

07 June 2018

RE: [Acta Med Port] Decisão editorial

Avaliação Psicométrica do exame final de especialidade em Portugal
/ Psychometric evaluation of the medical board exam in Portugal

Caro Editor da *Acta Médica Portuguesa*,

Agradecemos a possibilidade de rever e submeter o manuscrito “Avaliação Psicométrica do exame final de especialidade em Portugal” e vimos por este meio enviar a versão revista do mesmo. Agradecemos também todas as revisões e acreditamos que o manuscrito beneficiou significativamente com as alterações sugeridas.

Seguem abaixo os comentários dos revisores bem como a forma como os mesmos foram abordados (em itálico).

Notas do editor

1. Com o objectivo de otimizar a legibilidade do seu artigo e assim incrementar potencialmente as citações do mesmo, recomendamos que o conteúdo redigido em inglês sejam revistos por um "native speaker", tradutor qualificado ou empresa especializada em serviços de "language polishing";

Resposta: *Agradecemos desde já a sugestão apresentada e concordamos que é fundamental este trabalho de forma a que quer a qualidade quer futuras citações sejam aumentadas. O primeiro autor do artigo é investigador do Imperial College em Londres onde reside desde 2006, sendo a sua capacidade de escrita científica equivalente à de um native speaker. Também a co-autora Inês Lains é investigadora da Universidade de Harvard desde 2014 e apresenta também capacidades de escrita científica equiparadas às de um native speaker. Assim, é*

nossa opinião que o inglês apresentado no abstract é agora semelhante ao de qualquer outro artigo científico de uma revista internacional de língua inglesa.

2. O resumo e o abstract deverão reflectir a estrutura do artigo, com parágrafos independentes para cada um dos capítulos: Introdução, Material e Métodos, Resultados, Discussão, Conclusão;

Resposta: *Agradecemos a sugestão e pedimos as nossas desculpas por não o termos efetuado anteriormente. Procedemos à alteração do resumo e do abstract para conter os parágrafos mencionados. Retirámos desta forma o parágrafo 'objetivos'.*

3. Ao longo do corpo do manuscrito, as referências deverão ser identificadas por numerais em expoente/sobrescrito (p. ex "Por definição, a psicometria permite avaliar as propriedades de um exame, sendo que idealmente se entende que um exame deve não só avaliar os conhecimentos que se pretendem apreendidos, mas também a sua correta aplicação prática.¹ (em expoente)" em vez de "Por definição, a psicometria permite avaliar as propriedades de um exame, sendo que idealmente se entende que um exame deve não só avaliar os conhecimentos que se pretendem apreendidos, mas também a sua correta aplicação prática (Lineberry & Matthew Ritter, 2017)."

Resposta: *Agradecemos o alerta em relação às referências bibliográficas que, devido a um erro no gestor de referências bibliográficas, não se encontravam em conformidade com as Normas de Publicação. Alterámos a forma de apresentar as referências bibliográficas quer no texto quer na lista de referências de forma a ficar de acordo com as Normas de Publicação.*

4. As tabelas e as figuras (e bem assim os respectivos títulos e notas) deverão ser inseridas no final do ficheiro, após as referências. Ao longo do manuscrito, nos locais aproximados onde desejem que as mesmas sejam inseridas na fase de paginação, deverá constar apenas "Tabela 1" ou "Fig. 1";

Resposta: *Colocámos as tabelas e figuras no final do documento como solicitado, assinalando ao longo do documento o local de inserção das tabelas e figuras.*

5. A listagem de referências final necessita de uma remodelação completa, encontrando-se em inconformidade com as Normas de Publicação (nomeadamente, devem ser indicadas pela ordem cronológica em que surgem no corpo do manuscrito e formatadas conforme detalhado nas Normas).

