



UNIVERSIDADE DA BEIRA INTERIOR
Ciências da Saúde

Perguntas de Escolha Múltipla: dificuldades dos professores na sua construção e impacto no desempenho dos alunos

Gonçalo José Pereira Carneiro

Dissertação para obtenção do Grau de Mestre em
Medicina
(Ciclo de Estudos Integrado)

Orientadora: Professora Doutora Maria Isabel Neto

Covilhã, maio de 2019

Dedicatória

A toda a minha família, em especial aos que já partiram.

Agradecimentos

Aos meus pais e irmã, por terem dado tanto das suas vidas para que eu pudesse ter esta oportunidade, e por me terem ensinado a ter força e Fé em todos os momentos;

À minha Inês;

Aos meus amigos: aos que eternizaram na minha vida estes tão fugazes 6 anos e aos que me acompanham desde sempre;

À minha orientadora, Professora Doutora Isabel Neto, pela ajuda e sábios conselhos no decurso deste trabalho;

À Faculdade de Ciências da Saúde, por ter sido o berço da minha vida futura.

À Tuna-MUs, por me ter ensinado o que “não vem nos livros”;

À Covilhã, por me ter mostrado que Casa são as pessoas e não um sítio;

A tudo o que vivi até agora e que de uma forma ou de outra me permitiu ser o que sou hoje.

Prefácio

Título: Perguntas de Escolha Múltipla: dificuldades dos professores na sua construção e impacto no desempenho dos alunos.

Autor(a): Gonçalo José Pereira Carneiro.

Área(s) científica(s): Educação Médica.

Orientação científica: Professora Doutora Maria Isabel Neto.

Esta dissertação é resultado de trabalho e investigação individual, independente, com conteúdo original. As fontes consultadas encontram-se devidamente explícitas ao longo do desenvolvimento do texto e nas referências bibliográficas.

Resumo

Introdução: As perguntas de escolha múltipla (PEM) são um método de avaliação de conhecimentos extremamente disseminado e popular. No Mestrado Integrado em Medicina (MIM) da FCS-UBI, uma grande parte destas avaliações é feita através de PEM. Contudo, há entre os alunos descontentamento com a sua qualidade. Por outro lado, da parte dos docentes existem várias limitações que prejudicam uma boa construção de PEM (sobretudo tempo e necessidade de redigir uma grande quantidade de perguntas). Assim sendo, os principais objetivos deste trabalho foram perceber a perceção que os docentes têm sobre erros presentes nas PEM e o impacto negativo que os alunos atribuem a esses mesmos erros.

Material e Métodos: Foram aplicados a alunos de todos os anos e aos docentes de MIM questionários desenvolvidos pelos autores deste estudo, tendo por base os critérios definidos na literatura para a construção de PEM de qualidade. Foi depois realizada análise estatística (análise de consistência interna dos questionários, estatística descritiva e teste U de *Mann-Whitney*), tendo sido considerados como estatisticamente significativos valores de $p < 0.05$.

Resultados e Discussão: Dos 945 alunos matriculados em MIM, 728 responderam ao questionário (77,0%). Do lado dos docentes, dos 130 utilizadores da plataforma de submissão de perguntas, obtiveram-se 25 respostas (19,2%). Os alunos afirmaram que o não cumprimento de 14 dos 15 critérios enunciados nos questionários tem impacto negativo nas suas classificações, estando os principais relacionados com o desacordo das perguntas com os objetivos de aprendizagem e com formulações confusas/ambíguas. O critério que os estudantes consideraram ter menos impacto no seu desempenho foi “Pergunta baseada em casos clínicos ou num problema”. Verificou-se também que os alunos souberam tirar partido de erros/brechas de construção para acertarem em questões sobre as quais não têm conhecimentos, exceto quando essas pistas são fornecidas por convergência. Já os docentes afirmaram não ter dificuldade em respeitar 17 dos 19 critérios enunciados, tendo sido evidenciada maior facilidade em preparar perguntas que avaliam memorização do que perguntas que avaliam raciocínio e resolução de problemas. Este é, inclusive, uma das maiores dificuldades dos docentes, a par da construção de perguntas que não estejam na negativa. Em relação à capacidade de identificarem erros de construção nas PEM, os resultados indicam maior dificuldade em identificar distratores imprecisos/não rigorosamente incorretos e formulações desnecessariamente complicadas e capazes de induzir os alunos em erro. Por outro lado, indicam também mais facilidade em identificar perguntas construídas na negativa ou com duplas negações, além de pistas/brechas de construção capazes de sugerir a resposta correta.

Conclusão: Os alunos consideram que perguntas com erros têm um impacto considerável nos seus resultados académicos. Por outro lado, os docentes manifestam relativa facilidade em

cumprir a grande maioria das *guidelines* para a construção de PEM. Ainda assim, algumas das dificuldades manifestadas pelos docentes coincidem não só com alguns dos fatores aos quais os alunos atribuem mais impacto negativo, mas também com fatores dos quais os alunos se aproveitam para responderem a PEM sem terem conhecimentos para isso. Isto exige, portanto, uma reflexão sobre estes fatores em particular - como evitar escrever perguntas com estas falhas, como reconhecer estas falhas nas perguntas já existentes, como tornar as avaliações mais justas e fiáveis tendo em conta que eliminar totalmente as falhas é impossível.

Palavras-Chave: Avaliação; Perguntas de escolha múltipla; Critérios; *Testwise*; Dificuldade; Discriminação.

Abstract

Introduction: Multiple choice questions (MCQ) are an extremely popular method for knowledge assessments. At FCS-UBI's medical course, most of these assessments are carried out using MCQ. However, students quite often complain about their quality. On the other hand, teachers who write these questions are limited by several factors (mostly the lack of time and the need to write a considerable amount of questions). Therefore, the aim of this work was to evaluate teacher's perception about flaws in MCQ and the negative impact that students attribute to them.

Material and Methods: The authors of this work developed questionnaires based on published criteria on how to write good quality MCQ. These were then given to students from every year and medical teachers. Then, a statistical analysis was conducted (analysis of internal consistency of the applied questionnaires, descriptive statistics and Mann-Whitney U test). P values < 0.05 were considered as statistically significant.

Results and Discussion: Of the total 945 students enrolled in the medical course, 728 replied to the questionnaire (77,0%). Of the 130 teachers that use the MCQ submission platform, 25 replied to the questionnaire (19,2%). Students stated that 14 out of the 15 available criteria, when not satisfied, have a considerable negative impact on their responses to MCQ, the more problematic ones being related to discrepancy between content of questions and learning objectives, and ambiguous/confused formulation. The criterion students stated that has less negative impact on their performance is "Question based on clinical vignettes or a problem". It was also verified that students can take advantage from flaws in MCQ to correctly answer them even if they don't have the necessary knowledge to do so, except when those flaws are related to word convergence. As for teachers, they claimed no difficulties respecting 17 of the 19 stated criteria. They found easier writing memorization questions than logic and clinical problem questions. This is also one of teacher's biggest difficulties, just like writing non-negative phrased questions. Regarding their capability to identify flaws in MCQ, teachers had more trouble identifying imprecise/non strictly incorrect distracters and unnecessarily complicated formulations that induce students in error. On the other side, they identified negatively phrased questions, double negatives and clues that may suggest the correct answer with more ease.

Conclusion: Students considered that flawed questions have a considerably negative effect on their performances. On the other side, teachers showed relative ease in fulfilling the majority of criteria. Nevertheless, some of teachers' biggest difficulties match not only with some of the factors which students attribute more negative impact, but also with factors that students use to take unfair advantage of to answer to questions. That demands a reflection

about those particular factors - how to avoid them, how to recognize them in pre-existing questions and how to make fairer and more reliable assessments, bearing in mind that writing totally unflawed questions is impossible.

Keywords: Assessment; Multiple choice questions; Criteria; Testwise; Difficulty; Discrimination.

Índice

Dedicatória	iii
Agradecimentos	v
Prefácio	vii
Resumo	ix
Abstract	xi
Índice	xiii
Lista de Tabelas	xiii
Lista de Figuras.....	xv
Lista de Acrónimos	xvii
1. Introdução	1
1.1. Contextualização	1
1.2. - Questões da investigação:	3
1.3. - Objetivos:	4
2. Metodologia de investigação	5
2.1 Tipo de Estudo:	5
2.2 Definição da população e amostra em estudo:.....	5
2.3 Questionários (ver anexos 3-11):	5
2.4 Recolha dos dados	6
2.5 Análise estatística dos dados	6
3. Resultados	7
3.1. Caracterização da amostra:	7
3.2. Análise da consistência interna das escalas utilizadas.....	9
3.3. Identificar os problemas na construção de PEM que os alunos classificam como tendo mais influência negativa nos resultados obtidos:	11
3.4. Identificar que características das PEM os docentes têm mais dificuldades em respeitar, aquando da sua construção:	13
3.5. Avaliar em que medida os estudantes podem tirar partido de problemas na construção de PEM para melhorar os resultados obtidos:	17
3.6. Perceber em que medida os docentes são capazes de identificar os erros mais comuns na construção de PEM:.....	18
4. Discussão dos resultados	19
4.1. Perceção dos alunos em relação ao impacto negativo causado por erros de construção das PEM.....	20
4.2. Comparação da perceção entre alunos de anos básicos e anos clínicos	22
4.3. Perceção dos docentes quanto à dificuldade em respeitar critérios para a construção de PEM ²³	
4.4. Análise da capacidade <i>Testwise</i> dos alunos.....	25
4.5. Análise à capacidade dos docentes identificarem erros de construção comuns em PEM	26
4.6. Comparação entre a perceção de alunos e docentes quanto aos erros de construção das PEM	27

5. Limitações do estudo	29
6. Conclusão.....	31
7. Referências Bibliográficas	33
8. Anexos	35

Lista de Tabelas

Tabela 1 - Número e percentagem de alunos por ano curricular	18
Tabela 2 - Número e percentagem de alunos por subgrupo	18
Tabela 3 - Número e percentagem de docentes, entre anos básicos e clínicos	19
Tabela 4 - Correlação item-total e efeito da eliminação de cada item constante do questionário dos alunos.....	20
Tabela 5 - Correlação item-total e efeito da eliminação de cada item constante do questionário dos docentes	21
Tabela 6 - Respostas dos alunos (agrupadas por níveis e por subgrupo) à primeira secção do questionário e relação entre as respostas dos alunos de anos básicos e clínicos	23
Tabela 7 - Respostas agrupadas dos docentes à primeira secção do questionário, por subgrupo	25

Lista de Figuras

Figura 1 - Respostas certas (%) dos alunos nas PEM, utilizando erros de construção nas perguntas.....	17
Figura 2 - % de docentes que identificou a falha pretendida	18

Lista de Acrónimos

FCS	Faculdade de Ciências da Saúde
UBI	Universidade da Beira Interior
MIM	Mestrado Integrado em Medicina
PEM	Perguntas de escolha múltipla
ID	Índice de dificuldade
IDM	Índice de discriminação
HR	Hipóteses de resposta
SBA	Single Best Answer

1. Introdução

1.1. Contextualização

As perguntas de escolha múltipla (PEM) são um método de avaliação de conhecimentos extremamente disseminado e popular. Consistindo num enunciado com a questão seguido de um conjunto de hipóteses de resposta (HR), onde se encontram os distratores e a resposta considerada correta(1,2), apresentam como maiores vantagens a correção fácil, possibilidade de encurtar o tempo de exame e de avaliar uma grande quantidade de conteúdos num menor espaço de tempo(3). Permitem ainda avaliar várias capacidades intelectuais dos alunos - das mais elementares como memorização até às mais complexas como raciocínio clínico(4,5). Permitem também uma análise pós-exame completa e precisa sobre a fiabilidade com que avaliaram os alunos, permitindo-lhes também perceber onde erraram e de que forma podem colmatar as suas falhas. Contudo, construir PEM de elevada qualidade requer um processo de redação muito trabalhoso, norteado por *guidelines* credíveis e que não cessa apenas com a escrita, pois é fulcral revê-las pós-exame. Reside aqui a principal desvantagem das PEM: os recursos que consomem.

