

CARACTERÍSTICAS MÉTRICAS DEL EXAMEN ÚNICO DE INGRESO A RESIDENCIAS MÉDICAS 2020: LA DISTORSIÓN DEL PROMEDIO DE LA CARRERA COMO PARÁMETRO.

METRIC CHARACTERISTICS OF THE 2020 SINGLE ENTRANCE EXAM FOR MEDICAL RESIDENCES: THE DISTORTION OF THE CAREER AVERAGE AS A PARAMETER.

Autores:

Pablo Badr, Magister en Salud Colectiva, Universidad Nacional del Sur, Argentina.

Pedro Silberman, Magister en Salud Colectiva, Universidad Nacional del Sur, Argentina.

Martín Silberman, Doctor en Salud Pública, Universidad Nacional Arturo Jauretche, Hospital de Alta Complejidad en Red, El Cruce, Néstor Kirchner, Argentina.

Resumen

INTRODUCCIÓN:

Las residencias del equipo de salud son la alternativa más elegida entre quienes se gradúan de las carreras de medicina para realizar su etapa de especialización. En la Argentina en 2019 se constituye el Examen Único de Ingreso a Residencias (EU) con la participación de 23 provincias (menos Córdoba). Previo al 2019 se realizaban exámenes por jurisdicción (CABA, PBA, etc.) con instrumentos heterogéneos en términos de confiabilidad y de dificultad. Este tipo de exámenes tienen efectos en distintos entornos con complejas interacciones en las esferas política, económica, social, humana y educativa. El EU consiste en 100 preguntas de selección múltiple y está elaborado a partir de una estructura por área de incumbencia médica. El objetivo del presente trabajo es analizar el EU haciendo énfasis en el índice de dificultad del examen, índice de discriminación, análisis de confiabilidad, validez y la correlación mono rasgo/heterométrico.

METODOLOGÍA:

Se utilizó la base de datos correspondiente a las respuestas del examen único del año 2020, la cual incluye 5243 aspirantes a residencias médicas quienes se agruparon en 3 categorías: egresados/as de universidades nacionales públicas, universidades nacionales privadas y universidades extranjeras. Las preguntas del examen fueron agrupadas en 4 bloques temáticos: Salud de la Mujer (23 preguntas), Salud del Adulto (46 preguntas), Salud del Niño/a (20 preguntas) y Salud Pública (11 preguntas). Se cuantificaron para este trabajo los siguientes

parámetros: Índice de Dificultad del examen, Índice de Discriminación, Análisis de Confiabilidad y Análisis de Validez: Correlación mono-rasgo/heterométrico.

RESULTADOS:

El examen fue realizado por 5243 postulantes, 1564 provienen de universidades extranjeras (29,8%), 813 de universidades nacionales privadas (15,5%) y 2866 de universidades públicas (54,6%). El 82% de las preguntas tiene una dificultad aceptable, y el 13 % una dificultad óptima. En cuanto a la discriminación, 63 % de los ítems tiene excelente o buen grado de discriminación, mientras que el 6 % tiene un grado de discriminación muy malo. Para el análisis de Confiabilidad se utilizó el coeficiente de fiabilidad según la fórmula 21 de Kuder-Richardson arrojando un valor de 0.85. El promedio de las Universidades privadas fue más alto que las Públicas (7,12 versus 7,02); mientras que el desempeño en el examen fue inverso (63,3 para las públicas contra 48,03 en las privadas).

CONCLUSIONES:

En conclusión, el instrumento tiene propiedades adecuadas para discriminar un orden de mérito: Alta confiabilidad, validez y capacidad de discriminación. La utilización del promedio para los concursos desconociendo los parámetros que los determinan, introduce un sesgo en los resultados del mismo a favor, en este caso, de los/las egresados/as de las universidades de gestión privada.