Resposta: *Agradecemos o alerta em relação às referências bibliográficas que, devido a um erro no gestor de referências bibliográficas, não se encontravam em conformidade com as Normas de Publicação. Alterámos a forma de apresentar as referências bibliográficas quer no texto quer na lista de referências de forma a ficar de acordo com as Normas de Publicação.*

Revisor A

1. Nesse sentido, como facilmente se compreenderá pelo que segue, julgo mesmo que seria de considerar a humildade de remover do TÍTULO o termo psicométrica, uma vez que os requisitos pressupostos por uma análise psicométrica, se considerados em parte na presente “Avaliação do exame final de especialidade em Portugal”, em boa verdade só o estão de um modo algo parcimonioso.

Resposta: *Aceitamos o comentário do revisor e lamentamos não termos todos os dados necessários para efetuar uma completa análise psicométrica das notas no exame. De acordo com o sugerido, e de forma a não induzirmos em erro os leitores deste artigo, alterámos o título do manuscrito para “Avaliação do exame final de especialidade em Portugal”.*

2. Quando se fazem afirmações como “Em Portugal o formato da avaliação final (...) não é suficientemente válida ou discriminativa.”, necessário se torna fazer referência ao autor citado.

Resposta: *Agradecemos o comentário e tendo em conta a inexistência de dados publicados em relação à validade ou discriminação do modelo de exame final actualmente em vigor retirámos a frase supracitada. Acrescentámos a frase:*

(Página 6)

“No entanto, em Portugal até ao momento nunca foram publicados dados referentes à validade ou capacidade discriminativa do modelo de exame final actualmente em vigor.”

3. Veja-se por exemplo o caso da flagrante ausência de referenciação quando se afirma, na INTRODUÇÃO (pág. 5, linha 20): “Noutros países é comum os exames de avaliação (...) os resultados são regularmente publicados e disponíveis para análise.

Resposta: *Agradecemos a chamada de atenção para a falta de referência em algumas afirmações (que corrigimos e vamos fazendo referência ao longo da resposta ao revisor – ver pontos seguintes). De facto, o acrescento de referências nas afirmações apontadas*

pele revisor beneficiou esta versão revista do artigo. Na frase citada acrescentámos as seguintes referências, que foram também adicionadas à lista de referências:

5. Unwin E, Potts HWW, 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(1):70.
6. Farooq S. High Failure rate in Postgraduate Medical Examinations - Sign of a widespread Disease? *J Pak Med Assoc.* 2005;55(5):214–7.
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(1):38.
8. Dewhurst NG, McManus C, Mollon J, Dacre JE, Vale AJ. Performance in the MRCP(UK) Examination 2003–4: analysis of pass rates of UK graduates in relation to self-declared ethnicity and gender. *BMC Med.* 2007;3;5(1):8.
9. 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(2):124–6.
10. Membership of the Royal Colleges of Physicians of the united Kingdom. Pass rates for MRCP(UK) Diploma and Specialty Certificate Examinations. 2018 [cited 2018 Jun 4]. Available from: <https://www.mrcpuk.org/mrcpuk-examinations/results/exam-pass-rates>
11. American Board of Internal Medicine. Residency Program Pass Rates 2015 – 2017. 2017. Available from: http://www.abim.org/~media/ABIM_Public/Files/pdf/statistics-data/residency-program-pass-rates.pdf

4. Nos OBJECTIVOS, em coerência com o título, também “as propriedades psicométricas” deviam cair, em favor da reformulação da hipótese alternativa; ficando algo como: “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...”, assim consideradas de algum modo enquanto medidas externas independentes; embora o não sejam de facto. Por outro lado, sendo as estatísticas descritivas um recurso para sumariar os dados, quando aqui consideradas como um método de análise, julgo que seria adequado afirmar ser essa análise baseada em tais estatísticas. Será desde logo o caso em MATERIAIS E MÉTODOS do RESUMO/ABSTRACT: “...foram obtidas e analisadas com recurso a medidas de tendência central e variabilidade, as notas...”. E ainda: “tendo em vista a sua validação cruzada, foram também avaliadas as correlações com a média...”. Adiante, nos RESULTADOS: “Das medidas de tendência central e variabilidade, e consequentes medidas de formato, resulta que a distribuição das pontuações do exame final de especialidade é de reduzida amplitude e apresenta-se com uma forma manifestamente assimétrica.” De notar que aqui seria adequado incluir igualmente uma consideração sobre a curtose. No parágrafo que segue (pág. 2), da CONCLUSÃO, “exame final” surge repetido de modo redundante, podendo ser substituído na segunda frase por:

