RM. A practical guide for medical teachers. 2nd edition. Philadelphia: Elsevier Limited, 2005.

Mckinstry BH, Cameron HS, Elton RA, Riley SC. Leniency and halo effects in marking undergraduate short research projects. BMC Med Educ 2004; 4:28 [on-line] [acesso em abr. 2008]. Disponível em <http://www.biomedcentral.com/1472-6920/4/28>

McKinstry B. A systematic review of the literature on the effectiveness of self-assessment in clinical education. BEME 10, 2007 [on line] [acesso em set. 2008]. Disponível em <http://www.bemecollaboration.org/beme/pages/published.html>

Mcmullan M, Endacott R, Gray MA, Jasper M, Miller CML, Schole SJ, Webb C. Portfolios and assessment of competence: a review of the literature. J Adv Nurs 2003; 41:283-94.

Menachery EP, Knight AM, Kolodner K, Wright SM. Physician characteristics associated with proficiency in feedback skills. J Gen Intern Med 2006; 21:440-6.

Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990; Suppl 9:S63-7.

Misch DA. Evaluating physicians' professionalism and humanism: the case for humanism "connoisseurs". Acad Med 2002; 77:489-95.

Moore DA, Small DA. Error and bias in comparative judgment: on being both better and worse than we think we are. J Pers Soc Psychol 2007; 92:972-89.

Morgan PF, Cleave-Hogg D, Guest CB. A comparison of global ratings and checklist scores from an undergraduate assessment using an anesthesia simulator. Acad Med 2001; 76:1053-5.

- Morton JB, Macbeth WAAG. Correlations between staff, peer and self assessments of fourth-year students in surgery. *Med Educ* 1977; 11:167-70.
- Moss AC, Dugal T, Silke B. Attitudes to peer review as a competence assurance structure – results of a survey of Irish physicians. *Ir J Med Sci* 2005; 174:43-6.
- Murray E, Gruppen L, Catton P, Hays R, Woolliscroft JO. The accountability of clinical education: its definition and assessment. *Med Educ* 2000; 34:871-9.
- Myers IB, MacCaulley M, Queen N, Hammer A. *MBTI Manual: a guide in the development and use of the Myers-Briggs Type Indicator*. Palo Alto. California: Consulting Psychologist Press Inc. 1998.
- Nahum GG, Facs F. Evaluating medical student obstetrics and gynecology clerkship performance: which assessment tools are most reliable? *Am J Obst Gyn.* 2004; 191:1762-71.
- Nendaz MR, Perrier A, Simonet ML, Huber P, Junod A, Vu NV. Appraisal of clinical competence during clerkship: how knowledgeable in curriculum and assessment development should a physician-examiner be? *Acad Med* 2001; 76:S99-101.
- Newble DI. Assessment of clinical competence – Editorial II. *Br J Anaesth* 2000; 84:432-3.
- Newble DI, Jaeger K. The effect of assessment on the learning of medical students. *Med Educ* 1983; 17:165-71.
- Noel GL, Herbers JE, Caplow MP, Cooper GS, Pangaro LN, Harvey J. How well do internal medicine faculty members evaluate the clinical skills of residents? *Ann Inter Med* 1992; 117:757-65.
- Norcini JJ. Peer assessment of competence. *Med Educ* 2003; 37:539-43.

Norcini JJ, Boulet J. Methodological issues in the use of standardized patients for assessment. *Teach Learn Med* 2003; 15:293-7.

Norcini JJ. Current perspectives in assessment: the assessment of performance at work. *Med Educ* 2005; 39:880-9.

Norcini JJ, Mckinley DW. Assessment methods in medical education. *Teaching and Teacher Education* 2007; 23:239-50.

Norgaard K, Ringsted C, Domans D. Validation of a checklist to assess ward round performance of internal medicine. *Med Educ* 2004; 38:700-7.

Norman GR, Shannon SI, Marrin ML. The need for needs assessment in continuing medical education. *BMJ* 2004; 328: 999-1001.

Nuovo J, Bertakis KD, Azari R. Assessing resident's knowledge and communication skills using four different evaluation tools. *Med Educ* 2006; 40:630-6.

Ozuah PO. Undergraduate medical education: thoughts on the future. *BMC Med Educ* 2, 2002. [on line] [acesso em mai 2008]. Disponível em <http://www.biomedcentral.com/1472-6920/2/8>.

Papa FJ, Aldrich D, Schumacker RE. The effect of immediate online feedback upon diagnostic performance. *Acad Med* 1999; 74:S16-8.

Papinczak T, Young L, Groves M, Haynes M. An analysis of peer, self and tutor assessment in a problem-based learning tutorials. *Med Teach* 2007; 29:e122-32.

Park B, Judd CM. Measures and models of perceived group variability. *J Pers Soc Psychol* 1990; 59,173-91.

Pendleton D, Scofield T, Tate P, Havelock P. *The consultation: an approach to learning and teaching*. Oxford; Oxford University Press, 1984.

Prideaux D, Alexander H, Bower A, Dacre J, Haisr S, Jolly B, et al. Clinical teaching: maintaining an educational role for doctors in the new health care environment. *Med Educ* 2000, 34:820-6.

Pulito AR, Donnelly MB, Plymale M. Factors in faculty evaluation of medical students' performance. *Med Educ* 2007; 41:667-75.

Ramos MN. Qualificação, competências e certificação: visão educacional. *Formação* 2001; 2:17-26.

Ramsey PG, Wenrich MD, Carline JD, Inui TS, Larson EB, Logerfo JP. Use of peer ratings to evaluate physician performance. *JAMA* 1993; 269:1655-60.

Ramsey PG, Carline JD, Blank LL, Wenrich MD. Feasibility of hospital-based use of peer ratings to evaluate the performances of practicing physicians. *Acad Med* 1996; 71:364-70.

Ramsey PG, Wenrich MD. Peer Ratings: an assessment tool whose time has come. *J Gen Intern Med* 1999; 14:581-2.

Ramsey W, Owen C. Is there a role for peer review in performance appraisal of medical students? *Med Educ* 2006; 40:95-6.

Rees C, Shepherd M. Students' and assessors' attitudes towards students' self-assessment of their personal and professional behaviors. *Med Educ* 2005; 39:30-9.

Regehr G, Macrae H, Reznik RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance of an OSCE-format examination. *Acad Med* 1998; 73:993-7.

Rethans JJ, Norcini JJ, Barón-Maldonado M, Blackmore D, Jolly BC, Laduca T, Lew S, Page GG, Southgate LH. The relationship between competence and performance: implications for assessing practice performance. *Med Educ* 2002; 36:901-9.

Ribeiro C. Metacognição: um apoio ao processo de aprendizagem. *Psicologia: Reflexão e Crítica* 2003; 16:109-16.

Ribeiro, MMF, Amaral, CFS. Medicina centrada no paciente e ensino médico: a importância do cuidado com a pessoa e o poder do médico. *Rev Bras Edu Med*. 2008; 32:90-7.

Risucci DA, Tortolani AJ, Ward RJ. Ratings of surgical residents by self, supervisors and peers. *Surg Gynecol Obstet* 1989; 169:519-26.

Rodgers KG, Manifold C. 360-degree feedback: possibilities for assessment of the ACGME core competencies for emergency medicine residents. *Acad Emerg Med* 2002; 9:1300-4.

Rolfe IE, Sanson-Fisher RW. Translating learning principles into practice: a new strategy for learning clinical skills. *Med Educ* 2002; 36:345-52.

Rotenberg BW, Woodhouse RA, Gilbert M, Hutchison CR. A needs of assessment of surgical residents as teachers. *CJS* 2000; 43:295-300.

Rudy DW, Fejfar MC, Griffith CH, Wilson JF. Self- and peer assessment in a first-year communication and interviewing course. *Eval Health Prof* 2001; 24:436-45.

Sande GN, Goethals GR, Radloff CE. Perceiving one's own traits and others': the multifaceted self. *J Pers Soc Psychol* 1988; 54:13-20.

