

SERIE: THE CHALLENGES IN UNDERGRADUATE RADIOLOGY EDUCATION

Analysis and technical aspects of image-related questions in the exam for Medical Internal Residence in Spain, 2022



E. Murias Quintana^{a,b,*}, S.M. Costilla García^a, J.J. Curbelo García^b, J. Calvo Blanco^a, J. Vega Villar^b, J. Baladrón Romero^b

^a Área de Radiología y Medicina Física, Universidad de Oviedo, Oviedo, Spain

^b Academia Curso Intensivo MIR Asturias, Oviedo, Spain

Received 25 April 2023; accepted 9 June 2023

Available online 19 September 2024

KEYWORDS

Psychometrics;
Radiology;
Difficulty level;
Medical entrance
examination;
Discrimination
(Psychology);
Multiple-choice
questions

Abstract The examination for the Medical Intern Resident (MIR) is a multiple-choice test aimed at ranking candidates for specialized medical training positions in Spain. The objective of this study is to provide an objective analysis of this test in its 2022 edition as an evaluative tool for discrimination, with a particular focus on the field of radiology and nuclear medicine. The clinical cases associated with radiology images or nuclear medicine pose greater difficulty compared to the rest of the MIR exam questions. Out of the 14 questions related to radiological or nuclear medicine images, six of them exhibit high difficulty, and only 5 out of the 14 questions demonstrate good or excellent discriminatory capacity. While the MIR exam proves to be an excellent discriminatory tool in psychometric terms, the image-related questions show a significant potential for improvement. In order for the image-associated question to exhibit appropriate discrimination, it is essential to minimize irrelevant information, ensure that it complements the clinical information provided in the text without contradicting it, represent the characteristic imaging finding of the disease, utilize the appropriate imaging modality, maintain a moderate difficulty level for the questions, and ensure that the distractors are clearly false.

© 2023 SERAM. Published by Elsevier España, S.L.U. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

* Corresponding author.

E-mail address: emuriass@hotmail.com (E. Murias Quintana).

PALABRAS CLAVE

Psicometría;
Radiología;
Nivel de dificultad;
Examen de acceso
médico;
Discriminación
(Psicología);
Preguntas de opción
múltiple

Análisis y aspectos técnicos de las preguntas asociadas a imágenes radiológicas en el examen para médico interno residente en España, convocatoria 2022

Resumen El examen para médico interno residente (MIR) es una prueba de respuesta múltiple que busca ordenar a los aspirantes para optar a una plaza de formación médica especializada en España. El objetivo de este trabajo es un análisis objetivo de dicha prueba en su convocatoria 2022 como herramienta evaluadora de discriminación, con especial interés en el área de la radiología y medicina nuclear. Los casos clínicos asociados a imágenes de radiología o de medicina nuclear presentan una dificultad mayor que el resto de las preguntas del examen MIR. De las 14 preguntas vinculadas a imágenes radiológicas o de medicina nuclear, seis de ellas presentan una alta dificultad y solo 5 de las 14 preguntas tienen una capacidad de discriminación buena o excelente. A pesar de que este examen MIR es una herramienta excelente de discriminación en términos psicométricos, las preguntas asociadas a imagen presentan una significativa capacidad de mejora. Para que la pregunta con imagen asociada presente una adecuada discriminación se debe reducir al mínimo la información inútil, debe complementar la información clínica del texto sin contradecirla, debe representar el hallazgo característico de imagen de la enfermedad y con la prueba de imagen adecuada, la dificultad de las preguntas no debe ser alta y los distractores deben ser claramente falsos.

© 2023 SERAM. Publicado por Elsevier España, S.L.U. Se reservan todos los derechos, incluidos los de minería de texto y datos, entrenamiento de IA y tecnologías similares.