No Mestrado Integrado em Medicina (MIM) da FCS-UBI, uma grande parte das avaliações de conhecimentos é feita através de testes constituídos por PEM. Contudo, há entre os alunos descontentamento com a qualidade delas. Por outro lado temos os docentes, cuja carga de trabalho, necessidade de redigir perguntas para vários anos e vários cursos, além das suas próprias dificuldades neste processo, limitam o tempo disponível não só para redigir, mas sobretudo para a revisão e análise pós-exame das perguntas. Isto pode originar PEM deturpadas por fatores extra-conhecimento que podem ter impacto negativo nos resultados académicos dos alunos(6,7).

1.1.1. - Tipos de PEM e recomendações de construção

Não existem regras fixas/imutáveis para a construção de PEM, mas antes diversas recomendações(1) que devem ser seguidas. Cabe depois às instituições adaptar a construção de perguntas à sua metodologia de ensino e objetivos próprios.

Quanto aos diferentes tipos de PEM(1), podemos agrupá-los em duas famílias - *single best answer* (SBA) e verdadeiro/falso (V/F), sendo que no primeiro caso é suposto escolher a HR mais adequada e no segundo a única possível. Isto obriga a que, além das recomendações transversais às duas famílias (como enunciado e opções claros e concisos, evicção de vocabulário vago como “talvez” ou absoluto como “sempre” e ainda HR contidas numa única dimensão do conhecimento - ver anexo 1), existam recomendações particulares. No caso das SBA, a hipótese correta deve ser “mais verdadeira” que as restantes e os distratores podem

ser parcialmente ou totalmente falsos. Em adição, o enunciado deve permitir que o aluno encontre a resposta correta sem ter de olhar para as HR (“*cover the options rule*”). No caso das V/F, todas as HR devem ser **totalmente** verdadeiras ou falsas, não sendo permitidas “*Shades of grey*” (ver anexo 2).

O incumprimento destas recomendações pode adicionar dificuldade artificial (dificuldade irrelevante para a avaliação de conhecimentos), fornecer pistas que denunciam a resposta correta ou até não ter impacto nenhum(8). Deste modo, dada a imprevisibilidade(8,9) que a presença destes erros origina na qualidade das perguntas, é essencial respeitar ao máximo as recomendações existentes, ainda que estas não sejam regras obrigatórias e universais.

1.1.2. - Análise estatística pós-exame

É a mais poderosa ferramenta para melhorar a qualidade das PEM(10). Para essa análise usamos cálculos como o índice de dificuldade (ID) e o índice de discriminação (IDM).

ID corresponde ao rácio entre os alunos que acertaram uma determinada questão e o número total de respostas(1,10), variando entre 0 (mais difícil) e 1 (mais fácil)(10).

Já o IDM reflete a capacidade da pergunta distinguir alunos que têm ou não conhecimentos para a responder(1,10). Pode medir-se pela correlação entre a resposta dos alunos para uma dada pergunta e a sua classificação no global do exame, ou pela diferença entre a taxa de respostas certas do grupo forte e do grupo fraco (grupo de alunos com melhores e piores resultados no global do exame, respetivamente)(1,11). Varia entre -1 e 1, sendo que valores positivos e próximos de 1 indicam que alunos que responderam corretamente à questão tendem a ter bons resultados no global da prova, ao passo que alunos que responderam erradamente tendem a ter piores performances globais. Valores próximos de 0 indicam que não há relação entre a performance dos alunos naquela pergunta e no total do exame, ou seja, a pergunta não discrimina entre bons e maus alunos(10,12). Valores negativos indicam que a pergunta foi mais acertada por alunos com piores resultados no global da prova, o que é indesejável(1,10).

Posto isto, podemos facilmente perceber que estes índices estão intimamente ligados. Uma pergunta com ID elevado(1,10) foi acertada por grande parte dos alunos. É pois de esperar que tanto alunos bons como menos bons a tenham acertado. O mesmo para perguntas com ID baixo(1,10). Significa isto que perguntas com ID muito baixos ou muito altos correlacionam-se menos com os resultados no global do exame, possuindo um IDM baixo.

1.2. - Questões da investigação:

- 1) Qual a perceção dos estudantes sobre a influência de erros na formulação de perguntas de escolha múltipla no seu desempenho académico?

- 2) Qual a perceção dos professores sobre as suas dificuldades na construção de perguntas de escolha múltipla de qualidade?

1.3. - Objetivos:

- 1) Identificar os problemas na construção de PEM que os alunos classificam como tendo mais influência negativa nos resultados obtidos;
- 2) Avaliar em que medida os estudantes podem tirar partido de problemas na construção de PEM para melhorar os resultados obtidos;
- 3) Identificar que características das PEM os docentes têm mais dificuldades em respeitar, aquando da sua construção;
- 4) Perceber em que medida os docentes são capazes de identificar os erros mais comuns na construção de PEM;
- 5) Perceber se há relação entre o que os alunos identificam como mais prejudicial aos seus resultados e o que os docentes classificam como mais difícil de respeitar.

2. Metodologia de investigação

2.1 Tipo de Estudo:

Este estudo é observacional e transversal.

2.2 Definição da população e amostra em estudo:

Este estudo teve como população alvo todos os estudantes do MIM da UBI (N=945) e todos os docentes que preparam perguntas de escolha múltipla para integrarem as avaliações de conhecimentos do MIM (N=130).

Para se calcular a amostra mínima representativa da população para obtermos um intervalo de confiança de 95%, foi utilizada a calculadora *online* da Raosoft® disponível no portal <http://www.raosoft.com/samplesize.html>.

2.3 Questionários (ver anexos 3-11):

A construção dos questionários destinados a ser respondidos por alunos e professores baseou-se nos critérios de qualidade para a construção de perguntas de escolha múltipla presentes na literatura(1-4,6-8,13,14).

O questionário entregue aos alunos era composto por duas secções. Na primeira estavam referidos 15 daqueles critérios de qualidade, na forma de afirmações, relativamente às quais os alunos deveriam indicar em que medida o seu não cumprimento poderia influenciar negativamente a sua resposta, utilizando para isso uma escala de Likert de 5 pontos: 1 - Não influencia; 2 - Influencia pouco; 3 - Indiferente; 4 - Influencia; 5 - Influencia muito. Na secção seguinte, por forma a testar a habilidade dos alunos em responder a PEM com erros de construção (capacidade *testwise*), foi-lhes pedido que respondessem a 7 exemplos de PEM (retiradas do portal <http://www.essentialedits.ca/tests/take/mc/how2.htm>) onde era possível chegar à resposta correta sem qualquer conhecimento sobre o que se estava a perguntar.

O questionário entregue aos docentes, tal como os dos alunos, era composto por duas secções. Na primeira secção estavam expostos 19 critérios para a construção de PEM de boa qualidade, sendo pedido que indicassem o grau de dificuldade em respeitar esses critérios. Para isso foi utilizada uma escala de Likert de 5 pontos: 1 - Muito difícil; 2 - Pouco difícil; 3 - Indiferente; 4 - Fácil; 5 - Muito fácil. Na segunda secção, com o objetivo de avaliar a sua capacidade de identificar erros na construção de perguntas de escolha múltipla, foi pedido aos docentes que identificassem erros em perguntas específicas(1), por seleção na lista de critérios de qualidade anteriormente referidos.

2.4 Recolha dos dados

A aplicação dos questionários aos alunos e docentes de MIM da FCS-UBI foi feita através de um link para a plataforma *Google Forms*®, enviado por endereço de correio electrónico. A recolha foi feita entre Novembro e Dezembro de 2018.

2.5 Análise estatística dos dados

Foi utilizado o *software* informático SPSS® (Statistics Package for the Social Sciences), versão 25.0, para o sistema operativo *Microsoft Windows*®.

Através deste *software* foi construída uma base de dados, a partir da qual foram calculados parâmetros de estatística descritiva (frequências e mediana). Para fazer a comparação entre grupos foram usados testes não-paramétricos (*Mann-Whitney U*), pois segundo *Spencer E. Harpe*(15), a aplicação destes testes está indicada quando os dados recolhidos correspondem a categorias de uma escala de Likert, ao invés de um intervalo contínuo de valores, devendo por isso ser descritos por percentagens ou medianas.

Para as análises realizadas foram consideradas as seguintes hipóteses:

- Hipótese nula: Não há diferença estatisticamente significativa entre as respostas dos alunos pertencentes a anos clínicos e os pertencentes a anos básicos.
- H1: Há diferença estatisticamente significativa entre as respostas dos alunos pertencentes a anos clínicos e os pertencentes a anos básicos.

Considerou-se como estatisticamente significativo um valor de prova inferior a 5% ($P < 0,05$).

Foi também analisada a consistência interna de ambos os questionários, através do cálculo do Alfa de Cronbach (mede a fiabilidade de um instrumento de avaliação, sendo este tanto mais fiável quanto mais este valor se aproximar de 1)(12,16,17).

3. Resultados

3.1. Caracterização da amostra:

Através da metodologia descrita em cima, calculou-se que a amostra mínima representativa para o 1º ano é 132, para o 2º é 105, para o 3º ano é 114, para o 4º ano é 103, para o 5º ano é 109 e para o 6º ano é 109. Para os docentes a amostra mínima é 98.

3.1.1. Alunos:

Foram obtidas 728 respostas por parte dos alunos, estando a sua distribuição por anos e a percentagem em relação à população de cada ano apresentadas na tabela 1. A amostra mínima apenas não foi obtida no 5º ano. Foram depois criados, para efeitos de tratamento estatístico, dois subgrupos: anos básicos (B) e anos clínicos (C), cujo número e percentagem de alunos em relação à população estão apresentados na tabela 2. O objetivo da criação destes grupos é comparar as respostas de alunos que não tiveram prática clínica (anos básicos) com as respostas de alunos que já a tiveram (anos clínicos).

