

# ANÁLISE ESTRUTURAL DOS EXAMES DE INGRESSO NOS INTERNATOS COMPLEMENTARES DE 1994 A 1996

VITAL MORGADO, A. GOMES DA COSTA, ANA CORREIA\*, ANTÓNIO PEREIRA, A. CRUZ PINHO, A. TELES DE ARAÚJO, FERNANDO RAMALHO, JOAQUIM GOUVEIA, JORGE QUININHA, J. BRAZ NOGUEIRA, J. MANUEL ROMÃOZINHO, R. AMARAL MARQUES, RUI PATO  
 Júri Nacional do Exame de Ingresso.

## RESUMO

O principal instrumento de seriação dos candidatos para ingresso nos Internatos Complementares, é um exame escrito, tipo escolha-múltipla. Apesar de entendermos que é urgente repensar esta metodologia, é possível que este tipo de exame se mantenha, como parte da avaliação dos conhecimentos, à semelhança do que se passa noutros países. Não podemos, no entanto, continuar a elaborá-los sem que sejam criadas condições que permitam dispôr de um suporte lógico e científico, que resolva os problemas com que se debatem os médicos chamados a integrar os júris e que, por outro lado, constituem justas críticas dos médicos concorrentes. O trabalho que publicamos permite-nos avaliar os exames que elaborámos nos anos de 1994 a 1996. Julgamos assim contribuir para uma reflexão sobre este assunto, não constituído por afirmações teóricas ou de intenção, sempre fáceis de produzir, mas expondo um trabalho de três anos, acompanhado do cálculo de coeficiente de fiabilidade dos exames e dos índices de dificuldade e de discriminação das perguntas e dos exames no seu conjunto. Vamos publicar em Janeiro de 1997, em colaboração com o Departamento de Recursos Humanos da Saúde, uma compilação destes exames e uma análise pormenorizada dos mesmos.

## SUMMARY

### Structural analysis of the examinations for specialist medical training 1994-96

In order to obtain specialist training in Portugal, doctors must pass a multiple-choice examination. The aim of this article is to present a structural and mathematical analysis of the examinations in the last three years. We based our work on the calculation of reliability coefficient of the examinations, and the difficulty and discrimination index of the questions and the examinations as a whole. A detailed analysis of the examinations, including that of each of the three hundred questions, will be published by the Department of Health Manpower in January, 1997.

\* Falecida

A formação especializada em Medicina constitui um momento decisivo da vida dos médicos.

Em Portugal o ingresso nos Internatos Complementares é feito, desde 1970, após a realização de um exame escrito, tipo escolha-múltipla.

Os exames deste tipo pretendem medir os conhecimentos de forma objectiva e igual para todos os candidatos, ultrapassando os problemas da equidade e reprodutibilidade de outro tipo de exames, apesar do descrédito que progressivamente os foi atingindo, muitas vezes pelo pouco cuidado posto na sua elaboração e pela má utilização que deles foi feita<sup>1-3</sup>.

A organização destes exames exige um conhecimento

da metodologia de elaboração das questões e uma capacidade de adequação do tipo e da dificuldade das perguntas aos objectivos propostos que, no caso em apreço, é a avaliação dos conhecimentos exigíveis a médicos não-especialistas e a consequente discriminação dos concorrentes.

Entre nós esse desiderato não tem sido fácil de atingir nomeadamente pela inadequação das infra-estruturas às exigências metodológicas e ao impacto deste tipo de provas na vida dos médicos.

Aceitámos integrar o júri dos exames de 1994 a 1996 e isso pôs-nos perante as dificuldades e a responsabilidade de elaborarmos um exame adequado aos objectivos

propostos. Decidimos, agora, avaliar qualitativamente o nosso trabalho para ter uma visão crítica do mesmo e poder sugerir algumas modificações aos procedimentos actuais.

O trabalho que apresentamos tem como metas fundamentais<sup>2</sup>:

- Obter uma análise qualitativa dos testes efectuados.
- Avaliar explicitamente as características das perguntas inadequadas de forma a permitir uma informação fundamentada para a elaboração dos questionários futuros
- Proporcionar uma discussão da metodologia de elaboração deste tipo de exames.
- Contribuir para que os médicos que, futuramente, forem chamados à elaboração deste tipo de testes, disponham de material que os ajude na elaboração de um trabalho com um impacto profissional tão importante.
- Propor modificações aos procedimentos actuais tendo em vista melhorar a qualidade e rigor da elaboração das questões, introduzir alterações nas matérias e no tipo de perguntas por forma a privilegiar o raciocínio e as deduções e associações, essenciais à prática médica.

