

Avaliação do Exame Final do Internato Médico em Portugal

Evaluation of The Medical Board Exam in Portugal



Tiago Reis MARQUES✉*¹, Inês LAÍNS*^{2,3,4}, Maria João MARTINS^{5,6}, Francisco GOIANA-DA-SILVA⁷, Filipa SAMPAIO^{8,9}, Inês PESSANHA¹⁰, Diogo Hipólito FERNANDES¹¹, Mariana BRANDÃO^{12,13}, Pedro PINTO TEIXEIRA¹⁴, Manuel de OLIVEIRA SANTOS^{4,15}, João Carlos SILVA¹⁶, João Carlos RIBEIRO^{4,17,18}
Acta Med Port 2018 Nov;31(11):670-679 • <https://doi.org/10.20344/amp.10646>

RESUMO

Introdução: Existe uma elevada heterogeneidade na estrutura da avaliação da formação médica pós-graduada a nível mundial. No entanto, contrastando com outros países, não existem estudos científicos em Portugal que tenham avaliado o modelo da avaliação final da especialidade. O presente estudo pretendeu avaliar a adequação do exame do final da especialidade aos seus propósitos; aí incluída a sua validade enquanto consubstanciada na relação com a prova nacional de seriação e média final de curso de medicina.

Material e Métodos: Estudo transversal, observacional. Foram analisadas com recurso a medidas de tendência central e variabilidade, as notas na avaliação final da especialidade de 2439 médicos, de 47 especialidades, que terminaram a sua formação em 2016 e 2017. Tendo em vista a sua validação cruzada, foram também avaliadas as correlações com a média final de curso e a nota na prova nacional de seriação.

Resultados: Das medidas de tendência central e variabilidade, e consequentes medidas de formato, resulta que a distribuição das pontuações do exame final de especialidade se apresenta com uma forma manifestamente assimétrica negativa e leptocúrtica. No geral, verificou-se a existência de uma associação positiva entre a avaliação final da especialidade e a média de curso e a prova nacional de seriação.

Conclusão: Apesar de positivamente associado, no geral, com a média de curso e a prova nacional de seriação, o que indica a sua potencial validade, os resultados demonstram que a avaliação final de especialidade não apresenta uma capacidade discriminativa satisfatória. Deste modo, existe oportunidade para melhoria do modelo atual, nomeadamente através da alteração ao seu sistema de classificação e considerando outros modelos de exame.

Palavras-chave: Avaliação Educacional; Educação Médica Pós-Graduada; Escolas Médicas; Internato Médico; Modelos Educacionais

ABSTRACT

Introduction: There is a high heterogeneity in the structure of postgraduate medical training evaluation worldwide. However, in contrast to other countries, there have been no scientific studies of the final medical board examination, in Portugal. The present study aimed to evaluate the adequacy of the medical board examination including its validity as measured by its association with medical school grade average and national seriation examination.

Material and Methods: Cross-sectional, observational study. We analyzed the final results on the medical board examination of 2439 physicians, across 47 specialties, who completed their training in 2016 and 2017, using measures of central tendency and variability. We assessed the association between these grades and the national exam to initiate residency, and the grade average in Medical School.

Results: Measures of central tendency and variability, and consequent shape measures, revealed that the distribution of the scores of the final medical board exam is extremely negatively asymmetric and leptokurtic. A positive association was also found between the results in this exam and the score on national exam to initiate residency, and the grade average in Medical School.

Conclusion: Although the medical board examination was, in general, positively associated with scores on the national exam to initiate residency, and the mean final Medical School grades, thus indicating its potential validity, our results demonstrate that this exam

* Co-Primeiros Autores

1. Department of Psychosis Studies. Institute of Psychiatry, Psychology and Neuroscience. King's College. London. United Kingdom.
2. Massachusetts Eye and Ear. Harvard Medical School. Boston. United States.
3. Centro Hospitalar e Universitário de Coimbra. Coimbra. Portugal.
4. Faculdade de Medicina. Universidade de Coimbra. Coimbra. Portugal.
5. Centro de Investigação em Neuropsicologia e Intervenção Cognitivo-Comportamental. Faculdade de Psicologia e Ciências da Educação. Universidade de Coimbra. Coimbra. Portugal.
6. Instituto de Psicologia Médica. Faculdade de Medicina. Universidade de Coimbra. Coimbra. Portugal.
7. Department of Surgery and Cancer. Imperial College Medical School. London. United Kingdom.
8. Serviço de Oftalmologia. Hospital Pedro Hispano. Senhora da Hora. Portugal.
9. Unidade Local de Saúde de Matosinhos. Matosinhos. Portugal.
10. Serviço de Cirurgia Pediátrica. Centro Hospitalar e Universitário de Coimbra. Coimbra. Portugal.
11. Serviço de Oftalmologia. Centro Hospitalar de Lisboa Central. Lisboa. Portugal.
12. Institut Jules Bordet et L'Université Libre de Bruxelles (U.L.B.). Brussels. Belgium.
13. Instituto de Saúde Pública. Universidade do Porto. Porto. Portugal.
14. Serviço de Cardiologia. Centro Hospitalar de Lisboa Central. Lisboa. Portugal.
15. Serviço de Cardiologia. Centro Hospitalar e Universitário de Coimbra. Coimbra. Portugal.
16. Serviço de Gastroenterologia. Centro Hospitalar Vila Nova de Gaia/Espinho. Porto. Portugal.
17. Serviço de Otorrinolaringologia. Centro Hospitalar e Universitário de Coimbra. Coimbra. Portugal.
18. Coimbra Institute for Clinical and Biomedical Research (ICBR). Coimbra. Portugal.

✉ Autor correspondente: Tiago Reis Marques. tiago.marques@kcl.ac.uk

Recebido: 11 de abril de 2018 - Aceite: 07 de novembro de 2018 | Copyright © Ordem dos Médicos 2018



presents no satisfactory discriminative capacity. Therefore, there is room to improve the actual postgraduate medical examination model, including changes in its classification system and potentially consider other assessment models.

Keywords: Education, Medical, Graduate; Educational Measurement; Internship and Residency; Models, Educational; Schools, Medical

INTRODUÇÃO

A estrutura da formação médica pós-graduada e o seu formato de avaliação são muito heterogêneos a nível mundial. Em geral, nestes modelos formativos observa-se uma crescente utilização da educação baseada nas competências,¹ bem como da necessidade de avaliações frequentes do conhecimento, pois estas promovem uma melhor retenção da informação (*test enhanced learning*).² Em Portugal, a avaliação final do internato médico serve simultaneamente como prova de certificação de uma especialidade e secundariamente como medida de seriação tendo em vista a contratação dos recém-especialistas em concursos para vagas no serviço nacional de saúde. Desta forma, este exame tem um impacto significativo nas perspetivas profissionais dos médicos que exercem em Portugal.