Tabela 1 - Número e percentagem de alunos por ano curricular

	População	Amostra	% da população	% da amostra
1º Ano	199	150	75,4%	20,6%
2º Ano	144	112	77,8%	15,4%
3º Ano	161	138	85,7%	19,0%
4º Ano	140	110	78,6%	15,1%
5º Ano	150	107	71,3%	14,7%
6º Ano	151	111	73,5%	15,2%
Total	945	728	77,0%	100,0%

Tabela 2 - Número e percentagem de alunos por subgrupo

	População	Amostra	% da população	% da amostra
Anos Básicos	504	400	79,4%	54,9%
Anos Clínicos	441	328	74,4%	45,1%
Total	945	728	77,0%	100,0%

3.1.2. Docentes:

Foram obtidas 25 respostas por parte dos docentes, não se tendo obtido a amostra mínima representativa da população. A distribuição por subgrupo e a percentagem em relação à população estão descritas na tabela 3. O N relativo aos docentes que redigem PEM apenas para anos básicos (B) e apenas para anos clínicos (C) é difícil de determinar, pelo que esses campos (além das respetivas percentagens) surgem na referida tabela como N.D. (não determinado).

Tabela 3 - Número e percentagem de docentes, entre anos básicos e clínicos

	População	Amostra	% da população	% da amostra
Anos Básicos	N.D.	14	N.D.	56,0%
Anos Clínicos	N.D.	11	N.D.	44,0%
Total	130	25	19,2%	100,0%

3.2. Análise da consistência interna das escalas utilizadas

3.2.1. Primeira secção do questionário dos alunos

A análise da consistência interna deste questionário revelou um Alfa de Cronbach de 0,830. Podemos por isso dizer que os *items* deste questionário contribuem todos para medir de forma fiável uma única dimensão: a perceção dos alunos quanto ao impacto negativo destes *items* na sua resposta a PEM.

Testes adicionais indicam que não existem *items* negativamente correlacionados com a escala, nem que aumentem significativamente o valor de Alfa.

Tabela 4 - Correlação item-total e efeito da eliminação de cada item constante do questionário dos alunos

	Correlação de item total	Alfa de Cronbach
Pergunta não está de acordo com os objetivos de aprendizagem	0,398	0,823
Pergunta baseada em casos clínicos ou num problema	0,196	0,839
A pergunta está confusa não se percebendo o que se está a perguntar	0,509	0,820
A pergunta e/ou hipóteses de resposta estão mal formuladas	0,536	0,818
São utilizados termos não compreensíveis tendo em conta o ano frequentado	0,528	0,814
São usados termos inúteis/irrelevantes/redundantes para a compreensão da pergunta	0,487	0,817
A pergunta (enunciado e hipóteses) é demasiado longa e exige muito tempo de leitura	0,484	0,817
O vocabulário usado na pergunta é complicado e com imprecisões (ex: raramente, quase sempre, maioria das vezes)	0,602	0,810
Hipóteses de resposta não estão de acordo com o enunciado da pergunta	0,454	0,820
Pergunta não cumpre as regras gramaticais e de pontuação	0,430	0,822
Pergunta está construída de forma induzir os alunos em erro (ex: utilização de charadas ou de combinações de letras ou números)	0,604	0,811
O enunciado da pergunta está na negativa	0,414	0,823
A pergunta tem duplas negações (enunciado e hipóteses de resposta na negativa)	0,480	0,818
Não há distratores (respostas erradas) claramente incorretos	0,389	0,824
Há ambiguidades relativamente à resposta correta	0,509	0,818

3.2.2. Primeira secção do questionário dos docentes

A análise da consistência interna deste questionário revelou um Alfa de Cronbach de 0,853. Podemos por isso dizer que os *items* deste questionário contribuem todos para medir de forma fiável uma única dimensão: a perceção dos docentes quanto à dificuldade em respeitar os critérios enunciados.

Testes adicionais indicam que não existem *items* negativamente correlacionados com a escala, nem que aumentem significativamente o valor de Alfa.

Tabela 5 - Correlação item-total e efeito da eliminação de cada item constante do questionário dos docentes

	Correlação de item total	Alfa de Cronbach
Pergunta de acordo com os objetivos de aprendizagem.	0,619	0,834
Pergunta que avalia a capacidade de o estudante aplicar conhecimento ou	0,261	0,855
Pergunta que avalia a capacidade do estudante memorizar factos ou um	0,485	0,844
Pergunta que pode ser respondida apenas pela leitura do enunciado.	0,250	0,856
Formulação da pergunta está completa, bem como as hipóteses de resposta.	0,489	0,844
Os termos utilizados na pergunta são compreensíveis para o estudante tendo em conta o nível em que se encontra.	0,213	0,854
Não há termos inúteis na pergunta.	0,617	0,839
Está claramente identificado o que se quer perguntar.	0,759	0,833
O vocabulário usado na pergunta é simples e preciso.	0,779	0,832
Todas as hipóteses de resposta estão de acordo com o enunciado.	0,447	0,846
Respeito pelas regras gramaticais e de pontuação.	0,600	0,839
Pergunta construída de forma a não induzir os alunos em erro.	0,462	0,845
A pergunta não está construída na negativa.	0,249	0,855
A pergunta não contém dupla(s) negação ou negações (além de enunciado, contem também hipóteses de resposta na negativa).	0,467	0,845
A pergunta não contém pistas/brechas de construção que podem sugerir a resposta correta.	0,197	0,856
Homogeneidade na forma e conteúdo das hipóteses de resposta.	0,145	0,858
Hipóteses de resposta lógicas e plausíveis.	0,518	0,843
Os distratores (respostas erradas) são rigorosamente incorretos.	0,523	0,843
Não há qualquer ambiguidade sobre a validade da resposta correta.	0,460	0,845

3.3. Identificar os problemas na construção de PEM que os alunos classificam como tendo mais influência negativa nos resultados obtidos:

Na tabela 4 podemos encontrar os resultados das respostas dos alunos à primeira secção do questionário. Verifica-se que para 14 dos 15 critérios enunciados as respostas “influencia” e “influencia muito” (níveis 4+5) foram as predominantes (percentagem total superior a 50%). Em 4 desses critérios a percentagem total dos níveis 4+5 superou os 90%, constatando-se portanto que o que os alunos consideram que causa mais impacto negativo na sua resposta a PEM se relaciona com o desacordo das perguntas com os objetivos de aprendizagem e com formulações confusas/ambíguas.

Por outro lado, os alunos consideraram que perguntas baseadas em casos clínicos ou problemas não influenciam negativamente ou influenciam pouco as suas respostas a PEM. Existiu também diferença estatisticamente significativa entre os dois subgrupos, com os alunos de anos clínicos a atribuir menos impacto negativo que os de anos básicos (70,7% e 41,8% de respostas, respetivamente). De salientar ainda que, nesta questão, uma elevada percentagem de alunos dos anos básicos (32,2%) considera ser indiferente a utilização de vinhetas clínicas quanto ao impacto negativo na sua resposta a este tipo de perguntas.

Verifica-se também diferença estatisticamente significativa entre as respostas dadas por alunos de anos básicos e de anos clínicos relativamente à sua opinião sobre perguntas feitas na negativa e com duplas negações. Para ambas as situações, foram os alunos de anos clínicos quem considerou haver uma maior influência negativa, embora ambos os grupos considerem haver um maior impacto nas perguntas com duplas negações.

De destacar ainda que 3 dos 15 critérios tiveram taxas de resposta nos níveis 1+2 na ordem dos 20%, tendo em comum o facto de se referirem a situações onde a gramática/semântica prejudica a interpretação da pergunta.

Apesar de na tabela 4 não serem mostrados os resultados relativos ao nível 3 (“indiferente”), os valores registados nunca ultrapassaram os 25% da totalidade das respostas, tendo apenas chegado próximo deste valor em 4 critérios (onde se incluem os 3 critérios referido em cima).

Tabela 6 - Respostas dos alunos, agrupadas por níveis e por subgrupo (B - anos básicos; C- anos clínicos; T - total) à primeira secção do questionário e relação entre as suas respostas (P)

		1+2		4+5		p
		Nº	%	Nº	%	
Pergunta não está de acordo com os objetivos de aprendizagem	B	16	4,0%	364	91,0%	0,931
	C	12	3,7%	299	91,2%	
	T	28	3,8%	663	91,1%	
Pergunta baseada em casos clínicos ou num problema	B	167	41,8%	104	26,0%	0,000
	C	232	70,7%	46	14,0%	
	T	399	54,8%	150	20,6%	
A pergunta está confusa não se percebendo o que se está a perguntar	B	4	1,0%	388	97,0%	0,784
	C	4	1,2%	317	96,6%	
	T	8	1,1%	705	96,8%	
A pergunta e/ou hipóteses de resposta estão mal formuladas	B	6	1,5%	384	96,0%	0,347
	C	6	1,8%	310	94,5%	
	T	12	1,6%	694	95,3%	
São utilizados termos não compreensíveis tendo em conta o ano frequentado	B	28	7,0%	325	81,3%	0,092
	C	36	11,0%	251	76,5%	
	T	64	8,8%	576	79,1%	
São usados termos inúteis/irrelevantes/redundantes para a compreensão da pergunta	B	69	17,3%	234	58,5%	0,222
	C	68	20,7%	179	54,6%	
	T	137	18,8%	413	56,7%	
A pergunta (enunciado e hipóteses) é demasiado longa e exige muito tempo de leitura	B	35	8,8%	313	78,3%	0,753
	C	24	7,3%	252	76,8%	
	T	59	8,1%	565	77,6%	
O vocabulário usado na pergunta é complicado e com imprecisões (ex: raramente, quase sempre, maioria das vezes)	B	32	8,0%	287	71,8%	0,345
	C	31	9,5%	248	75,6%	
	T	63	8,7%	535	73,5%	
Hipóteses de resposta não estão de acordo com o enunciado da pergunta	B	21	5,3%	347	86,8%	0,801
	C	15	4,6%	282	86,0%	
	T	36	4,9%	629	86,4%	
Pergunta não cumpre as regras gramaticais e de pontuação	B	75	18,8%	226	56,5%	0,186
	C	79	24,1%	174	53,0%	
	T	154	21,2%	400	54,9%	
Pergunta está construída de forma induzir os alunos em erro (ex: utilização de charadas ou de combinações de letras ou números)	B	20	5,0%	351	87,8%	0,248
	C	12	3,7%	277	84,5%	
	T	32	4,4%	628	86,3%	
O enunciado da pergunta está na negativa	B	108	27,0%	197	49,3%	0,000
	C	48	14,6%	217	66,2%	
	T	156	21,4%	414	56,9%	
A pergunta tem duplas negações (enunciado e hipóteses de resposta na negativa)	B	45	11,3%	295	73,8%	0,000
	C	8	2,4%	294	89,6%	
	T	53	7,3%	589	80,9%	
Não há distratores (respostas erradas) claramente incorretos	B	48	12,0%	293	73,3%	0,259
	C	43	13,1%	227	69,2%	
	T	91	12,5%	520	71,4%	
Há ambiguidades relativamente à resposta correta	B	9	2,3%	377	94,3%	0,865
	C	5	1,5%	308	93,9%	
	T	14	1,9%	685	94,1%	

1+2 - Não tem influência + Influencia pouco

4+5 - Influencia + Influencia muito

3.4. Identificar que características das PEM os docentes têm mais dificuldades em respeitar, aquando da sua construção:

Na tabela 5 podemos encontrar as respostas dos docentes à primeira secção do questionário.