Introduction

The medical specialty training entrance exam (MIR from its acronym in Spanish) is a multiple-choice test that ranks applicants, determining the order in which successful candidates will select a specialisation training post in Spain. Their final score combines the result of the exam (90%) and their academic grades (10%). This article concerns the MIR exam announced in 2022 and held on 21 January 2023.^{1,2}

Spain's Ministries of Health and Education have staged the MIR exam every year since 1978. Students across the country sit the same exam on the same day at the same time. From 2009 to 2018, candidates were given five hours to answer 225 multiple-choice questions and 10 multiple-choice reserve questions on any field of medicine. Each correct answer scored three points and each incorrect answer deducted one point. In 2019 and 2020, the number of questions was reduced to 175 plus 10 reserve questions, and candidates were given four hours to complete the exam. In the exams announced in 2021 and 2022, candidates were given four and a half hours to answer 200 questions and 10 reserve questions.

Successful candidates who obtain scores that exceed the cut-off mark—the minimum score required for access to a specialised medical training post—are then able to choose the speciality and hospital where they will undergo their specialty training. Yet not everyone who scores above the cut-off mark chooses to select a place. In the first adjudication process in 2020, the last vacancy chosen was taken by the candidate ranked 9,845th, and in 2021, the last post chosen was elected by the candidate ranked 9,931st, with 218 vacancies remaining unfilled, of which 200 were in general practitioner posts. In the second round of adjudication, the results did not fulfil expectations and only 124 of the 217

vacancies were filled, i.e., 93 posts were left vacant, all in general practice. In 2021, 272 diagnostic radiology posts were offered: the first person to select this specialty had been ranked 12th and the last person to select it had been ranked 4,409th.

While MIR exams have always included questions that directly relate to concepts in radiology, post-2009 MIR exams have included questions featuring one or more images. This image may be considered radiological—if it features image-based diagnostic tests—or non-radiological (if it features other images that include clinical photographs, histological images, diagrams, spirometry graphs, electrocardiograms, etc.)

For the purpose of this analysis, 'radiological concept' is defined as imaging test-based knowledge, from the fields of radiology or nuclear medicine, required by the candidate if they are to select the correct answer. This knowledge therefore includes indications for the tests, identification of anatomical structures, detection of abnormalities, assessment of severity, characterisation of lesions and any other relevant information obtained through imaging tests. This definition is considered to include radiological signs as well as anatomical and diagnostic knowledge.

The aim of this paper is to analyse the 2022–2023 MIR exam as an objective tool for knowledge assessment, with a special focus on the field of radiology and nuclear medicine. It is not the aim of this paper to analyse the method of study of the universities or MIR preparation centres, nor to consider its pros and cons within the Spanish training system, and the authors express no opinion on these topics. Neither is it the aim of this paper to analyse differences in medical training among universities or the different ways they prepare their students for the MIR exams.

Analysis of the 2022–2023 exam

The MIR exam announced in 2022 and held on 21 January 2023, had a total of 200 multiple-choice questions, each with four options, of which only one was considered to be correct, and 10 reserve questions. It consisted of 118,809 characters, an increase of 6,928 over the exam announced in 2021, which had the same number of questions.

With regard to the areas of knowledge evaluated, questions mostly referred to the five areas of endocrinology; neurology and neurosurgery; preventive medicine and biostatistics; rheumatology; and cardiology. In recent years, the number of questions per area has been highly uneven, with significant variations in the number of questions allocated to each area of knowledge. This means that the exam does not directly relate to the weight of each subject in the university curriculum, as it did in the early years of the exam.

The distribution of question types was similar to that of other years: 18.6% were typical test questions, 49.5% were clinical cases, 11.9% were image-based clinical cases and 20% were negative questions (by which we mean that the candidate had to identify the incorrect answer from among a set of answers). Ultimately, whether or not they included imaging, the bulk of the exam questions, a total of 129 out of 210, were clinical cases.

If we examine the items according to the concept type, the questions refer to the treatment of diseases (31.4%), diagnostic methods (23.8%), the clinical presentation of diseases (21%), pathophysiology (7.1%), aetiology (5.7%), and other concepts (11%).

