

Descrição e análise de um sistema de avaliação das aprendizagens numa Faculdade de Medicina

Description and analysis of an assessment system in a College of Medicine

Leonor Ramalho*

Universidade Católica Portuguesa

Resumo

Neste estudo descreve-se o sistema de avaliação das aprendizagens na disciplina de Clínica Geral da FCML e analisa-se também na perspectiva dos resultados produzidos.

Verificou-se que os resultados da avaliação de desempenho são superiores aos da avaliação de conhecimentos; não há variabilidade dos resultados dos alunos em cinco anos lectivos; as alunas têm, na avaliação de conhecimentos, classificações superiores à dos alunos. Na avaliação de desempenho, não se demonstrou influência da interação do género do tutor e do aluno. A classificação final da disciplina de Clínica Geral da FCML reflecte os resultados da avaliação sumativa.

Para aumentar o peso da avaliação formativa, face à sumativa, sugere-se que os tutores tenham formação pedagógica e em técnicas de supervisão, bem como treino na utilização dos instrumentos de avaliação.

Palavras-chave: Educação Médica, estudantes de medicina, disciplina de Clínica Geral, sistema de avaliação, instrumentos de avaliação, avaliação sumativa, avaliação formativa, avaliação de desempenho, avaliação de conhecimentos, classificação final. ◀◀

Introdução

A avaliação desempenha um papel muito importante na vida dos estudantes de Medicina e também para a Sociedade em geral, pela sua responsabilidade na certificação de médicos competentes para cuidar da população (1, 2).

Desde há muito tempo que as faculdades sentem necessidade de introduzir novos métodos de ensino, de forma a proporcionar aos alunos experiências de aprendizagem aplicáveis às situações da vida real (3) o que determina a necessidade de novas formas de avaliar quer os alunos, quer os programas, quer ainda as próprias metodologias (4). Os programas

Abstract

The main concern of this research is to describe and analyse the system of assessment in the General Practice Course of the FCML.

We found out that outcomes of performance assessment are higher than outcomes of knowledge assessment; there is no variability in the students' outcomes through five years; in knowledge assessment, female students have higher outcomes than male students. In performance assessment, we didn't find out any interaction of tutor gender and student gender. The final classification reflects, mainly, summative assessment outcomes.

In order to increase the weight of the formative assessment, over the summative, we suggest that tutors should have preparation in pedagogic and supervising techniques as well as training in the use of assessment instruments.

Keywords: Medical Education, medical students, General Practice Course, assessment system, assessment instruments, summative assessment, formative assessment, performance assessment, knowledge assessment, final classification. ◀◀

de ensino médico são muito complexos e o ensino dispersa-se por várias disciplinas, em vários locais (hospitais, centros de saúde, aulas, laboratórios, etc.) e durante vários anos. São usados muitos tipos e estilos de ensino e de aprendizagem e os alunos têm experiências variadas durante o seu treino, por isso, torna-se necessária escolha eclética de métodos e instrumentos de avaliação (5).

Os instrumentos de avaliação e a forma como são aplicados estão intimamente associados com a forma como os indivíduos aprendem e o desenvolvimento das estratégias da avaliação é uma pedra-chave para o desenvolvimento do *curriculum* (6).

Em Portugal, apesar da relevância do tema, não há divulgação de estudos relativos à avaliação dos estudantes de Medicina, ao contrário do que acontece em relação a outros níveis de ensino (7) e do que

* leonoramalho@hotmail.com

se passa noutros países. Apesar de existirem alguns relatórios relacionados com a avaliação institucional das Faculdades de Medicina (8, 9) com conclusões relativamente optimistas em relação aos aspectos organizativos e pedagógicos, bem como às competências dos alunos, surgem, entre os docentes, algumas vozes discordantes. Cita-se a propósito o Prof. Jorge Soares:

“Os modelos de ensino actuais fazem (o que não seria inesperado que acontecesse!) os alunos felizes. E porquê? Porque prolongam o formato securitário do ensino liceal no qual obtiveram indiscutível sucesso (a admissão ao curso!), centram na “nota da cadeira” o objectivo maior de aprendizagem (que os leva a preocuparem-se mais com a justiça da nota da disciplina, por comparação “com a do colega do lado”, do que com as condições sociais do doente e da sua situação de sofrimento), promovem a “metrificação” das classificações (os estudantes constituíram-se “adaptativamente”, em insuperáveis “profissionais de fazer exames!”), restringem o seu espaço de liberdade (e a grande maioria dos estudantes não a deseja porque não sabe o que fazer com ela!)” (10) .

É do conhecimento geral que o percurso académico dos alunos de Medicina decorre num ambiente de grande competitividade que encontra a sua expressão máxima nas classificações obtidas, que nem sempre são a tradução efectiva da aquisição de competências humanísticas e éticas ou de conhecimentos técnico-científicos, essenciais no exercício da profissão médica. A cultura da classificação e da nota elevada está bem patente na inflação que se tem verificado desde há anos nas classificações dos estudantes de Medicina e que não se observa noutros cursos.

Este artigo apresenta um estudo que teve por objecto a avaliação das aprendizagens dos estudantes de Medicina e que foi realizado no âmbito do Mestrado em Educação Médica do ICS/UCP.

O estudo, de tipo exploratório, descreve um sistema de avaliação das aprendizagens dos estudantes de Medicina, na área da Clínica Geral, analisa os instrumentos utilizados e os resultados obtidos, procurando preencher a lacuna existente nos trabalhos realizados, no âmbito da Educação Médica, em Portugal. Tendo em conta estes objectivos formularam-se quatro hipóteses, nomeadamente com o intuito de saber se:

1. Os resultados da avaliação de desempenho (Centro de Saúde Urbano, Centro de Saúde Rural), avaliação de conhecimentos e classificação final serão equivalentes;
2. Existirá variação nos resultados das avaliações de desempenho, de conhecimentos e da classificação final em cinco anos lectivos;
3. Os resultados da avaliação serão influenciados pelo género do aluno,
4. Os resultados da avaliação de desempenho serão influenciados pelo facto de o aluno e o tutor serem do mesmo género.

Face ao sistema de avaliação existente, considerou-se que um relato empiricamente fundamentado seria um contributo para a reflexão sobre a necessidade de mudança do sistema actual de avaliação das aprendizagens, tendo em conta as pressões decorrentes da evolução quer do ensino médico no mundo, quer da do ensino superior na Europa.

Material e métodos

Na descrição interna do sistema de avaliação, utilizaram-se métodos qualitativos, nomeadamente a pesquisa bibliográfica e a análise documental, foi usada informação disponível no *site* da Faculdade de Ciências Médicas de Lisboa, foram também utilizadas entrevistas informais com alguns dos elementos intervenientes no processo de avaliação dos estudantes, quer assistentes da disciplina quer tutores, e ainda com alguns estudantes, bem como análise dos comentários do *item* Análise Crítica do Estágio que consta no relatório. Foi utilizada esta metodologia numa perspectiva de triangulação da informação através das fontes.

Foram também utilizados métodos quantitativos, no sentido de identificar associações com significado estatístico nas classificações atribuídas aos estudantes.