Verifica-se que para 17 dos 19 critérios, as respostas “fácil” e “muito fácil” (níveis 4+5) foram as predominantes, tendo 14 deles obtido mais de 50% do total de respostas (enquanto os restantes 3 critérios obtiveram a maioria de respostas nos níveis 4+5, mas com uma percentagem total inferior a 50%). Podemos portanto dizer que os docentes revelaram ter mais facilidade em preparar perguntas que avaliam memorização (84,0% de respostas nos níveis 4+5) do que perguntas que avaliam capacidade de raciocínio e de resolução de problemas (40,0%). Inclusive, este é um dos critérios que os docentes consideraram como mais difíceis de cumprir (44,0% de respostas nos níveis 1+2), tendo docentes de anos básicos e clínicos relatado um grau de dificuldade semelhante (42,9% e 45,5%, respetivamente). Os resultados também indicam que existe uma elevada dificuldade dos docentes em construir perguntas que não estejam na negativa (44,0% de respostas nos níveis 1+2).

Em 9 dos 17 critérios referidos anteriormente, a taxa de respostas apesar de superior a 50% nos níveis 4+5, foi também superior a 20% nos níveis 1+2. Desses critérios, 5 relacionam-se com a adição de dificuldade artificial, sendo eles “Pergunta está construída de forma a não induzir os alunos em erro” (24,0 %), “A pergunta não contém dupla(s) negação ou negações (além do enunciado, contém também hipóteses de resposta na negativa)” (20,0%), “Hipóteses de resposta lógicas e plausíveis” (20,0%), “Os distratores (respostas erradas) são rigorosamente incorretos (28,0%) e “Não há qualquer ambiguidade sobre a validade da resposta correta” (36,0%). Um deles relaciona-se com a capacidade *testwise* dos alunos, sendo ele “A pergunta não contém pistas/brechas de construção que podem sugerir a resposta correta (24,0%). O critério “Homogeneidade na forma e conteúdo das hipóteses de resposta” (24,0%) pode relacionar-se tanto com a adição de dificuldade artificial como com a capacidade *testwise* dos alunos. Os restantes 2 são “Pergunta de acordo com os objetivos de aprendizagem” (20,0%) e “Pergunta que pode ser respondida apenas pela leitura do enunciado” (20,0%), não se enquadrando em nenhuma das categorias mencionadas anteriormente.

Por fim, de destacar ainda a existência de 11 critérios em que os docentes revelaram indiferença relativamente à dificuldade ou facilidade no seu cumprimento superior a 20%, sendo eles “Pergunta que pode ser respondida apenas pela leitura do enunciado” (36,0%), “Os termos utilizados na pergunta são compreensíveis para o estudante tendo em conta o nível em que se encontra” (20,0%), “Não há termos inúteis na pergunta” (20,0%), “Pergunta construída de forma a não induzir os alunos em erro” (24,0%), “A pergunta não está construída na negativa” (20,0%), “A pergunta não contém dupla(s) negação ou negações (além do enunciado, contém também hipóteses de resposta na negativa)” (24,0%), “A pergunta não

Perguntas de Escolha Múltipla: dificuldades dos professores na sua construção e impacto no desempenho dos alunos contém pistas/brechas de construção que podem sugerir a resposta correta” (28,0%), “Homogeneidade na forma e conteúdo das hipóteses de resposta” (20,0%), “Hipóteses lógicas e plausíveis” (24,0%), “Os distratores (respostas erradas) são rigorosamente incorretos” (20,0%) e “Não há qualquer ambiguidade sobre a validade da resposta correta” (20,0%).

Tabela 7 - Respostas agrupadas dos docentes à primeira secção do questionário, por subgrupo (B - anos básicos, C - anos clínicos, T - total)

		1+2		4+5	
		Nº	%	Nº	%
Pergunta de acordo com os objetivos de aprendizagem.	B	3	21,4%	11	78,6%
	C	2	18,2%	8	72,7%
	T	5	20,0%	19	76,0%
Pergunta que avalia a capacidade do estudante aplicar conhecimento ou resolver um problema.	B	6	42,9%	4	28,6%
	C	5	45,5%	6	54,5%
	T	11	44,0%	10	40,0%
Pergunta que avalia a capacidade do estudante memorizar factos ou um conjunto de matéria.	B	1	7,1%	13	92,9%
	C	2	18,2%	8	72,7%
	T	3	12,0%	21	84,0%
Pergunta que pode ser respondida apenas pela leitura do enunciado.	B	3	21,4%	5	35,7%
	C	2	18,2%	6	54,5%
	T	5	20,0%	11	44,0%
Formulação da pergunta está completa, bem como as hipóteses de resposta.	B	2	14,3%	10	71,4%
	C	2	18,2%	8	72,7%
	T	4	16,0%	18	72,0%
Os termos utilizados na pergunta são compreensíveis para o estudante tendo em conta o nível em que se encontra.	B	0	0,0%	11	78,6%
	C	1	9,1%	8	72,7%
	T	1	4,0%	19	76,0%
Não há termos inúteis na pergunta.	B	2	14,3%	9	64,3%
	C	2	18,2%	7	63,6%
	T	4	16,0%	16	64,0%
Está claramente identificado o que se quer perguntar.	B	2	14,3%	11	78,6%
	C	1	9,1%	8	72,7%
	T	3	12,0%	19	76,0%
O vocabulário usado na pergunta é simples e preciso.	B	1	7,1%	11	78,6%
	C	2	18,2%	9	81,8%
	T	3	12,0%	20	80,0%
Todas as hipóteses de resposta estão de acordo com o enunciado.	B	2	14,3%	12	85,7%
	C	2	18,2%	6	54,5%
	T	4	16,0%	18	72,0%
Respeito pelas regras gramaticais e de pontuação.	B	1	7,1%	11	78,6%
	C	2	18,2%	8	72,7%
	T	3	12,0%	19	76,0%
Pergunta construída de forma a não induzir os alunos em erro.	B	4	28,6%	6	42,9%
	C	2	18,2%	7	63,6%
	T	6	24,0%	13	52,0%
A pergunta não está construída na negativa.	B	6	42,9%	5	35,7%
	C	5	45,5%	4	36,4%
	T	11	44,0%	9	36,0%

1+2 - Muito difícil + Pouco difícil

4+5 - Fácil + Muito fácil

Tabela 7 (continuação) - Respostas agrupadas dos docentes à primeira secção do questionário, por subgrupo (B - anos básicos, C - anos clínicos, T - total)

		1+2		4+5	
		Nº	%	Nº	%
A pergunta não contém dupla(s) negação ou negações (além de enunciado, contém também hipóteses de resposta na negativa).	B	1	7,1%	10	71,4%
	C	4	36,4%	4	36,4%
	T	5	20,0%	14	56,0%
A pergunta não contém pistas/brechas de construção que podem sugerir a resposta correta.	B	4	28,6%	8	57,1%
	C	2	18,2%	4	36,4%
	T	6	24,0%	12	48,0%
Homogeneidade na forma e conteúdo das hipóteses de resposta.	B	4	28,6%	8	57,1%
	C	2	18,2%	6	54,5%
	T	6	24,0%	14	56,0%
Hipóteses de resposta lógicas e plausíveis.	B	3	21,4%	7	50,0%
	C	2	18,2%	7	63,6%
	T	5	20,0%	14	56,0%
Os distratores (respostas erradas) são rigorosamente incorretos.	B	4	28,6%	7	50,0%
	C	3	27,3%	6	54,5%
	T	7	28,0%	13	52,0%
Não há qualquer ambiguidade sobre a validade da resposta correta.	B	5	35,7%	7	50,0%
	C	4	36,4%	4	36,4%
	T	9	36,0%	11	44,0%

1+2 - Muito difícil + Pouco difícil

4+5 - Fácil + Muito fácil

3.5. Avaliar em que medida os estudantes podem tirar partido de problemas na construção de PEM para melhorar os resultados obtidos:

Nesta secção era pedido aos alunos que respondessem a exemplos de PEM com diversas falhas, para avaliarmos a sua capacidade *testwise* - chegar à resposta correta através das falhas de construção da pergunta. As perguntas encontram-se em anexo. Na figura 1 encontra-se a percentagem de alunos que, em cada pergunta, escolheu a resposta considerada certa.

Assim, analisando os resultados constantes da figura, na **1ª pergunta** 58,9% dos alunos acertaram, sendo a resposta correta identificável por lógica (distratores pertencem ao mesmo campo lexical). Na **2ª pergunta** houve 52,1% de acerto, sendo novamente a resposta correta identificável por lógica (todos os distratores têm uma característica comum evidente). Na **3ª pergunta** houve 90,2% de acerto, sendo a resposta correta identificável por pistas gramaticais. Na **4ª pergunta** houve 69,9% de acerto, sendo que aqui a resposta correta era destacadamente mais longa que as restantes. Na **5ª pergunta** houve 67,9% de acerto, sendo a resposta correta identificável por pistas no enunciado (*clang clues*). Na **6ª pergunta** houve 68,1% de acerto, sendo a resposta correta identificável por falta de coerência com as restantes alternativas de resposta. Na **7ª e última pergunta** houve 9,3% de acerto, sendo a resposta correta identificável por convergência.

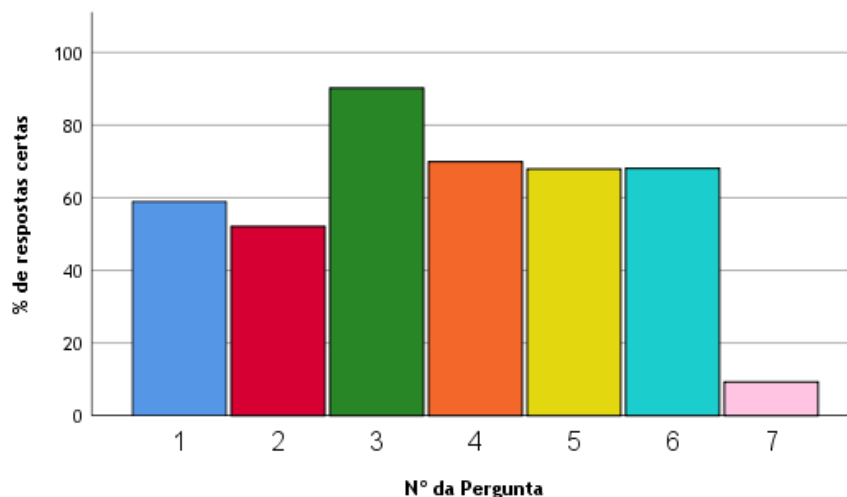


Figura 1 - Respostas certas (%) dos alunos nas PEM, utilizando erros de construção nas perguntas

3.6. Perceber em que medida os docentes são capazes de identificar os erros mais comuns na construção de PEM:

Nesta secção eram mostrados exemplos de PEM com falhas na sua construção, pedindo-se aos docentes que, através da lista de critérios fornecidos na primeira secção, identificassem o que é que estava em falha. Uma vez que a mesma pergunta pode conter mais do que uma falha, os autores deste estudo definiram em conjunto aquelas que, na sua opinião, são as mais relevantes em cada caso, passando então essas a ser consideradas como as respostas corretas. O gráfico 2 mostra a percentagem de docentes que identificou a falha pretendida em cada questão, e as PEM utilizadas neste questionário encontram-se em anexo.