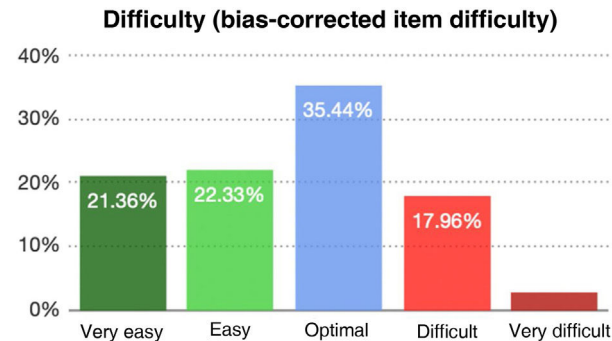
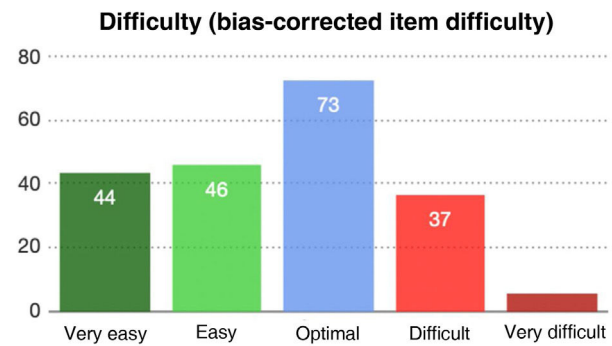
No bias towards concealment of the correct answer was found in this test. In other words, normally the two central options are the most common positions for the correct answer. In the 2022–2023 exam, a computer system was probably used to allocate the four options in a random and equal way, whereby each option was allocated 25% of the correct answers and approved by the marking commissions. Four questions were invalidated after the exam.

Analysis of questions linked to imaging

A total of 25 questions linked to imaging were included in this exam, appearing at the beginning of the exam paper. They can be divided into the following groups:

- 1 Eight CT images.
- 2 Five anatomical pathology and haematology images, including a blood smear and macro- and micropathology images.
- 3 Four clinical images.
- 4 Two questions with associated electrocardiogram images.
- 5 Two plain radiograph images: one of the thorax and one of bone.
- 6 One abdominal ultrasound image.
- 7 One magnetic resonance image.
- 8 One scintigraphy image.
- 9 One PET-CT image.

With regard to questions that contain images and radiological or nuclear medicine concepts, questions were asked on the following images:



Scale, summary by items, absolute and relative value				
Very easy	≥ 0.80	44	21.36%	
Easy	> 0.66-0.80	46	22.33%	
Optimal	> 0.33-0.66	73	35.44%	
Difficult	> 0.00-0.33	37	17.96%	
Very difficult	≤ 0.00	6	2.91%	
Total		206	100.00%	

Figure 1 Analysis of the difficulty of the 2022–2023 MIR exam according to classical test theory's bias-corrected item difficulty on the distribution of difficulty among the questions. Around 20% of the questions are difficult or very difficult and 35% have an optimal level of difficulty for an exam of this type.

- 1 Computed tomography for the diagnosis of a cerebral abscess located in the posterior fossa
- 2 Computed tomography for the diagnosis of supratentorial amyloid haemorrhage
- 3 Plain radiograph for the diagnosis of a fracture of the humeral head
- 4 Chest CT scan for the diagnosis of bronchiectasis
- 5 Abdominal ultrasound for the diagnosis of acute cholecystitis
- 6 Computed tomography for the diagnosis of urothelial neoplasia in the bladder
- 7 Positron emission tomography and computed tomography for staging primary lung neoplasia
- 8 Bone scintigraphy to diagnose the extent of a costal Ewing's sarcoma
- 9 Magnetic resonance image for the diagnosis of pituitary macroadenoma
- 10 CT scan for the diagnosis of acute appendicitis