População

A população-alvo refere-se ao universo constituído pelos resultados das avaliações dos estudantes do 6º ano de Medicina na disciplina de Clínica Geral da Faculdade de Ciências Médicas de Lisboa (FCML), durante um período de 5 anos lectivos (2000 a 2004):

- Avaliação de Desempenho do Centro de Saúde (CS) rural – com pontuação de 8 a 40;

- Avaliação de Desempenho CS urbano – com pontuação de 8 a 40;
- Avaliação de Desempenho (MAD) – média aritmética das duas anteriores, com conversão para a pontuação de 0 a 20 valores;
- Avaliação de conhecimentos/Avaliação da Discussão do Relatório (AC) – com pontuação de 0 a 20 valores;
- Classificação Final (CF) – calculada pela fórmula $\frac{MAD + 2AC}{3}$ no ano lectivo 2000-2001; e pela fórmula $\frac{MAD + 3AC}{4}$ nos anos lectivos seguintes.

Tratamento estatístico

Elaborou-se uma base de dados no programa SPSS (*Statistical Package for the Social Sciences*) que também foi utilizado para o tratamento estatístico. Utilizou-se ainda o programa EXCEL para a realização de alguns gráficos.

Para a descrição das *variáveis dependentes* (Avaliação de desempenho CS rural; Avaliação de desempenho CS urbano; Avaliação de desempenho; Avaliação de conhecimentos e Classificação final) utilizaram-se medidas descritivas: percentagens, medidas de tendência central e medidas de dispersão. Aplicaram-se ainda testes de normalidade (Kolmogorov-Smirnov) a todas as variáveis dependentes, tendo-se constatado que em nenhum dos casos se verificou uma distribuição normal. Contudo, tendo em conta o Teorema do Limite Central (11, 12) aceita-se que a sua distribuição se aproxima da normal.

Como expressão gráfica para a comparação entre as diversas variáveis numéricas, utilizou-se o gráfico de erro na comparação entre as médias.

Na aplicação dos testes estatísticos procedeu-se à verificação dos pressupostos dos testes paramétricos (distribuição normal, homocedasticidade, observações independentes). Para a interpretação dos resultados de todos os testes utilizados considerou-se sempre o nível de significância de 0,05.

Para verificar a primeira hipótese, optou-se pelo teste *t* de Student para amostras emparelhadas. Procedeu-se ainda à análise das variáveis, com recurso a estatísticas descritivas, nomeadamente o coeficiente de correlação de Pearson e o coeficiente de determinação (13). No que diz respeito à segunda hipótese, aplicou-se o teste *F* da ANOVA, análise da variância com um factor. O Teste de Levene mostrou que não existia homogeneidade das variâncias na

variável Avaliação de Desempenho do CS urbano. Apesar de alguns autores considerarem que este teste paramétrico é robusto a violações da homocedasticidade quando o número de observações é igual ou aproximadamente igual (14), optou-se por uma alternativa não paramétrica (15), o teste Kruskal Wallis, para testar aquela variável. Em relação à terceira hipótese, aplicou-se o teste *t* de Student para duas amostras independentes, igualdade de variâncias não assumida (14) por o teste de Levene não confirmar a homogeneidade das variâncias. Optou-se pelo teste ANOVA, análise da variância com dois factores para estudar a interacção entre o género do tutor e o do aluno. Não se tendo verificado interacção entre os factores, interpretaram-se os efeitos dos factores isoladamente.

Descrição e análise do sistema de avaliação das aprendizagens da disciplina de Clínica Geral da FCML

A disciplina de Clínica Geral é leccionada na Faculdade de Ciências Médicas de Lisboa (FCML) desde o ano lectivo 2000-2001 (16).

O ensino desta disciplina decorre sob a responsabilidade do Departamento Universitário de Clínica Geral, em Centros de Saúde (CS) adstritos à Faculdade por protocolos celebrados com a Administração de Saúde de Lisboa e Vale do Tejo e com a Administração Regional de Saúde do Alentejo.

O pessoal docente é constituído pelo Director do Departamento e Regente da disciplina, por oito assistentes convidados, sendo que eram apenas três quando se iniciou o ensino desta disciplina (16-18) e por tutores que exercem a sua actividade profissional como médicos de família em CS rurais ou urbanos.

Nos anos lectivos analisados, os tutores totalizaram 170, sendo que 91 trabalham em CS urbanos e 79 em CS rurais (Figura 1). Verificou-se, em ambos os

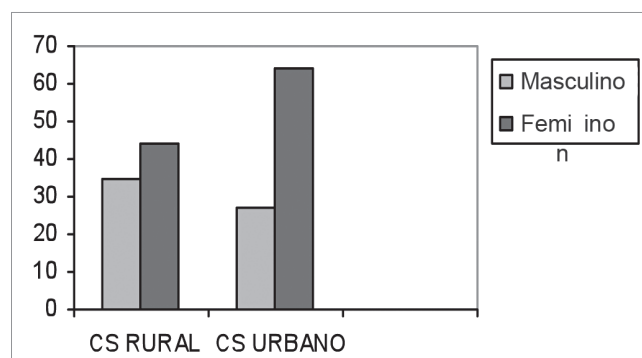


Figura 1 – Distribuição dos tutores segundo o género

grupos, uma predominância do género feminino, mais acentuada nos CS urbanos.

Em relação ao número de alunos orientados (Figura 2), constata-se que nos CS urbanos é mais elevada a proporção de tutores que tiveram entre um e três alunos.

No que diz respeito aos 593 discentes, (Figura 3), existe um predomínio do género feminino em todos os anos lectivos estudados.

Organização Pedagógica da Disciplina

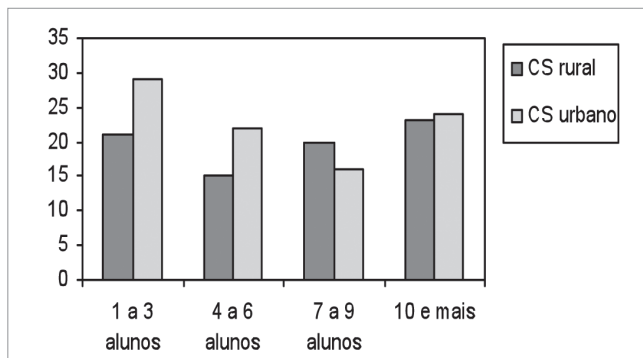


Figura 2 – Distribuição dos alunos pelos tutores

O ensino/ aprendizagem da disciplina inicia-se com um curso teórico de dois dias que corresponde a cerca de 15 horas lectivas, a que se segue uma com-

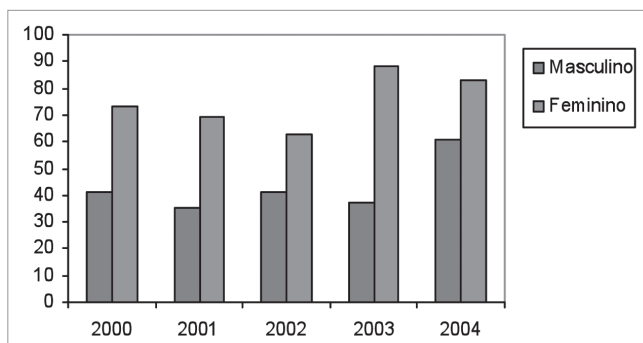


Figura 3 – Distribuição dos alunos segundo o género pelos anos lectivos

ponente prática, denominada de Exercício Orientado (EO). Esta tem duração de quatro semanas, sendo que duas são em CS urbanos ou suburbanos e duas em CS rurais. Durante o período de estágio decorrem ainda duas sessões teórico-práticas em grupo.