Atualmente a avaliação final do internato médico é definida pelo decreto-lei nº 13/2018, regulamentada pela portaria 79 de 16 de março de 2018. Após a realização das avaliações de desempenho e de conhecimento positivas em cada estágio do internato, o médico interno é aceite para uma avaliação final de todo o processo formativo. Esta avaliação compreende três provas públicas e eliminatórias: discussão curricular, provas prática e teórica. A prova de discussão curricular consiste na apreciação e discussão do *curriculum vitae*. Habitualmente o colégio da Ordem dos Médicos da respetiva especialidade recomenda uma grelha para apreciação curricular do mesmo pelo júri das provas finais. A prova prática avalia a capacidade do médico interno para resolver problemas e atuar, assim como reagir em situações do âmbito da especialidade. Inclui habitualmente a observação de um doente, a elaboração de história clínica e sua discussão ou análise de casos, com elaboração de relatório e sua discussão. A prova teórica destina-se a avaliar o nível de conhecimentos do candidato. A prova é habitualmente oral, mas em algumas especialidades é substituída por uma prova escrita ou teste de escolha múltipla. Quando isto acontece, reveste-se de carácter nacional e é realizada por todos os candidatos em simultâneo. Por fim, a classificação final é a média aritmética das classificações obtidas nas provas curricular, prática e teórica, traduzida quantitativamente numa escala de 0 a 20 valores. Atualmente este exame decorre em duas épocas anuais, sendo que em cada época o exame é realizado por diferentes júris.

Idealmente entende-se que um exame deve não só avaliar conhecimentos, mas também a sua correta aplicação prática.³ De facto, o formato da avaliação final da especialidade tem sido amplamente debatido dentro da classe médica e tem sido empiricamente criticado pela sua subjetividade, pelas elevadas notas obtidas pela maioria dos candidatos e por fim pela quase ausência de reprovações. No entanto, até ao momento nunca foram publicados dados referentes à validade ou capacidade discriminativa deste

exame.

Na avaliação dos resultados de um teste de conhecimentos deve-se incluir a análise:

- das principais medidas de tendência central - média, moda e mediana- que fornecem informação sobre a distribuição dos dados, bem como o valor/pontuação mais representativo ou central. É defendido que um teste de avaliação de conhecimentos deve seguir uma distribuição normal (i.e. curva simétrica bilateralmente, dividida em dois, em que cada metade tem 50% dos dados; com a média, mediana e moda a coincidir com o centro da distribuição)⁴;
- das medidas de variabilidade - mínimo e máximo, variância, distribuição por quartis e desvio-padrão que permitem descrever a distribuição dos dados pelos valores possíveis. A variabilidade afere a utilidade de um teste, sendo que quanto maior a variabilidade melhor a capacidade de um teste distinguir entre sujeitos. A evidência para validade de um teste pode ser mensurada tendo em conta várias fontes, uma das quais consiste na exploração de padrões de convergência e divergência. Esta evidência de validade é obtida pela análise das associações entre os resultados de um teste e os resultados em testes que pretendem medir o mesmo construto ou construtos semelhantes.⁴

Internacionalmente, é comum os exames de avaliação da formação médica pós-graduada serem analisados de acordo com estas propriedades e os resultados são regularmente publicados e disponíveis para análise.⁵⁻¹¹ Estudos internacionais têm também analisado as associações entre as notas no final da especialidade e os exames de acesso aos colégios de especialidade.¹² Dentro de cada especialidade tem-se também analisado, quais os preditores de sucesso na admissão aos respetivos colégios¹³: incluindo as classificações obtidas em momentos de avaliação anteriores ao exame.¹⁴⁻¹⁶

No entanto, em Portugal, só um estudo relacionou diferentes momentos de avaliação ao longo da formação médica, mostrando associações positivas entre a média de curso e a nota na prova nacional de seriação.¹⁷ Assim, até ao momento não existem estudos que analisem a relação entre os resultados obtidos em diferentes momentos avaliativos da formação médica.

O presente estudo tem como objetivos avaliar a precisão/fiabilidade das classificações obtidas na avaliação final da especialidade, através da avaliação da distribuição das notas, assim como a sua validade, através da sua associação com outras medidas de desempenho disponíveis para consulta pública, nomeadamente com as classificações nas provas final de seriação e média final de curso de medicina.

MATERIAL E MÉTODOS

Estudo transversal e observacional.

Participantes

Foram identificados os médicos que terminaram a formação específica em 2016 e 2017 em Portugal e suas respetivas notas na avaliação final da especialidade, através das listas publicadas oficialmente pela Administração Central do Sistema de Saúde (ACSS). Estas notas são as finais, incluindo a média aritmética das provas que a compreendem. Considerando que estas notas só foram publicadas a partir de 2016, só foi possível incluir neste estudo os médicos que fizeram o exame final desde essa data. Os restantes dados, incluindo a média de curso e nota na Prova Nacional de Seriação (i.e. o exame de acesso à especialidade) estão igualmente disponíveis publicamente e foram obtidos após consulta de listagens oficiais publicadas no site da ACSS.

Assim, neste estudo foram considerados os participantes para os quais se dispunha de informação acerca das três variáveis necessárias para os objetivos definidos: nota da avaliação final da especialidade, média de curso e nota na prova nacional de seriação.

Procedimento e análise estatística

A análise estatística foi realizada com recurso ao software SPSS versão 24.0, considerando-se um valor de $p < 0,05$ como estatisticamente significativo.

Para estimar a precisão/fiabilidade das notas no exame da especialidade foram utilizadas: a) medidas de tendência central (média, mediana e moda); b) medidas de variabilidade (amplitude, desvio padrão, variância e divisão por quartis); e c) medidas de forma (curtose e assimetria). Para a interpretação dos valores de curtose e assimetria foram considerando os critérios de Field (2005)¹⁸ em que a distribuição dos dados segue uma distribuição normal quando os valores de z (calculados através dos valores de assimetria e curtose a dividir pelos respetivos erros padrão) se encontram entre -1,96 e 1,96. Foi também calculado o teste Kolmogorov–Smirnov que permite perceber se a distribuição dos dados se distingue de forma estatisticamente significativa de uma distribuição normal. Histogramas com a respetiva curva da normalidade são apresentados para a amostra total e por especialidade (Apêndice 1: https://www.actamedicaportuguesa.com/revista/index.php/amp/article/view/10646/Apendice_01.pdf). São ainda apresentados diagramas de extremos e quartis (*boxplot*) para ilustrar a variabilidade das notas por especialidade (Apêndice 2: https://www.actamedicaportuguesa.com/revista/index.php/amp/article/view/10646/Apendice_02.pdf).