Salerno MP, O'Malley PG, Pangaro LN, Wheeler GA, Moores KL, Jackson LJ. Faculty development seminars based on the one-minute preceptor improve feedback in the ambulatory setting. *J Gen Inter Med* 2002; 17:779-87.

Scheffé H. *The Analysis of Variance*. New York: Wiley, Chapters 5 and 7, 1959.

Schultz KW, Kirby J, Delva D, Godwin M, Verma S, Birtwhistle R, Knapper C, Seguin R. Medical students' and residents' preferred site characteristics and preceptor behaviours for learning in the ambulatory setting: a cross-sectional survey. BMC Med Educ [on-line] [acesso em abr. 2006]. Disponível em <http://www.biomedcentral.com/1472-6920/4/12>.

Schuwirth LWT, van der Vleuten CPM. Changing Education, changing assessment, changing research? Med Educ 2004; 38:805-12.

Searle J. Defining competency – the role of standard setting. Med Educ 2000; 34:363-6.

Shue CK, Arnold L, Stern DT. Maximizing participation in peer assessment of professionalism: the students speak. Acad Med 2005; 80:51-5.

Silber GC, Nasca TJ, Paskin DI, Eiger G, Robeson M, Veloski JJ. Do global rating forms enable program directors to assess the ACGME competencies? Acad Med 2004; 79:549-56.

Sliwa JA, Kowalske KJ. Assessing resident clinical competence. Am J Phys Med Rehabil 2000; 79: 468-73.

Smee S. Skill based assessment. BMJ 2003; 326:703-6.

Smith ST, Dollase RH, Boss JA. Assessing students' performances in a competency-based curriculum. Acad Med 2003; 78:97-107.

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; 5:131-40.

Southgate L. Professional competence in medicine. Hosp Med 1999; 60:203-5.

Southgate L, Hays RB, Norcini J, Mulholland H, Ayres B, Woolliscroft J, Cusimano M, Mcavoy P, Ainsworth M, Haist S, Campbell M. Setting performance standards for medical practice: a theoretical framework. *Med Educ* 2001; 35:474-81.

Souza MPG, Rangel M. Avaliação: um Impasse na Educação Médica. *Rev Bras Edu Med* 2003; 2:213-22.

Speer AJ, Solomon DJ, Fincher RM. Grade inflation in internal medicine clerkship: results of a national survey. *Teach Learn Med.* 2000; 12:112-6.

Srinivasan M, Hauer KE, Der-Martirosian C, Wilkes M, Gesundheit N. Does feedback matter? Practice-based learning for medical students after a multi-institutional clinical performance examination. *Med Educ* 2007; 41:857-65.

Staggers N, Gassert CA, Skiba D. Health professionals' views of informatics education. *J Am Med Inform Assoc* 2000; 7:550-8.

Stedile NLR, Friendlander MR. Metacognição e o ensino de enfermagem: uma combinação possível? *Rev Latino-am Enfermagem* 2003; 11:792-9.

St George I, Kaigas T, Mcavoy P. Assessing the competence of practicing physicians in New Zealand, Canada and the United Kingdom: progress and problems. *Fam Pract* 2004; 36:172-7.

Sturmberg JP, Atkinson K, Farmer EA. Attainment and maintenance of professional capabilities. *Aust Fam Physician* 2005; 34:371-3.

Sullivan ME, Hitchcock MA, Dunnington MD. Peer and self-assessment during problem-based tutorials. *American Journal of Surgery* 1999; 177:266-9.

Suls J, Lemos K, Stewart L. Self-esteem, construal, and comparison with self, friends and peers. *Journal of Personality and Social Psychology* 2002; 82:252-61.

Swick HM. Toward a normative definition of medical professionalism. *Acad Med* 2000; 75:612-6.

Swick S, Hall S, Beresin E. Assessing the ACGME competencies in psychiatry training programs. *Academic Psychiatry* 2006, 30:330-51.

Thomas PA, Gebo KA & Hellmann DB. A pilot study of peer review in residency training. *J Gen Inter Med* 1999; 14:551-4.

Thorndike EL. *Mental and Social Measurements*. Teachers College, Columbia University: New York, 1912.

Troncon LEA. Significance of experts' overall ratings for medical student competence in relation to history-taking. *São Paulo Med J* 2006; 124:101-4.

Troncon LEA. Carta ao leitor. *Rev Bras Educ Med* 2008; 32:132-3.

Tukey JW. *Exploratory Data Analysis*, Addison-Wesley: Reading MA, 1970.

van Bernardin C, Orban JA, Carlyle JJ. Performance ratings as a function of trust in appraisal and rater individual differences. *Academy of Management Proceeding* 1981; 311-5.

van der Vleuten CPM. Validity of final examinations in undergraduate medical training. *BMJ* 2000; 321:1217-9.

van der Vleuten CPM, Schuwirth LWT. Assessing professional competence: from methods to programmes. *Med Educ* 2005; 39:309-17.

van Rosendaal GMAV, Jennett PA. Comparing peer and faculty evaluations in an internal medicine residency. *Acad Med* 1994; 69:299-303.

Veloski J, Boex JR, Grasberger MJ, Evans A, Wolfson DB. Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No. 7. *Med Teach* 2006; 28:117-28.

Violato C, Lockyer J. Self and peer assessment of pediatricians, psychiatrists and medicine specialists: implications for self-directed learning. *Adv Health Sci Edu* 2006; 11:235-44.

Vozenilek J, Huff JS, Reznek M, Gordon JA. See one, do one, teach one: advanced technology in medical education. *Acad Emerg Med* 2004; 11:1149-54.

Wang E, Sanders AB. Global assessment tool for emergency medicine – specific core competency evaluation. *Acad Emerg Med* 2004; 11:1370-1.

Ward M, Gruppen L, Regehr G. Measuring self-assessment: current state of art. *Adv in Health Sciences Education* 2002; 7:63-80.

Wass V, van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *The Lancet* 2001; 357:945-9.

Wass V, van der Vleuten C. The long case. *Med Educ* 2004; 38: 1176-80.

Wendling A, Hoekstra L. Interactive peer review: an innovative resident evaluation tool. *Fam Med* 2002; 34:738-43.

Wheeler DW, Whittlestone KD, Smith HL, Gupta AK, Menon DK. A web-based system for teaching, assessment and examination of the undergraduate peri-operative medicine curriculum. *Anaesthesia* 2003; 58:1079-86.

Whelan GP, Boulet JR, Mckinley DW, Norcini JJ, Van Zantem M, Hambleton RK, Burdik WP, Peitzman SJ. Scoring standardized patient examinations: lessons learned from the development and administration of ECFMG clinical skills assessment (CSA). *Med Teach* 2005; 27:200-6.

Wilkinson TJ, Frampton CM. Assessing performance in final year medical students. Can a postgraduate measure be used in an undergraduate setting? Med Educ 2003; 37:233-40.

Wilkinson TJ, Frampton CM. Comprehensive undergraduate medical assessments improve prediction of clinical performance. Med Educ 2004; 38:1111-6.

Wooliscroft JO, Tenhaken J, Smith J, Calhoun JD. Medical students' clinical self-assessments: comparison with external measures of performance and the students' self-assessments of overall performance and effort. Acad Med 1993; 68: 285-94.

World Federation for Medical Education. [Basic Medical Education, WFME Global Standards for Quality Improvement](#), 2001. [acesso em ago 2008]. Disponível em [http: www.wfme.org](http://www.wfme.org).

8. Anexos

8.1. Anexo 1 – Objetivos do Módulo Atenção Integral à Saúde

Objetivo geral:

Contribuir para a formação geral do médico, pelo desenvolvimento do raciocínio clínico, da compreensão do processo diagnóstico e terapêutico e da prática da relação médico-paciente, em situações de atendimento primário em saúde da criança e do adolescente, da mulher, do adulto e do idoso(a).