Nos CS e durante cada período de duas semanas o aluno é acompanhado por um tutor que é responsável pela programação das actividades, os tutores dispõem de um instrumento de trabalho, o Guia do Tutor, onde constam informações, con-

sideradas úteis no acompanhamento dos alunos. Os alunos cumprem em média 25 horas semanais de permanência no CS das quais, pelo menos, 12 horas na consulta com o seu tutor.

Sistema de avaliação da disciplina de Clínica Geral

A avaliação dos estudantes (16) assume dois aspectos, um formativo e outro sumativo. A avaliação formativa é realizada diariamente pelo tutor em função das prestações do aluno nas diferentes tarefas que lhe são propostas e tem como finalidade corrigir as deficiências apresentadas, assim como salientar os aspectos mais positivos dos objectivos atingidos. No final das duas semanas do EO, o tutor preenche a ficha de avaliação de desempenho correspondente. Recomenda-se que seja dada ao aluno informação de retorno.

A avaliação sumativa (classificação final da disciplina) resulta da média ponderada, das avaliações de desempenho atribuídas pelos tutores e de uma avaliação de conhecimentos/ discussão do relatório, realizada por um júri de dois docentes que ocorre no último dia útil da 4ª semana do estágio. O relatório apresentado pelo aluno é elaborado de acordo com as especificações dadas no “Diário do Exercício Orientado/Relatório”. Na discussão do relatório, que não é superior a 30 minutos, os examinadores procuram apreciar o nível de conhecimentos do aluno, incluindo o dos valores éticos e deontológicos.

Para efeitos de classificação final é obrigatório que o aluno tenha avaliação positiva (≥ 10) em qualquer das avaliações realizadas (16).

Os alunos, no início das actividades lectivas, têm acesso ao programa da disciplina, bem como informação sobre o seu sistema de avaliação, respectivos instrumentos e critérios de classificação disponíveis (16).

Instrumentos de avaliação

Avaliação de desempenho

Existem dois impressos, com formato idêntico, para cada um dos locais de estágio, em que constam oito indicadores de competência: *Identificação das necessidades; Exame físico; Diagnóstico e Gestão de problemas; Sistemas de informação; Comunicação; Relações interpessoais e Interinstitucionais; Implicação no Estágio e Auto-avaliação*. Cada indicador tem três sub-itens que são classificados numa escala de Likert de 1 a 5 (1= não satisfaz; 2= precisa de muita

supervisão; 3= satisfaz; 4= bom; 5= muito bom), a cotação de cada indicador varia entre 1 e 5 e resulta da média do total dos sub-itens. A classificação de desempenho resulta do somatório de todos os indicadores e pode variar entre 8 e 40.

Avaliação de conhecimentos/discussão de relatório

O Diário do Exercício Orientado (DEO), integra um conjunto de conteúdos obrigatórios e previamente definidos, que visam facilitar o preenchimento por parte do aluno. De facto, a diversidade dos locais de estágio, a sua dispersão geográfica e as características do próprio estágio, condicionam o tempo disponível para a sua redacção. O relatório (DEO) procura demonstrar as experiências de aprendizagem vividas, o que se fez, onde, quando e quanto se fez; assim como permite saber se, das tarefas efectuadas, o aluno sabe justificar porque se faz aquilo que diz ter feito (sustentação científica da prática) (19).

No que diz respeito à estrutura e conteúdos, o relatório está organizado por áreas, nomeadamente *Necessidades, Expectativas e Receios*, onde se pede que o aluno identifique os objectivos específicos do seu estágio; *Registo de Morbilidade*, motivos de consulta e problemas de saúde de doentes observados; *Análise de Situações*, descrição e análise de uma situação vivida durante o estágio e segundo um modelo pré-definido; *Apresentação de um Caso*, onde se espera que o aluno demonstre as suas aptidões diagnósticas e terapêuticas; *Identificação de Necessidades de Aprendizagem; Análise Crítica do Estágio*.

As perguntas efectuadas pelos examinadores, na discussão do relatório, são anotadas e variam entre 5 a 9, são formuladas a partir das descrições que constam no relatório e resultam do cruzamento dos conteúdos do relatório com uma listagem de temas que são do conhecimento do aluno; a cada resposta corresponde uma pontuação de 0 a 5 (0= não respondeu; 1= resposta insuficiente; 2= resposta suficiente; 3= resposta boa; 4= resposta muito boa; 5= resposta excelente). Com esta cotação faz-se um cálculo do nível médio das respostas e posteriormente, por consenso é feita a sua conversão para a escala de 0 a 20 valores.

Classificações dos alunos

A média das avaliações de desempenho dos alunos nos CS rurais foi de 36,7 sendo a mediana de 37,0 e o desvio padrão de 2,92. O valor mínimo foi de 20 e o máximo de 40.

A média das avaliações de desempenho dos alunos nos CS urbanos foi de 36,9 sendo a mediana de 38,0 e o desvio padrão de 2,98. O valor mínimo foi de 24 e o máximo de 40.

A média da avaliação de conhecimentos dos alunos foi de 16,1 sendo a mediana 17 e o desvio padrão de 2,32. O valor mínimo foi de 8,5 e o máximo de 20.

A média da classificação final dos alunos foi de 16,7 sendo a mediana de 17 e o desvio padrão de 1,83. O valor mínimo foi de 9 e o máximo de 20.

A taxa de reprovações foi de 0,5% no total dos cinco anos lectivos estudados.

Hipóteses

1. Os resultados da avaliação de desempenho (CS urbano, CS rural), avaliação de conhecimentos e classificação final são equivalentes

Para medir a intensidade da associação entre as variáveis agrupadas duas a duas aplicou-se o Teste de Correlação de Pearson (Tabela 1).

Tabela 1- Coeficiente de correlação de Pearson

	Avaliação CS rural	Avaliação Conhecimentos	Classificação Final
Avaliação CS urbano	r = 0,254*	-	-
Avaliação Desempenho	-	r = 0,229*	r = 0,369*
Avaliação Conhecimentos	-	-	r = 0,970**

* p< 0,001 ** p<0,05

Existe uma correlação positiva e fraca entre as variáveis estudadas, excepto para *Avaliação de Conhecimentos e Classificação final*, em que se verificou existir uma correlação positiva forte (r= 0,97, p<0.05). O coeficiente de determinação R² é 0,94 o que significa que aproximadamente 94% dos valores da Classificação Final são explicados pela variância dos valores da Avaliação de Conhecimentos.

A Tabela 2 mostra as diferenças entre as médias das várias avaliações, e a significância do teste t de Student para amostras emparelhadas.

A análise dos resultados obtidos permite constatar que não existe diferença, com significado estatístico, entre as médias das avaliações de desempenho no CS rural e urbano (p= 0,187). No que diz respeito à comparação das médias das outras três variáveis agrupadas duas a duas, verifica-se que as diferen-

ças observadas são estatisticamente significativas ($p < 0,05$).