Para verificar a validade das notas obtidas no exame final da especialidade recorreu-se ao estudo da correlação das mesmas com a média de curso e os resultados na prova nacional de seriação, através do coeficiente de correlação de Spearman (tendo em conta que as variáveis em análise não seguem a distribuição normal) com a amostra total e

por especialidade. Considerando a variabilidade no número de participantes por especialidade e de forma a garantir o rigor estatístico e a representatividade dos resultados, foram realizadas análises apenas para as especialidades para as quais estavam disponíveis dados sobre 10 ou mais indivíduos. Para a avaliação da magnitude das correlações foi utilizado o critério de Cohen (1988)¹⁹ que propõe que valores de 0,10 correspondem a um efeito pequeno, 0,30 a um efeito médio e 0,50 a um efeito grande. Desta forma, foram considerados os seguintes intervalos: de 0,10 (inclusive) a 0,30 (exclusive) corresponde a uma correlação fraca, de 0,30 (inclusive) a 0,50 (exclusive) corresponde a uma correlação moderada e acima de 0,50 (inclusive) corresponde a uma correlação forte. De forma a avaliar a validade do exame da especialidade, pretendeu-se encontrar correlações positivas moderadas a fortes entre as notas no mesmo e as notas em avaliações anteriores (média do final do curso e prova nacional de seriação). Gráficos de dispersão destas relações são apresentados como material complementar (amostra total).

RESULTADOS

Amostra

Neste estudo foram incluídos dados de 2439 médicos. A distribuição dos mesmos por especialidade pode ser consultada na Tabela 1.

Variabilidade de notas no exame final de especialidade

As notas no exame final de especialidade variaram entre 10,8 e 20, com uma mediana de 18,9 (Fig. 1). Observando a curva da distribuição das notas de todos os sujeitos independentemente da especialidade, verifica-se que a mesma é assimétrica com a maioria dos sujeitos a apresentarem notas elevadas, ou seja apresentando uma assimetria negativa, e leptocúrtica. No entanto, quando se observam as curvas de distribuição por especialidade (material suplementar) verifica-se que embora a maioria siga este padrão, há especialidades que apresentam uma curva de distribuição aproximadamente simétrica (como por exemplo as que representam a distribuição de notas referentes à Gastreenterologia, a Neurologia e a Saúde Pública).

A análise da distribuição e normalidade destes dados é apresentada na Tabela 2. As medidas de tendência central (média = 18,5; moda = 19,5 e mediana = 18,9) revelaram-se próximas do valor máximo (20). As medidas de variabilidade (desvio padrão = 1,3; variância = 1,6) mostraram a baixa variabilidade dos resultados. As medidas de forma e as análises de normalidade demonstraram que a distribuição das notas no exame final da especialidade não segue uma distribuição gaussiana normal ($p < 0,001$).

Para ilustrar a variação das notas no exame da especialidade por especialidade (apenas foram analisadas as especialidades com mais de dez sujeitos), na Fig. 2 podem ser observados os diagramas de extremos e quartis (*boxplot*). Verificam-se diferenças entre os diagramas de diferentes especialidades.

Tabela 1 – Número e percentagem de médicos considerados por especialidade (n = 2439)

	2016	2017	Total	Total
	n	n	n	%*
Anatomia Patológica	11	18	29	1,2
Anestesiologia	65	60	125	5,1
Cardiologia	25	23	48	2,0
Cardiologia Pediátrica	3	3	6	0,2
Cirurgia Geral	33	59	92	3,8
Cirurgia Maxilo-Facial	5	3	8	0,3
Cirurgia Pediátrica	4	2	6	0,2
Cirurgia Plástica, Estética e Reconstructiva	7	10	17	0,7
Cirurgia Torácica	7	7	14	0,6
Cirurgia Vasculiar / Angiologia	7	8	15	0,6
Dermatovenereologia	9	5	14	0,6
Endocrinologia	13	13	26	1,1
Estomatologia	6	5	11	0,5
Gastrenterologia	22	25	47	1,9
Genética Médica	2	3	5	0,2
Ginecologia / Obstetrícia	42	54	96	3,9
Hematologia Clínica	10	15	25	1,0
Imuno-Hemoterapia	5	12	17	0,7
Imunoalergologia	1	9	10	0,4
Infeciologia	10	14	24	1,0
Medicina Desportiva	0	2	2	0,1
Medicina do Trabalho	0	4	4	0,2
Medicina Física e Reabilitação	24	3	27	1,1
Medicina Geral e Familiar	356	362	718	29,4
Medicina Interna	147	156	303	12,4
Medicina Legal	5	6	11	0,5
Medicina Nuclear	2	3	5	0,2
Nefrologia	13	18	31	1,3
Neurocirurgia	6	8	14	0,6
Neurologia	21	12	33	1,4
Neurorradiologia	7	9	16	0,7
Oftalmologia	33	20	53	2,2
Oncologia Médica	30	30	60	2,5
Ortopedia	28	41	69	2,8
Otorrinolaringologia	18	26	44	1,8
Patologia Clínica	11	13	24	1,0
Pediatria Médica	71	45	116	4,8
Pedopsiquiatria / Psiquiatria da Infância e Adolescência	9	15	24	1,0
Pneumologia	16	16	32	1,3
Psiquiatria	45	36	81	3,3
Radiologia	24	25	49	2,0
Radioterapia	9	4	13	0,5
Reumatologia	10	11	21	0,9
Saúde Pública	10	19	29	1,2
Urologia	12	13	25	1,0
Total	1194	1245	2439	

* Esta percentagem diz respeito aos médicos considerados para este estudo (que fizeram exame da especialidade em 2016 e 2017). Não dispomos atualmente de dados nacionais fidedignos acerca da percentagem de médicos em cada uma das especialidades a nível nacional