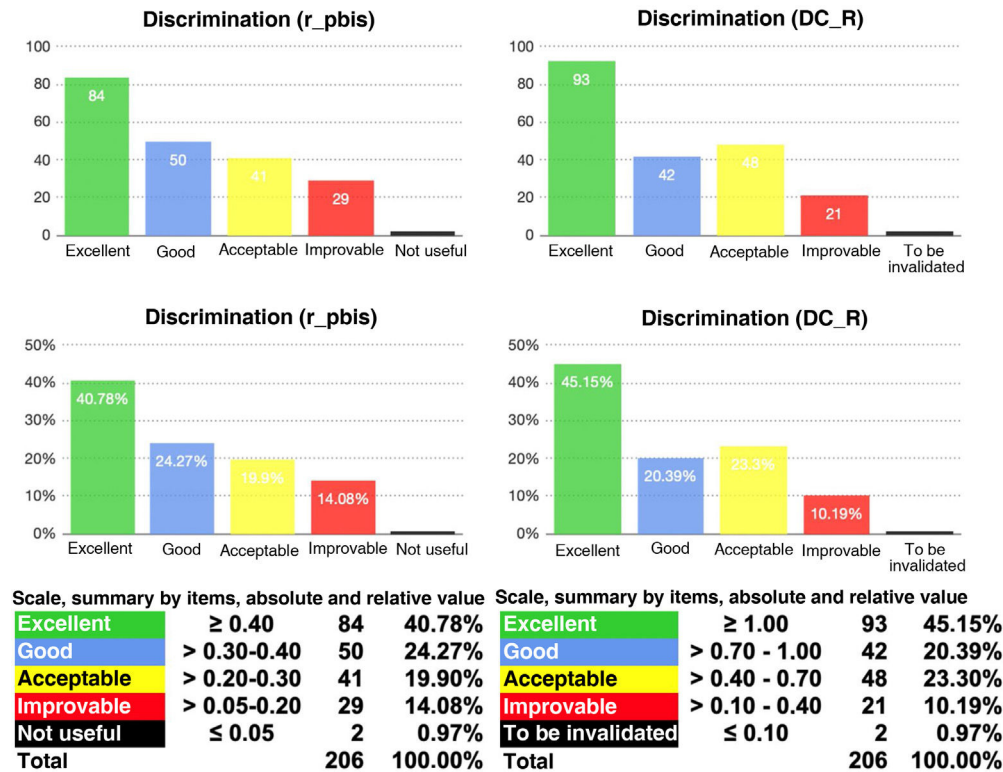


Figure 2 Distribution of discrimination capacity in the 2022–2023 MIR exam according to r_{pbis} (classical test theory) and DC.R discrimination (from item response theory). The MIR exam is a very good tool, fulfilling its discrimination function with only two questions flagged for replacement by the psychometric and 21 flagged for improvements.

- 11 Computed tomography for the diagnosis of chronic pancreatitis
- 12 Computed tomography for the diagnosis of a cervical spine metastasis
- 13 Computed tomography for the diagnosis of oesophageal achalasia
- 14 Plain chest X-ray for the diagnosis of pulmonary atelectasis

In addition to these 14 questions in the area of radiology and nuclear medicine, image-based diagnostic concepts were referred to throughout the exam in the remaining text-only questions. If we analyse these questions, there were 24 questions in which a key radiological concept was instrumental in finding the correct answer (11.4%) and approximately eight percent of the exam questions could only be answered with an understanding of the radiological concept. In the remaining questions, radiology and nuclear medicine complemented other clinical data in the question statement.

Looking at the 24 examination questions in which the radiological concept was essential to reach a diagnosis, the key concept in four questions was directly related to imaging signs, six questions were direct questions on the indication of a test based on the test characteristics, and 14 questions focused on the detection of an abnormality and its severity or extent, or the imaging features of a lesion.

Psychometric analysis of the 2022–2023 MIR exam and comparison with the two previous years

Psychometrics are the methods, techniques and theories involved in measuring and quantifying psychological variables of the human psyche. Psychometrics encompass the theory and construction of valid and reliable tests, examinations and other measurement procedures. This area includes the development and application of statistical procedures to determine whether or not a test is valid for the measurement of a previously defined psychological variable.

As in other works on the MIR exam published in *RADIOLOGÍA*,^{3,4} we have used measurement methods derived from both classical test theory (CTT) and item response theory (IRT), with formulas already used and contrasted in previous works analysing the MIR exam. These include the quantification of the difficulty of a question by correcting for the influence of random chance (bias-corrected item difficulty, from classical test theory), the discrimination ability of the question according to the point biserial correlation coefficient (r_{pbis} , from classical test theory) and the DC.R discrimination and difficulty measured according to item response theory (2 PL).^{5–9}

The psychometric analysis of a sample of 2022–2023 MIR exam answers revealed that the level of difficulty was average in relation to the previous four examinations. A total of 35.44% of the questions were of optimal difficulty, 2.91% were very difficult and 17.96% were difficult (Fig. 1).