Assim sendo, na 1ª pergunta 8,0% dos docentes identificaram a falha correta, na 2ª pergunta 28,0%, na 3ª pergunta 32,0%, na 4ª pergunta 16,0%, na 5ª pergunta 8,0%, na 6ª pergunta novamente 8,0% dos docentes, na 7ª pergunta 24,0% e na 8ª e última pergunta 52,0%. Podemos portanto dizer que os docentes tiveram mais dificuldade em identificar distratores imprecisos/não rigorosamente incorretos e formulações desnecessariamente complicadas e capazes de induzir os alunos em erro. Por outro lado, tiveram mais facilidade em identificar pistas/brechas capazes de denunciar a resposta correta e perguntas construídas na negativa ou com duplas negações.

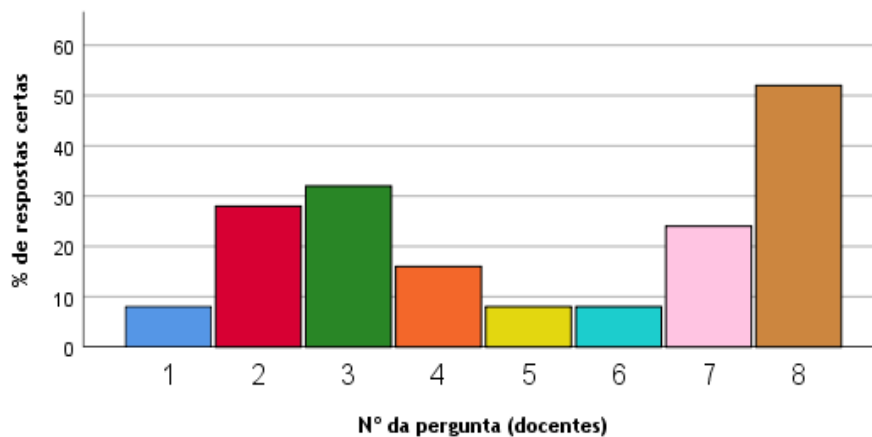


Figura 2 - % de docentes que identificou a falha pretendida

4. Discussão dos resultados

O presente trabalho tinha como objetivo analisar a perspetiva de alunos e docentes em relação às PEM (que compõem a maioria das avaliações de conhecimentos no MIM). Adicionalmente, pretendia-se perceber se os alunos eram capazes de tirar partido de perguntas mal construídas para identificarem a resposta correta sem terem os conhecimentos necessários para isso, bem como a perceção dos docentes sobre erros presentes em exemplos de PEM mal construídas.

Foram para isso aplicados questionários desenvolvidos pelos autores deste estudo, tendo-se obtido uma boa participação por parte dos alunos (728 em 945 inscritos, o que corresponde a 77%), mas uma menor participação por parte dos docentes (dos 130 docentes que preparam perguntas de escolha múltipla para avaliação apenas 25 responderam, o que corresponde a 19,2%). Sendo as PEM um tema muito sensível para os alunos (uma vez que definem grande parte do seu percurso académico e são um alvo frequente de críticas e reclamações), seria de esperar a elevada taxa de respostas. Por outro lado, também os docentes têm interesse por este tema, mas a menor taxa de respostas prende-se com uma maior dificuldade em abordá-los e com menor disponibilidade por parte dos mesmos.

4.1. Perceção dos alunos em relação ao impacto negativo causado por erros de construção das PEM

Os alunos consideraram que perguntas baseadas em problemas ou casos clínicos é o fator que menos impacto tem nos seus resultados académicos. Isto já era expectável, pois ao contrário dos outros critérios este não corresponde a uma regra formal/obrigatória que deva ser cumprida mas apenas a uma opção de redação que até é recomendada(1,4,5,18). Contudo, este tipo de perguntas é por norma mais desafiante (pois avalia capacidade de raciocínio e integração de matérias, em vez de apenas memorização)(5) pelo que é normal que alguns alunos lhe atribuam alguma dificuldade. Ainda assim, apenas 20,6% do total de alunos considera que estas perguntas influenciam negativamente a sua resposta, pelo que praticamente 80% dos alunos lida bem com este tipo de formulação.

Em relação aos critérios que obtiveram a maioria de respostas nos níveis 4+5, há 3 aos quais os alunos atribuem uma menor dificuldade que os restantes (mais de 20% de respostas nos níveis 1+2).

O primeiro relaciona-se com o não cumprimento de regras gramaticais e de pontuação (o que pode originar enunciados sem sentido ou desconexos das HR, prejudicando a compreensão da pergunta). Contudo, os alunos consideraram que perguntas confusas onde não se percebe o que está a ser perguntado têm mais impacto negativo (96,8% de respostas nos níveis 4+5). Seria de esperar taxas de resposta semelhantes para estes dois fatores (já que ambos se relacionam com a compreensão do objetivo da pergunta), mas no primeiro critério coloca-se em foco a questão gramatical/pontuação, ao passo que no segundo não se foca nenhum fator particularmente responsável pela má compreensão da pergunta. Isto poderá indicar que existem outros fatores capazes de adulterar mais o sentido das perguntas do que as questões gramaticais ou de pontuação, contudo pelos resultados obtidos não conseguimos saber quais. Em adição, não cumprir regras gramaticais pode dar pistas sobre a resposta correta. De facto, na segunda secção do questionário havia uma pergunta com estas características, tendo sido acertada por 90,2% dos alunos. Isto permite-nos concluir que não cumprir a gramática pode tanto facilitar como dificultar as PEM. Deve então valorizar-se particularmente este aspeto, dado o potencial impacto negativo que tem no ID das perguntas(1,6,8,13).

O segundo critério relaciona-se com o impacto causado pela presença de termos não necessários e que aumentam o tempo de leitura, retirando tempo de raciocínio. Verificamos depois que critérios onde se aborda o impacto da não compreensão de vocabulário relevante para a estrutura da pergunta (como termos desadequados para o ano em questão ou vocabulário complicado e impreciso) têm taxas de resposta nos níveis 4+5 consideravelmente superiores (cerca de 20%). Isto indica que os alunos atribuem mais influência à presença de termos importantes para a compreensão da pergunta mas que eles não conseguem descortinar, do que à presença de termos que consomem tempo mas não prejudicam a

Perguntas de Escolha Múltipla: dificuldades dos professores na sua construção e impacto no desempenho dos alunos interpretação da mesma. De notar ainda que os alunos consideraram que perguntas (enunciado e/ou hipóteses) demasiado longas têm um impacto negativo considerável (77,6% de respostas nos níveis 4+5). Sabemos que há PEM (como as que possuem vinheta clínica) em que o enunciado é mais longo e detalhado mas onde praticamente toda a informação é importante. Isto é mais um indício que os alunos se preocupam bastante com vocabulário que consideram importante (e com o tempo que a sua leitura/compreensão exige), mas lidam relativamente bem com a existência de vocabulário irrelevante. Ainda assim, relembramos que em todos os casos aqui abordados a taxa total de respostas nos níveis 4+5 foi sempre superior a 50%, significando que apesar de atribuírem pesos diferentes, os alunos reconhecem uma importância considerável à questão vocabulário.

O terceiro critério relaciona-se com enunciados fraseados na negativa, algo relativamente comum apesar de desaconselhado por vários autores(1,19), que consideram que complica desnecessariamente a interpretação da pergunta. Contudo, não é uma formulação proibida, requer apenas mais cautela na redação. Em relação à existência de duplas negações (enunciado e HR na negativa), constatamos que os alunos atribuíram um impacto negativo bastante superior a este fator (80,9% de respostas nos níveis 4+5, contra 56,9%). Isto diz-nos que eles lidam pior com esta formulação, o que já era expectável uma vez que a dupla negação traz uma complexidade superior à negação simples.

Por fim, realçamos que os alunos atribuíram mais impacto negativo à presença de ambiguidades relativamente à resposta correta do que à inexistência de distratores claramente incorretos (94,1% de respostas nos níveis 4+5 contra 71,4%). Isto indica que os alunos consideram mais prejudicial uma resposta correta mais difícil de identificar do que distratores não obviamente errados. Mesmo estes não tendo a qualidade desejada, existirá à partida uma resposta correta facilmente identificável, ao passo que uma resposta correta com ambiguidades facilmente se confunde com os distratores, conduzindo alunos que dominam os conteúdos a escolher outra opção além desta, por não satisfazer a correção científica que seria esperada. Estes assuntos tomam particular importância nas questões de tipo *Single Best Answer*, onde as hipóteses de resposta não costumam ser totalmente verdadeiras ou falsas, pertencendo antes a um “espectro de cinzentos”(14)(ver anexo 2).

4.2. Comparação da percepção entre alunos de anos básicos e anos clínicos

Apenas 3 critérios revelaram significância estatística entre os alunos de anos básicos e os de anos clínicos.

Para o critério “Pergunta baseada em casos clínicos ou num problema”, 26,0% dos alunos de anos básicos admitem que influencia negativamente a sua resposta, contra apenas 14,0% dos alunos de anos clínicos. Era expectável haver diferenças neste campo, uma vez que os alunos de anos clínicos têm logicamente mais experiência neste tipo de questões.

Quanto aos critérios “O enunciado da pergunta está na negativa” e “A pergunta tem duplas negações (enunciado e hipóteses de resposta na negativa)”, de notar que foram os alunos de anos clínicos que atribuíram mais influência negativa a estes fatores. Não existe uma justificação clara para este facto. Contudo, é sabido que quanto mais experientes os alunos são, mais exigentes tendem a ser com a qualidade das PEM(8). Isto pode levar a que as perguntas na negativa obriguem a maior esforço na sua interpretação, ao passo que os alunos de anos básicos perdem menos tempo a analisar a influência que a negativa tem no sentido das frases. Ainda assim, seria de esperar então que houvesse significância em mais critérios relacionados com a semântica e a interpretação das perguntas, o que não aconteceu.

4.3. Perceção dos docentes quanto à dificuldade em respeitar critérios para a construção de PEM

Os docentes admitiram ser particularmente difícil (maioria das respostas nos níveis 1+2) redigir perguntas que avaliam a capacidade do estudante aplicar conhecimentos ou resolver problemas bem como redigir perguntas que não estejam na negativa. No primeiro caso, entende-se a dificuldade manifestada, pois testar lógica e capacidade de raciocínio clínico através de PEM exige uma construção mais trabalhosa e criteriosa(1,4,5). Em relação ao segundo caso, se por vezes é mais simples redigir HR na negativa do que na afirmativa(19) a literatura mostra que perguntas fraseadas na afirmativa são as mais frequentes(19), pelo que não seria de esperar a existência de grandes dificuldades neste campo.

Dentro dos critérios que obtiveram a maioria das respostas nos níveis 4+5 (fácil + muito fácil), os docentes consideraram alguns como sendo mais difíceis de respeitar que os restantes. Destes, destacam-se 3 relacionados com a construção de HR e 1 com a capacidade *testwise* dos alunos.