PALABRAS CLAVES:

Internado y Residencia; Capacitación Profesional

Disponible en:
<http://repositorio.hospitalelcruce.org/xmlui/handle/123456789/1301>

Abstract

INTRODUCTION:

The residences of the health team are the most chosen alternative among those who graduate from medical careers to carry out their specialization stage. In Argentina, in 2019, the Unique Residence Entrance Exam (EU) was established with the participation of 23 provinces (except Córdoba). Prior to 2019, exams were carried out by jurisdiction (CABA, PBA, etc.) with heterogeneous instruments in terms of reliability and difficulty. These types of exams have effects in different environments with complex interactions in the political, economic, social, human and educational spheres.

OBJECTIVE:

The EU consists of 100 multiple-choice questions and is prepared from a structure by area of medical concern. The objective of this paper is to analyze the EU, emphasizing the test difficulty index, discrimination index, reliability analysis, validity, and the monotrait/heteromethod correlation.

METHODS:

The database corresponding to the responses of the single exam of the year 2020 was used, which includes 5243 applicants for medical residences who were grouped into 3 categories: graduates of national public universities, national private universities and foreign universities.

The exam questions were grouped into 4 thematic blocks: Women's Health (23 questions), Adult Health (46 questions), Child

Health (20 questions) and Public Health (11 questions).

The following parameters were quantified for this work: Test Difficulty Index, Discrimination Index, Reliability Analysis and Validity Analysis: Monotrait/heteromethod correlation.

RESULTS:

The exam was taken by 5,243 applicants, 1,564 from foreign universities (29.8%), 813 from national private universities (15.5%) and 2,866 from public universities (54.6%). 82% of the questions have an acceptable difficulty, and 13% an optimal difficulty. Regarding discrimination, 63% of the items have an excellent or good degree of discrimination, while 6% have a very bad degree of discrimination. For the Reliability analysis, the reliability coefficient was used according to the Kuder-Richardson formula, yielding a value of 0.85. The average of the private Universities was higher than the Public ones (7.12 versus 7.02); while the performance in the exam was inverse (63.3 for the public ones against 48.03 in the private ones).

CONCLUSIONS:

The instrument has adequate properties to discriminate an order of merit: High reliability, validity and discrimination capacity. The use of the average for the competitions, ignoring the parameters that determine it, introduces a bias in the results of the same in favor, in this case, of the graduates of private management universities.

KEYWORD:

Internship and residency; Professional Training

INTRODUCCIÓN

Las residencias del equipo de salud son la alternativa más elegida entre quienes se gradúan de las carreras de medicina para realizar su etapa de especialización (1). Se trata de un sistema destinado a graduados recientes, rentado y con dedicación exclusiva, al que se accede por concurso abierto.

El sistema de residencias, en particular las residencias médicas—en sus distintas modalidades nacionales, universitarias, provinciales, municipales y privadas—ofrece en conjunto cerca de 6500 vacantes para unos 7500 médicos/as inscriptos/as anuales (2).

En la Argentina, en el año 2019 y producto de un proceso comenzado en 2011, se constituye el Examen Único de Ingreso a Residencias (EU) (Resol.186/2019) con la participación de 23 provincias (con la excepción de Córdoba¹) y la constitución de un comité técnico federal².

El EU puede ser rendido por aspirantes de cualquier provincia, incluso bajo diferentes modalidades, por ejemplo, en el año 2020 debido a la pandemia COVID³ y dado el aislamiento obligatorio decretado por el gobierno nacional (Decreto DNU 297 / 2020), el examen se realizó en formato digital y pudo ser rendido desde los respectivos hogares(2).

Previo al 2019 se realizaban exámenes por jurisdicción (CABA, PBA, etc.) con instrumentos heterogéneos en términos de confiabilidad y de dificultad. A partir de dicho año se organiza

un único examen (EU) que tiene como objetivo disminuir la heterogeneidad, proveer una mayor cobertura de vacantes y hacer más eficiente la gestión¹(3)(4).

Los exámenes, como el EU, para acceder a la formación en residencias médicas, y en escalas nacionales, son conocidos como “exámenes de alto impacto” y “exámenes de gran escala”⁴ (5)(6). Este tipo de exámenes tienen efectos en distintos entornos con complejas interacciones en las esferas política, económica, social, humana y educativa (7)(8)(9)(10).