Tabela 2 – Teste *t* de Student para amostras emparelhadas

	Diferença entre as Médias	<i>p</i>
Av. CS rural – Av. CS urbano	-0,195	0,187
Av. Desempenho – Av. Conhecimentos	2,278	0,000
Av. Conhecimentos – Classificação Final	-0,596	0,000
Av. Desempenho – Classificação Final	1,682	0,000

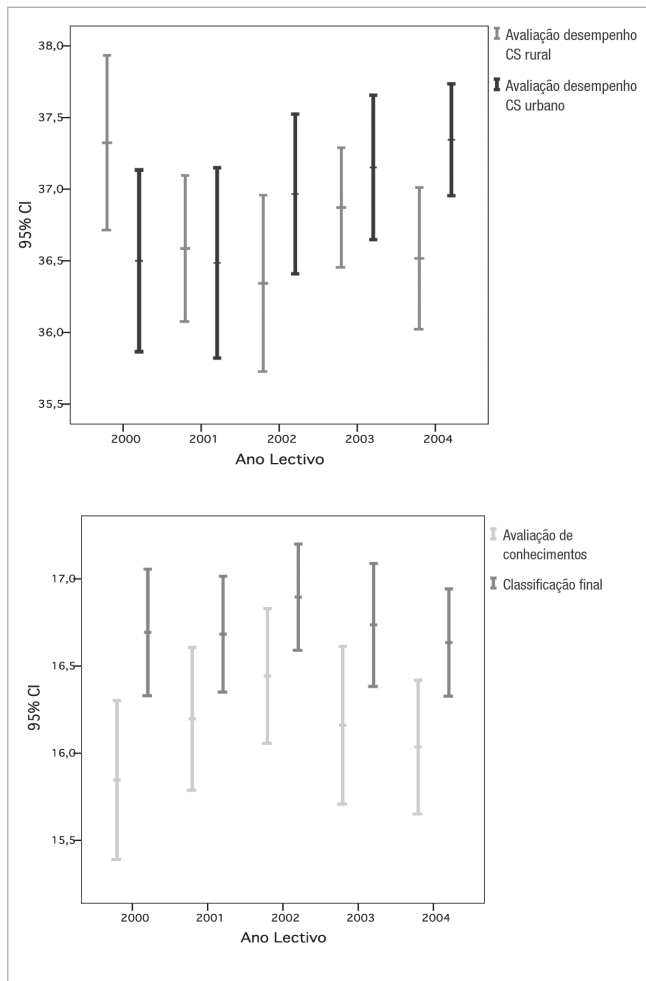


Figura 4 – Gráficos de erro Avaliação de Desempenho, Avaliação de Conhecimentos e Classificação final

2. Existe variação nos resultados das avaliações de desempenho, de conhecimentos e da classificação final, nos cinco anos lectivos

A Figura 4 mostra a evolução e as diferenças entre as médias da Avaliação dos CS urbanos e rurais, da Avaliação de Conhecimentos e da Classificação Final.

Na Tabela 3 constam as significâncias do teste *F* da ANOVA, análise da variância com um factor e do teste Kruskal Wallis que comparam as médias das avaliações nos cinco anos lectivos.

Tabela 3 – Testes *F* da ANOVA * e Kruskal Wallis **

	<i>p</i>
Avaliação CS rural *	0,095
Avaliação Conhecimentos *	0,409
Classificação Final *	0,854
Avaliação CS urbano **	0,434

Não se observou variabilidade com significado estatístico, nas médias da Avaliação de Desempenho do CS rural, da Avaliação de Conhecimentos, da Classificação Final e da Avaliação de Desempenho do CS urbano em cinco anos lectivos.

3. Os resultados da avaliação são influenciados pelo género do aluno

Na Tabela 4 pode verificar-se que as médias das avaliações são sempre mais elevadas no género feminino.

Tabela 4 – Média das avaliações segundo o género do aluno

	Género aluno	Média
Avaliação de Desempenho	Masculino	18,325
	Feminino	18,453
Avaliação de Conhecimentos	Masculino	15,752
	Feminino	16,339
Classificação Final	Masculino	16,432
	Feminino	16,888

A Tabela 5 mostra a significância do teste *t* de Student para duas amostras independentes, que compara as médias das três avaliações segundo o género do aluno, assumida a não homogeneidade de variâncias (teste de Levene).

Tabela 5 – Teste *t* de Student para duas amostras independentes

	<i>p</i>
Avaliação Desempenho	0,226
Avaliação Conhecimentos	0,004
Classificação Final	0,005

Em relação à Avaliação de Desempenho constata-se que as diferenças observadas entre os dois géneros, não têm significado estatístico ($p > 0,05$). No que diz respeito à Avaliação de Conhecimentos e à Classificação Final, as alunas têm em média mais 0.5 valores, do que os alunos, e essa diferença tem significado estatístico ($p < 0,05$).

4. Os resultados da avaliação de desempenho são influenciados pelo facto de o aluno e o tutor serem do mesmo género

A análise das médias das classificações dos tutores, segundo o seu género e o género dos alunos (Figura 5) mostra que nos CS urbanos, os tutores do género masculino classificaram os alunos com médias mais elevadas do que os tutores do género feminino. Nos CS urbanos e rurais os tutores do género masculino

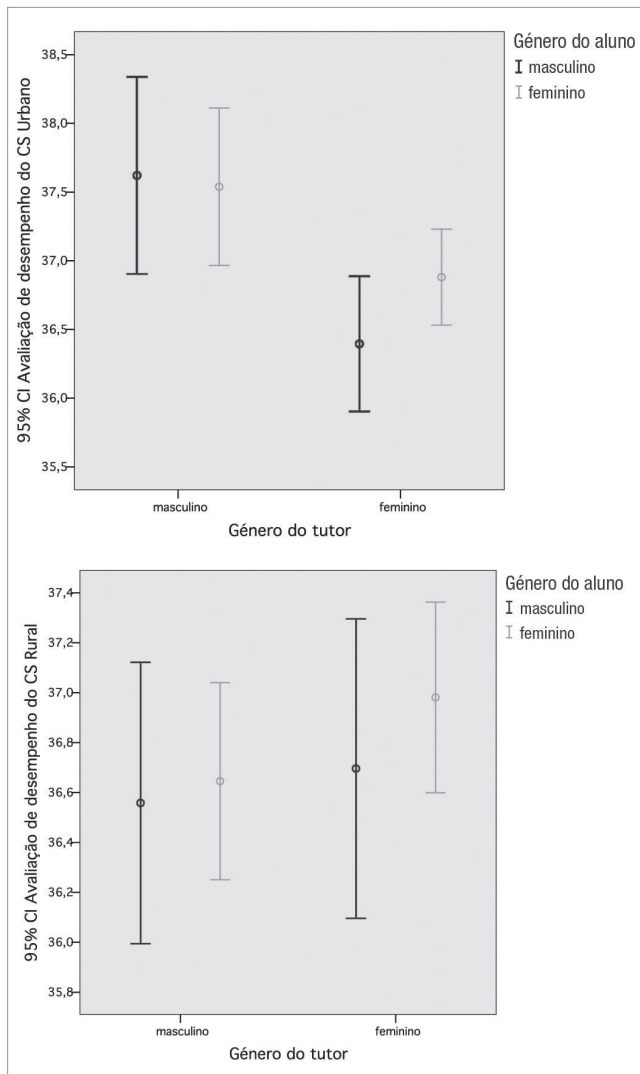


Figura 5 – Gráfico de erro para as classificações dos tutores dos CS urbanos e rurais segundo o seu género e o género dos alunos

classificaram os alunos do género masculino com média semelhante aos alunos do género feminino. Os tutores do género feminino, nos CS urbanos e rurais, classificaram, os alunos do género feminino com médias mais elevadas que as dos alunos do género masculino.