Assim sendo, os resultados evidenciam alguma dificuldade em construir HR lógicas e plausíveis e também em redigir distratores rigorosamente incorretos. Isto é compreensível, pois escrever distratores inquestionavelmente errados mas simultaneamente plausíveis pode ser muito complicado, ocorrendo muitas vezes utilização de expressões imprecisas que tanto podem colocar em causa a sua falsidade como podem torna-los obviamente implausíveis e facilmente eliminados pelos alunos(1,3,13,18).

Em adição, evidenciou-se também alguma dificuldade em redigir HR que sejam corretas sem qualquer tipo de ambiguidade. Tal como é difícil redigir distratores plausíveis, é igualmente desafiante redigir opções corretas que não sejam tão óbvias que por si só descredibilizem as restantes HR. Para evitar que isto ocorra, por vezes são utilizados termos vagos e imprecisos que causam dúvidas quanto à validade da opção correta (ainda que os distratores sejam rigorosamente incorretos, uma alternativa de resposta correta ambígua prejudica a fiabilidade das PEM(1)).

No seguimento disto, verificou-se também alguma dificuldade em redigir HR homogêneas em forma e conteúdo. De facto, por vezes a hipótese certa é mais pormenorizada que os distratores, da mesma forma que estes são mais ambíguos e imprecisos(1,14). Isto ocorre na tentativa de torna-los indiscutivelmente corretos ou incorretos, respetivamente. Contudo, acaba por resultar em HR díspares e incoerentes. Pelo que já foi dito anteriormente, percebe-se que respeitar este critério seja desafiante para os docentes. Além disso, não respeitá-lo pode ter efeitos imprevisíveis na avaliação, pois tanto pode dificultar como facilitar a resposta.

Foi também evidenciada alguma dificuldade em redigir perguntas sem fornecer pistas que

Perguntas de Escolha Múltipla: dificuldades dos professores na sua construção e impacto no desempenho dos alunos sugiram a resposta correta. É natural existir alguma dificuldade neste campo, não só pela variedade de fatores capazes de fornecer pistas (desde *clang clues*, tempos verbais e até a construção gramatical da pergunta)(1,14), mas também porque tendo em conta o que já foi referido anteriormente, muitas vezes estas surgem da necessidade de tornar a hipótese certa como inquestionavelmente correta ou os distratores como inquestionavelmente errados, tornando-se difícil evitar fornece-las.

Por fim, as dificuldades evidenciadas nestes critérios demonstram o equilíbrio ténue que existe entre opções de resposta rigorosamente corretas, incorretas e “intermédias”, sendo isto especialmente relevante nas questões do tipo *single best answer*(14) (ver anexo 2).

4.4. Análise da capacidade *Testwise* dos alunos

A taxa de acerto foi superior a 50% em 6 das 7 perguntas, o que confirma que a capacidade *testwise* dos alunos tem de facto um peso considerável.

Ao ordenarmos as perguntas por percentagem de respostas certas, verificamos que a resposta por convergência foi a que obteve uma menor taxa (tendo inclusive sido a única com uma taxa inferior a 50%), depois surgem as duas questões que podiam ser resolvidas através de lógica, seguindo-se as que podiam ser resolvidas através da homogeneidade da forma e conteúdo das HR e por fim a resposta por coerência gramatical (que obteve a maior percentagem de acerto).

Podemos então inferir que a coerência/coesão gramatical é um dos fatores que mais chama a atenção dos alunos. A comprovar isto temos não só o exemplo da pergunta 3, mas também o da pergunta 5 onde a repetição das mesmas palavras denuncia a resposta correta.

Olhando para as questões que podiam ser resolvidas por lógica, ambas possuíam distratores com características comuns entre si mas ausentes na hipótese correta. Como se verifica pelos resultados, a presença de associações lógicas entre os distratores permite identifica-los e eliminá-los sem ser preciso estabelecer juízos quanto à sua validade. Quanto à pergunta 4, fica bem evidente a importância que a homogeneidade das HR tem, especialmente em questões do tipo *single best answer*, onde a hipótese correta é a que se adapta melhor ao que é dito no enunciado. Hipóteses mais detalhadas e completas, em oposição a outras mais curtas, geralmente cumprem melhor esse requisito.

Por fim, destacar a baixa percentagem de acertos na pergunta 7, onde eram fornecidos conjuntos de elementos pretendendo-se a identificação do conjunto correto. O intuito era avaliar a capacidade de responder por convergência - perceber que a resposta correta era a que continha os elementos mais repetidos ao longo de todas as HR. Contudo, esta pergunta tem uma particularidade. Um dos conjuntos é formado por elementos que têm em comum o facto de serem frutos, ao passo que na resposta considerada correta isso não acontecia. Curiosamente, essa resposta foi selecionada por 75,5% dos alunos. Isto pode indicar-nos que os alunos identificam mais facilmente e/ou são mais atraídos pela presença de características comuns/relações entre as hipóteses de resposta, do que a quantidade de vezes que elas surgem.

4.5. Análise da capacidade dos docentes identificarem erros de construção comuns em PEM

Esta secção fornece-nos dados quanto à capacidade dos docentes identificarem falhas nas perguntas e permite-nos também inferir quanto à sua perceção sobre as mesmas.

Tendo em conta o que foi dito nos resultados, podemos ver que os docentes identificam melhor fatores que podem denunciar a resposta correta (como pistas/brechas - questões 2 e 4; e falta de homogeneidade nas HR - questão 3) do que fatores que podem dificultar a sua identificação (como distratores imprecisos, não rigorosamente corretos e por isso mais difíceis de eliminar - questão 1). Sendo objetivo que as PEM sejam respondidas apenas por quem tem os conhecimentos necessários, é compreensível que haja mais preocupação em não denunciar a resposta correta do que em torná-la evidentemente mais correta que os distratores (ou neste caso, os distratores evidentemente mais errados).

Por outro lado, os docentes também tiveram mais dificuldades em identificar perguntas formuladas de forma a induzir os alunos em erro (questão 5) do que perguntas que não identificam de forma clara o que se quer perguntar (questão 7). Uma pergunta imprecisa não orienta o aluno e pode desvirtuar a validade das HR(1). Já uma pergunta que pela sua complexidade induz o aluno em erro não altera a validade das HR, mas antes propicia erros de leitura/interpretação que podem levar o aluno a escolher uma opção errada (ainda que ele saiba responder corretamente)(1). Ambos os casos se relacionam com a adição de dificuldade artificial, mas um deles tem claramente um efeito mais nefasto para a estrutura da pergunta. Nesse sentido, acreditamos ser compreensível que os docentes identifiquem e se preocupem mais com esse fator.

Por fim, os docentes tiveram poucas dificuldades em identificar a presença de negativa/duplas negações (questão 8). Como já foi dito anteriormente(19), por vezes os docentes redigem perguntas na negativa como forma de aumentar a sua dificuldade. Por isso mesmo, é extramente importante saber quando é que o uso da negativa complica demais a questão, a ponto de prejudicar a sua qualidade. A julgar pelos resultados, os docentes mostraram ser capazes de identificar essas situações.

4.6. Comparação entre a perceção de alunos e docentes quanto aos erros de construção das PEM

Como seria de esperar, os alunos consideram que praticamente todos os problemas enunciados têm impacto negativo nos seus resultados académicos, ao passo que os docentes admitem ter poucos problemas em respeitá-los. Existem algumas correspondências entre aquilo a que os alunos atribuem mais impacto negativo e o que os docentes identificam como mais difícil de respeitar, mas vamos abordar apenas aquela que mais se destaca - ambiguidades relativamente à resposta correta. De facto, como já foi discutido, parece que os alunos sentem mais dificuldades quando a resposta correta não é identificável do que quando os distratores não são inquestionavelmente errados. Também já foi discutido que os docentes tendem a preocupar-se bastante com quão óbvia a resposta correta é. Dissimular a resposta correta entre os distratores muitas vezes coloca em causa a sua validade. Isto resulta numa interação viciosa, onde os docentes na tentativa de preservar a validade da pergunta acabam por redigi-la de uma forma que os alunos consideram prejudicial aos seus resultados.

5. Limitações do estudo

Uma das grandes limitações deste estudo é o tamanho reduzido da amostra de docentes, impossibilitando conclusões representativas e fidedignas da perceção deles.

Outra grande limitação são os questionários. Estes foram desenvolvidos pelos autores deste estudo, carecendo de validação e revisão, pelo que os seus resultados podem não ser os mais fidedignos.

Em relação à segunda secção do questionário dos alunos, por forma a evitar que os alunos conseguissem responder às perguntas com base em conhecimento, foram utilizadas perguntas genéricas, algumas delas com enunciados sem nenhuma interpretação possível. Isto fez com que as falhas que estavam a ser testadas se tornassem demasiado óbvias, o que pode ter condicionado a validade dos resultados obtidos (o ideal seria testar estas falhas em perguntas idênticas às presentes nas avaliações, contudo uma vez que o questionário era o mesmo para todos os anos, seria difícil incluir perguntas onde todos os alunos estivessem em pé de igualdade para responder).

Em relação à segunda secção do questionário dos docentes: num questionário de reduzidas dimensões, como é o caso (8 exemplos apenas), não só é impossível colocar exemplos de todas as falhas existentes, como também é impossível confrontar as diversas falhas umas com as outras. O confronto entre as falhas é extremamente importante, pois é aí que conseguimos ter a verdadeira noção de como é que os docentes as percebem. Além disso, neste questionário existem falhas que estão mais presentes do que outras (ex: há 3 perguntas onde se considerou que existem pistas/brechas que sugerem a resposta correta). Deste modo, temos de ter cuidado quando dizemos que os docentes identificam melhor uma falha do que outra, pois confrontá-los mais vezes com ela propicia a identifica-la mais. Em adição, numa investigação futura seria importante perceber que falhas foram mais identificadas pelos docentes no total de respostas às questões, para percebermos se existem falhas que apesar de não serem as que mais prejudicam a qualidade de uma dada questão, pela sua presença no global das questões acabam por ter um impacto considerável na fiabilidade da avaliação.

Por fim, estes resultados refletem apenas a opinião de quem respondeu aos questionários. É pois necessário, com base nisto, identificar os problemas que originam estas opiniões e desenvolver soluções para eles, sendo essencial reavaliar a opinião de docentes e alunos após a resolução destas questões.

6. Conclusão

O grande objetivo deste trabalho foi perceber a opinião de alunos e docentes face às dificuldades que as PEM (e os seus erros - que são mais frequentes do que pensamos(6)) representam para eles.

Como se constatou, os alunos consideram que perguntas com problemas na sua construção têm um impacto considerável no seu desempenho académico. Por outro lado, os docentes manifestam relativa facilidade em cumprir a grande maioria das orientações para a construção de PEM. Ainda assim, algumas das dificuldades manifestadas pelos docentes coincidem não só com alguns dos fatores aos quais os alunos atribuem mais impacto negativo, mas também com fatores dos quais os alunos se aproveitam para responder a PEM sem terem conhecimentos para isso. Isto exige, portanto, uma reflexão sobre estes fatores em particular - como evitar escrever perguntas com estas falhas, como reconhecer essas falhas nas perguntas já existentes, como tornar as avaliações mais justas e fiáveis tendo em conta que eliminar totalmente as falhas é impossível.