Objetivos específicos:

Ao final do estágio, o aluno deve estar apto a:

1. Realizar anamnese e exame físico completo do paciente, registrando as informações de modo claro e ordenado e valorizando os dados relevantes para cada caso;
2. Indicar condutas diagnósticas e terapêuticas para as condutas mais comuns na prática médica;
3. Realizar prescrições sob supervisão, reconhecendo a importância da adequação da dieta e das doses dos medicamentos à idade e à massa corporal do paciente;
4. Integrar os processos diagnóstico, terapêutico e propostas de seguimento para os casos que acompanhe;
5. Capacitar atendimento domiciliar e elaboração de projeto terapêutico interprofissional;

6. Acompanhar e registrar a evolução de uma criança normal (puericultura) ou doente, reconhecendo quais as informações que devem constar das anotações para o acompanhamento do caso;
7. Habilitar o aluno a realizar propedêutica específica ginecológica, mamária e obstétrica, incluindo interpretação dos exames subsidiários mais utilizados na especialidade;
8. Habilitar o aluno a exercitar o raciocínio clínico em ginecologia e obstetrícia, utilizando elementos de anamnese, exame clínico e subsidiários e considerando as inter-relações com o ambiente onde a mulher está inserida;
9. Habilitar o estudante para a atenção integral à mulher, incluindo educação para a saúde, nas situações ou síndromes clínicas mais prevalentes em atenção primária à saúde: gestante de baixo risco, mulher na menopausa ou perimenopausa e mulher com queixas ginecológicas ou buscando assistência em saúde sexual e reprodutiva.

8.2. Anexo 2 – Modelo de Portfólio

PORTFÓLIO

1- Incluir as planilhas de atendimento

2- Relatório do Caso Clínico (1 para GO, 1 para CM e 1 para Ped)

DATA DO ATENDIMENTO _____/_____/_____

DOCENTE E/OU TUTOR COM QUEM DISCUTIU _____

INICIAIS DO PACIENTE _____ **FF/FI** _____

QUEIXA PRINCIPAL E DURAÇÃO _____

Transcrever o atendimento escolhido para ser objeto do portfólio (anamnese completa, com exame físico, HD e conduta). Ressaltar os aspectos relevantes. Discutir como foram realizadas todas as HD com argumentos baseados na literatura.

Fonte bibliográfica(s) consultada(s).

Faça uma avaliação crítica sobre sua anamnese, exame físico, hipóteses diagnósticas e conduta proposta baseada em leitura de livro-texto, salientando o que poderia ter sido melhor abordado ou suprimido e por quê.

Auto-crítica de aprendizado.

Referente ao universo de casos atendidos (planilhas de registro de casos)

Referente às dificuldades enfrentadas e os avanços alcançados

Identificação de carências e propostas de resolução (incluindo procedimentos propedêuticos clínicos)

Avaliação do portfólio

Data ____/____/____

	CM	GO	Ped	
Organização, apresentação (0-1)				
Completude (0-1)				
Apresentação dos 3 casos clínicos (0-6)				
Auto-crítica do aprendizado (0-2)				
Nota				MÉDIA
Docente				

Comentários para orientação:**Aos docentes e tutores da unidade**

Ao aluno

8.3. Anexo 3 – Avaliação Estruturada GO

ALUNO: _____

RA: _____

DOCENTE: _____

UBS: _____

DATA: ____/____/____

CONSULTA: () Caso novo () Retorno INÍCIO: _____

TÉRMINO: _____

Como é o desempenho /atitude do aluno nesta consulta em relação a: (NA = Não se aplica)

1. Aparência pessoal: adequação da vestimenta, demonstração de higiene e organização pessoal.	0 1 2 3 4 5 6 7 8 9 10 ruim excelente	NA
Comentários, justificativas e sugestões de melhoria:		
2. Rapport: cumprimentou paciente e acompanhantes, apresentou-se pelo nome, tratou paciente pelo nome.	0 1 2 3 4 5 6 7 8 9 10 ruim excelente	NA
Comentários, justificativas e sugestões de melhoria:		
3. Início da consulta: tranquilizou o paciente, adotou postura acolhedora, deixou paciente à vontade.	0 1 2 3 4 5 6 7 8 9 10 ruim excelente	NA
Comentários, justificativas e sugestões de melhoria:		
ANAMNESE		
4. História: permitiu paciente e/ou acompanhante contar história com as próprias palavras, antes de intervir e orientar a anamnese?	0 1 2 3 4 5 6 7 8 9 10 inadequada adequada	NA
Comentários, justificativas e sugestões de melhoria:		
5. Conteúdo da história clínica - tópicos importantes:	SIM	NÃO
data da última menstruação, menopausa		
nº gestações, partos vaginais, abortos e sectárias		
data da menarca		
método anticoncepcional utilizado atualmente		
número de parceiros sexuais durante a vida?		
antecedentes de doenças sexualmente transmissíveis, inclusive HIV/AIDS, da paciente e seus parceiros?		
quando paciente realizou último Papanicolaou (procurou confirmar?)		
antecedentes clínicos ou cirúrgicos		
queixas mamárias		
data da última mamografia		
resultado da última mamografia		
antecedentes familiares de câncer, em especial da mama e ovários?		
Nota geral para qualidade das perguntas feitas acima.	0 1 2 3 4 5 6 7 8 9 10 ruim excelente	NA
Comentários, justificativas e sugestões de melhoria:		

Em caso de climatério, anticoncepção ou alterações menstruais, complementar com avaliação específica

EXAME FÍSICO												
6. Orientações ao paciente: tranquilizou a paciente para o exame ginecológico.	0	1	2	3	4	5	6	7	8	9	10	NA
	ruim excelente											
	Comentários, justificativas e sugestões de melhoria:											
7. Verificação de: PA, ausculta cardíaca, palpação abdominal.	0	1	2	3	4	5	6	7	8	9	10	NA
	ruim excelente											
	Comentários, justificativas e sugestões de melhoria:											
8. Inspeção estática das mamas.	0	1	2	3	4	5	6	7	8	9	10	NA
	inadequada adequada											
	Comentários, justificativas e sugestões de melhoria:											
9. Inspeção dinâmica das mamas.	0	1	2	3	4	5	6	7	8	9	10	NA
	inadequada adequada											
	Comentários, justificativas e sugestões de melhoria:											
10. Palpação das axilas.	0	1	2	3	4	5	6	7	8	9	10	NA
	inadequada adequada											
	Comentários, justificativas e sugestões de melhoria:											
11. Palpação das mamas.	0	1	2	3	4	5	6	7	8	9	10	NA
	inadequada adequada											
	Comentários, justificativas e sugestões de melhoria:											
12. Colocou paciente em posição ginecológica.	0	1	2	3	4	5	6	7	8	9	10	NA
	inadequada adequada											
	Comentários, justificativas e sugestões de melhoria:											
13. Inspeção dos genitais externos.	0	1	2	3	4	5	6	7	8	9	10	NA
	inadequada adequada											
	Comentários, justificativas e sugestões de melhoria:											
14. Introdução do espéculo.	0	1	2	3	4	5	6	7	8	9	10	NA
	inadequada adequada											
	Comentários, justificativas e sugestões de melhoria:											
15. Coleta do exame de citologia da ectocérvice e endocérvice.	0	1	2	3	4	5	6	7	8	9	10	NA
	inadequada adequada											
	Comentários, justificativas e sugestões de melhoria:											
16. Colocação do fixador na lâmina.	0	1	2	3	4	5	6	7	8	9	10	NA
	inadequada adequada											
	Comentários, justificativas e sugestões de melhoria:											