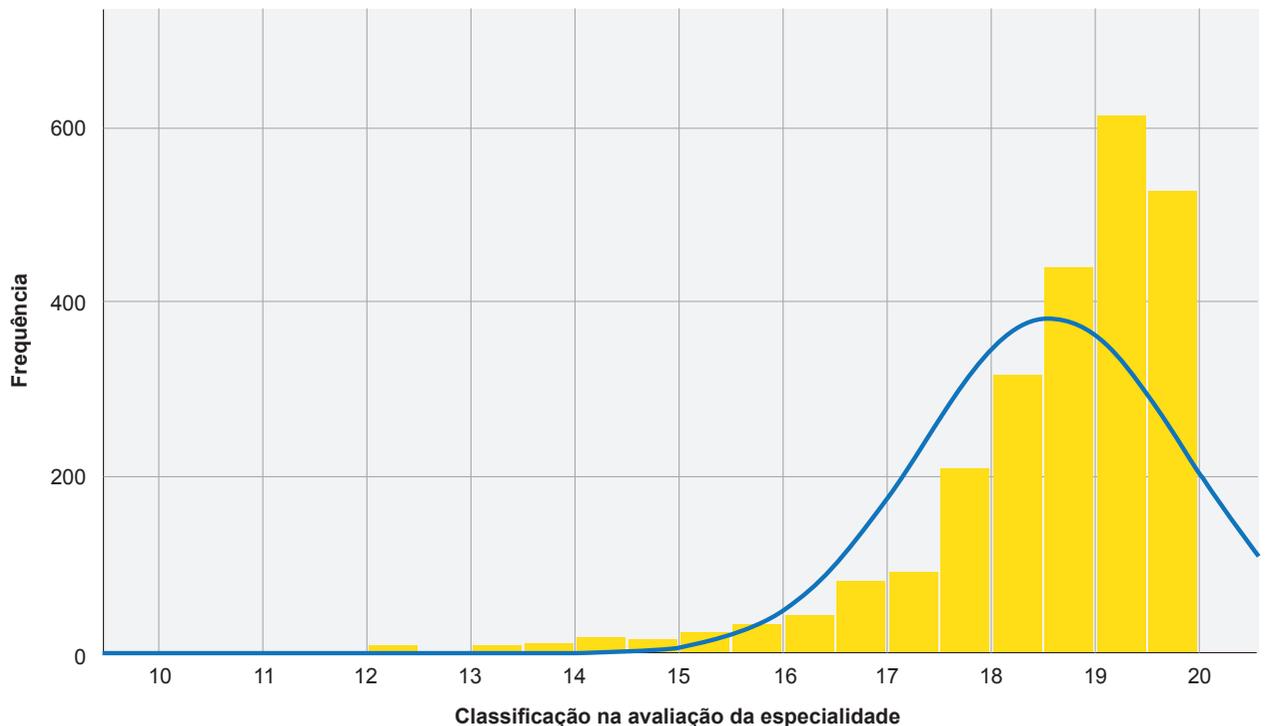


Figura 1 – Distribuição da classificação na avaliação da especialidade referente à amostra completa (n = 2439)

Relação entre a média final de curso, nota na PNS e nota no exame final da especialidade

Quando considerada a amostra completa, a avaliação da associação entre a média de final de curso e a nota do exame final da especialidade revelou a presença de uma correlação significativa, positiva e de magnitude moderada ($r_s = 0,42$, $p < 0,001$). De modo semelhante, verificou-se uma correlação significativa, positiva e de magnitude elevada entre a nota na prova nacional de seriação e a nota no exame final da especialidade ($r_s = 0,59$, $p < 0,001$). A mesma avaliação foi feita por especialidade, como apresentado na Tabela 3. Quando se analisam os resultados por especialidade verifica-se que existe uma grande variabilidade consoante a especialidade em análise. Encontramos especialidades como a Anatomia Patológica, a Cirurgia Vascular/Angiologia ou a Ginecologia/Obstetrícia que apresentam correlações significativas em ambas as relações estudadas. A Anestesiologia, Pediatria Médica e a Psiquiatria apresentaram correlações significativas apenas na relação entre o exame final da especialidade e a média

final de curso. Especialidades como a Cardiologia, Cirurgia Geral ou a Patologia Clínica só apresentam relações significativas entre o exame final da especialidade e a prova nacional de seriação. Por outro lado, os resultados mostram que há especialidades como a Cirurgia Plástica, Estética e Reconstructiva, a Cirurgia Torácica, Dermatovenereologia ou a Medicina Legal que não demonstraram atingir o nível de significância em nenhuma das associações estudadas. De referir que estas especialidades são as que têm amostras menores ($n < 30$, com exceção da Otorrinolaringologia e Pneumologia). Mesmo tendo em conta as especialidades que apresentaram associações significativas nas mesmas relações, a magnitude dessas correlações é muito variável. Por exemplo, na relação entre a média de final de curso e o exame final da especialidade há associações de magnitude fraca (ex.: Pediatria Médica), moderada (ex.: Medicina Interna) e forte (ex.: Cirurgia Vascular/Angiologia). O mesmo acontece na relação entre o exame final da especialidade e a PNS, com efeitos fracos (ex.: Ortopedia), moderados (ex.: Medicina Interna) e fortes (ex.: Imuno-Hemoterapia)

Tabela 2 – Medidas de tendência central, variabilidade, forma e análise da normalidade da distribuição da variável avaliação da especialidade

Medidas de tendência central			Medidas de variabilidade				Medidas de forma				Análise da normalidade
Média	Moda	Mediana	DP	Mínimo-Máximo	Variância	Divisão por quartis (25; 50; 75)	Assimetria (EP)	z	Curtose (EP)	z	K-S
18,52	19,50	18,90	1,27	10,8 - 20	1,61	18,10 18,90 19,40	-2,047 (0,05)	-40,94	5550 (0,10)	55,5	0,144***