Em adição, a fiabilidade de qualquer instrumento de avaliação consiste em este dar os mesmos resultados, desde que utilizado nas mesmas circunstâncias. Temos portanto que ter em conta que os alunos não são todos iguais, pelo que perguntas que numa determinada avaliação obtiveram bons resultados (a nível de índice de discriminação, índice de dificuldade, entre outros(10)), poderão já ser desadequadas se aplicadas a alunos diferentes(20). É por isso importante uma análise constante e intensiva dos índices que nos indicam a qualidade das perguntas, uma revisão atenta das mesmas quando não cumprirem estes parâmetros(20) e um cuidado acrescido na uniformização do ensino proporcionado aos alunos.

Por fim, a avaliação de conhecimentos utilizando perguntas de escolha múltipla tem várias utilidades, mas não substitui de forma alguma qualquer outro método de avaliação. A integração de várias metodologias permite avaliações mais precisas e consistentes. Por isso mesmo, é importante haver mais estudos como este, onde se tenta perceber o impacto das perguntas de escolha múltipla e suas limitações, para que possamos colmatá-los e proporcionar aos alunos avaliações melhores e mais justas.

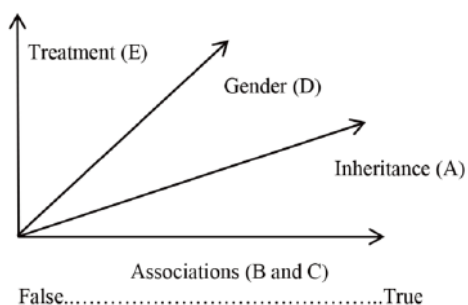
7. Referências Bibliográficas

1. National Board of Medical Examiners. Constructing Written Test Questions For the Basic and Clinical Sciences: Third Edition (Revised). 2002;181.
2. Considine J, Botti M, Thomas S. Design, format, validity and reliability of multiple choice questions for use in nursing research and education. Collegian [Internet]. 2005;12(1):19-24. Available from:
<http://www.sciencedirect.com/science/article/pii/S1322769608604783>
3. McCoubrie P. Improving the fairness of multiple-choice question1. McCoubrie P. Improving the fairness of multiple-choice questions: A literature review. Med Teach. 2004;26(8):709-12. s: A literature review. Med Teach. 2004;26(8):709-12.
4. Buckles S, Siegfried JJ. Using multiple-choice questions to evaluate in-depth learning of economics. J Econ Educ. 2006;37(1):48-57.
5. Palmer EJ, Devitt PG. Assessment of higher order cognitive skills in undergraduate education: Modified essay or multiple choice questions? Research paper. BMC Med Educ. 2007;7:1-7.
6. Downing SM. The effects of violating standard item writing principles on tests and students: The consequences of using flawed test items on achievement examinations in medical education. Adv Heal Sci Educ. 2005;10(2):133-43.
7. Downing SM. Construct-irrelevant variance and flawed test questions: Do multiple-choice item-writing principles make any difference? Acad Med. 2002;77(10 Suppl):S103-4.
8. Pham H, Besanko J, Devitt P. Examining the impact of specific types of item-writing flaws on student performance and psychometric properties of the multiple choice question. MedEdPublish. 2018;7(4):1-16.
9. Tarrant M, Ware J. Impact of item-writing flaws in multiple-choice questions on student achievement in high-stakes nursing assessments. Med Educ. 2008;42(2):198-206.
10. DiBattista D, Kurzawa L. Examination of the Quality of Multiple-choice Items on Classroom Tests. Can J Scholarsh Teach Learn. 2011;2(2):1-23.
11. Silveira FL da. Considerações sobre o índice de discriminação de itens em testes educacionais. Educ Seleção. 1983;7(July).
12. Cronback LJ. Coefficient alpha and the internal structure of tests. Psychometrika [Internet]. 1951;16(3):297-334. Available from:
<https://link.springer.com/content/pdf/10.1007/BF02310555.pdf>
13. Ali SH, Carr PA, Ruit KG. Validity and Reliability of Scores Obtained on Multiple-Choice Questions: Why Functioning Distractors Matter. J Scholarsh Teach Learn. 2016;16(1):1.
14. Board of Medical Examiners N. National Board of Medical Examiners Writing Multiple Choice Questions: An Introductory Tutorial. Available from:

- http://download.usmle.org/IWTutorial/MCQs_Intro_Tutorial.pdf
15. Harpe SE. How to analyze Likert and other rating scale data. *Curr Pharm Teach Learn* [Internet]. 2015;7(6):836-50. Available from:
<http://dx.doi.org/10.1016/j.cptl.2015.08.001>
 16. Taber KS. The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Res Sci Educ*. 2018;48(6):1273-96.
 17. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ*. 2011;2:53-5.
 18. Questions T. A Short Guide to Writing Effective Test Questions. Available from:
<https://www.k-state.edu/ksde/alp/resources/Handout-Module6.pdf>
 19. Karegar Maher MH, Barzegar M, Gasempour M. The Relationship between Negative Stem and Taxonomy of Multiple-Choice Questions in Residency Pre-Board and Board Exams. *Res Dev Med Educ* [Internet]. 2016;5(1):32-5. Available from:
<http://dx.doi.org/10.15171/rdme.2016.007>
 20. Haladyna TM, Downing SM, Rodriguez C. Applied Measurement in Education A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. *Appl Meas Educ*. 2002;15(3):309-333.

8. Anexos

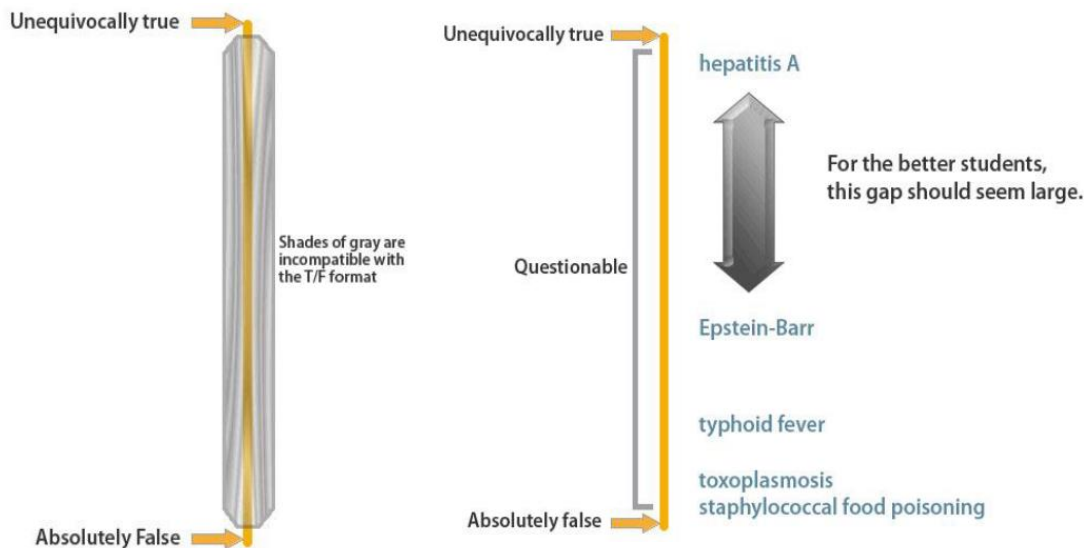
Anexo 1 - Exemplo de PEM retirada do manual do NBME e gráfico elucidativo das dimensões do conhecimento das alternativas de resposta



Which of the following is true about pseudogout?

- A. *It is clearly hereditary in most cases*
- B. *It is seldom associated with acute pain in a joint*
- C. *It may be associated with a finding of chondrocalcinosis*
- D. *It occurs frequently in women*
- E. *It responds well to treatment with allopurinol*

Anexo 2 - “Axis of truth”, disponível no tutorial online do NBME (ref)



Anexo 3 - Questionário dos alunos

Perguntas de Escolha Múltipla: Limites dos professores na sua construção e impacto no desempenho dos alunos.

O meu nome é Gonçalo Carneiro, sou aluno do 6ºano de Medicina e estou a realizar uma tese de mestrado sobre perguntas de escolha múltipla (PEM). O objetivo desta tese é avaliar em que medida os erros associados à construção (e não ao conteúdo científico) destas perguntas podem ter influência nas respostas dos alunos. O questionário é anónimo e confidencial. Os dados recolhidos serão apenas usados no presente trabalho e serão apresentados de forma completamente anónima e confidencial em apresentações públicas, congressos e publicações científicas.

Qualquer dúvida sobre o questionário ou o projeto de investigação pode ser colocada à orientadora do trabalho, Profª Doutora Isabel Neto (ineto@fcsaude.ubi.pt), ou ao investigador, Gonçalo Carneiro (a32552@fcsaude.ubi.pt).

Consentimento informado:

Li e compreendi a informação fornecida sobre o questionário que integra a presente investigação.

***Obrigatório**

1. Concordo em responder voluntariamente a este questionário. *

Ao selecionar a opção "Concordo" estará a indicar que tem 18 anos ou mais, que leu integralmente o presente termo de aceitação, que compreendeu os procedimentos/condições de participação e que participa de livre vontade neste questionário.
Marcar tudo o que for aplicável.

Concordo

2. Que ano curricular frequenta: *

Marcar apenas uma oval.

- 1º ano
 2º ano
 3º ano
 4º ano
 5º ano
 6º ano

Anexo 4 - Questionário dos alunos (cont.)

3. Em que nível sente que os seguintes fatores influenciam negativamente a resposta a perguntas de escolha múltipla? *

1-não tem influência; 2-influencia pouco; 3-indiferente; 4-influencia; 5-influencia muito
 Marcar apenas uma oval por linha.

	1	2	3	4	5
Pergunta não está de acordo com os objetivos de aprendizagem.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pergunta baseada em casos clínicos ou num problema.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
A pergunta está confusa não se percebendo o que se está a perguntar.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
A pergunta e/ou hipóteses de resposta estão mal formuladas.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
São utilizados termos não compreensíveis tendo em conta o ano frequentado.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
São usados termos inúteis/irrelevantes/redundantes para a compreensão da pergunta.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
A pergunta (enunciado e hipóteses) é demasiado longa e exige muito tempo de leitura.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
O vocabulário usado na pergunta é complicado e com imprecisões (ex: raramente, quase sempre, maioria das vezes).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hipóteses de resposta não estão de acordo com o enunciado da pergunta.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pergunta não cumpre as regras gramaticais e de pontuação.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pergunta está construída de forma induzir os alunos em erro (ex: utilização de charadas ou de combinações de letras ou números).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
O enunciado da pergunta está na negativa.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
A pergunta tem duplas negações (enunciado e hipóteses de resposta na negativa).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Não há distratores (respostas erradas) claramente incorretos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Há ambiguidades relativamente à resposta correta.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Perguntas de Escolha Múltipla: Limites dos professores na sua construção e impacto no desempenho dos alunos.

Responder a perguntas de escolha múltipla requer conhecimento científico, mas também experiência neste tipo de exames. Alunos mais experientes podem, por vezes, responder corretamente a questões sobre as quais não têm conhecimento, mas que apresentam erros de construção que dão indicação da resposta certa.