“...torna-se necessário reformulá-lo, redesenhando o mesmo; nomeadamente através...”

Resposta: *O resumo foi reformulado tendo em conta a revisão proposta por este revisor e as sugestões relativas à revisão da versão de língua inglesa. Nesta revisão do resumo considerámos também todas as revisões sugeridas e incluídas ao longo do texto. Incluímos também a informação solicitada sobre a análise da curtose. Transcrevemos abaixo a versão final do resumo que contempla todas estas alterações:*

“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 dados científicos que tenham avaliado o modelo de avaliação do exame 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 no exame 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 é de reduzida amplitude e apresenta-se com uma forma manifestamente assimétrica e leptocúrtica. No geral, verificou-se a existência de uma associação positiva entre a nota no exame final da especialidade e a média de curso e a prova nacional de seriação.

Conclusão: Os resultados demonstram que o exame 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.”

5. No segundo parágrafo da INTRODUÇÃO onde se lê “discussão curricular, prática e teórica” (pág. 4, linha 18), deve estar “discussão curricular, provas prática e teórica”. Depois, a concluir esse mesmo parágrafo (pág. 4) seria de realçar (acrescentando) que: “Desta ausência de critérios uniformes e sistemas de cotação pouco claros, resulta uma avaliação aparentemente impregnada de subjectividade.” Porque daqui resulta a hipótese de trabalho (objectivo). E na sequência, o parágrafo seguinte (a iniciar a pág. 5) deve iniciar agora a locução por meio de: “De facto, o formato do exame final...”.

Resposta: *Foram incluídas todas as alterações sugeridas na secção Introdução (Página 5 e início da página 6)*

“Desta ausência de critérios uniformes e sistemas de cotação pouco claros, resulta uma avaliação aparentemente impregnada de subjetividade. De facto, o formato do exame final da especialidade [...]”

6. E no parágrafo final da INTRODUÇÃO (pág. 6), devem agora os OBJECTIVOS vir reformulados do seguinte modo: “avaliar até onde possível a) a precisão/fiabilidade das notas da avaliação obtidas através do exame final da especialidade nos seus moldes actuais, b) bem assim como a sua validade.” Remetendo a referência às “relações entre... a nota no exame da especialidade” para a respectiva secção dos MÉTODOS.

Resposta: *Os autores incluíram todas as sugestões acima descritas na versão revista do manuscrito.*

(Página 6)

“O presente estudo tem como objetivos avaliar, dentro dos parâmetros adequados e apesar de não existirem dados que permitam uma análise psicométrica do atual exame, a precisão/fiabilidade das notas da avaliação obtidas através do exame final da especialidade nos seus moldes atuais, assim como a sua validade.”

7. Das observações anteriores resultam de algum modo as que se repetem agora ao longo das secções de Métodos, Resultados, Discussão e Conclusões. Nos MÉTODOS (pág. 7) seria assim de substituir “a análise da variabilidade das notas no exame...” (linha 18) por “estimar a precisão/fiabilidade das notas no exame...”; e de igual modo o “valor mínimo e máximo” (linha 20) por “amplitude”. E adiante (linha 29) onde se lê “Para o estudo das associações entre a nota no exame...”, será de considerar antes “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...”.

Resposta: *Foram acrescentadas todas as alterações sugeridas.*

(Página 7)

“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).”