QNo.	Answer	Subject	QuestionType	Bias-corrected item difficulty	r_pbis	DC_R
1	3	Otorhinolaryngology	CC with image	0.6008	0.1929	Poor
2	3	Neurology	CC with image	0.3751	0.0793	Poor
3	4	Traumatology	CC with image	0.5595	0.2664	Fair
7	4	Pulmonology	CC with image	0.7409	0.4170	Excellent
8	1	Gastroenterology	CC with image	0.5267	0.2494	Fair
9	4	Nephrology	CC with image	0.3231	0.1614	Poor
18	2	Pulmonology	CC with image	0.6046	0.3521	Good
19	4	Traumatology	CC with image	0.3048	0.1692	Poor
20	2	Pulmonology	CC with image	0.7042	0.2294	Fair
21	2	Gastroenterology	CC with image	0.4057	0.1794	Poor
22	4	Gastroenterology	CC with image	0.7801	0.3613	Good
23	3	Nephrology	CC with image	0.4088	0.2325	Fair
24	1	Gastroenterology	CC with image	0.2610	0.3204	Good
25	1	Pulmonology	CC with image	0.3428	0.3501	Good

Figure 3 The calculation of the corrected degree of difficulty (cDD) represents the correction of the percentage of students who get a question correct by random chance factor. Questions with values from -0.33 to 0 are considered to be very difficult; values between 0 and 0.33 are considered to be difficult; values between 0.33 and 0.66 are considered to be optimal; values between 0.66 and 0.80 are considered to be easy; and values between 0.80 and 1 are considered to be very easy. The point biserial correlation index (*r_pbis*) measures the discriminatory quality of the questions. The higher the *r_pbis* value, the greater the relationship between getting a high score on the test and having answered that particular question correctly. This index allows the discrimination capacity of questions to be classified as excellent (greater than or equal to 0.40), good (greater than or equal to 0.30 and less than 0.40), fair (greater than or equal to 0.20 and less than 0.30), poor (greater than or equal to 0 and less than 0.20), or terrible (negative).

The purpose of the MIR examination is not to grade the candidates based on their level of knowledge, but rather to rank them. Furthermore, the order of this ranking is then used for the adjudication of training posts. As such, discrimination between knowledge levels is one of the primary goals of the test, regardless of which questions are asked. Criticism of the exam suggests this idea is not clear and, in our opinion, is a cause of unwarranted controversy.

The discrimination of the 2022–2023 examination was good, confirming objectively that it fulfilled its function. Eighty-five percent of the test questions show excellent, good or acceptable *r_pbis* discrimination, with approximately 14% of questions flagged for revisions and only two of the 206 questions flagged for replacement by the psychometric, though not eventually disqualified by the marking committee (Fig. 2).

The purpose of the MIR examination is to distinguish between candidates and rank them to provide an order for the selection of specialty doctor training positions. The reduction in the number of MIR questions since the 2019–2020 examination, and modifications to the academic grade measurement scale, have multiplied the number of ties in total scores (exam + academic grade). In this respect, the current test design produces less discrimination between candidates than in the past.

One point in the academic grade is equivalent to around two net questions in the 2022–2023 exam. In the area of

maximum density of net scores in the distribution of results (between 120 and 129.67 net questions in the past MIR), a difference of one point in academic grade can lead to a difference of 276 places among candidates with the same exam score.

Psychometric analysis of radiological and nuclear medicine questions

Clinical cases involving radiology or nuclear medicine images are more difficult than other questions in the MIR exam and more difficult than clinical cases that do not contain images. Of the 14 questions related to radiological or nuclear medicine imaging, the bias-corrected difficulty of six items were classed as difficult or very difficult. Analysing the quality of the question according to its discriminating ability (*r_pbis*), only 5 of the 14 questions have a good or excellent discriminating capacity and thereby adequately fulfil their function in the test. Of the 31 questions with improvable discrimination or no discrimination, five (16%) contained radiological or nuclear medicine images (Fig. 3).