Verifica-se (Tabela 6) que não existe interacção entre o género do tutor e o género do aluno na Avaliação de Desempenho dos CS urbano e rural, o efeito do género do tutor sobre a Avaliação de Desempenho é o mesmo para os alunos do género masculino e feminino ($p > 0,05$).

Tabela 6 – ANOVA, análise da variância com dois factores fixos

	<i>p</i>
Género tutor CS urbano	0,001
Género aluno	0,486
Género tutor CS urbano × Género aluno	0,327
Género tutor CS rural	0,342
Género aluno	0,455
Género tutor CS rural × Género do aluno	0,691

Uma vez que não existe interacção entre os factores, analisou-se o efeito principal do género do tutor considerado isoladamente, sobre a avaliação de desempenho dos alunos. No caso do CS urbano ($p = 0.001$) constata-se que o género do tutor do CS urbano tem influência na média da avaliação de desempenho dos alunos, o que não acontece no CS rural ($p = 0.455$). Para o factor género do aluno, e em ambos os casos (CS urbano e rural), o nível de significância é > 0.05 , confirma-se, portanto, que as médias da Avaliação de Desempenho são iguais nos dois géneros.

Discussão

Durante o período estudado verifica-se que o número de alunos sofre um lento mas progressivo aumento decorrente das sucessivas intervenções a nível dos *numerus clausus*, que condicionam o acesso às Faculdades de Medicina (20). No que diz respeito ao género verifica-se, em todos os anos estudados, um predomínio do género feminino. Desde há alguns anos que se vem observando um aumento progressivo do acesso das mulheres ao ensino superior (21-23) em geral, e às Faculdades de Medicina em particular (24, 25).

Em relação aos tutores, predomina também o género feminino, a este facto não é certamente

alheia a chamada “feminização da medicina” (23, 26). O facto de existir, nos CS urbanos, uma grande proporção de tutores que orientaram entre 1 e 3 alunos em cinco anos, mostra uma maior rotatividade daqueles relativamente aos tutores dos CS rurais.

A organização pedagógica da disciplina inspira-se numa perspectiva do ensino baseado em competências (27), em que se privilegiam as áreas do conhecimento, da perícia e das atitudes no contexto da prática médica real. O facto de, no desenho da avaliação, estarem previstas vertentes teóricas e práticas e se utilizarem vários instrumentos, faz pressupor que existe coerência entre o ensino e a avaliação.

O instrumento utilizado na avaliação de desempenho apresenta algumas das características da Lista de Verificação (*Checklist Evaluation*), considerada útil na avaliação de competências e dos seus componentes; na avaliação das perícias relacionadas com o doente (história clínica, exame físico e procedimentos); nas relacionadas com a comunicação e relações interpessoais e ainda na auto-avaliação (28). De facto, os indicadores previstos na avaliação de desempenho contemplam estes aspectos e, em relação ao modelo de Miller (29) correspondem aos dois últimos níveis da pirâmide da avaliação da aprendizagem (mostra como se faz e faz).

Este instrumento de avaliação não foi testado quanto à sua validade e fiabilidade, desconhece-se também o seu impacto na aprendizagem, o que parece acontecer frequentemente em Educação Médica (30).

Como também acontece noutros países europeus, (31, 32) nem todos os tutores tiveram formação pedagógica e não foram treinados no uso do instrumento de avaliação. Embora a boa definição dos critérios possa contribuir para a uniformidade da aplicação do instrumento, o facto de este ser aplicado por muitos tutores e se saber que as escalas de Likert são influenciadas pela personalidade individual (33) pode condicionar alguns enviesamentos. Por outro lado, o estatuto voluntário e não remunerado dos tutores pode condicionar o seu investimento numa actividade que não lhes traz benefícios. Finalmente, o facto de todo o processo de ensino-aprendizagem e de avaliação ocorrer em simultâneo com a prática clínica diária normal do tutor, limita a sua disponibilidade para o aluno (...) *se continua a consultar 15-20 utentes por dia, não consegue ter tempo para explicar o que faz, o que prescreve, inteirar o aluno daquela pessoa e dos seus problemas específicos.*” N. O., in *Análise Crítica do Estágio* 14/9/2006

Cabe aqui realçar que a avaliação de conhecimentos nesta disciplina tem sido, nalguns aspectos, verdadeiramente inovadora. De facto, desde o ano lectivo 2000-2001 que utiliza um relatório que tem o formato e algumas características do *portfolio* (34).

Embora a avaliação final tenha como base a discussão do relatório, não estão definidos critérios explícitos para a avaliação dos seus conteúdos, como se prevê (35, 36) na aplicação deste tipo de instrumento. Dado que em última análise se avaliam os conhecimentos do aluno, esta avaliação tem muitas das características de um exame oral, não standardizado (37). Apesar de o exame oral ser afectado por muitos factores e ser difícil estimar a sua fiabilidade, ainda é considerado por muitas instituições como uma dimensão ímpar, importante e válida nos seus sistemas de avaliação (38).

Em relação às outras Faculdades de Medicina do país, a divulgação da forma como se faz a avaliação da disciplina de Clínica Geral é muito variável. Também varia, nas diversas faculdades, a denominação desta disciplina e o ano em que é leccionada, embora de uma forma geral os objectivos gerais e específicos sejam muito semelhantes entre elas. Para facilitar a comparação entre os vários sistemas de avaliação vigentes na altura em que foi realizado o estudo, (39-42) elaborou-se um quadro-resumo.

Nas faculdades analisadas, a avaliação incide sobretudo no desempenho, excepto na Faculdade de Medicina de Coimbra em que apenas são avaliados os conhecimentos, a avaliação sumativa ainda é a modalidade predominante. No que diz respeito aos instrumentos utilizados, de uma forma geral, parecem adequar-se à modalidade de avaliação escolhida, excepto na ECSUM em que, na altura, se preconizavam Exame Escrito e Trabalho de Grupo para avaliar o desempenho e os conhecimentos dos estudantes.

Quando se faz a análise do sistema de avaliação da FCML na perspectiva dos resultados que gerou, verifica-se que a taxa de reprovações é muito baixa e que as classificações são muito elevadas. No sentido de verificar se este é um fenómeno isolado ou ocorre também nas outras disciplinas do estágio do 6º ano, solicitaram-se à FCML todas as classificações obtidas pelos alunos nos anos estudados. Constatou-se que, quer nos estágios médicos, quer nos cirúrgicos, a distribuição das classificações segue o mesmo padrão (43-49). Nas duas Faculdades de Medicina que disponibilizam os resultados das suas avaliações encontram-se resultados também muito semelhantes (39, 50). Alguns autores têm referido que os alunos

admitidos em Medicina são relativamente homogéneos e que a competição, durante o curso, tende a traduzir-se por classificações académicas muito elevadas (51).