(Página 8)

“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”.

8. Nos RESULTADOS, onde na leitura da Figura 1/Quadro 2, se referem as medidas de forma (pág.11, linhas 2-3), deveriam ser mencionadas a assimetria e a curtose da curva obtida.

Resposta: *Foram acrescentadas as informações sugeridas acerca da assimetria e da curtose da curva obtida:*

(Página 9)

“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”.

9. O início da DISCUSSÃO (pág. 14) deve pautar-se pelo que anteriormente ficou dito, sendo reformulado de acordo com os OBJECTIVOS, da seguinte forma: “Pretendeu-se com este estudo avaliar a fiabilidade do exame final de especialidade através da análise da distribuição das pontuações obtidas nessa avaliação.” Adiante (linha 7), onde se refere “a existência de uma associação...”, deve ela ser caracterizada; seja o caso, como “associação positiva”. Independentemente do uso corrente que se lhe reconhece, também será de evitar o recurso a anglicismos como “consistência interna” (linha 18), usando preferivelmente, sempre que se possa, o vernáculo; neste caso “coerência interna”. No parágrafo seguinte (linha 23) onde se lê “avaliar as propriedades de um exame... variabilidade.”, será de considerar em seu lugar “caracterizar a precisão de um teste de avaliação é através das medidas de tendência central e variabilidade da distribuição das pontuações com ele obtidas.” E de seguida (linha 25) deverá o “distinguem-se” ser substituído por “se distinguem”. E mais adiante (linha 31), onde se lê “capacidade discriminatória do actual exame, pode ser

necessário efectuar adaptações...”, deve ler-se “capacidade discriminativa do actual exame, é necessário efectuar alterações...”.

Resposta: *Todas as sugestões foram incluídas no manuscrito.*

(Página 11)

“Pretendeu-se com este estudo avaliar a fiabilidade do exame final de especialidade através da análise da distribuição das pontuações obtidas nessa avaliação.”

“Para além disso, quer considerando a amostra total, quer para 23 especialidades em concreto, verificou-se a existência de uma associação positiva entre a nota no exame final da especialidade e a média de curso e/ou a prova nacional de seriação.”

“A fidedignidade é também uma característica importante dos testes com boas qualidades psicométricas, sendo usado, por exemplo, o índice de coerência interna [...]”

“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 pontuações com ele obtidas.”

“Considerando estes resultados, nomeadamente, a pobre capacidade discriminatória do actual exame, é necessário efetuar adaptações ao mesmo”

- 10.** Uma das questões já mencionadas, mas que não será demais repetir, é a de que afirmações que não estejam devidamente referenciadas não valem mais do que meras opiniões. Assim, quando se diz, por exemplo, que “Estas são, na verdade, um dos principais parâmetros considerados nos concursos públicos...” (pág. 15, linhas 4-5), necessário se torna referenciar o regulamento citado, permitindo assim confirmar o pressuposto. E logo de seguida o mesmo se pode dizer em relação à origem da afirmação sobre as “boas práticas já implementadas em outros países” (linha 6).

Resposta: *Agradecemos mais uma vez a chamada de atenção para a falta de referências em algumas afirmações. Em relação ao primeiro ponto, não existe um regulamento único sendo que a maior parte dos concursos de admissão têm em conta a nota obtida no exame final. Por exemplo, no último concurso de abertura para médicos especialistas a nota de exame final é a variável de desempate entre candidatos (Diário da República – Aviso n.º 1347/2017, 2017). Adicionámos esta explicação bem como a respectiva referência ao manuscrito.*

(Página 12):

Não existindo um regulamento único, a maior parte dos concursos de admissão têm em conta a nota obtida no exame final. Por exemplo, no último concurso de abertura para

médicos especialistas a nota de exame final foi a variável de desempate entre candidatos²⁰.