17. Aplicação do ácido acético no colo uterino e paredes vaginais.	0 1 2 3 4 5 6 7 8 9 10 inadequada adequada	NA
	Comentários, justificativas e sugestões de melhoria:	
18. Realização do teste de Schiller no colo e nas paredes vaginais.	0 1 2 3 4 5 6 7 8 9 10 inadequada adequada	NA
	Comentários, justificativas e sugestões de melhoria:	
19. Retirada do espécuro.	0 1 2 3 4 5 6 7 8 9 10 inadequada adequada	NA
	Comentários, justificativas e sugestões de melhoria:	
20. Toque genital bi manual.	0 1 2 3 4 5 6 7 8 9 10 Inadequado adequado	NA
	Comentários, justificativas e sugestões de melhoria:	
PREENCHIMENTO DOS FORMULÁRIOS E ORIENTAÇÕES AO PACIENTE		
21. Preenchimento do prontuário.	0 1 2 3 4 5 6 7 8 9 10 ruim excelente	NA
	Comentários, justificativas e sugestões de melhoria:	
22. Preenchimento da planilha de atendimento?	0 1 2 3 4 5 6 7 8 9 10 ruim excelente	NA
	Comentários, justificativas e sugestões de melhoria:	
23. Realização da hipótese diagnóstica.	0 1 2 3 4 5 6 7 8 9 10 inadequada adequada	NA
	Comentários, justificativas e sugestões de melhoria:	
24. Conduta proposta.	0 1 2 3 4 5 6 7 8 9 10 inadequada adequada	NA
	Comentários, justificativas e sugestões de melhoria:	
25. Explicação das hipóteses à paciente	0 1 2 3 4 5 6 7 8 9 10 ruim excelente	NA
	Comentários, justificativas e sugestões de melhoria:	
26. Fornecimento e explicação da receitas, pedidos de exames ou avaliações.	0 1 2 3 4 5 6 7 8 9 10 inadequado adequado	NA
	Comentários, justificativas e sugestões de melhoria:	
27. Linguagem usada nas orientações quanto à HD e em dúvidas ou perguntas feitas pela paciente.	0 1 2 3 4 5 6 7 8 9 10 inadequada adequada	NA
	Comentários, justificativas e sugestões de melhoria:	

28. Orientação sobre o uso de Conduz p/ prevenção de HIV/DST?	<u>0 1 2 3 4 5 6 7 8 9 10</u> inadequada adequada Comentários, justificativas e sugestões de melhoria:	NA
29. Abertura dada à paciente para fazer perguntas sobre suas dúvidas.	<u>0 1 2 3 4 5 6 7 8 9 10</u> ruim excelente Comentários, justificativas e sugestões de melhoria:	NA
30. Empatia: valorização das preocupações, atitudes e reações emocionais da paciente em relação ao problema que a fez buscar o ginecologista.	<u>0 1 2 3 4 5 6 7 8 9 10</u> inadequada adequada Comentários, justificativas e sugestões de melhoria:	NA
31. Identificação de possíveis dificuldades para que a paciente siga a conduta.	<u>0 1 2 3 4 5 6 7 8 9 10</u> ruim excelente Comentários, justificativas e sugestões de melhoria:	NA
32. Término da consulta: orientou sobre retorno agendado, despediu-se do paciente e acompanhantes.	<u>0 1 2 3 4 5 6 7 8 9 10</u> inadequada adequada Comentários, justificativas e sugestões de melhoria:	NA

8.4. Anexo 4 – Avaliação Estruturada CM

ALUNO: _____

RA: _____

DOCENTE: _____

UBS: _____

DATA: ____/____/____

CONSULTA: () Caso novo () Retorno INÍCIO: _____

TÉRMINO: _____

Como você avalia o desempenho do aluno nesta consulta em relação a: (NA = Não se aplica)

1. Aparência pessoal: adequação da vestimenta, demonstração de higiene e postura.	0 1 2 3 4 5 6 7 8 9 10	NA	
	ruim excelente Comentários, justificativas e sugestões de melhoria		
2. Início da consulta: cumprimentou paciente e acompanhantes, apresentou-se pelo nome, tratou paciente pelo nome, tranquilizou o paciente, adotou postura acolhedora, deixou paciente à vontade.	0 1 2 3 4 5 6 7 8 9 10	NA	
	ruim excelente Comentários, justificativas e sugestões de melhoria		
ANAMNESE			
3. História: permitiu paciente e/ou acompanhante contar história com as próprias palavras, antes de intervir e orientar a anamnese, colheu a história de forma sistematizada (QPD, HPMA, IS)?	0 1 2 3 4 5 6 7 8 9 10	NA	
	inadequada adequada Comentários, justificativas e sugestões de melhoria		
4. Conteúdo da história clínica - tópicos importantes:	SIM	NÃO	NA
Antecedentes pessoais			
Medicação em uso			
Antecedentes familiares			
Revisão dos resultados de exames laboratoriais			
Nota geral para qualidade da história	0 1 2 3 4 5 6 7 8 9 10	NA	
	ruim excelente Comentários, justificativas e sugestões de melhoria		

EXAME FÍSICO																							
5. Orientações ao paciente: explicou ao paciente o exame físico	<table border="1"> <tr> <td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td> </tr> <tr> <td>ruim</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>excelente</td> </tr> </table> Comentários, justificativas e sugestões de melhoria	0	1	2	3	4	5	6	7	8	9	10	ruim										excelente
0	1	2	3	4	5	6	7	8	9	10													
ruim										excelente													
6. Verificação de Peso/ altura, PA, FC, PR	<table border="1"> <tr> <td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td> </tr> <tr> <td>ruim</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>excelente</td> </tr> </table> Comentários, justificativas e sugestões de melhoria	0	1	2	3	4	5	6	7	8	9	10	ruim										excelente
0	1	2	3	4	5	6	7	8	9	10													
ruim										excelente													
7. Exame físico geral.	<table border="1"> <tr> <td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td> </tr> <tr> <td>Inadequado</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>adequado</td> </tr> </table> Comentários, justificativas e sugestões de melhoria	0	1	2	3	4	5	6	7	8	9	10	Inadequado										adequado
0	1	2	3	4	5	6	7	8	9	10													
Inadequado										adequado													
8. Exame físico CV.	<table border="1"> <tr> <td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td> </tr> <tr> <td>inadequado</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>adequado</td> </tr> </table> Sugestões de melhoria:	0	1	2	3	4	5	6	7	8	9	10	inadequado										adequado
0	1	2	3	4	5	6	7	8	9	10													
inadequado										adequado													
9. Exame físico respiratório	<table border="1"> <tr> <td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td> </tr> <tr> <td>inadequado</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>adequado</td> </tr> </table> Comentários, justificativas e sugestões de melhoria	0	1	2	3	4	5	6	7	8	9	10	inadequado										adequado
0	1	2	3	4	5	6	7	8	9	10													
inadequado										adequado													
10. Exame físico abdominal	<table border="1"> <tr> <td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td> </tr> <tr> <td>inadequado</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>adequado</td> </tr> </table> Sugestões de melhoria:	0	1	2	3	4	5	6	7	8	9	10	inadequado										adequado
0	1	2	3	4	5	6	7	8	9	10													
inadequado										adequado													
11. Exame de pulsos e edemas	<table border="1"> <tr> <td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td> </tr> <tr> <td>inadequado</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>adequado</td> </tr> </table> Comentários, justificativas e sugestões de melhoria:	0	1	2	3	4	5	6	7	8	9	10	inadequado										adequado
0	1	2	3	4	5	6	7	8	9	10													
inadequado										adequado													
12. Exame físico de outros sistemas relacionados a queixas clínicas	<table border="1"> <tr> <td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td> </tr> <tr> <td>inadequado</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>adequado</td> </tr> </table> Comentários, justificativas e sugestões de melhoria:	0	1	2	3	4	5	6	7	8	9	10	inadequado										adequado
0	1	2	3	4	5	6	7	8	9	10													
inadequado										adequado													
PREENCHIMENTO DOS FORMULÁRIOS E ORIENTAÇÕES AO PACIENTE																							
13. Preenchimento do prontuário	<table border="1"> <tr> <td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td> </tr> <tr> <td>ruim</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>excelente</td> </tr> </table> Comentários, justificativas e sugestões de melhoria:	0	1	2	3	4	5	6	7	8	9	10	ruim										excelente
0	1	2	3	4	5	6	7	8	9	10													
ruim										excelente													
14. Preenchimento da planilha de atendimento.	<table border="1"> <tr> <td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td> </tr> <tr> <td>ruim</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>excelente</td> </tr> </table> Comentários, justificativas e sugestões de melhoria:	0	1	2	3	4	5	6	7	8	9	10	ruim										excelente
0	1	2	3	4	5	6	7	8	9	10													
ruim										excelente													
15. Realização da hipótese diagnóstica.	<table border="1"> <tr> <td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td> </tr> <tr> <td>inadequada</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>adequada</td> </tr> </table> Comentários, justificativas e sugestões de melhoria:	0	1	2	3	4	5	6	7	8	9	10	inadequada										adequada
0	1	2	3	4	5	6	7	8	9	10													
inadequada										adequada													
16. Conduta proposta.	<table border="1"> <tr> <td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td> </tr> <tr> <td>inadequada</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>adequada</td> </tr> </table> Comentários, justificativas e sugestões de melhoria:	0	1	2	3	4	5	6	7	8	9	10	inadequada										adequada
0	1	2	3	4	5	6	7	8	9	10													
inadequada										adequada													
17. Explicação das hipóteses à paciente	<table border="1"> <tr> <td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td> </tr> <tr> <td>ruim</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>excelente</td> </tr> </table> Comentários, justificativas e sugestões de melhoria:	0	1	2	3	4	5	6	7	8	9	10	ruim										excelente
0	1	2	3	4	5	6	7	8	9	10													
ruim										excelente													