## MATERIAL E MÉTODOS

O concurso de ingresso nos Internatos Complementares consiste num exame escrito, tipo escolha-múltipla, com 100 perguntas, cada uma com 5 respostas alternativas, das quais uma é a mais adequada. As perguntas limitam-se a 5 áreas da medicina: a gastroenterologia, a pneumologia, a cardiologia, a hematologia e a nefrologia. Não há pontuações negativas para respostas incorrectas.

Não estão definidas, oficialmente, obras de referência para a elaboração e validação do exame. A chave de respostas correctas é da responsabilidade do júri que elabora as perguntas.

A metodologia que vamos seguir é a descrita por Guilbert<sup>3</sup>, Ebel<sup>4</sup>, Pieron<sup>5</sup> e Nunnally<sup>6</sup>, aplicada a exames de escolha-múltipla.

Para avaliar a qualidade de um exame, independentemente da sua natureza, é importante estudar algumas características, nomeadamente a validade, a fiabilidade, a objectividade e a pertinência. As três primeiras dizem respeito à validade estrutural do exame e ao estudo individualizado de cada pergunta. A quarta característica refere-se à adequabilidade do conteúdo das perguntas<sup>3</sup>.

1 - A validade aprecia o grau com que um teste mede o que se pretende medir. O seu cálculo é difícil e existem duas possibilidades. O primeiro é validar o conteúdo em função dos conhecimentos que os médicos obtiveram na Faculdade de Medicina, quer dizer, do que os concorrentes deveriam saber. A segunda é validar o conteúdo em função do objectivo que se pretende alcançar com o exame, neste caso, o ordenamento dos concorrentes em função dos seus conhecimentos. A validade é influenciada por uma série de características como o grau de dificuldade, o poder de discriminação e a adequabilidade das perguntas.

2 - A fiabilidade é um parâmetro mais facilmente objectivável e representa a consistência com que um ins-

trumento mede uma determinada variável. Em termos operacionais é um conceito estatístico que mede o erro aleatório dos resultados que, naturalmente, varia na razão inversa da fiabilidade

3 - A objectividade é definida como o grau de consenso, entre diferentes examinadores, do que é uma resposta correcta. É mais facilmente garantida num teste de escolha-múltipla do que noutro tipo de provas.

4 - A pertinência avalia se o que se pergunta ou avalia no exame corresponde ao que se devia perguntar ao avaliar. Idealmente o exame deveria ser capaz de identificar os concorrentes que têm os melhores níveis de conhecimentos.

Neste contexto e para avaliar as características dos exames de 1994 a 1996 iremos calcular a fiabilidade, a dificuldade e o poder de discriminação do exame, assim como os índices de dificuldade e de discriminação de cada uma das 100 perguntas, e de cada um dos 5 grupos de matérias, com o objectivo de apreciar a qualidade dos exames realizados.

### 1 - Fiabilidade

Há vários métodos para o estudo da fiabilidade de um teste sem que pareça haver vantagens de um sobre os outros. Para calcular o coeficiente de fiabilidade vamos utilizar a fórmula de Kuder-Richardson<sup>5</sup>:

$$\text{Coeficiente de fiabilidade} = \frac{K(1 - M(K-M))}{K-1 KS2}$$

K = número de perguntas do exame

M = média de respostas correctas obtidas pelos examinandos

S2 = variância do número de respostas correctas

Considera-se como valor aceitável um coeficiente de fiabilidade superior a 0,90.

### 2 - Dificuldade

O índice de dificuldade de uma pergunta permite determinar, se ela é difícil ou fácil para o grupo de examinandos.

O índice de dificuldade (tal como o índice de discriminação) calcula-se comparando as respostas correctas do grupo de concorrentes melhor classificados (grupo forte) com as obtidas pelos concorrentes pior classificados (grupo fraco). A composição dos grupos forte e fraco pode obter-se de várias formas mas Ebel sugere que se utilizem os primeiros 27 % (grupo forte) e os últimos 27 % (grupo fraco), segundo a ordenação dos concorrentes.

Calcula-se, para cada pergunta, através da seguinte fórmula:

$$\text{Índice de dificuldade} = \frac{100(F + D)}{N1 + N2}$$

F = número de respostas correctas no grupo forte

D = número de respostas correctas no grupo fraco

N1 = número de concorrentes do grupo forte que responderam à pergunta

$N_2$  = número de concorrentes do grupo fraco que responderam à pergunta

Os valores que este índice pode ter variam de 0 a 100; os índices elevados, correspondem às perguntas fáceis, enquanto que as perguntas difíceis apresentam índices mais baixos. O valor óptimo situa-se entre 50 e 60; os índices maiores que 30 e menores que 80 consideram-se aceitáveis. Uma pergunta com um índice  $< 30$  considera-se difícil e com um índice  $> 80$  considera-se fácil. O índice de dificuldade de um exame calcula-se pela média dos índices de dificuldade das perguntas.