(Lista de referências)

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

Em relação ao segundo ponto referido, as boas práticas existentes em outros países reflectem-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. Acrescentámos também esta discussão no manuscrito, adicionando as respetivas referências.

(Página 12)

As boas práticas existentes em outros países reflectem-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 literatura²¹⁻²³.

(Lista de referências)

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;19(3):308–18.

22. Musick D. A Conceptual Model for Program Evaluation in Graduate Medical Education. *Acad Med.* 2006;81(8):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(5):Doc63.

11. Ainda na DISCUSSÃO ao serem consideradas as limitações do estudo, talvez entre “o primeiro” e o “segundo lugar”, será de acrescentar a limitação inerente à própria natureza transversal do estudo; e nomeadamente de que as associações encontradas não permitem, per se, inferir causalidade, antes tendo de se presumir por princípio formal, serem devidas a covariáveis não consideradas no âmbito do presente estudo.

Resposta: *Agradecemos que o revisor nos tenha chamado a atenção para esta limitação que realmente é importante na interpretação e generalização dos resultados obtidos. A referida limitação foi acrescentada a esta versão do manuscrito na secção da discussão (Páginas 13 e 14):*

“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.”

- 12.** Finalmente quanto às CONCLUSÕES (pág. 17), onde se lê “não apresenta as propriedades psicométricas desejáveis”, talvez fosse de considerar antes “não apresenta propriedades psicométricas indispensáveis”. E a frase seguinte “...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.”. E na última linha, a “métodos de avaliação alternativos” acrescentaria “mais objectivos.”

Resposta: *Foram incluídas todas as alterações sugeridas.*

(Página 15)

“O atual exame final da especialidade em Portugal não apresenta as propriedades psicométricas indispensáveis, não possuindo dessa forma uma capacidade discriminativa satisfatória.”

“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.”

“Os resultados deste estudo sugerem que o atual exame final da especialidade necessita de ser repensado, seja através da sua re-estruturação e/ou da implementação de métodos de avaliação mais objetivos.”

Revisor B

1. Os autores conseguiram responder de forma adequada a praticamente todos os pontos levantados na primeira revisão. O único ponto que foi menos elaborado e a meu ver deveria ser alvo de uma avaliação mais aprofundada é sobre as especialidades que já têm formas de exame mais “avançadas”. Não tenho conhecimento aprofundado sobre cada uma, mas além da Medicina Geral e Familiar (que além do exame nacional de escolha múltipla, tem casos clínicos “virtuais”), existem outras que já tentaram dar um passo em frente (julgo que a Radiologia por exemplo). Esse esforço deve ser identificado e as várias opções tomadas recolhidas, para se perceber se existe um fio condutor comum, ou se cada colégio está a avançar para onde julga ser mais adequado de forma unilateral.

Resposta: *Agradecemos desde já o comentário efetuado pelo revisor. Concordamos que em algumas (poucas) especialidades já existem exames finais com uma estrutura diferente da normalmente utilizadas pelas diferentes especialidades. Discutimos este assunto com mais detalhe na discussão deste artigo.*

Empiricamente é opinião dos autores deste artigo que não existe nenhum fio condutor comum e que cada colégio tem tido uma abordagem unilateral. Concordamos que seria importante a existência de guidelines centrais para a elaboração de um exame que avaliasse os conhecimentos necessários e cujo sistema classificativo fosse seriativo. Acrescentamos estes pontos na discussão do artigo.

(Página 15)

“Os resultados obtidos e discutidos neste trabalho, em conjunto com a observação empírica de que cada especialidade parece ter uma abordagem unilateral à avaliação final, sugerem a ausência de uma estrutura de avaliação comum às diferentes especialidades, de acordo com guidelines internacionais, que permita a avaliação de conhecimentos e seriação de candidatos”.