As perguntas seguintes não exigem qualquer conhecimento científico mas contêm erros de construção que permitem responder corretamente. Em relação a cada uma dela deverá selecionar a resposta que parece mais correta:

Anexo 5 - Questionário dos alunos (cont.)

1. Se houvesse aqui uma questão, qual seria a hipótese correta? *

Marcar apenas uma oval.

- Banido.
- Vitorioso.
- Conquistado.
- Derrotado.

2. A que país se refere esta questão? *

Marcar apenas uma oval.

- Nepal.
- Canadá.
- Grã-Bretanha.
- EUA.

3. A resposta correta aqui seria a: *

Marcar apenas uma oval.

- Soberano.
- Glicose.
- Vagabundo.
- Masculino.

4. Hixek norfolken piffle? *

Marcar apenas uma oval.

- Hiku nippon ibitus.
- Efil Yadlan Ruoj.
- Yokon Gnithol.
- Ikkek zippo unkerzotz notiaplan hipposlump enslife yolent.

5. Em que pifflerock se encontra zorkrans inkle? *

Marcar apenas uma oval.

- Em Hien.
- Em Inkle.
- Na Frankel.
- No Fronks.

6. A resposta ERRADA aqui é: *

Marcar apenas uma oval.

- Alguns indivíduos selecionados de forma randomizada.
- Alguns burocratas destacados.
- Alguns oficiais eleitos.
- O público em geral.

Anexo 6 - Questionário dos alunos (cont.)

7. Das seguintes, escolha a combinação correta: *

Marcar apenas uma oval.

- Melão, alho e maçã.
 - Limão, abóbora e pimento.
 - Alho, melão e laranja.
 - Melão, laranja e limão.
 - Alho, ananás e banana.
-

Anexo 7 - Questionário dos docentes

Perguntas de Escolha Múltipla: Limites dos professores na sua construção e impacto no desempenho dos alunos.

O meu nome é Gonçalo Carneiro, sou aluno do 6ºano de Medicina e estou a realizar uma tese de mestrado sobre perguntas de escolha múltipla (PEM). O objetivo desta tese é avaliar as dificuldades dos docentes na sua construção e os principais erros por eles cometidos. O questionário é anónimo e confidencial. Os dados recolhidos serão apenas usados no presente trabalho e serão apresentados de forma completamente anónima e confidencial em apresentações públicas, congressos e publicações científicas.

Qualquer dúvida sobre o questionário ou o projeto de investigação pode ser colocada à orientadora do trabalho, Profª Doutora Isabel Neto (ineto@fcsaude.ubi.pt.) ou ao investigador, Gonçalo Carneiro (a32552@fcsaude.ubi.pt).

Consentimento informado:

Li e compreendi a informação fornecida sobre o questionário que integra a presente investigação.

***Obrigatório**

1. Concordo em responder voluntariamente a este questionário. *

Ao selecionar a opção "Concordo" estará a indicar que tem 18 anos ou mais, que leu integralmente o presente termo de aceitação, que compreendeu os procedimentos/condições de participação e que participa de livre vontade neste questionário.

Marcar tudo o que for aplicável.

Concordo.

2. Elabora questões, maioritariamente, para que fase do curso de medicina? *

Marcar apenas uma oval.

Anos básicos (1º, 2º e 3º anos).

Anos clínicos (4º, 5º e 6º anos).

Anexo 8 - Questionário dos docentes (cont.)

3. Elaborar perguntas de escolha múltipla implica respeitar critérios e normas rigorosas. As mais importantes dessas regras encontram-se listadas em baixo. Indique, para cada uma delas, o grau de dificuldade que sente em respeitá-las, aquando da elaboração de uma PEM. *

1 – Muito difícil; 2 - pouco difícil; 3 – indiferente; 4 – fácil; 5 – Muito fácil

Marcar apenas uma oval por linha.

	1	2	3	4	5
A) Pergunta de acordo com os objetivos de aprendizagem.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
B) Pergunta que avalia a capacidade de o estudante aplicar conhecimento ou resolver um problema.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
C) Pergunta que avalia a capacidade do estudante memorizar factos ou um conjunto de matéria.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
D) Pergunta que pode ser respondida apenas pela leitura do enunciado.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
E) Formulação da pergunta está completa, bem como as hipóteses de resposta.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
F) Os termos utilizados na pergunta são compreensíveis para o estudante tendo em conta o nível em que se encontra.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
G) Não há termos inúteis na pergunta.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
H) Está claramente identificado o que se quer perguntar.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I) O vocabulário usado na pergunta é simples e preciso.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
J) Todas as hipóteses de resposta estão de acordo com o enunciado.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
K) Respeito pelas regras gramaticais e de pontuação.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
L) Pergunta construída de forma a não induzir os alunos em erro.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
M) A pergunta não está construída na negativa.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
N) A pergunta não contém dupla(s) negação ou negações (além de enunciado, contém também hipóteses de resposta na negativa).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
O) A pergunta não contém pistas/brechas de construção que podem sugerir a resposta correta.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
P) Homogeneidade na forma e conteúdo das hipóteses de resposta.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Q) Hipóteses de resposta lógicas e plausíveis.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
R) Os distratores (respostas erradas) são rigorosamente incorretos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
S) Não há qualquer ambiguidade sobre a validade da resposta correta.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Cada uma das perguntas de escolha múltipla que se seguem apresenta pelo menos uma falha principal/major na sua construção. Da lista abaixo, identifique o critério que

Anexo 9 - Questionário dos docentes (cont.)

considera NÃO estar a ser cumprido, colocando a respetiva letra no quadrado.

1. Em relação à Fibrose Quística, escolha a verdadeira: _____

- a) É uma doença autossómica recessiva;
- b) Crianças com esta doença geralmente morrem na adolescência;
- c) Homens com esta doença são inférteis;
- d) A incidência é 1:2000.

2. “Secondary gain” é: _____

- a) Uma complicação de uma variedade de doenças e que tende a prolongar muitas das mesmas;
- b) Um problema frequente nos doentes com POC;
- c) Um problema que nunca ocorre em doenças cerebrais;
- d) Sinónimo de queixas/sintomas fingidos.

3. A administração de furosemida resulta em: _____

- a) Diminuição do potássio urinário;
- b) Aumento do potássio urinário;
- c) Maior controlo da glicose nos doentes com diabetes tipo 2;
- d) Não altera o potássio urinário;
- e) Requer diminuição da dose na insuficiência renal.

Homem de 54 anos, alcoólico crónico e com historial de doenças psiquiátrica, está confuso e agitado. Diz que sente

4. que o mundo é “irreal”. Este sintoma chama-se: _____

- a) Despersonalização;
- b) Descarrilamento;
- c) Desrealização;
- d) Défice focal de memória;
- e) Ansiedade.

Anexo 10 - Questionário dos docentes (cont.)

Em relação às seguintes vacinas, tendo em conta a altura da administração da ÚLTIMA dose, ordene-as por sequência temporal (desde a que tem a sua última dose mais cedo, até à que a tem mais tarde):

I: BCG;
II: VHB;
III: Hib;
IV: DTPa;
V: Pn13.

- a) I, II, V, III, IV;
- b) I, II, III, IV, V;
- c) II, I, IV, III, V;
- d) V, I, II, IV, III;
- e) I, III, IV, II, V.

Paracelso continua a aceitar a existência da teoria dos quatro elementos, mas afirma que cada um é formado por três princípios:

- a) Mercúrio, enxofre, magnésio;
- b) Sódio, terra, éter;
- c) Enxofre, mercúrio, sal;
- d) Mercúrio, sal, água;
- e) Enxofre, zinco, potássio.

7. Qual das seguintes afirmações é correta:

- a) A hipófise anterior é constituída por axónios dos neurónios hipotalâmicos;
- b) A insulina é segregada pelas células A do pâncreas;
- c) O cortisol é uma hormona esteroide;
- d) A adrenalina atua no córtex supra-renal;
- e) A medula supra-renal é formada por três camadas histologicamente diferentes.

8. Sobre doenças genéticas, assinale a resposta ERRADA:

- a) Nenhuma delas surge de forma espontânea do portador;
- b) Algumas afetam mais homens do que mulheres;
- c) Nem todas podem ser passadas à descendência;
- d) Nem todas são dominantes;
- e) Algumas estão ligadas ao cromossoma Y.

Anexo 11 - Questionário dos docentes (cont.)

Marcar apenas uma oval.

- Pergunta de acordo com os objetivos de aprendizagem.
- Pergunta que avalia a capacidade de o estudante aplicar conhecimento ou resolver um problema.
- Pergunta que avalia a capacidade do estudante memorizar factos ou um conjunto de matéria.
- Pergunta que pode ser respondida apenas pela leitura do enunciado.
- Formulação da pergunta está completa, bem como as hipóteses de resposta.
- Os termos utilizados na pergunta são compreensíveis para o estudante tendo em conta o nível em que se encontra.
- Não há termos inúteis na pergunta.
- Está claramente identificado o que se quer perguntar.
- O vocabulário usado na pergunta é simples e preciso.
- Todas as hipóteses de resposta estão de acordo com o enunciado.
- Respeito pelas regras gramaticais e de pontuação.
- Pergunta construída de forma a não induzir os alunos em erro.
- A pergunta não está construída na negativa.
- A pergunta não contém duplas negações (além de enunciado, contem hipóteses de resposta na negativa).
- A pergunta não contém pistas/brechas de construção que sugiram a resposta correta.
- Homogeneidade na forma e conteúdo das hipóteses de resposta.
- Hipóteses de resposta lógicas e plausíveis.
- Os distratores (respostas erradas) são rigorosamente incorretos.
- Não há qualquer ambiguidade sobre a validade da resposta correta.

Anexo 12 - Chave da segunda secção do questionário dos alunos

Questão 1 - Vitorioso;

Questão 2 - Nepal;

Questão 3 - Glicose;

Questão 4 - Ikkek zippo unkerzotz notiaplan hipposlump enslife yolent;

Questão 5 - Em Inkle.

Questão 6 - O público em geral;

Questão 7 -Alho, melão e laranja.

Anexo 13 - Chave da segunda secção do questionário dos docentes

Questão 1 - “Os distratores (respostas erradas) são rigorosamente incorretos”;

Questão 2 - “Hipóteses de resposta lógicas e plausíveis” + “A pergunta não contém pistas/brechas de construção que podem sugerir a resposta correta”;

Questão 3 - “Homogeneidade na forma e conteúdo das hipóteses de resposta” + “O vocabulário usado na pergunta é simples e preciso”;

Questão 4 - “A pergunta não contém pistas/brechas de construção que sugiram a resposta correta”;

Questão 5 - “Pergunta construída de forma a não induzir os alunos em erro”;

Questão 6 - “A pergunta não contém pistas/brechas de construção que podem sugerir a resposta correta”;

Questão 7 - “Está claramente indicado o que se quer perguntar”;

Questão 8 - “A pergunta está construída na negativa” + “A pergunta não contém dupla(s) negação ou negações (além do enunciado, contém também hipóteses de resposta na negativa)”.