Adding an image to a clinical case usually increases the difficulty of the question, and incorporates multiple variables that may generate confusion among candidates. This is the main reason for a decrease in discrimination capacity.

It goes without saying that this does not mean that questions should not contain radiological or nuclear medicine images. The psychometric implication is that the technique for generating a good question that contains a radiological or nuclear medicine image is complicated and involves several technical factors that must be taken into account when writing these questions.¹⁰

The first recommendation is to *minimise unnecessary information*. The radiological or nuclear medicine image should not cause the candidate confusion by being irrelevant to the diagnosis when the diagnosis has already been inferred from the text.

Secondly, *the image should complement the clinical information in the text and not contradict it*. Each question has an objective, towards which the text and answers should lean and which the image should not contradict. The questions should present ideal patient clinical cases and the images should align with the text.

The third recommendation is to present a concrete image of the disease and, in cases of several pathological findings, confusion should be reduced by clarifying in the question statement which pathological findings are mistaken or unimportant. If a disease is principally diagnosed with a certain diagnostic test, this is the one that should be referred to in the question rather than another test that is used for differential diagnosis or to rule out other diseases. Thus *the radiological image should represent the characteristic imaging finding of the disease with the appropriate imaging test*. Otherwise, it generates confusion that does not favour the strongest candidates (those with scores in the top 27%), generates noise in scoring and impairs the discrimination capacity of the question.

On this point, it might be said that the MIR exam should reflect the day-to-day problems encountered by doctors, for example, in interpreting ambiguous, imprecise or contradictory findings. Although this may be the thinking of some

Question (11991): (2) Question containing image 2. Eighty-year-old patient, with dyslipidaemia, with no other known vascular risk factors and a one-year history of cognitive deterioration who suddenly presented with right hemiparesis on the day of admission. Code stroke was activated and the patient was transferred to the reference hospital, where on arrival a neurological examination was performed revealing right faciobrachiorural weakness (NIHSS scale 8) and a cranial CT scan was carried out, as shown in the image. Which of the following is the most likely diagnosis?



- 1 Cerebral infarction secondary to cocaine-type drug consumption.
- 2 Deep intraparenchymal haemorrhage associated with undiagnosed hypertension
- 3 Left parietal lobar haemorrhage associated with a possible amyloid angiopathy.
- 4 Intraparenchymal haemorrhage associated with a subarachnoid haemorrhage due to a possible traumatic brain injury.

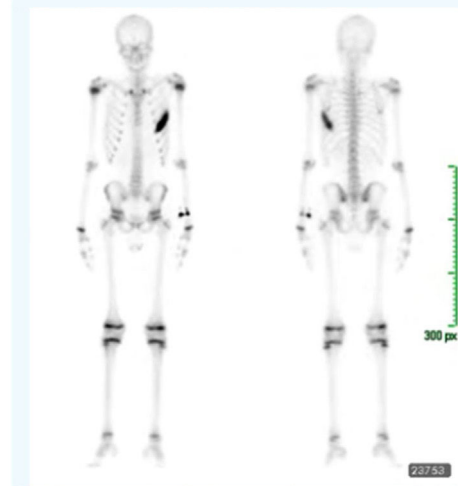
Bias-corrected item difficulty	0.3751	Optimal
R_PBIS	0.0793	Improvable
DC_R	0.01577	Improvable

Figure 4 Question 2, MIR 2022–2023 exam, correct answer 3. An example of a poorly differentiating and highly difficult question that does not serve to properly rank candidates according to their level of medical knowledge and only generates confusion and noise in the test scores. The question contains appropriate wording and a radiological image typical of the disease in question, and the aim of the question is to distinguish between ischaemic and haemorrhagic stroke and, more specifically, between the different types of haemorrhagic stroke. The possible error in question design that is likely the cause of its poor discrimination is an attempt to increase the difficulty of the answers and the distractors. The weak group (candidates whose scores are among the lowest 27%) interprets the image as a brain haemorrhage, but the strong group (candidates whose scores are among the highest 27% of the test) assesses the location of the image, and when the option selected as correct includes a location, it should be on no way ambiguous, and certainly not erroneous, in order to not generate errors that lead to low discrimination (r_{pbis} 0.08).

examiners, from a psychometric point of view, this only leads to noise in the distribution of marks and a lack of discrimination. This type of knowledge is more related to the ‘art’ of medicine and probably cannot be evaluated accurately with the MIR exam methodology.