Revisor C

1. Os autores devem rever as noções de distribuição normal, em particular no que se refere aos pontos de corte para valores de assimetria e curtose, pois não se procura que sejam elevados, mas sim baixos. Além disso penso que devem estar a usar critérios para Modelação de Equações Estruturais, o que não é correto, pois este estudo não versa nisso. Devem rever os critérios de Kline (2005). As referências ao teste K-S e observação do histograma estão corretas.

Resposta: *Agradecemos o comentário do revisor uma vez que a referência estava realmente incorreta pois referia-se, por engano, a valores critério para modelos de equações estruturais. Alterámos os valores de referência para estatísticas univariadas de Field (2005). No manuscrito foi acrescentada a seguinte informação*

(Página 7):

“Para a interpretação dos valores de curtose e assimetria foram considerando os critérios de Field (2005)¹¹ 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.”

2. O símbolo r para correlação de Spearman deve ser substituído por rs. (O "s" deve estar 3 ptos inferior à linha).

Resposta: *Alterámos o símbolo como requerido no texto e na Tabela 3.*

3. Os autores devem acrescentar na Tabela 2 uma coluna com M (DP) relativos a todas as especialidades. Como complemento podem acrescentar a Mediana e os percentis P25 e P75. Para ganhar espaço os autores podem remover as colunas do p-valor e substituir pela notação *p<.05;**p<.01;***p<.001 na coluna do rs. O valor exato de p não é imprescindível quando se avalia a correlação, uma vez que os coeficientes de correlação são em si mesmo medidas de effect-size.

Resposta: *Fizemos as alterações sugeridas à Tabela 3 tendo acrescentado média e desvio-padrão, mediana e percentis 25 e 75 para todas as especialidades.*

4. Relativamente aos resultados da Tabela 3 sugere-se que os autores expliquem melhor o porquê de algumas especialidades apresentarem correlações muito baixas, o que pode ter a ver não só com a ausência efetiva de relação entre a média de curso/ PNS com o resultado da nota da especialidade, mas com o reduzido espectro de variabilidade desta classificação.

Resposta: *Consideramos pertinente esta hipótese explicativa relativamente à ausência de correlação em algumas das especialidades e desta forma a mesma foi acrescentada ao manuscrito na secção discussão.*

(Página 13)

“De realçar que, tendo em conta a análise já realizada em termos da variabilidade das notas, a ausência de correlações significativas ou a existência de correlações baixas, poderá dever-se ao reduzido espectro de variabilidade desta classificação, não traduzindo por isso necessariamente à ausência efetiva de relação”.

5. Os autores podem encarar a hipótese de calcular Z-scores para as notas de cada uma das especialidades para que as possam comparar entre si. Com este método podem comparar as notas obtidas nas especialidades para perceber se existem disparidades entre as mesmas.

Resposta: *Os autores agradecem o comentário do revisor. No entanto gostaríamos de esclarecer o revisor C que devido a limitações da amostra e do tamanho do manuscrito, não pretendemos analisar ou comparar as diferentes especialidades entre si. Para que este objectivo se torne mais explícito, optámos por acrescentar um parágrafo na discussão*

(Página 14):

“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 esta possa ser feita por futuros estudos.”

Revisor D

1. Os autores continuam a dedicar grande parte do texto da introdução e da discussão ao tema da avaliação psicométrica, quando os próprios demonstram que não é possível efetuá-la. A agravar os autores mantêm no título do manuscrito os termos “avaliação psicométrica/psychometric evaluation”.

Resposta: *Agradecemos os comentários e sugestão efetuada pelo revisor e efetuámos as alterações sugeridas. Como sugerido, e tendo em conta as limitações inerentes ao formato das notas no exame da especialidade para efetuar uma análise psicométrica completa, alterámos o título do manuscrito para “Avaliação do exame final de especialidade em Portugal”.*

Ao longo do texto procurou-se eliminar as referências à psicometria, apesar de se referir que foram utilizados alguns métodos comumente utilizados neste tipo de análises.