DP: desvio-padrão; EP: erro padrão; K-S: Kolmogorov-Smirnov; *** $p < 0,001$

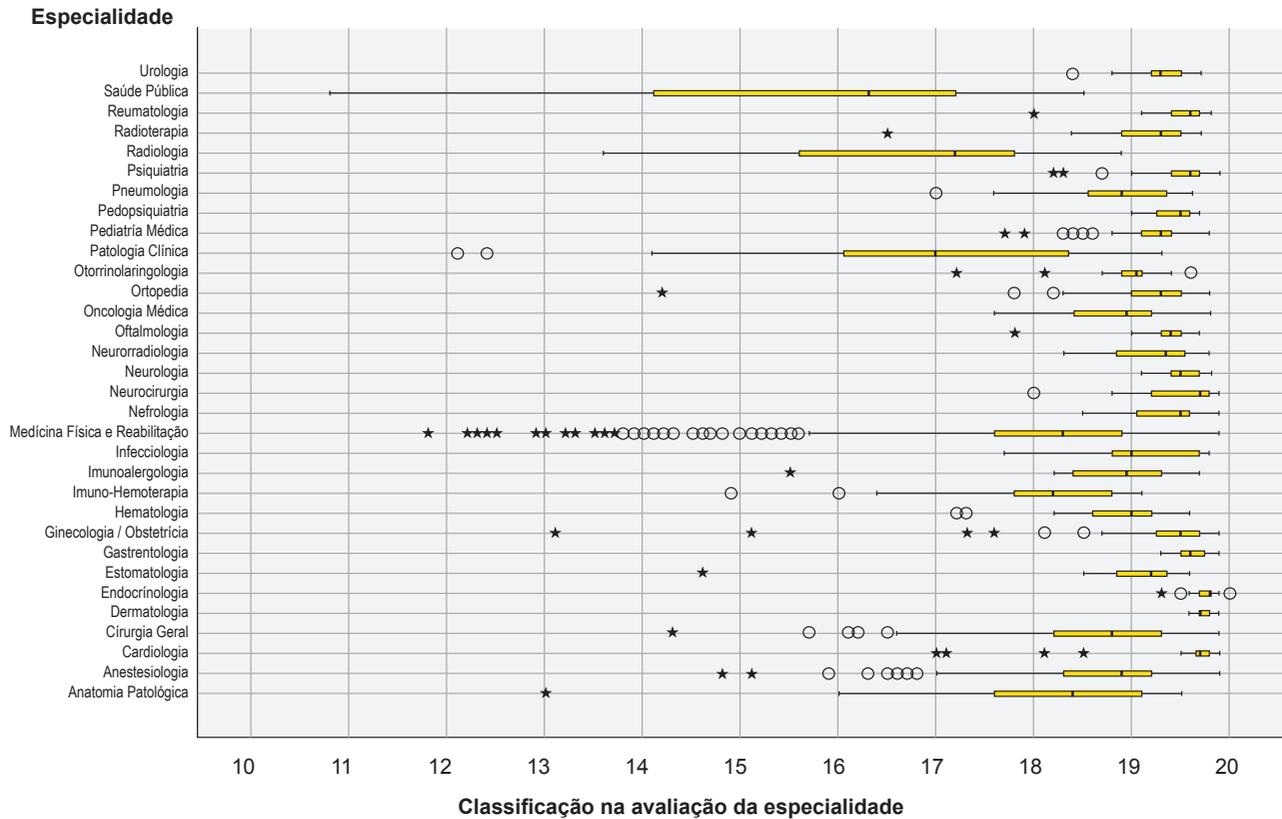


Figura 2 – Diagramas de extremos e quartis referentes à classificação na avaliação da especialidade, por especialidade

a serem detetados. Verifica-se que, nesta última relação, existem mais efeitos significativos e, de forma, geral as magnitudes encontradas são mais elevadas.

DISCUSSÃO

A formação médica pós-graduada é um pilar essencial da formação médica e um garante de qualidade. Uma correta avaliação dos conhecimentos é um elemento necessário e essencial deste processo, sendo para tal necessários exames e métodos de avaliação do mesmo, baseados em evidência científica. Testes com boas propriedades psicométricas revelam-se válidos tendo em conta vários fatores: conteúdo (representatividade dos itens do teste em relação ao domínio a ser testado), relações com outras variáveis (validade convergente, validade divergente, validade preditiva), entre outros.²⁰ A fidedignidade é também uma característica importante dos testes com boas qualidades psicométricas, sendo usado, por exemplo, o índice de coerência interna (que avalia se os vários itens de um teste estão a medir o mesmo constructo). No entanto, para a avaliação completa destas variáveis (validade e fidedignidade) são necessários múltiplos dados avaliativos ao longo da formação médica, bem como dados de outra natureza (por exemplo questionários com múltiplos itens). Este estudo avaliou a fiabilidade do exame final de especialidade através da análise da distribuição das pontuações obtidas nessa avaliação, bem como a validade do mesmo através da análise da associação dos resultados neste exame com outras medidas de desempenho (média de curso e a prova

nacional de seriação).

Como mencionado, uma das formas de caracterizar a precisão de um teste de avaliação é através das medidas de tendência central e variabilidade da distribuição das suas pontuações. Os nossos resultados demonstraram que as notas do exame final da especialidade não apresentam uma distribuição normal e as medidas de forma indicam uma assimetria negativa, ou seja, a maioria dos sujeitos obtêm notas elevadas. As medidas de tendência central (média, moda e mediada) encontram-se muito próximas do valor máximo possível (20 valores), com um desvio-padrão extremamente baixo, o que indica que as notas se encontram muito próximas da média (18,5).

Estes resultados têm importantes implicações para a forma como se interpretam as notas do exame final da especialidade e para possíveis alterações futuras deste. Considerando estes resultados, nomeadamente, a pobre capacidade discriminatória do atual exame, é necessário efetuar adaptações ao mesmo para que a maioria dos resultados esteja no centro da distribuição, nomeadamente o seu sistema de classificação pode ter de ser alterado ou outros modelos de exame devem ser ponderados. Tal torna-se particularmente importante considerando que a progressão dos médicos nas carreiras profissionais no serviço nacional de saúde (SNS) em Portugal está atualmente muito dependente das notas obtidas no mesmo. Estas são, na verdade, um dos principais parâmetros considerados nos concursos públicos para aqueles que pretendem ingressar o SNS como especialistas.²⁰