18. Fornecimento e explicação da receitas, pedidos de exames ou avaliações. Linguagem usada nas orientações quanto à HD e em dúvidas ou perguntas feitas pela paciente.	<u>0 1 2 3 4 5 6 7 8 9 10</u> inadequado adequado	NA
19. Abertura dada à paciente para fazer perguntas sobre suas dúvidas.	<u>0 1 2 3 4 5 6 7 8 9 10</u> ruim excelente	NA
20. Empatia: valorização das preocupações, atitudes e reações emocionais da paciente em relação ao seu problema que a fez buscar o médico.	<u>0 1 2 3 4 5 6 7 8 9 10</u> inadequada adequada	NA
21. Identificação de possíveis dificuldades para que a paciente siga a conduta.	<u>0 1 2 3 4 5 6 7 8 9 10</u> ruim excelente	NA
22. Término da consulta: orientou sobre retorno agendado, despediu-se do paciente e acompanhantes.	<u>0 1 2 3 4 5 6 7 8 9 10</u> inadequado adequado	NA
Comentários, justificativas e sugestões de melhoria:		

8.5. Anexo 5 – Avaliação Estruturada Ped

ALUNO: _____

RA: _____

DOCENTE: _____

UBS: _____

DATA: ____/____/____

CONSULTA: () Caso novo () Retorno INÍCIO: _____

TÉRMINO: _____

Como é o desempenho /atitude do aluno nesta consulta em relação a: (NA = Não se aplica)

1. Aparência pessoal: adequação da vestimenta, demonstração de higiene e postura.	0 1 2 3 4 5 6 7 8 9 10 ruim excelente	NA
	Comentários, justificativas e sugestões de melhoria:	
Início da consulta: cumprimentou paciente e acompanhantes, apresentou-se pelo nome, tratou paciente pelo nome, tranqüilizou o paciente, adotou postura acolhedora, deixou paciente à vontade.	0 1 2 3 4 5 6 7 8 9 10 ruim excelente	NA
	Comentários, justificativas e sugestões de melhoria::	
3. História: permitiu paciente e/ou acompanhante contar história com as próprias palavras, antes de intervir e orientar a anamnese?	0 1 2 3 4 5 6 7 8 9 10 inadequada adequada	NA
	Comentários, justificativas e sugestões de melhoria	
4. Queixa principal: identificou a queixa principal e conduziu a HPMA.	0 1 2 3 4 5 6 7 8 9 10 inadequada adequada	NA
	Comentários, justificativas e sugestões de melhoria:	
5. Revisão do IDA.	0 1 2 3 4 5 6 7 8 9 10 inadequada adequada	NA
	Comentários, justificativas e sugestões de melhoria:	
6. Revisão DNPM.	0 1 2 3 4 5 6 7 8 9 10 inadequada adequada	NA
	Comentários, justificativas e sugestões de melhoria:	
7. Revisão da vacinação.	0 1 2 3 4 5 6 7 8 9 10 inadequada adequada	NA
	Comentários, justificativas e sugestões de melhoria:	
8. Revisão da alimentação.	0 1 2 3 4 5 6 7 8 9 10 inadequada adequada	NA
	Comentários, justificativas e sugestões de melhoria:	
9. Revisão dos medicamentos em uso: vitaminas, ferro profilático; verificação de banho de sol.	0 1 2 3 4 5 6 7 8 9 10 inadequada adequada	NA
	Comentários, justificativas e sugestões de melhoria:	

10. Revisão dos antecedentes pessoais e familiares.	0 1 2 3 4 5 6 7 8 9 10 inadequada adequada	NA
	Comentários, justificativas e sugestões de melhoria:	
11. Revisão da dinâmica familiar e identificação de possíveis fatores de risco.	0 1 2 3 4 5 6 7 8 9 10 inadequada adequada	NA
	Comentários, justificativas e sugestões de melhoria:	
12. Explicação do exame físico: explicou ao paciente e acompanhante os procedimentos do exame físico, utilizando.	0 1 2 3 4 5 6 7 8 9 10 inadequada adequada	NA
	Comentários, justificativas e sugestões de melhoria:	
13. Realização do exame físico.	0 1 2 3 4 5 6 7 8 9 10 incorreto correto	NA
	Comentários, justificativas e sugestões de melhoria:	
14. Semiotécnica.	0 1 2 3 4 5 6 7 8 9 10 inadequada adequada	NA
	Sugestões de melhoria:	
15. Realização da antropometria (P, C, PC).	0 1 2 3 4 5 6 7 8 9 10 inadequada adequada	NA
	Comentários, justificativas e sugestões de melhoria:	
16. Preenchimento das curvas de referência.	0 1 2 3 4 5 6 7 8 9 10 inadequado adequado	NA
	Comentários, justificativas e sugestões de melhoria:	
17. Preenchimento do prontuário.	0 1 2 3 4 5 6 7 8 9 10 Inadequado adequado	NA
	Comentários, justificativas e sugestões de melhoria:	
18. Registro dos dados no cartão.	0 1 2 3 4 5 6 7 8 9 10 inadequado adequado	NA
	Comentários, justificativas e sugestões de melhoria:	
19. Formulação da hipótese diagnóstica.	0 1 2 3 4 5 6 7 8 9 10 incorreta correta	NA
	Comentários, justificativas e sugestões de melhoria:	
20. Qualidade da conduta proposta.	0 1 2 3 4 5 6 7 8 9 10 ruim excelente	NA
	Comentários, justificativas e sugestões de melhoria:	
21. Orientações ao paciente: orientou quanto à HD, usando linguagem acessível.	0 1 2 3 4 5 6 7 8 9 10 ruim excelente	NA
	Comentários, justificativas e sugestões de melhoria:	

22. Receita e exames: forneceu receita e explicou seu conteúdo e o pedido dos exames e/ou avaliações.	<u>0 1 2 3 4 5 6 7 8 9 10</u> ruim excelente	NA
Comentários, justificativas e sugestões de melhoria:		
23. Participação do paciente: possibilitou que o paciente ou acompanhante fizesse perguntas e/ou esclarecesse dúvidas.	<u>0 1 2 3 4 5 6 7 8 9 10</u> inadequada adequada	NA
Comentários, justificativas e sugestões de melhoria:		
24. Orientações: forneceu orientações e esclarecimentos, usando linguagem acessível; apontou possíveis impedimentos para que a conduta seja seguida.	<u>0 1 2 3 4 5 6 7 8 9 10</u> inadequada adequada	NA
Comentários, justificativas e sugestões de melhoria:		
25. Empatia: valorizou preocupações, atitudes e reações emocionais da mãe em relação à criança.	<u>0 1 2 3 4 5 6 7 8 9 10</u> inadequada adequada	NA
Comentários, justificativas e sugestões de melhoria:		
26. Término da consulta: orientou sobre retorno agendado, despediu-se do paciente e acompanhantes.	<u>0 1 2 3 4 5 6 7 8 9 10</u> inadequada adequada	NA
Comentários, justificativas e sugestões de melhoria:		

8.6. Anexo 6 – Formulário de Avaliação do Conceito Geral para Docentes

Esse formulário faz parte do processo de avaliação dos alunos no MD 856. Por favor, complete os dados abaixo a seu respeito e em relação ao aluno, cujo nome você deverá preencher no espaço correspondente.