(Página 7)

“O presente estudo tem como objetivos avaliar, dentro dos parâmetros adequados e apesar de não existirem dados que permitam uma análise psicométrica do atual exame, a precisão/fiabilidade das notas da avaliação obtidas através do exame final da especialidade nos seus moldes atuais, assim como a sua validade.”

(Página 11)

Pretendeu-se com este estudo avaliar a fiabilidade do exame final de especialidade através da análise da distribuição das pontuações obtidas nessa avaliação.

2. A informação obtida por especialidade médica continua a ser praticamente ignorada, apesar de terem sido efetuados por especialidade histogramas das notas no exame (Figuras 2 a 46 – ficheiro suplementares), assim como uma tabela (Tabela 3) em que são apresentados também por especialidade os coeficientes de correlação da nota no exame de especialidade com a nota do curso e com a nota da prova nacional de seriação. Apesar da diversidade que é possível observar entre especialidades, a nota do exame de especialidade continua a ser apresentada e discutida no manuscrito como se se comportasse de forma uniforme em todas as especialidades. A nota no exame final de especialidade é apresentada e discutida no texto como tendo uma assimetria negativa, sendo considerado apenas o padrão observado para o conjunto total das notas. Nas figuras 2 a 46 é no entanto

apresentada a distribuição das classificações por especialidade, verificando-se em algumas especialidades uma distribuição aproximadamente simétrica (ver por exemplo Figura 15 que representa a Gastreenterologia, Figura 31 que representa a Neurologia ou Figura 45 que representa a Saúde Pública). Na especialidade de Saúde Pública além da distribuição aproximadamente simétrica, a gama de notas no exame de especialidade e o centro da distribuição difere muito das restantes especialidades. As figuras 2 a 46 são apenas referidas na secção de Métodos como incluídas como material suplementar. Na secção de Resultados continua a não ser feita qualquer leitura das mesmas, mas apenas a leitura da figura que inclui o histograma com todas as notas do exame independentemente da especialidade que representam. Na secção de Discussão a variabilidade das notas no exame de especialidade continua a considerar apenas a distribuição global, ignorando as diferentes especialidades.

Resposta: *Agradecemos a análise detalhada que o revisor fez aos resultados por especialidade. Consideramos realmente interessante a diferença encontrada nas distribuições consoante a especialidade em análise. No entanto, neste trabalho, o objectivo dos autores centra-se na análise das notas como um todo, sendo que as análises por especialidade são apresentadas de forma a ilustrar que eventualmente estas diferenças existirão consoante a especialidade. No entanto, considerámos pertinente a sugestão do revisor e incluímos na secção dos resultados alguma informação referente a estas diferenças*

(Página 9)

“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. 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).”

3. Na secção de Resultados a Tabela 3 é apenas apresentada como “A mesma avaliação foi feita por especialidade, como apresentado na Tabela 3”, não havendo qualquer texto com a leitura dos coeficientes de correlação apresentados da nota no exame de especialidade com a nota do curso e com a nota da prova nacional de seriação. No primeiro parágrafo da secção de Discussão é sumariada a informação que seria possível retirar da análise da Tabela 3: “... quer considerando a amostra total, quer para 23 especialidades em concreto,

verificou-se a existência de uma associação entre a nota no exame final da especialidade e a média de curso e/ou a prova nacional de seriação”. Esta afirmação está no entanto errada pois está em desacordo com os valores apresentados na Tabela 3, que revela em várias especialidades ausência de associação significativa ou associações muito baixas entre a nota no exame final da especialidade e a média de curso e/ou a prova nacional de seriação.