El EU consiste en 100 preguntas de selección múltiple y está elaborado a partir de una estructura por área de incumbencia médica (Tabla de especificaciones) (11).

La nota resultante del mismo, junto al promedio individual de la carrera de grado, genera un orden que define la prioridad del/la egresado/a para la elección de especialidad y lugar donde realizarla, en el caso del Concurso Unificado⁵ (Resol 250/2019 Anexo I). Para el resto de las jurisdicciones que no están dentro del Concurso Unificado, se incorporan diferentes requisitos⁶.

La inclusión del promedio individual de la carrera de grado como elemento en la definición del orden de prioridad es motivo de debate

4 *Se define un examen de alto impacto como aquel cuyo resultado tiene consecuencias importantes y directas para los individuos, programas o instituciones involucrados en el examen*

5 *El CU es el concurso que integra las instituciones nacionales, de la provincia de Buenos Aires, y Ciudad Autónoma de Buenos Aires, y te permite acceder a las residencias de salud de especialidades básicas y posbásicas (y concurrencias, en el caso de CABA) de las tres jurisdicciones. <https://www.argentina.gob.ar/salud/residencias/ingreso-residencias/concurso-unificado#1>*

6 *<https://www.argentina.gob.ar/salud/residencias/ingreso/requisitos-y-componentes-de-los-concursos-de-residencias>*

1 *Córdoba realizó su examen antes del inicio del ASPO decretado a partir de la Pandemia Covid-19 y quedó desfasado del examen único que se postergó a agosto del año 2020*

2 *<https://www.argentina.gob.ar/normativa/nacional/resoluci%C3%B3n-186-2019-319575/texto>*

3 *<https://www.who.int/es/news/item/29-06-2020-covidtimeline>*

dada su variabilidad y su potencial sesgo.

Es importante que el EU pueda discriminar entre los aspirantes, de manera que permita identificar los mejores desempeños; que sea válido, de manera que mida aquello que se propone medir (conocimientos por áreas) y que tenga un nivel de dificultad razonable, es decir, que tenga relación con los conocimientos de un egresado reciente (12).

Existe escasa bibliografía en nuestro medio acerca de la métrica de este tipo de exámenes y son de bajo nivel de evidencia (13). Los trabajos al respecto muestran defectos en este sistema de exámenes que comprometerían la confiabilidad, validez y legitimidad, generando inequidades en el ingreso a las residencias médicas, por ejemplo, el análisis realizado sobre el ENARM⁷ en México (14).

Por ello, el objetivo del presente trabajo es analizar el EU haciendo énfasis en el índice de dificultad del examen, índice de discriminación, análisis de confiabilidad, validez y la correlación mono rasgo/heterométrico.

METODOLOGÍA

Se utilizó la base de datos correspondiente a las respuestas del examen único del año 2020, la cual incluye 5243 graduados y graduadas que aspiraron a residencias médicas básicas. Los y las aspirantes se agruparon según la universidad en la que finalizaron su carrera de grado en 3 categorías: egresados/as de universidades nacionales públicas, universidades nacionales privadas y universidades extranjeras.

Las preguntas del examen fueron agrupadas en 4 bloques temáticos: Salud de la Mujer (23 preguntas),

Salud del Adulto (46 preguntas), Salud del Niño/a (20 preguntas) y Salud Pública (11 preguntas).

Se cuantificaron para este trabajo los siguientes parámetros:

- Índice de Dificultad del examen
- Índice de Discriminación
- Análisis de Confiabilidad
- Análisis de Validez: Correlación mono-rasgo/heterométrico

El **Índice de Dificultad** de un ítem es la proporción del total de sujetos que aciertan del total que intentan resolverlo. Para realizar este cálculo, y con el objetivo de poder realizar comparaciones con trabajos similares, se realizó el análisis comparativo de grupos de candidatos alto-bajo (A-B), comparando el grupo alto como el ubicado en 27% superior del desempeño del examen, con el grupo bajo, que incluye al 27% inferior (15).