Data: ____/____/____

Nome do Aluno Avaliado: _____

Nome da Unidade Básica de Saúde: _____

Dados do avaliador: (fazer um X ou pintar de amarelo)

Docente	Especialidade (clicar)	Sexo (clicar)	Idade (clicar)
Escrever número da matrícula: _____	GO	(M) Masculino	< 25 anos
	P	(F) Feminino	entre 25 e 30 anos
	CM		> entre 30 e 40 anos
			> 40 anos

Por favor, avalie o aluno de forma geral, assinalando a nota que julgar apropriada para o desempenho dele no MD 856, considerando que 0 é o valor mais baixo e 10 o mais alto.

CONCEITO GERAL

0	1	2	3	4	5	6	7	8	9	10
O aluno é muito ruim									O aluno é exemplar	

Comentários: as linhas abaixo estão reservadas para observações que você julgar relevantes:

8.7. Anexo 7 – Formulário de Auto-avaliação do Conceito Geral

Esse formulário faz parte do seu processo de avaliação no MD 856.

Por favor, preencha os dados abaixo a seu respeito.

Nome: _____

Idade: _____

Nome da Unidade Básica de Saúde: _____

Data: ____/____/____

Por favor, faça uma auto-avaliação em relação ao seu desempenho no MD 856, assinalando a nota que julgar apropriada, considerando que 0 é o valor mais baixo e 10 o mais alto.

CONCEITO GERAL

0	1	2	3	4	5	6	7	8	9	10
O aluno é muito ruim									O aluno é exemplar	

Comentários: as linhas abaixo estão reservadas para observações que você julgar relevantes:

8.8. Anexo 8 – Formulário de Avaliação do Conceito Global Itemizado – Docentes

Esse formulário faz parte do processo de avaliação dos alunos no MD 856. Por favor, preencha os dados abaixo a seu respeito e a respeito do(a) aluno(a) cujo nome você deverá escrever no espaço correspondente:

Data: ____/____/____

Nome do Aluno Avaliado: _____

Nome da Unidade Básica de Saúde: _____

Dados do avaliador:

Docente	Especialidade (clicar)	Sexo (clicar)	Idade (clicar)
Escrever número da matrícula: _____	GO	(M) Masculino	< 25 anos
	P	(F) Feminino	entre 25 e 30 anos
	CM		> entre 30 e 40 anos
			> 40 anos

CONCEITO GLOBAL ITEMIZADO

Por favor, avalie o aluno assinalando a nota que julgar apropriada, em cada um dos atributos abaixo, levando em consideração que 0 é o valor mais baixo e 10 o mais alto.

Parte 1 – COMPETÊNCIA TÉCNICA

1. Qualidade da história

0	1	2	3	4	5	6	7	8	9	10
Faltam informações críticas; anotações pobres e ilegíveis									História sempre completa; anotações bem organizadas e legíveis	

Comentários adicionais: _____

2. Exame clínico

0	1	2	3	4	5	6	7	8	9	10
Desempenha de forma incompetente (ou não desempenha) os passos do exame clínico e/ou procedimentos considerados essenciais.									Desempenha de forma competente os passos do exame clínico e/ou procedimentos considerados essenciais.	

Comentários adicionais: _____

3. Conhecimento médico

0	1	2	3	4	5	6	7	8	9	10
Conhecimento teórico limitado e fragmentado									Conhecimento teórico extensivo e bem integrado	

Comentários adicionais: _____

4. Julgamento clínico

0	1	2	3	4	5	6	7	8	9	10
Tem dificuldade em integrar fatos médicos com dados clínicos; raramente faz diagnóstico correto; desenvolve plano terapêutico inapropriado									Tem excelente habilidade para integrar fatos médicos com dados clínicos; sempre faz diagnóstico correto; desenvolve plano terapêutico apropriado.	

Comentários adicionais: _____

5. Solução de Problemas

0	1	2	3	4	5	6	7	8	9	10
Falha em avaliar informações, riscos e questões essenciais									Avalia criticamente informações, riscos e questões essenciais	
Comentários adicionais: _____										

6. Hábitos de Trabalho

0	1	2	3	4	5	6	7	8	9	10
Não é pontual; não completa os registros adequadamente, não adere a princípios éticos.									É pontual; completa os registros adequadamente, adere a princípios éticos.	
Comentários adicionais: _____										

Parte 2 – COMPETÊNCIA HUMANÍSTICA

7. Comunicação e relacionamento com paciente e familiares

0	1	2	3	4	5	6	7	8	9	10
Comunica-se de forma confusa; usa palavras difíceis, não dá explicações claras, não sabe ouvir; desenvolve envolvimento emocional inapropriado.									Comunica-se com objetividade, usa palavras fáceis, dá explicações claras; ouve atentamente; estabelece envolvimento emocional apropriado.	
Comentários adicionais: _____										

8. Respeito

0	1	2	3	4	5	6	7	8	9	10
Demonstra falta de respeito, de atenção e atitude acolhedora aos paciente e familiares.									Demonstra muito respeito, atenção e atitude acolhedora ao paciente e familiares.	
Comentários adicionais: _____										

9. Capacidade de auto-reflexão

0	1	2	3	4	5	6	7	8	9	10
Não assume responsabilidade pelas próprias ações e decisões; culpa os outros quando comete erros, demonstra atitude negativa em relação às críticas.									Assume totalmente a responsabilidade por suas próprias ações e decisões, demonstra atitude positiva em relação às críticas.	
Comentários adicionais: _____										

10. Percepção do contexto

0	1	2	3	4	5	6	7	8	9	10
Mostra consideração inadequada e insensibilidade em relação às necessidades de conforto e ajuda aos pacientes e familiares.									Leva em consideração e demonstra sensibilidade em relação às necessidades de conforto e ajuda dos pacientes e familiares.	
Comentários adicionais: _____										

11. Interação com colegas

0	1	2	3	4	5	6	7	8	9	10
Não se relaciona bem com colegas; não apresenta atitudes de cooperação e ajuda.									Relaciona-se muito bem com colegas; apresenta atitudes de cooperação e ajuda.	
Comentários adicionais: _____										

12. Interação com docentes

0	1	2	3	4	5	6	7	8	9	10
Não pede orientação; não verifica entendimento.									Pede orientação; verifica entendimento.	
Comentários adicionais: _____										

13. Interação com demais profissionais

0	1	2	3	4	5	6	7	8	9	10
Não se demonstra cooperativo com outros profissionais; apresenta dificuldade para trabalhar em equipe.									Demonstra-se cooperativo com outros profissionais; apresenta dificuldade para trabalhar em equipe.	
Comentários adicionais: _____										

8.9. Anexo 9 – Formulário de Avaliação do Conceito Global Itemizado – Auto-avaliação

Esse formulário faz parte do seu processo de avaliação no MD 856. Por favor, preencha os itens abaixo com seus dados.