### 3 - Discriminação

O poder da discriminação é a capacidade de distinguir um candidato bom de um candidato mau, em função das respostas correctas às perguntas do exame. Uma discriminação adequada é um elemento absolutamente crucial para a validade deste tipo de exames.

Para o cálculo do índice de discriminação de uma pergunta, usa-se a seguinte fórmula:

$$\text{Índice de discriminação} = \frac{2(F - D)}{N_1 + N_2}$$

F = número de respostas correctas no grupo forte

D = número de respostas correctas no grupo fraco

$N_1$  = número de concorrentes do grupo forte que responderam à pergunta

$N_2$  = número de concorrentes do grupo fraco que responderam à pergunta

O índice de discriminação varia entre -1 e +1. Quanto mais elevado é o índice discriminativo de uma pergunta, maior é a sua capacidade para distinguir um concorrente forte de um concorrente fraco.

Com respeito a este índice, as perguntas podem classificar-se da seguinte forma:

$> 0,35$	pergunta excelente
$0,34 - 0,25$	pergunta boa
$0,24 - 0,15$	pergunta limite
$0,14 - 0$	pergunta má
$< 0$	pergunta muito má

Tal como para o índice de dificuldade, o índice de discriminação do exame obtêm-se pela média aritmética dos índices de discriminação das perguntas. Classifica-se como bom, um exame com um índice discriminativo global superior a 0,25.

Um exame com um índice de dificuldade entre 50 e 60 e um índice discriminativo igual ou superior a 0,25 tem grandes probabilidades de ser fiável e válido.

Relacionando os índices de dificuldade e de discriminação é possível identificar as perguntas inadequadas ou sejam as que têm índices de discriminação maus ou muito maus e as que têm uma dificuldade inadequada com um poder de discriminação limite. A análise das características destas perguntas tem enorme importância na elaboração destes testes.

## RESULTADOS

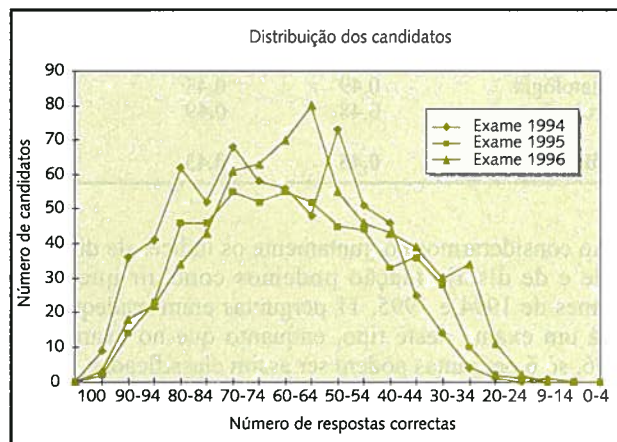
Apresentamos os resultados da análise que fizemos dos exames de 1994, 1995 e 1996, que foram elaborados pelo mesmo júri.

No *quadro I* encontram-se alguns dados estatísticos referentes aos três exames:

*Quadro I*

Ano	Classific. Máx.	Classific. Mín.	Média	Desvio Padrão
1994	98	9	60,74	17,03
1995	96	14	60,62	17,04
1996	98	17	57,89	17,59

A *figura 1* mostra a distribuição dos candidatos nos três exames, de acordo com o número de respostas correctas.



*Fig. 1*

### 1 - Fiabilidade dos exames

Nos exames que estamos a avaliar, os coeficientes de fiabilidade foram sempre superiores a 0,90 (1994 - 0,93; 1995 - 0,93; 1996 - 0,92) o que, de acordo com a metodologia utilizada, garantem a consistência dos resultados obtidos.

### 2 - Dificuldade dos exames

No *quadro II* encontram-se os índices de dificuldade, por áreas, dos exames dos três últimos anos.

De acentuar que os valores médios permitem classificar as provas de 1994 e de 1995 como aceitáveis, enquanto que, o exame de 1996, com um índice de dificuldade de 57,70 se situa no intervalo ideal (50-60).

De acordo com a metodologia utilizada verifica-se que nos dois primeiros exames, 3 perguntas eram difíceis, 83 apresentavam dificuldade adequada e 14 eram fáceis. No exame de 1996, 6 perguntas eram difíceis, 89 tinham dificuldade adequada e 5 eram fáceis.

Quadro II

	1994	1995	1996
Gastrenterologia	71,87	55,30	56,12
Pneumologia	70,31	72,29	65,41
Cardiologia	59,40	60,24	58,03
Hematologia	59,51	57,91	52,22
Nefrologia	60,25	56,10	56,69
<b>Média geral</b>	<b>64,27</b>	<b>60,37</b>	<b>57,70</b>

### 3 - Discriminação do exame

Todos os exames tiveram óptimos índices de discriminação, que se apresentam no *Quadro III*.