The fourth technical factor to consider is difficulty. When the aim of the image-based question is to be as difficult as

Question (12008): (19) Question associated with image 19. Thirteen-year-old male with a painful tumour on left side. Chest CT scan confirms a lesion in the left anterior margin of the fifth rib with soft tissue mass and radiological data indicating aggressiveness. Laboratory tests showed no significant alterations. An ultrasound-guided biopsy was performed, compatible with a malignant rib tumour and its immunohistochemical profile was congruent with Ewing's sarcoma. A bone scintigraphy was performed to study the extent of the tumour, as seen in the image. Select the correct statement:



- 1 It is a normal bone scintigraphy.
- 2 Bone scintigraphy is not indicated in the follow-up study, as Ewing sarcoma does not develop bone metastases.
- 3 There is disseminated metastatic bone involvement
- 4 The only abnormal finding is the pathological accumulation of radiotracer in the known rib tumour.

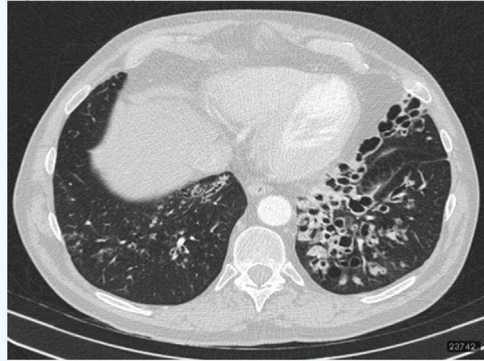
Bias-corrected item difficulty	0.3048	Difficult
r_{pbis}	0.1692	Improvable
DC_R	0.3270	Improvable

Figure 5 Question 19, MIR 2022–2023 exam, correct answer 4. Example of a question with suboptimal discrimination (r_{pbis} 0.17) and high difficulty. Again the wording is correct and the objective of the question is clearly defined: to get the candidate to identify the staging of a neoplasm using scintigraphy and the diagnosis of tracer accumulation in growing bones. The possible technical error in the question may be related to the confusion caused by the presence of tracer uptake at the puncture site in the image. This pseudo-uptake should have been mentioned and discarded in the statement, as it may cause confusion in the strong group that has identified the image on the left wrist. In addition, some of the answer options are long, not concrete and do not refer directly to the text or the image, but rather include a general reference to the disease in an attempt to increase the difficulty or create a confusing distractor. The overall consequence is that the question has a poor discrimination capacity.

possible, differences in discrimination are not progressive or linear, as a critical point is reached from which discrimination drops significantly. Thus, *the difficulty of the questions should not be high*, and it is better to opt for other tools to increase the discrimination capacity of the test, such as increasing the number of questions.

The final thing to consider is the distractors. We do not discriminate effectively by tweaking distractors so they differ only slightly from the correct answer. In other words, the distractors must be clearly false. With this technique, the

Question (11996): (7) Question associated with image 7. Sixty-year-old male, active smoker, presenting with long-term daily cough and expectoration, occasionally haemoptotic. In recent months, progressive dyspnoea has been added. He had nail clubbing while pulmonary auscultation revealed rhonchi and end-inspiratory wheezing in the left lung. The high-resolution lung CT scan is shown in the attached image. What is the most likely diagnosis?



1. Cystic carcinoma
2. Pulmonary emphysema
3. Cavitated tuberculosis
4. Bronchiectasis

Bias-corrected item difficulty	0.7409	Easy
r_pbis	0.4170	Excellent
DC_R	1.1247	Excellent

Figure 6 Question 7, MIR 2022–2023 exam, correct answer 4. Example of a good question in technical terms. The question wording talks about the typical clinical manifestations of the disease. The text defines the test performed and its indication in concrete terms. The image is typical and appropriate for the disease. The distractors are correct, concrete and devoid of pathologies with confusingly similar imaging. The result is a question with a good discrimination capacity.

strongest candidates understand these nuances as plausible possibilities and this causes them to answer the item incorrectly more than the weaker group, leading to a decrease in the discrimination capacity of the question (Figs. 4–6). Formulating an image-based question for the MIR exam is not an easy task, and questions need to be written with the right technique to fulfil their purpose. The MIR is not a teaching or knowledge assessment exam, nor does it provide candidates with any feedback. Its sole function is to rank the candidates.