Tabela 3 – Médias, desvios-padrão, medianas, percentis (25 e 75) e avaliação da correlação entre a média de curso e a nota na PNS com a nota no exame final da especialidade, por especialidade

	n	M (DP)	Mediana	P25	P75	Média curso e exame final	PNS e exame final
						r_s	r_s
Anatomia Patológica	29	18,01 (1,37)	18,40	17,45	19,10	0,44*	0,69***
Anestesiologia	125	18,61 (0,95)	18,90	18,30	19,20	0,23*	0,05 ^{ns}
Cardiologia	48	19,56 (0,61)	19,70	19,63	19,80	0,14 ^{ns}	0,49***
Cirurgia Geral	92	18,68 (0,89)	18,85	18,20	19,30	0,14 ^{ns}	0,25*
Cirurgia Plástica, Estética e Reconstructiva	17	18,61 (0,67)	18,50	18,00	19,40	-0,11 ^{ns}	0,18 ^{ns}
Cirurgia Torácica	14	18,13 (1,05)	18,30	17,80	18,83	0,47 ^{ns}	0,33 ^{ns}
Cirurgia Vasculiar / Angiologia	15	19,17 (0,49)	19,30	19,00	19,50	0,77**	0,61*
Dermatovenereologia	14	19,74 (0,09)	10,70	19,70	19,80	-0,05 ^{ns}	0,05 ^{ns}
Endocrinologia	26	19,77 (0,15)	19,80	19,70	19,83	0,04 ^{ns}	0,23 ^{ns}
Estomatologia	11	18,75 (1,42)	19,20	18,50	19,50	0,45 ^{ns}	0,68*
Gastroenterologia	47	19,64 (0,15)	19,60	19,50	19,80	0,11 ^{ns}	0,32*
Ginecologia / Obstetrícia	96	19,30 (0,89)	19,50	19,23	19,70	0,30**	0,33**
Hematologia Clínica	25	18,84 (0,61)	19,00	18,50	19,30	0,17 ^{ns}	0,40*
Imuno-Hemoterapia	17	17,94 (1,16)	18,20	17,60	18,80	0,69**	0,78***
Imunoalergologia	10	18,63 (1,20)	18,95	18,35	19,33	-0,07 ^{ns}	-0,69*
Infeciologia	24	19,07 (0,63)	19,00	18,80	19,70	0,39 ^{ns}	0,27 ^{ns}
Medicina Física e Reabilitação	27	18,94 (0,58)	19,10	18,60	19,30	0,52**	0,49*
Medicina Geral e Familiar	718	17,83 (1,28)	18,20	17,50	18,60	0,48***	0,54***
Medicina Interna	303	18,49 (1,13)	18,80	18,00	19,30	0,35***	0,45***
Medicina Legal	11	17,25 (1,29)	17,80	16,50	18,10	-0,16 ^{ns}	0,33 ^{ns}
Nefrologia	31	19,32 (0,44)	19,50	19,00	19,60	0,56**	0,46**
Neurocirurgia	14	19,45 (0,56)	19,70	19,10	19,83	0,39 ^{ns}	0,59*
Neurologia	33	19,50 (0,19)	19,50	19,40	19,70	0,45**	0,51**
Neurorradiologia	16	19,23 (0,48)	19,35	19,78	19,58	0,24 ^{ns}	0,01 ^{ns}
Oftalmologia	53	19,33 (0,34)	19,40	19,30	19,50	0,40**	0,31*
Oncologia Médica	60	18,84 (0,57)	18,95	18,40	19,20	0,25 ^{ns}	0,50***
Ortopedia	69	19,10 (0,73)	19,30	19,00	19,50	0,08 ^{ns}	0,27*
Otorrinolaringologia	44	18,97 (0,38)	19,05	18,90	19,10	0,09 ^{ns}	-0,08 ^{ns}
Patologia Clínica	24	16,86 (1,97)	17,00	16,03	18,38	0,32 ^{ns}	0,52**
Pediatria Médica	116	19,22 (0,35)	19,30	19,10	19,40	0,22*	0,17 ^{ns}
Pedopsiquiatria / Psiquiatria da Infância e Adolescência	24	19,41 (0,23)	19,50	19,23	19,60	0,16 ^{ns}	0,24 ^{ns}
Pneumologia	32	18,86 (0,61)	18,90	18,53	19,38	0,21 ^{ns}	0,31 ^{ns}
Psiquiatria	81	19,49 (0,35)	19,60	19,40	19,70	0,33**	0,09 ^{ns}
Radiologia	49	16,79 (1,46)	17,20	15,45	17,85	0,26 ^{ns}	0,10 ^{ns}
Radioterapia	13	19,00 (0,84)	19,30	18,85	19,50	0,66*	0,39 ^{ns}
Reumatologia	21	19,49 (0,39)	19,60	19,40	19,70	0,36 ^{ns}	0,15 ^{ns}
Saúde Pública	29	15,56 (2,13)	16,30	14,05	17,30	0,32 ^{ns}	0,20 ^{ns}
Urologia	25	19,29 (0,28)	19,30	19,20	19,50	0,23 ^{ns}	0,32 ^{ns}

n: número; M: média; DP: desvio-padrão; r_s : coeficiente de correlação de Spearman; * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$; ^{ns}: valor não significativo; valores de $p < 0,05$ são apresentados a negrito

As boas práticas existentes em outros países refletem-se na existência de estruturas que orientam e certificam os programas de internato, incluindo os métodos de avaliação (*Accreditation Council for Graduate Medical Education* nos Estados Unidos da América; *General Medical Council* no Reino Unido). Estes métodos têm sido amplamente discutidos na literatura.²¹⁻²³ De acordo com estas práticas, várias são as sugestões passíveis de melhorar a avaliação da formação pós-graduada em Portugal. Os testes de resposta selecionada (ex.: questões de escolha-múltipla) têm sido apontados como uma forma eficaz, versátil e direta de avaliação da obtenção de conhecimentos devido às suas vantagens em termos de objetividade de cotação e propriedades psicométricas (ex.: validade, coerência interna).²⁴ No entanto, podem apresentar importantes limitações na avaliação de conhecimentos práticos. Desta forma poderá ser necessário considerar também a utilização de exames com casos clínicos estandardizados, comuns para todos os candidatos da referida especialidade.²⁵ Uma eventual mudança na abordagem atual de aquisição de competências baseadas em tempo, para uma abordagem futura baseada na aquisição de competências poderá eventualmente ajudar a uma melhoria na curva de aprendizagem geral e individual de cada médico especialista.²⁶ Isto pode aumentar quer a qualidade média do internato médico, quer um aumento na capacidade formativa global para especialidades mais carenciadas. Para além disso, outras opções incluem a criação de uma estrutura de avaliação que supervisione a avaliação médica pós-graduada; a publicação de documentação de apoio e material formativo; a reformulação do método de avaliação e por fim a criação de uma rede de avaliadores, sujeitos a formação contínua e com competências na criação de exames. Por outro lado, uma avaliação final de especialidade pode compreender uma série de competências que vão muito para além do conhecimento académico. Como tal, a avaliação prática que se realiza, de modo não sistemático, sem uma avaliação por lista seriada, pode não avaliar suficientemente a qualidade de um especialista.²⁷ Cada especialidade deve propor as competências mínimas que se pretende sejam apreendidas por um recém-especialista. De acordo com essas características, vários modelos de avaliação podem ser diferenciadamente propostos e a sua efetividade avaliada incluindo doentes estandardizados e listas de verificação.