Quadro III

	1994	1995	1996
Gastrenterologia	0,38	0,42	0,41
Pneumologia	0,34	0,37	0,45
Cardiologia	0,45	0,44	0,44
Hematologia	0,49	0,43	0,42
Nefrologia	0,48	0,49	0,50
<b>Média geral</b>	<b>0,45</b>	<b>0,43</b>	<b>0,44</b>

Ao considerarmos conjuntamente os índices de dificuldade e de discriminação podemos concluir que, nos exames de 1994 e 1995, 11 perguntas eram inadequadas para um exame deste tipo, enquanto que no exame de 1996, só 6 perguntas podem ser assim classificadas.

Consideram-se inadequadas ou incorrectas as perguntas com índice discriminativo mau ou muito mau e as de dificuldade inadequada (difíceis ou fáceis) com discriminação limite.

### DISCUSSÃO

Os resultados que apresentamos permitem-nos afirmar que os exames estavam elaborados de forma adequada quanto à fiabilidade, dificuldade e discriminação, para a população a que se dirigia.

É fundamental que estes objectivos sejam sempre alcançados, mas para isso é necessário rotinar determinados procedimentos, que diminuirão ao mínimo factores individuais, de grupo ou circunstanciais que influenciarão negativamente não só o instrumento como os objectivos de selecção propostos.

Assim julgamos indispensável a criação de uma base de perguntas, necessidade aliás já referida pelo júri que nos precedeu. Estas perguntas seriam elaboradas por médicos com responsabilidades no ensino médico e

segundo regras previamente definidas e a partir da qual se elaboraria, em cada ano o exame. Isto permitiria ultrapassar as dificuldades actuais, eliminando alguns dos erros atribuídos aos exames dos últimos anos.

A existência de vários tipos de perguntas, como é preconizado para este tipo de exames<sup>3,7,8</sup>, nomeadamente as que incidissem sobre situações concretas de casos clínicos e suas implicações diagnósticas e terapêuticas, levariam os candidatos a privilegiar o desenvolvimento de um conhecimento raciocinado e não meramente repetitivo e a ter a necessidade de estabelecer associações e correlações de entre as informações colhidas na prática clínica.

O alargamento das áreas do conhecimento médico a incluir no exame é indispensável, de forma a incluir entidades nosológicas tão importantes como as doenças infecciosas, as doenças cerebro-vasculares, algumas doenças endócrinas, entre outras, assim como se deveria ponderar o impacto do aumento do número de perguntas na capacidade discriminativa do exame

A existência de um júri, não envolvido na selecção das perguntas, e que, no dia do exame, o validasse, daria a este maior rigor e permitiria corrigir eventuais problemas com perguntas incorrectamente formuladas ou de resposta dúbia.

Julgamos ainda que deve ser introduzido um sistema de ponderação das classificações em cada ano, que anule as desvantagens introduzidas por inevitáveis variações da dificuldade do exame.

Naturalmente que estas modificações teriam que ser implementadas progressivamente e sobretudo após um esclarecimento atempado, nomeadamente por uma ampla divulgação, entre os candidatos, dos objectivos, metodologia, tipos de perguntas, inclusive com uma larga divulgação de exemplo dos exames a implementar.

Sem querer entrar, neste momento, na polémica do método de selecção para acesso aos Internatos Complementares julgamos urgente tomar algumas medidas que permitam melhorar a qualidade e dignificar o exame actual.

### BIBLIOGRAFIA

1. PINTO-CORREIA J, MATOS LC, GIL VM: Exames de Medicina com questões de escolha múltipla. Enunciado do problema e análise de uma experiência em Medicina Interna. *Jornal Soc. Ciências Médicas*, 1979;143:131-58
2. HERNANDEZ PJS, GONZALEZ RM: Validez y adecuacion estructural del examen nacional de acceso a la formacion medica especializada. Evaluacion de la prueba correspondiente a la convocatoria de 1988-89. *Rev San Hig Pub*, 1989;63:117-33
3. GUILBERT J J: Educational handbook for health personnel. Geneva: World Health Organization, 1992
4. EBEL R: Measuring Educational Achievement. New York: Prentice Hall, 1965
5. PIERON H: Examens et docimologie. Paris, PUF, 1989
6. NUNNALLY J C: Psychometric theory. New York, Mc Graw-Hill, 1978
7. PINTO-CORREIA J, GRIMA N, MATOS L: Análise dos exames finais de uma cadeira de Medicina Interna. *Jornal Soc. Ciências Médicas* 1984;148:270-77
8. National Board of Medical Examiners. General instructions, content description, and sample items, 1996