Conclusions

Overall, the discrimination capacity of the 2022–2023 MIR exam was good, which makes it a precise, useful and effective tool. In recent years, radiology and nuclear medicine questions have often been among the most difficult and least discriminating in the exam. The exam, as a whole, discriminates most heavily among those candidates with the lowest marks. For images to be suitable, they should keep unnecessary information to a minimum; complement the clinical information in the text without contradicting it; and represent the characteristic imaging finding of the disease with the appropriate imaging test. Moreover, the difficulty of the questions should not be high, and any distractors should be clearly false. One way of improving the discrimination

capacity of questions that contain radiological images in the MIR exam would be to include a greater number of this type of questions.

Ethical responsibilities

The data obtained for this study were obtained from public information published by the Ministry of Health and the Ministry of Education. Some of the data were obtained from the candidates themselves, and we sought consent to use such data. Notwithstanding, in no part of the article of study can any of the candidates be identified or located, and all the information has been treated in an anonymised form, with no personal data.

Funding

The authors have received no funding of any kind for the development of this study.

Author contributions

Eduardo Murias Quintana is responsible for the study concept and design, data collection, analysis and interpretation, and the drafting of the article.

The rest of the authors have critically reviewed the intellectual content and have given their final approval to the version presented here.

Conflicts of interest

The authors declare they have no conflicts of interest associated with the completion or publication of this study.

References

- Real Decreto 127/1984, de 11 de enero, por el que se regula la formación médica especializada y la obtención del título de médico especialista. BOE núm. 26, de 31 de enero de 1984; p. 2524–2528.
- Programas de Formación Sanitaria Especializada. Ministerio de Sanidad, Servicios Sociales e Igualdad. Available from: <https://fse.mscbs.gob.es/fseweb/view/index.xhtml> (access 24 de abril de 2023).
- Murias E, Sánchez-Lasheras F, Fernández-Somoano A, Romeo JM, Baladrón J. Análisis de la elección de la especialidad de radio-diagnóstico en el examen MIR desde el año 2006 hasta 2015. *Radiología*. 2017;59:232–46.
- Murias E, Sánchez F, Costilla SM, Cadenas M, Calvo J, Baladrón J. Psychometric analysis of questions associated with radiological images in the competitive examination for access to residency programs in Spain. *Radiología*. 2019;61:412–29.
- Rodríguez A, Martínez F. Aplicaciones informáticas de psicometría en investigación educativa. *Comunicar*. 2003;21:163–6.
- Baladrón J, Sánchez-Lasheras F, Romeo JM, Curbelo J, Fonseca P. Evolución de los parámetros dificultad y discriminación en el ejercicio de examen MIR. Análisis de las convocatorias de 2009 a 2017. *FEM*. 2018;21:181–93.
- Baladrón J, Sánchez-Lasheras F, Villacampa T, Romeo-Ladrero JM, Fernández-Somoano A. Propuesta metodológica para la detección de preguntas susceptibles de anulación en la

- prueba MIR. Aplicación a las convocatorias 2010 a 2015. FEM. 2017;20:161–75.
8. Baladron J, Sánchez-Lasheras F, Villacampa T, Romeo-Ladrero JM, Fernández-Somoano A. El examen MIR 2015 desde el punto de vista de la teoría de respuesta al ítem. FEM. 2017;20:29–38.
 9. Baladrón J, Curbelo J, Sánchez-Lasheras F, Romeo-Ladrero JM, Villacampa T, Fernández-Somoano A. El examen al examen MIR 2015: aproximación a la validez estructural a través de la teoría clásica de los tests. FEM. 2016;19:217–26.
 10. Soler Fernández R, Méndez Díaz C, Rodríguez García E. Continuing medical education: how to write multiple choice questions. Radiologia. 2013;55 Suppl 1:S28–36.