Para calcular el Índice de Dificultad, se utilizará la siguiente fórmula (15):

$$ID = \frac{A + B}{N_1 + N_2}$$

donde A es el número de aciertos del grupo alto, y B el número de aciertos del grupo bajo, N1 el total de intentos de acertar del grupo alto, y N2 el total de intentos de acertar del grupo bajo. Considerando el supuesto que el grupo bajo, al tener menos conocimientos, puede presentar menor cantidad de respuestas que los del grupo alto, se corrige el desequilibrio agregando a B el grupo de estudiantes del grupo bajo que no respondieron y responderían correctamente en el caso que hubieran contestado cada ítem, y cuya fórmula se detalla a continuación (16):

$$FC = (N_1 - N_2) \times \frac{B}{N_2}$$

⁷ ENARM: examen nacional para aspirantes a residencias medicas <https://enarm.com.mx/enarm.php>

En función del valor obtenido, cada ítem se clasifica en grado de dificultad con los puntos de corte descriptos en la tabla 1, a fin de comparar los resultados con investigaciones similares (16)

Parámetros del nivel de dificultad	
Porcentaje	Nivel
Menos de 10%	Muy difícil
10 a 30%	Difícil
30,1 a 50%	Relativamente difícil (aceptable)
50,1 a 60 %	Dificultad adecuada (óptima)
60,1 a 80%	Relativamente fácil (aceptable)
Mayor a 80,1%	Muy fácil

Tabla 1.
 Clasificación de ítems según índice de dificultad

El **índice de discriminación** se define como la correlación entre las puntuaciones de los sujetos en el ítem y sus puntuaciones en el test, y para el cual se utilizará la siguiente fórmula (16):

$$DS = 2 \times \frac{A - B}{N_1 + N_2}$$

El índice de discriminación, con esta fórmula, puede tomar valores entre -1 y +1. La clasificación de las preguntas según su índice de discriminación se llevará a cabo teniendo en cuenta los valores de la tabla 2 (16).

Índice de discriminación	
IC	Nivel
> 0,34	Excelente
> 0,24 a 0,34	Bueno
> 0,14 a 0,24	Revisable
0 a 0,14	Malo
< 0	Muy malo

Tabla 2.
 Clasificación de ítems según índice de discriminación

Al igual que en el caso del índice de dificultad, el índice de discriminación se ha calculado utilizando el factor de corrección.

La confiabilidad se define como la consistencia con la que el examen es capaz de medir una variable dada. Para analizar la confiabilidad en exámenes de respuestas dicotómicas, se utilizará la fórmula 21 de Kuder-Richardson (17):

$$KR_{21} = \frac{K}{K - 1} \times \left(1 - \frac{M \times (K - M)}{K \times S^2} \right)$$

donde K es el número de preguntas del test; M, la media aritmética de las puntuaciones obtenidas por todos los examinandos (en este caso se utilizará el número de respuestas acertadas por cada examinando); y S, la desviación estándar de la media.

Para el **análisis de validez**, se realizó un análisis multi rasgo/multi método, en este caso monorrasgo/heterométrico (18). Bajo el supuesto que las carreras de medicina presentan una formación en los contenidos que luego se evalúan en el EU, y que dicho aprendizaje se refleja en el promedio de cada graduado/a, se realizó un análisis de correlación entre el promedio del estudiantado y el desempeño en el EU. Considerando las diferencias observadas en promedios de los distintos grupos de carreras (públicas, privadas y extranjeras) la correlación mono-rasgo/heterométrico se realizó en cada uno de estos subgrupos (18) (19).

RESULTADOS

El examen único para el ingreso a residencias médicas del año 2020 estuvo compuesto por 100 preguntas, las cuales fueron agrupadas según bloque temático de acuerdo a la tabla de especificaciones.