Resposta: *Apesar de o objectivo específico deste trabalho estar relacionado com a análise das notas como um todo e não por especialidade, à semelhança do ponto anterior, incluímos referência às diferenças por especialidade na secção dos resultados. (Página 10)*

“Quando se analisam os resultados por especialidade verifica-se que existe uma grande variabilidade nas relações estudadas consoante a especialidade em análise. Encontramos especialidades (n = 10) como a Anatomia Patológica, a Cirurgia Vascular/Angiologia ou a Ginecologia/Obstetria 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 (n = 10) 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 (n = 10) 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 um menor número de participantes (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 pequena (ex.: Pediatria Médica), média (ex.: Medicina Interna) e grande (ex.: Cirurgia Vascular/Angiologia). O mesmo acontece na relação entre o exame final da especialidade e a PNS, com efeitos pequenos (ex.: Ortopedia), médios (ex.: Medicina Interna) e grandes (ex.: Imuno-Hemoterapia) a serem detetados. Verifica-se que, nesta última relação, existem mais efeitos significativos e, de forma, geral as magnitudes encontradas são mais elevadas.”

4. Na minha opinião todo o texto do artigo deverá ser reformulado, e não apenas revisto, tendo em conta o estudo efetuado: Introdução, Métodos, Resultados e

Discussão. A construção de novas tabelas e figuras possíveis de ser incluídas no texto é necessária para mostrar a diversidade existente entre as especialidades médicas. As figuras apresentadas como material suplementar além de serem num número excessivo, não evidenciam as diferenças ou semelhanças existentes entre as notas do exame de especialidade das diferentes especialidades. A formatação das tabelas e figuras deverá ser mais cuidada tendo em conta a sua possível publicação. Por exemplo no manuscrito submetido os valores de prova inferiores a 0.001 são apresentados quer como “<0.001” quer como “0.000”, os histogramas e diagramas de dispersão são apresentados tal como são obtidos automaticamente pelo programa SPSS, apesar do texto ser em português os números são apresentados com a formatação inglesa.

Resposta: *Agradecemos desde já os comentários efectuados pelo revisor. Por uma questão de espaço e de objetivo do estudo, não nos foi possível apresentar os resultados ou discutir todas as 47 especialidades existentes. Salientámos este facto na discussão do artigo alertando para a importância de, em estudos futuros, estas análises serem mais detalhadas.*

(Página 14)

“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 esta possa ser feita por futuros estudos”.

Alterámos também as figuras para uma formatação mais cuidada tendo em vista a publicação. Mantivemos a vírgula como separador decimal visto ser o separador mais comumente usado em língua Portuguesa, e uniformizámos o seu uso ao longo de todo o documento.

5. Uma análise descritiva da nota final do exame das especialidades médicas em Portugal e a sua associação com outras avaliações parece-me ser relevante, no entanto deve ser efetuada de forma a revelar as diversidades existentes entre especialidades. Sendo a nota do exame de especialidade estudada tendo em conta a nota do curso e a nota da prova nacional de seriação, será importante fornecer também uma descritiva destas notas no global e por especialidade. Se houver informação disponível, a instituição em que as especialidades médicas foram obtidas poderá ser outro fator a ter em conta na análise das notas obtidas nos exames de especialidade. O acesso a dados de diferentes especialidades parece-me ter

potencial para descrever e discutir de forma mais aprofundada a nota do exame final da especialidade em Portugal. As diferenças e semelhanças entre especialidades deverá ser discutida tendo em conta não só a análise estatística das notas, como também a diversidade e especificidade das avaliações efetuadas nas diferentes especialidades.

Resposta: *Agradecemos o comentário efectuado pelo revisor. No entanto, como referido anteriormente, não é possível num só artigo discutir as especificidades de todas as especialidades, nomeadamente as diferenças e semelhanças de notas e a diversidade e especificidade das avaliações efectuadas nas diferentes especialidades. Optámos por acrescentar este ponto nas limitações do artigo e sugerir que futuros estudos abordem estas diferenças*

(Página 14)

“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 esta possa ser feita por futuros estudos.”

Revisor E

1. No seguinte paragrafo incluir “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”

Resposta: *A Universidade da Beira Interior foi acrescentada à frase supracitada.*