Relativamente à validade, na medicina esta tem sido avaliada através da comparação com outros exames e/ou com resultados da prática clínica.²⁸ No entanto, do nosso conhecimento, não há estudos em Portugal que mostrem a associação entre vários momentos de avaliação ao longo do percurso médico académico e profissional. Neste estudo, foi efetuada uma análise da associação entre os resultados no exame do final da especialidade e os resultados obtidos anteriormente (média de final de curso e prova nacional de seriação). Esta revelou que, considerando a amostra total, as associações são positivas e de magnitude moderada, indicando assim, preliminarmente, uma validade adequada. No entanto, analisando os resultados por especialidade

verificamos que estes variam muito entre especialidades, o que tem sido encontrado, também, em estudos internacionais.²⁹ Tendo em conta a análise já realizada em termos da variabilidade das notas, a ausência de correlações significativas ou a existência de correlações baixas em algumas especialidades, poderá dever-se ao reduzido espectro de variabilidade desta classificação, não traduzindo por isso necessariamente uma ausência efetiva de relação. Devido à natureza do atual exame do final da especialidade e à escassez de dados publicados não nos foi possível neste estudo avaliar a sua fidedignidade e, de forma completa, a sua validade. Do mesmo modo, não podemos comentar a capacidade deste exame avaliar o conhecimento e a aquisição de competências ao longo do internato. No entanto, a baixa taxa de reprovação é surpreendente quando comparada com outros países, onde as taxas de aprovação variam entre os 60% e os 90%. Embora não possamos tirar conclusões definitivas em relação à validade da prova em análise, os resultados do estudo correlacional enfatizam a potencial validade da prova. Apesar de a variabilidade entre especialidades ser grande, sujeitos com notas mais elevadas na média final de curso e na prova nacional de seriação parecem ter também notas mais elevadas no exame final da especialidade.

O presente estudo apresenta algumas limitações que devem ser consideradas. Em primeiro lugar, apenas foram analisadas as notas no exame final da especialidade obtidas nos dois anos imediatamente anteriores à realização do presente trabalho. Desta forma, a generalização destes resultados pode ser questionável. No entanto, não existem dados que sugiram a ocorrência de alterações consideráveis nas classificações no exame final da especialidade durante o período de análise referido. Assim, os resultados globais esperam-se sobreponíveis aos obtidos em períodos anteriores. Em segundo lugar, a natureza transversal deste estudo não permite inferir causalidade a partir das associações encontradas. Desta forma, estas associações poderão ser devidas a covariáveis não consideradas no âmbito do presente estudo. Em terceiro lugar, devido ao tamanho limitado da amostra disponível, não foi possível efetuar uma análise dos dados por local de realização do exame do final da especialidade. Deverá ser também tida em conta a eventual variabilidade introduzida pelo facto de, mesmo dentro de uma mesma especialidade, na mesma época, o exame ser efetuado por diferentes júris examinadores. A influência da variabilidade inter-examinador em medicina é conhecida e estudada tendo em conta diferentes métodos de avaliação.³⁰ Em quarto lugar, o facto de o exame final de especialidade não ser exatamente igual para todas as especialidades pode incorrer num viés adicional. Por outro lado, o objetivo do presente estudo focou-se na avaliação das notas do exame de especialidade, ilustrando a sua distribuição (no geral e por especialidade) e não na comparação das notas nas diferentes especialidades entre si. Embora a comparação entre especialidades fosse importante, consideramos que estudos futuros com um maior número de examinandos poderão debruçar-se sobre

este assunto. Em quinto lugar, teremos sempre de salientar as limitações inerentes ao tipo de avaliação que é realizada. Sabe-se que exames com estas características não medem competências humanas, sociais e por vezes até técnicas dos candidatos a especialistas.³¹ Em sexto lugar, o facto de se utilizarem as classificações na prova final de seriação e média final de curso de medicina para estudar a validade da classificação no exame final de especialidade constitui-se como uma limitação pois estes medem constructos que podem estar relacionados mas são em última análise diferentes (e são medidos em pontos temporais diferentes). Finalmente, é importante referir que o facto de não ser possível o acesso pelos autores às notas obtidas em cada uma das três provas realizadas na avaliação da especialidade médica (usando apenas o resultado médio final) constitui-se em si mesmo como uma limitação do presente trabalho. Estudos futuros poderão explorar a fidedignidade e validade de cada um dos componentes de avaliação tendo assim uma avaliação mais completa do exame da especialidade uma vez que estes componentes avaliam aspetos relacionados, mas diferentes.

No futuro, será também interessante analisar a validade, fidedignidade, custo-efetividade e aceitabilidade do exame final da especialidade, caso seja possível obter elementos avaliativos múltiplos ao longo da formação pós-graduada. A sugestão frequentemente levantada sobre a falta de validade e discriminação relativamente à prova final levanta-se igualmente durante a formação pré-graduada, onde a diversidade é grande quanto aos métodos de avaliação e mesmo às unidades curriculares lecionadas. Neste momento, tanto quanto os autores conhecem, apenas a Faculdade de Medicina da Universidade do Porto, a Universidade do Minho e a Universidade da Beira Interior procedem à análise sistemática da qualidade das avaliações internas. A mudança que se prevê a breve prazo no modelo de exame de acesso à especialidade poderá eventualmente encontrar eco nos modelos de avaliação internos de cada uma das escolas médicas em Portugal, assim como no modelo de avaliação final de especialidade.