Nome: _____

Idade: _____

Nome da Unidade Básica de Saúde: _____

Data: ____/____/____

CONCEITO GLOBAL ITEMIZADO

Por favor, faça sua auto-avaliação assinalando a nota que julgar apropriada, em cada um dos atributos abaixo, levando em consideração que 0 é o valor mais baixo e 10 o mais alto.

Parte 1 – COMPETÊNCIA TÉCNICA

1. Qualidade da história

0	1	2	3	4	5	6	7	8	9	10
Faltam informações críticas; anotações pobres e ilegíveis									História sempre completa; anotações bem organizadas e legíveis	

Comentários adicionais: _____

2. Exame clínico

0	1	2	3	4	5	6	7	8	9	10
Desempenha de forma incompetente (ou não desempenha) os passos do exame clínico e/ou procedimentos considerados essenciais.									Desempenha de forma competente os passos do exame clínico e/ou procedimentos considerados essenciais.	

Comentários adicionais: _____

3. Conhecimento médico

0	1	2	3	4	5	6	7	8	9	10
Conhecimento teórico limitado e fragmentado									Conhecimento teórico extensivo e bem integrado	

Comentários adicionais: _____

4. Julgamento clínico

0	1	2	3	4	5	6	7	8	9	10
Tem dificuldade em integrar história com dados clínicos; raramente faz diagnóstico correto; desenvolve plano terapêutico inapropriado.									Tem excelente habilidade para integrar história com dados clínicos; sempre faz diagnóstico correto; desenvolve plano terapêutico apropriado.	

Comentários adicionais: _____

5. Solução de Problemas

0	1	2	3	4	5	6	7	8	9	10
Falha em avaliar informações, riscos e questões essenciais									Avalia criticamente informações, riscos e questões essenciais	
Comentários adicionais: _____										

6. Hábitos de Trabalho

0	1	2	3	4	5	6	7	8	9	10
Não é pontual; não completa os registros adequadamente, não adere a princípios éticos.									É pontual; completa os registros adequadamente, adere a princípios éticos.	
Comentários adicionais: _____										

Parte 2 – COMPETÊNCIA HUMANÍSTICA

7. Comunicação e relacionamento com paciente e familiares

0	1	2	3	4	5	6	7	8	9	10
Comunica-se de forma confusa; usa palavras difíceis, não dá explicações claras, não sabe ouvir; desenvolve envolvimento emocional inapropriado.									Comunica-se com objetividade, usa palavras fáceis, dá explicações claras; ouve atentamente; estabelece envolvimento emocional apropriado.	
Comentários adicionais: _____										

8. Respeito

0	1	2	3	4	5	6	7	8	9	10
Demonstra falta de respeito, de atenção e atitude acolhedora aos paciente e familiares.									Demonstra muito respeito, atenção e atitude acolhedora ao paciente e familiares.	
Comentários adicionais: _____										

9. Capacidade de auto-reflexão

0	1	2	3	4	5	6	7	8	9	10
Não assume responsabilidade pelas próprias ações e decisões; culpa os outros quando comete erros, demonstra atitude negativa em relação às críticas.									Assume totalmente a responsabilidade por suas próprias ações e decisões, demonstra atitude positiva em relação às críticas.	
Comentários adicionais: _____										

10. Percepção do contexto

0	1	2	3	4	5	6	7	8	9	10
Mostra consideração inadequada e insensibilidade em relação às necessidades de conforto e ajuda aos pacientes e familiares.									Leva em consideração e demonstra sensibilidade em relação às necessidades de conforto e ajuda dos pacientes e familiares.	
Comentários adicionais: _____										

11. Interação com colegas

0	1	2	3	4	5	6	7	8	9	10
Não se relaciona bem com colegas; não apresenta atitudes de cooperação e ajuda.									Relaciona-se muito bem com colegas; apresenta atitudes de cooperação e ajuda.	
Comentários adicionais: _____										

12. Interação com docentes

0	1	2	3	4	5	6	7	8	9	10
Não pede orientação; não verifica entendimento.									Pede orientação; verifica entendimento.	
Comentários adicionais: _____										

13. Interação com demais profissionais

0	1	2	3	4	5	6	7	8	9	10
Não se demonstra cooperativo com outros profissionais; apresenta dificuldade para trabalhar em equipe.									Demonstra-se cooperativo com outros profissionais; apresenta dificuldade para trabalhar em equipe.	
Comentários adicionais: _____										

8.10. Anexo 10 – Formulário de Avaliação do Conceito Global Itemizado – Pares

Esse formulário faz parte do processo de avaliação de seus colegas de grupo do MD856. Você deve avaliar todos os colegas de seu grupo de atividades de estágio clínico. Por favor, preencha o formulário abaixo com os dados relacionados ao colega de grupo que você está avaliando:

Nome do Aluno Avaliado: _____

Nome da Unidade Básica de Saúde: _____

Data: ____/____/____

Por favor, preencha os dados abaixo sobre você:

Dados do avaliador:

Sexo	Idade
(M) Masculino	() < 25 anos
(F) Feminino	() entre 25 e 30 anos
	() > 30 anos

CONCEITO GLOBAL ITEMIZADO

Por favor, avalie seu colega de grupo assinalando a nota que julgar apropriada, em cada um dos atributos abaixo, levando em consideração que 0 é o valor mais baixo e 10 o mais alto.

Parte 2 – COMPETÊNCIA HUMANÍSTICA

7. Comunicação e relacionamento com paciente e familiares

0	1	2	3	4	5	6	7	8	9	10
Comunica-se de forma confusa; usa palavras difíceis, não dá explicações claras, não sabe ouvir; desenvolve envolvimento emocional inapropriado.									Comunica-se com objetividade, usa palavras fáceis, dá explicações claras; ouve atentamente; estabelece envolvimento emocional apropriado.	
Comentários adicionais: _____										

8. Respeito

0	1	2	3	4	5	6	7	8	9	10
Demonstra falta de respeito, de atenção e atitude acolhedora aos paciente e familiares.									Demonstra muito respeito, atenção e atitude acolhedora ao paciente e familiares.	
Comentários adicionais: _____										

9. Capacidade de auto-reflexão

0	1	2	3	4	5	6	7	8	9	10
Não assume responsabilidade pelas próprias ações e decisões; culpa os outros quando comete erros, demonstra atitude negativa em relação às críticas.									Assume totalmente a responsabilidade por suas próprias ações e decisões, demonstra atitude positiva em relação às críticas.	
Comentários adicionais: _____										

10. Percepção do contexto

0	1	2	3	4	5	6	7	8	9	10
Mostra consideração inadequada e insensibilidade em relação às necessidades de conforto e ajuda aos pacientes e familiares.									Leva em consideração e demonstra sensibilidade em relação às necessidades de conforto e ajuda dos pacientes e familiares.	
Comentários adicionais: _____										

11. Interação com colegas

0	1	2	3	4	5	6	7	8	9	10
Não se relaciona bem com colegas; não apresenta atitudes de cooperação e ajuda.									Relaciona-se muito bem com colegas; apresenta atitudes de cooperação e ajuda.	
Comentários adicionais: _____										

12. Interação com docentes

0	1	2	3	4	5	6	7	8	9	10
Não pede orientação; não verifica entendimento.									Pede orientação; verifica entendimento.	
Comentários adicionais: _____										

13. Interação com demais profissionais

0	1	2	3	4	5	6	7	8	9	10
Não se demonstra cooperativo com outros profissionais; apresenta dificuldade para trabalhar em equipe.									Demonstra-se cooperativo com outros profissionais; apresenta dificuldade para trabalhar em equipe.	
Comentários adicionais: _____										

8.11. Anexo 11 – Questionário de Opinião em relação à auto-avaliação e avaliação de pares

Por favor, dê sua opinião em relação à experiência de auto-avaliação e de avaliação dos colegas fazendo um X na nota que julgar apropriada, em cada um dos itens abaixo, levando em consideração que 0 é o valor mais baixo e 10 o mais alto.