Para explicar estes resultados, aparentemente também verificados nas outras Faculdades de Medicina podem admitir-se algumas hipóteses: Inflação das notas que desde há muitos anos se verifica nas Faculdades de Medicina; tendo em conta a selecção a que foram sujeitos, os estudantes podem ter um elevado nível de conhecimentos no último ano do curso; o nível de exigência dos docentes pode ser reduzido, uma vez que no contexto de um estágio, para não prejudicar o aluno e/ou por atitudes defensivas, podem classificá-lo num nível mais elevado.

A média da avaliação de desempenho dos tutores dos CS urbanos é semelhante à dos tutores dos CS rurais. Tendo em conta as características do ensino-aprendizagem em que se pressupõe uma progressividade dos conhecimentos do aluno, poder-se-ia esperar que o seu desempenho melhorasse entre a primeira e a segunda avaliação. Teoricamente a avaliação formativa (35) acompanha o processo de aprendizagem, evidencia os progressos realizados e simultaneamente proporciona *feedback* (33). Os resultados obtidos na análise desta “avaliação formativa”, em dois momentos diferentes, parecem estar em contradição com aqueles pressupostos. Para explicar este facto pode admitir-se a influência de factores relacionados com a própria avaliação que pode ser configurada com algumas das características da avaliação sumativa. Cada um dos avaliadores ao aplicá-la sempre pela primeira vez, embora ao fim de duas semanas de observação, documenta apenas o ponto de chegada e nunca o ponto de partida

nem o que foi sendo adquirido durante o percurso. Por outro lado, o intervalo de duas semanas pode não ser suficiente para identificar mudanças no que diz respeito ao desempenho.

A média da avaliação de desempenho é mais elevada que a da avaliação de conhecimentos. Numa abordagem reflexiva e decorrente da experiência pessoal admite-se que a relação aluno/avaliador e o contexto em que decorrem as avaliações podem ter alguma influência na diferença entre estes resultados. Com o tutor o aluno tem uma relação pessoal de 1:1, neste contexto é relativamente comum a influência de factores de ordem afectiva; o tutor sente que a sua função é ajudar o aluno, a avaliação é desdramatizada e perde a sua conotação negativa. No que diz respeito à avaliação de conhecimentos, apesar de o aluno poder conhecer os elementos do júri, a relação que se estabelece é mais formal, mais curta e transporta inevitavelmente a carga negativa de um exame oral.

A evidência de uma relação linear forte e positiva entre a avaliação de conhecimentos e a classificação final é consequência da fórmula utilizada para o cálculo da classificação final do aluno. A média mais elevada da classificação final relativamente à média da avaliação de conhecimentos reflecte provavelmente a influência da avaliação de desempenho apesar de o seu coeficiente de ponderação ser de 1. De facto, a avaliação de desempenho tem pouca influência na classificação final e, com o tempo, tem vindo a ser desvalorizada. No ano lectivo de 2000-2001 o coeficiente de ponderação da avaliação de conhecimentos era 2, passou para 3 nos anos lectivos seguintes.

Nesta disciplina e na Educação Médica em geral (52), embora se admita a mudança de paradigma

Quadro I – Comparação dos sistemas de avaliação da disciplina de Clínica Geral nas Faculdades de Medicina

		FCML	FML	FMUP	FMC	ECSUM
Ano do curso		6º	6º	6º	5º	4º
Conteúdos da avaliação	Desempenho	X	X	X		X
	Conhecimentos	X			X	X
Modalidade de avaliação	Formativa	X				X
	Sumativa	X			X	X
	Contínua		X	X		
Instrumentos Utilizados		Lista de verificação Relatório Exame oral	Crítérios individuais	Formulário Trabalhos individuais	Exame escrito	Exame escrito Trabalho de grupo

FML – Faculdade de Medicina de Lisboa; FMUP – Faculdade de Medicina da Universidade do Porto; FMC – Faculdade de Medicina de Coimbra; ECSUM – Escola de Ciências da Saúde da Universidade do Minho

* Não há dados disponíveis sobre as Faculdades de Medicina dos Açores e Madeira, Faculdade de Ciências da Saúde da Universidade da Beira Interior nem do Instituto de Ciências Biomédicas Abel Salazar

em termos de cultura da avaliação no sentido de valorizar a avaliação formativa, a última palavra ainda continua a ser a da avaliação sumativa.

Não se observou variabilidade, nos cinco anos lectivos estudados, em nenhuma das médias das classificações dos alunos. Admite-se que, ao longo dos anos, se pudesse constatar alguma variação nas classificações da avaliação de desempenho. Em parte, por se poder esperar que a experiência dos tutores (53), factor de grande importância na avaliação formativa, (52) pudesse ter alguma repercussão nos resultados da avaliação. Contudo, os resultados obtidos vão de encontro ao que tem sido referido na literatura internacional (51) sobre a homogeneidade destes alunos. De facto, é possível que o processo de selecção que só permite a entrada dos melhores alunos, contribua para que os alunos de Medicina, de ano para ano, constituam coortes muito semelhantes.

Quando se analisaram os resultados segundo o género do aluno constatou-se que relativamente à avaliação de desempenho não verificaram diferenças. Contudo, as médias da avaliação de conhecimentos e as da classificação final são mais elevadas no género feminino. Ao contrário da avaliação de desempenho parece haver na avaliação de conhecimentos um maior empenho das alunas em relação aos alunos no sentido de obter classificações mais elevadas provavelmente porque ela é decisiva para a classificação final.

Os resultados obtidos são concordantes com estudos nacionais (21, 22) e internacionais (25, 54-56) que referem que, no ensino superior em geral e no caso da Medicina em particular, as alunas obtêm níveis de classificação superiores aos alunos. Na literatura internacional sobre Educação Médica, os estudos mais antigos, até aos anos 90, (57-59) ou não identificaram quaisquer diferenças relacionadas com o género do aluno ou observaram uma diferença marginal favorável ao género feminino (60) e particularmente em áreas como a ginecologia (61), a cirurgia (62) e a psiquiatria (63). Os estudos mais recentes (25, 54, 56) referem que de uma forma geral as alunas têm um melhor desempenho que os alunos. Um estudo irlandês (25) que analisou as diferenças nos resultados do exame final dos alunos de Medicina durante cinco anos, constatou que as alunas têm melhores resultados que os alunos em todas as áreas. Outro estudo (54) que analisou os resultados de duas faculdades de Medicina de Londres em duas coortes sucessivas, obteve resultados semelhantes.

Em relação à avaliação de desempenho não se demonstrou que exista interacção entre o género do tutor e o género do aluno, tanto em contexto urbano como rural. Constatou-se, no entanto, que nos CS urbanos os tutores do género masculino classificam os alunos, independentemente do seu género, com médias mais elevadas do que os tutores do género feminino. Há autores (55) que admitem que os tutores do género feminino têm critérios diferentes dos do género masculino quando avaliam os alunos, sendo que as mulheres são avaliadoras mais críticas que os homens. Na literatura relacionada com a Educação Médica, foi publicado um estudo (55) que analisa o efeito do género do aluno e do tutor num contexto de ambulatório. Não foi demonstrado que o género do tutor só por si tivesse alguma influência na avaliação. Contudo a interacção do género do aluno e do tutor pareceu afectar a avaliação, os tutores do género masculino classificaram as alunas com níveis mais elevados que os alunos.