REFERÊNCIAS

- Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. *Med Teach*. 2010;27;32:676–82.
- Larsen DP, Butler AC, Roediger HL. Test-enhanced learning in medical education. *Med Educ*. 2008;42:959–66.
- Lineberry M, Matthew Ritter E. Psychometric properties of the Fundamentals of Endoscopic Surgery (FES) skills examination. *Surg Endosc*. 2017;10;31:5219–27.
- Urbina S. *Essentials of psychological testing*. Vol. 53. New Jersey: John Wiley & Sons, Inc; 2014.
- Unwin E, Potts HW, Dacre J, Elder A, Woolf K. Passing MRCP (UK) PACES: a cross-sectional study examining the performance of doctors by sex and country. *BMC Med Educ*. 2018;6;18:70.
- Farooq S. High failure rate in postgraduate medical examinations - sign of a widespread disease? *J Pak Med Assoc*. 2005;55:214–7.
- Bowhay AR, Watmough SD. An evaluation of the performance in the UK Royal College of Anaesthetists primary examination by UK medical school and gender. *BMC Med Educ*. 2009;29;9:38.
- Dewhurst NG, McManus C, Mollon J, Dacre JE, Vale AJ. Performance in the MRCP(UK) Examination 2003–4: analysis of pass rates of UK

CONCLUSÃO

Os resultados do presente estudo parecem indicar que o atual exame final da especialidade em Portugal poderá não possuir uma capacidade discriminativa satisfatória, tendo em conta a baixa variabilidade das notas. No entanto, quer de um modo geral, quer em algumas especialidades, observou-se uma associação significativa entre a nota do exame final da especialidade e a média de curso, por um lado, tal como com a prova nacional de seriação, por outro. Estes resultados parecem indicar a potencial validade do exame da especialidade quando consideramos a sua relação com estas duas avaliações de desempenho prévias.

Os resultados deste estudo sugerem que o atual exame final da especialidade necessita de ser repensado, seja através da sua reestruturação e/ou da implementação de métodos de avaliação mais objetivos, que permitam uma melhor avaliação de conhecimentos e seriação de candidatos.

PROTECÇÃO DE PESSOAS E ANIMAIS

Os autores declaram que os procedimentos seguidos estavam de acordo com os regulamentos estabelecidos pelos responsáveis da Comissão de Investigação Clínica e Ética e de acordo com a Declaração de Helsínquia da Associação Médica Mundial.

CONFIDENCIALIDADE DOS DADOS

Os autores declaram ter seguido os protocolos do seu centro de trabalho acerca da publicação de dados.

CONFLITOS DE INTERESSE

Os autores declaram não terem qualquer conflito de interesse relativamente ao presente artigo.

FONTES DE FINANCIAMENTO

Os autores declaram não ter recebido subsídios ou bolsas para a elaboração do artigo.

- graduates in relation to self-declared ethnicity and gender. *BMC Med*. 2007;3;5:8.
- Rushd S, Landau AB, Lindow SW. An evaluation of the first time performance of international medical graduates in the MRCOG Part 1 and Part 2 written examinations. *Eur J Obstet Gynecol Reprod Biol*. 2013;166:124–6.
- Membership of the Royal Colleges of Physicians of the United Kingdom. Pass rates for MRCP(UK) Diploma and Specialty Certificate Examinations. [consultado 2018 jun 04]. Disponível em: <https://www.mrcpuk.org/mrcpuk-examinations/results/exam-pass-rates>.
- American Board of Internal Medicine. Residency Program Pass Rates 2015 – 2017. [consultado 2018 jun 04]. Disponível em: http://www.abim.org/~media/ABIM_Public/Files/pdf/statistics-data/residency-program-pass-rates.pdf.
- Baker K, Sun H, Harman A, Poon KT, Rathmell JP. Clinical performance scores are independently associated with the American board of anesthesiology certification examination scores. *Anesth Analg*. 2016;122:1992–9.
- Muensterer OJ, Bronstein ME, Mackenzie R, Snyder CW, Carachi R. Factors associated with passing the European Board of Paediatric

- Surgery Exam. *Pediatr Surg Int.* 2015;31:671–6.
14. Puscas L. Otolaryngology resident in-service examination scores predict passage of the written board examination. *Otolaryngol - Head Neck Surg.* 2012;147:256–60.
 15. McClintock JC, Gravlee GP. Predicting success on the Certification Examinations of the American Board of Anesthesiology. *Anesthesiology.* 2010;J112:212–9.
 16. Monteiro K, George P, Dollase R, Dumenco L. Predicting United States Medical Licensure Examination Step 2 clinical knowledge scores from previous academic indicators. *Adv Med Educ Pract.* 2017;8:385–91.
 17. Martins IP. Admissão ao internato complementar em Portugal: análise dos resultados do exame nacional de seriação entre 2006 e 2011. *Acta Med Port.* 2013;26:569–77.
 18. Field A. *Discovering statistics using SPSS.* London: Sage; 2005.
 19. Cohen J. *Statistical power analysis for the behavioral sciences.* New Jersey: Lawrence Erlbaum; 1988.
 20. Diário da República. Aviso n.º 1347/2017. [consultado 2018 jun 04]. Disponível em: http://www.acss.min-saude.pt/wp-content/uploads/2017/02/Aviso_1347_2017.pdf.
 21. Durning SJ, Hemmer P, Pangaro LN. The structure of program evaluation: an approach for evaluating a course, clerkship, or components of a residency or fellowship training program. *Teach Learn Med.* 2007;19:308–18.
 22. Musick D. A conceptual model for program evaluation in graduate medical education. *Acad Med.* 2006;81:1051–6.
 23. Weggemans MM, van Dijk B, van Dooijeweert B, Veenendaal AG, Ten Cate O. The postgraduate medical education pathway: an international comparison. *GMS J Med Educ.* 2017;34:Doc63.
 24. Downing SM. Validity: on the meaningful interpretation of assessment data. *Med Educ.* 2003;37:830–7.
 25. Downing SM. Assessment of knowledge with written test forms. In: Norman GR, van der Vleuten CP, Newble DI, Dolmans DH, Mann KV, Rothman A, et al, editors. *International handbook of research in medical education.* Dordrecht: Springer Netherlands; 2002. p. 647–72.
 26. Harden RM. Revisiting 'assessment of clinical competence using an objective structured clinical examination (OSCE).' *Med Educ.* 2016;50:376–9.
 27. Pusic MV, Boutis K, Hatala R, Cook DA. Learning curves in health professions education. *Acad Med.* 2015;90:1034–42.
 28. Taveira-Gomes I, Mota-Cardoso R, Figueiredo-Braga M. Communication skills in medical students – an exploratory study before and after clerkships. *Porto Biomedical Journal.* 2016;1:173–80.
 29. Stempień KL. Predictive validity of the examination for the Membership of the Royal Colleges of Physicians of the United Kingdom. London: University College London; 2014.
 30. Sharaf AA, AbdelAziz AM, El Meligy OA. Intra- and inter-examiner variability in evaluating preclinical pediatric dentistry operative procedures. *J Dent Educ.* 2007;71:540–4.
 31. Patterson F, Knight A, Dowell J, Nicholson S, Cousins F, Cleland J. How effective are selection methods in medical education? A systematic review. *Med Educ.* 2016;50:36–60.