1. Quão confortável você se sentiu realizando sua auto-avaliação?	<p style="text-align: center;">_0 1 2 3 4 5 6 7 8 9 10</p> <p>nada muito confortável confortável</p> <p>Comentários adicionais:</p>
2. Quão confortável você se sentiu realizando a avaliação de seus colegas?	<p style="text-align: center;">_0 1 2 3 4 5 6 7 8 9 10</p> <p>nada muito confortável confortável</p> <p>Comentários adicionais:</p>
3. Quão preparados para se auto-avaliar de maneira responsável você acredita que os alunos do 4º ano de Medicina estão?	<p style="text-align: center;">_0 1 2 3 4 5 6 7 8 9 10</p> <p>nada muito preparados preparados</p> <p>Comentários adicionais:</p>
4. Quão preparados para avaliar os colegas de maneira responsável você acredita que os alunos do 4º ano de Medicina estão?	<p style="text-align: center;">_0 1 2 3 4 5 6 7 8 9 10</p> <p>nada muito preparados preparados</p> <p>Comentários adicionais:</p>
5. Quão justa foi sua auto-avaliação?	<p style="text-align: center;">_0 1 2 3 4 5 6 7 8 9 10</p> <p>nada justa muito justa</p> <p>Comentários adicionais:</p>
6. Quão justa foi a avaliação que você fez de seus colegas?	<p style="text-align: center;">_0 1 2 3 4 5 6 7 8 9 10</p> <p>nada justa muito justa</p> <p>Comentários adicionais:</p>
7. Quanto você valoriza o exercício de auto-avaliação	<p style="text-align: center;">_0 1 2 3 4 5 6 7 8 9 10</p> <p>nada muito</p> <p>Comentários adicionais:</p>
8. Quanto você valoriza o exercício de avaliar os colegas?	<p style="text-align: center;">_0 1 2 3 4 5 6 7 8 9 10</p> <p>nada muito</p> <p>Comentários adicionais:</p>

8.12. Anexo 12 – Parecer Comitê de Ética e Pesquisa

**FACULDADE DE CIÊNCIAS MÉDICAS
COMITÊ DE ÉTICA EM PESQUISA**

Caixa Postal 6111, 13083-970 Campinas, SP

Fone (0_19) 3788-8936

FAX(0_19) 3788-7187

www.fcm.unicamp.br/pesquisa/etica/index.html

cep@fcm.unicamp.br

CEP, 25/10/05. (Grupo III)

PARECER PROJETO: N° 581/2005

CAAE: 1521.0.146.000-05

I-IDENTIFICAÇÃO:

PROJETO: "CONCEITO GLOBAL PARA ALUNOS DE MEDICINA EM ESTÁGIO CLÍNICO: COMPARAÇÃO DA AUTO-AVALIAÇÃO COM A AVALIAÇÃO DE DOCENTES E PARES".

PESQUISADOR RESPONSÁVEL: Rosângela Curvo Leite Domingues

INSTITUIÇÃO: FCM - UNICAMP **APRESENTAÇÃO AO CEP:** 15/10/05

II- OBJETIVOS

Verificar a correlação do conceito global itemizado com o conceito geral dos alunos do 4º ano de medicina, em estágio clínica, dentre três tipos de avaliadores (docentes, colegas e próprio aluno). Verificar se o conceito global itemizado se assemelha ao conceito geral de cada aluno dado pelos tipos de avaliadores: docente e próprio aluno. Comparar o valor do conceito global itemizado do aluno dado pelos docentes dentre três especialidades médicas: clínica, pediatria e ginecologia/obstetrícia. Comparar os conceitos globais itemizado que o aluno recebe de cada um dos três tipos de avaliadores (ele mesmo, docente, par). Verificar a homogeneidade do desempenho do aluno nos diferentes itens do conceito global itemizado.

III- SUMÁRIO

A pesquisadora pretende estabelecer um modelo de avaliação interativo para um grupo específico de alunos da medicina. Trata-se de um estudo transversal analítico para análise secundária de dados, serão incluídos todos os sujeitos (alunos do quarto ano) em estágio clínico na unidade básica de saúde (USB), os alunos serão subdivididos para atendimento nas áreas de clínica, obstetrícia e pediatria. Os critérios de inclusão e descontinuidade estão descritos e as variáveis dependentes são atreladas ao conceito global itemizado baseado nos parâmetros de observação definidos: competências clínicas e competências humanistas, enquanto as variáveis independentes são: o tipo de avaliador, especialidade, conceito geral, nota de provas cognitivas, nota de portfólio, nota de avaliação estruturada, haverá também as variáveis de controle: sexo do avaliador, do avaliado, idade do avaliador e do avaliado, e USB. Existe um formulário para coleta de opinião dos atores desse processo avaliatório.

IV - COMENTÁRIOS DOS RELATORES

O estudo oferece uma perspectiva inovadora para o processo de avaliação dos acadêmicos de medicina, está consistentemente referenciado, com objetivos bem definidos e metodologia claramente descrita, entretanto precisa ser revisto em alguns aspectos para estar de acordo com a legislação que norteia as decisões do Comitê de Ética em Pesquisa, uma vez que a pesquisadora solicita dispensa do Termo de Consentimento Livre e Esclarecido.

V - PARECER DO CEP

O Comitê de Ética em Pesquisa da Faculdade de Ciências Médicas da UNICAMP, após acatar os pareceres dos membros-relatores previamente designados para o presente caso e atendendo todos os dispositivos das Resoluções 196/96 e complementares, bem como ter aprovado todos os anexos incluídos na Pesquisa, resolve aprovar sem restrições o Protocolo de Pesquisa supracitado.

O conteúdo e as conclusões aqui apresentados são de responsabilidade exclusiva do CEP/FCM/UNICAMP e não representam a opinião da Universidade Estadual de Campinas nem a comprometem.

VI - INFORMAÇÕES COMPLEMENTARES

O sujeito da pesquisa tem a liberdade de recusar-se a participar ou de retirar seu consentimento em qualquer fase da pesquisa, sem penalização alguma e sem prejuízo ao seu cuidado (Rés. CNS 196/96 - Item IV. 1.f) e deve receber uma cópia do Termo de Consentimento Livre e Esclarecido, na íntegra, por ele assinado (Item IV.2.d).

Pesquisador deve desenvolver a pesquisa conforme delineada no protocolo aprovado e descontinuar o estudo somente após análise das razões da descontinuidade pelo CEP que o aprovou (Rés. CNS Item III. 1.z), exceto quando perceber risco ou dano não previsto ao sujeito participante ou quando constatar a superioridade do regime oferecido a um dos grupos de pesquisa (Item V.3.).

O CEP deve ser informado de todos os efeitos adversos ou fatos relevantes que alterem o curso normal do estudo (Rés. CNS Item V.4.). É papel do pesquisador assegurar medidas imediatas adequadas frente a evento adverso ocorrido (mesmo que tenha sido em outro centro) e enviar notificação ao CEP e à Agência Nacional de Vigilância Sanitária - ANVISA -junto com seu posicionamento.

Eventuais modificações ou emendas ao protocolo devem ser apresentadas ao CEP de forma clara e sucinta, identificando a parte do protocolo a ser modificada e suas justificativas. Em caso de projeto do Grupo I ou II apresentados anteriormente à ANVISA, o pesquisador ou patrocinador deve enviá-las também à mesma junto com o parecer aprovatório do CEP, para serem juntadas ao protocolo inicial (Rés. 251/97, Item IH.2.e). Relatórios parciais e final devem ser apresentados ao CEP, de acordo com os prazos estabelecidos na Resolução CNS-MS 196/96.

VII - DATA DA REUNIÃO

Homologado na X Reunião Ordinária do CEP/FCM, em 25 de outubro de 2005.



Prof. Dra. Carmem Silvia Bertuzzo
PRESIDENTE DO COMITÊ DE ÉTICA EM PESQUISA
FCM / UNICAMP