Conclusão

A disciplina de Clínica Geral é leccionada na FCML desde o ano lectivo 2000-2001 tendo sido analisados os cinco primeiros anos. Verificou-se que embora a avaliação de desempenho tenha a designação de formativa, no contexto real assume as características de uma avaliação sumativa e tem pouca importância, face à avaliação de conhecimentos, na classificação final. Não existem critérios explícitos para a avaliação dos conteúdos do relatório, e não é possível estimar a fiabilidade do exame oral porque as perguntas são diferentes para todos os alunos. Constatou-se que, as classificações dos alunos foram, de uma forma geral, muito elevadas e que a taxa de reprovação foi muito baixa. A análise dos resultados gerados pelo sistema de avaliação fez-se através da verificação quatro hipóteses de investigação.

Os resultados da avaliação de desempenho (CS urbano, CS rural), avaliação de conhecimentos e classificação final não são equivalentes. Tendo em conta o pressuposto de uma avaliação formativa e sequencial constatou-se que, ao contrário do esperado, as médias da avaliação de desempenho nos CS rurais são iguais às dos CS urbanos. No caso particular da comparação entre as médias da avaliação de desempenho e as da avaliação de conhecimentos, e apesar de não existirem referências na literatura, as diferenças encontradas foram atribuídas à possível influência de factores afectivos. A evidência de uma relação linear forte e positiva

entre a avaliação de conhecimentos e a classificação final é determinada pela maior influência daquela relativamente à avaliação de desempenho.

Nos cinco anos lectivos estudados não se verificou variabilidade com significado estatístico, nas médias das avaliações de desempenho, de conhecimentos e da classificação final. A forma como estes alunos são seleccionados pode levar à formação de coortes com comportamentos e resultados académicos muito semelhantes entre si.

Em relação à possível influência do género nas classificações dos estudantes, verificou-se que as médias da avaliação de desempenho são iguais nos dois géneros. O facto de se tratar de uma avaliação parcelar, que pouco contribui para a classificação final, pode minimizar a competição pelas notas elevadas. Em relação à avaliação de conhecimentos e à classificação final, as suas médias são mais elevadas no género feminino. Estes resultados são concordantes com estudos nacionais e internacionais, no âmbito da psicologia, sociologia e Educação Médica.

Não se identificaram interacções com significado estatístico entre o género do aluno e do tutor tanto no contexto rural como no urbano. Contudo, demonstrou-se que nos CS urbanos os tutores do género masculino classificam os alunos com médias mais elvada que os tutores do género feminino.

Esta conclusão de uma investigação empírica permite enunciar algumas sugestões relativamente ao sistema de avaliação analisado. Em primeiro lugar, sugere-se que se reforce a importância da avaliação formativa face à sumativa e que se inverta a tendência, observada nos últimos anos, de desvalorização da primeira. Um corpo mais estável de tutores com formação e treino nas áreas pedagógicas e avaliativas como forma de otimizar e uniformizar os seus procedimentos pode contribuir para a valorização daquela avaliação. Em segundo lugar, e relativamente à avaliação sumativa sugere-se que sejam explicitados os critérios para a avaliação do relatório. Desta forma, este instrumento pode complementar o exame oral para reforçar o alinhamento entre o ensino-aprendizagem e a avaliação dos conhecimentos e das competências adquiridas pelo aluno.

Para terminar, salienta-se que o facto de os alunos de Medicina obterem classificações muito elevadas não determina necessariamente que todos eles venham a ser excelentes médicos. A Medicina continua, em muitos aspectos, a ser uma arte e nem sempre os melhores da turma se tornam, na prática, os melhores profissionais. A valorização da

avaliação de desempenho dos alunos, em particular na disciplina de Clínica Geral, pode certamente contribuir para melhorar e adequar as suas atitudes e comportamentos na sua prática clínica futura independentemente da especialidade que escolherem.

Para concluir, formula-se uma pergunta dirigida aos leitores, eventuais investigadores em Educação Médica: – Será que a investigação académica na área da avaliação pode, no futuro, contribuir para uma avaliação mais justa, mais benéfica, mais efectiva, mais válida e mais transparente nas Faculdades de Medicina portuguesas?

Agradecimentos

Ao Professor Doutor Coelho Rosa e ao Mestre Dr. Vítor Alaíz pelas sugestões. À FCML e ao Professor Doutor Mário Bernardo que manifestou o seu apoio a este estudo e me disponibilizou as classificações dos alunos. Por último os meus agradecimentos à Dra. Eleonora Paixão pelas sugestões e pela revisão da estatística.

Bibliografia

1. Shumway JM, Harden RM. AMEE Guide N° 25: The Assessment of learning outcomes for the competent and reflective physician. *Medical Teacher*. 2003;25(6):569-84.
2. WFME. Educação Médica Pré-Graduada Standards Globais da WFME. *Journal [serial on the Internet]*. 2005 Date; 15(1): Disponível em: http://www.spem.pt/docs/WFME_EMB.pdf.
3. Burge SM. Undergraduate medical curricula: are students being trained to meet future service needs? *Clinical Medicine*. 2003;3(3):243-6.
4. Norman G. Research in medical education: three decads of progress. *BMJ*. 2002(324):1560-2.
5. Wilkies M, Bligh J. Evaluating educational interventions. *BMJ*. 1999(318):1296-72.
6. Fowell SL, Maudsley G, Maguire P, Lenister SJ, Bligh J. Report of findings: student assessment in undergraduate medical education in the United Kingdom 1998. *Medical Education*. 2000;34(Suppl. 1, September):1-78.
7. Barreira C, Pinto J. A investigação em Portugal sobre a avaliação das aprendizagens dos alunos (1990-2005). s.d. [updated s.d.; cited Dezembro 2006]; Disponível em: <http://www.edu.fc.ul.pt/docentes/msantos/spce.pdf>.
8. Jollie C, McKimm J. Relatório do Projecto Nacional de Avaliação das Capacidades dos Alunos que Concluem os Cursos de Licenciatura em Medicina. Lisboa; 2004.
9. FMCoimbra. Comissão de Avaliação Externa das Licenciaturas em Medicina. 2004.
10. Soares J. Educação Médica: Quo vadis? *Medicom*. 2005 15 Setembro 2005(95):12.
11. Melo FG. Probabilidades e Estatística, conceitos e métodos fundamentais. Lisboa: Escolar Editora; 1993.
12. Pereira A. SPSS Guia Prático de Utilização. 5.ª ed. Lisboa: Edições Sílabo; 2004.

13. Hill MM, Hill A. *Investigação por questionário*. 2 ed. Lisboa: Edições Sílabo; 2005.
14. Pestana MH, Gageiro JN. *Análise de Dados para Ciências Sociais*. 4ª ed. Lisboa: Edições Sílabo; 2005.
15. Melo FG. *Métodos Estatísticos em Estudos Comparativos*. Lisboa: ENSP; 1985.
16. FCML. Programa da Disciplina de Clínica Geral 2000-2001; 2000.
17. FCML. Programa da Disciplina de Clínica Geral 2006-2007; 2006.
18. FCML. Programa da Disciplina de Clínica Geral 2002-2003; 2002.
19. FCML. DEO-2006-2007: Instruções do seu preenchimento e outras informações 2006
20. ME. Guia de Candidatura 2006. Journal [serial on the Internet]. 2006 Date: Disponível em: www.acessoensinosuperior.pt.
21. Saavedra L. Escola, Género e Democracia: caminhos de igualdade e diferença. Journal [serial on the Internet]. s.d. Date.
22. Nóvoa (coord) A. Percursos Escolares dos estudantes da universidade de Lisboa, Relatório Preliminar nº1: Factores de sucesso e insucesso escolar na Universidade de Lisboa. Lisboa; 2005.
23. INE. *Mulheres e Homens em Portugal nos anos 90*. Lisboa; 2002.
24. Jorge NS. *Lógicas da Procura nos Cursos de Medicina em Portugal*. In: IV Congresso Português Sociologia, editor. Coimbra; 2000.
25. McDonough CM, Horgan A, Codd MB, Casey PR. Gender Differences in the results of the final medical examinations at University College Dublin. *Medical Education*. 2000 Jan.; 34(1):2-3.
26. Martins J, Biscaia A, Conceição C, Fronteira I, Hipólito F, Carrolo M, et al. Caracterização dos profissionais de saúde em Portugal: Parte I- Quantos somos e quem somos. *Revista Portuguesa de Clínica Geral*. 2003(19):513-7.
27. ACGME. Outcome Project Competency Perspective on Training. Journal [serial on the Internet]. 2006 Date: Disponível em: www.acgme.org.
28. ACGME, ABMS. Toolbox of Assessment Methods. Journal [serial on the Internet]. 2000 Date: Disponível em: www.acgme.org.
29. Miller GE. The Assessment of clinical skills/competence/performance. *Academic Medicine*. 1990(65Supl):563-7.
30. Newble D, Jolly B, Wakeford R. The certification and recertification of doctors: issues in the assessment of clinical competence. *BMJ*. 1994(309):1096.
31. Duroux G. Évaluation d'un SASPAS. Étude réalisée en Aquitaine. *La Revue du Praticien*. 2005;19(702/703):979-80.
32. Bireaud A. *Os métodos Pedagógicos no Ensino Superior*. Porto Editora; 1995.
33. Cotton J. *The Theory of Assessment*. London: Kogan Page; 2002.
34. Drissen E, vanTartwijk J, Vermunt JD, VanDerVeuten CP. Use of portfolios in early undergraduate medical training. *Medical Teacher* 2003;25(1):18-23.
35. Valadares J, Graça M. *Avaliando para Melhorar a Aprendizagem*. Lisboa: Plátano Edições Técnicas; 1998.
36. Roberts C, Newble D, O'Rourke A. Portfolio- based assessment in medical education: are they valid and reliable for sumative purposes? *Medical Education*. 2002(36):899-900.
37. Joughin G, Collom G. Oral assesement. Journal [serial on the Internet]. 2003 Date: Disponível em: <http://www.heacademy.ac.uk/resources>.
38. Wass V, Wakeford R, Neighbour R, Vleuten CVd. Achieving acceptable reliability in oral examinations: an analysis of the Royal College of General Practitioners membership examination's oral component. *Medical Education*. 2003(37):126-31.
39. FML. O 6º Ano em Estágio Clínico: Apreciação dos primeiros cinco anos. Lisboa; 2005 Contract No.: Document Number].
40. FMUPorto. *Medicina Comunitária (Prática Clínica)*. Journal [serial on the Internet]. 2006 Date: Disponível em: www.fmup.pt.
41. FMCoimbra. Programa da Disciplina de Clínica Geral. Journal [serial on the Internet]. 2006 Date: Disponível em: www.fmed.uc.pt.
42. EscolaCiênciasdaSaúde. *Residências em C.Saúde*. Journal [serial on the Internet]. 2006 Date: Disponível em: www.uminho.pt.
43. FCML. *Saúde Pública 2000, 2001, 2002, 2003, 2004, 2005*. xls. 2006.
44. FCML. *Clínica Cirúrgica 2000, 2001, 2002, 2003, 2004, 2005*. xls. 2006.
45. FCML. *Clínica Médica 2000, 2001, 2002, 2003, 2004, 2005*. xls. 2006.
46. FCML. *Clínica Obstétrica 2000, 2001, 2002, 2003, 2004, 2005*. xls. 2006.
47. FCML. *Clínica Pediátrica 2000, 2001, 2002, 2003, 2004, 2005*. xls. 2006.
48. FCML. *Especialidade 2000, 2001, 2002, 2003, 2004, 2005*. xls. 2006.
49. FCML. *Saúde Mental 2000, 2001, 2002, 2003, 2004, 2005*. xls. 2006.
50. EscolaCiênciasdaSaúde. *Relatório Anual 2004-2005 do Curso de Licenciatura em Medicina*. Journal [serial on the Internet]. 2006 Date: Disponível em: www.uminho.pt.
51. McLachlan JC, Whiten SC. Marks, scores and grades: scaling and aggregating student assessment outcomes. *Medical Education*. 2000(34):788-97.
52. Rushton A. Formative assessment a key to deep learning? *Medical Teacher*. 2005;27(6):509-13.
53. Mathers J, Parry J, Lewis S, Greenfield S. What impact will an increased number of teaching general practices have on patients, doctors and medical students. *Medical Education*. 2004;38(12):1219.
54. Haq I, Higham J, Morris R, Dacre J. Effect of ethnicity and gender on performance in undergraduate medical examinations. *Medical Education*. 2005(39):1126-8.
55. Wang-Cheng RM, Fulkerson PK, Barnas GP, Lawrence SL. Effect of Student and Perceptor Gender on Clinical Grades in an Ambulatory Care Clerkship. *Academic Medicine*. 1995;70(4):324-6.
56. Woodfield R, Earl-Novell S. Gender and performance: the impact of mode of assessment. Journal [serial on the Internet]. 2002 Date: Disponível em: <http://www.heacademy.ac.uk/resources>.
57. Holmes FF, Holmes GE, Hassanein R. Performance of male and female medical students in a medicine clerkship. *JAMA*. 1978;239(21):2259-62.
58. Weinburg E, Rooney JF. The academic performance of women students in medical school. *Medical Education*. 1973(48):240-7.
59. Harward DH, Lyons CM, Porter CQ, Hunter RS. Comparison of the performance of male and female medical students and residents. *Medical Education*. 1981(56):853-5.
60. Anderson JR, Lennox B, Low A. Medical students' performance. An analysis of selection procedures and examination marks in Glasgow. *Lancet*. 1964(1):96-100.
61. Plauché WC, Miller JM. Performances of female medical students in an obstetrics and gynecology clerkship. *Medical Education*. 1986(61):323-5.
62. Linn BS, Zeppa R. Sex and ethnicity in surgical clerkship performance. *Medical Education*. 1980(55):513-20.
63. Keitner GI, Baldwin LM, McKendall MJ. Gender and psychiatric clerkship performance. *Can J Psychiatry*. 1984(239):327-9.