El examen fue realizado por 5243 postulantes, de los cuales 1564 provienen de universidades

extranjeras (29,8%), 813 de universidades nacionales privadas (15,5%) y 2866 de universidades nacionales públicas (54,6%). El resultado del examen para las respuestas correctas fue el siguiente:

N	Válido	5243
Puntaje promedio		58,14
Mediana de puntaje obtenido		59,00
Moda del puntaje obtenido		60
Desv. Desviación		12,473
Varianza		155,564
Rango		70
Mínimo		20
Máximo		90

Tabla 3.
Variables estadísticas de los resultados obtenidos del EU

Índice de Dificultad:

La clasificación de las preguntas por grado de dificultad muestra que el 82% de las preguntas tiene una dificultad aceptable, y el 13 % una dificultad óptima. (Tabla 4)

Índice de discriminación:

En cuanto a la discriminación, 63 % de los ítems tiene excelente

o buen grado de discriminación, mientras que el 6 % tiene un grado de discriminación muy malo. (Tabla 5)

Confiabilidad:

Para el análisis de Confiabilidad se utilizó el coeficiente de fiabilidad según la fórmula 21 de Kuder-Richardson arrojando un valor de 0.85

Validez:

Correlación mono rasgo/hetero método.

Para considerar la correlación entre el promedio de la carrera y el desempeño en el examen se agrupó a los aspirantes en tres grupos según tipo de carrera en la que se graduaron. En los gráficos 1 y 2 puede observarse la distribución de ambas variables. En el caso del promedio de la carrera, es mayor en el caso de estudiantes de carreras nacionales privadas. Esto no se corresponde a un comportamiento similar en el desempeño del examen, donde el grupo de aspirantes de universidades nacionales públicas tiene una media, primer y tercer cuartil, superiores a los otros dos grupos. Para cada grupo de carreras

Grado de dificultad	Examen total	Bloque Salud del Adulto	Bloque salud de la mujer	Bloque Salud del Niño	Bloque Salud Pública
Muy difícil	3 %	2 %	4,3 %	4,5 %	–
Difícil	9 %	11 %	13 %	4,5 %	–
Relativamente difícil	31 %	26 %	43,5 %	36,4 %	27,3 %
Dificultad óptima	13 %	17 %	13,9 %	4,5 %	9,1 %
Relativamente fácil	38 %	37 %	26,1 %	45,5 %	45,5 %
Muy fácil	6 %	7 %	–	4,5 %	18,2 %
Media de Dificultad	0,53				

Tabla 4.
Clasificación de preguntas según índice de dificultad

Tabla 5.
Clasificación de preguntas según índice de discriminación

Grado de discriminación	Examen total	Bloque Salud del Adulto	Bloque salud de la mujer	Bloque Salud del Niño	Bloque Salud Pública
Excelente	41 %	37 %	56,5 %	40,0 %	27,3 %
Bueno	22 %	21,7 %	13,0 %	20,0 %	45,5 %
Revisable	19 %	23,9 %	13,0 %	20,0 %	9,1 %
Malo	12 %	15,2 %	4,3 %	15,0 %	9,1 %
Muy malo	6 %	2,2 %	13,0 %	5,0 %	9,2 %

se calculó la correlación entre promedio de carrera y desempeño del examen, mediante el coeficiente de correlación *r* de Pearson, siendo para los tres grupos moderadamente positiva. La correlación fue menor en el grupo de las universidades extranjeras, mientras que fue similar para los dos grupos de universidades nacionales (Tabla 6).

DISCUSIÓN:

Los resultados obtenidos en este estudio demuestran que el Examen Único presenta características psicométricas aceptables para una

0.56; en EU: 0.79 y 0.58, y en PBA: 0.83 y 0.55, respectivamente (4). Los resultados de este trabajo muestran similitudes en confiabilidad (0.85) y dificultad media (0.53). El examen MIR (ESPAÑA) presenta un índice de confiabilidad superior, de 0,96 y un índice de dificultad de 0,54 (16), similar al de nuestro análisis.

El índice de discriminación es bueno o excelente en el 63 % de los ítems, mientras que el MIR ha presentado un 49 % de ítems con estas categorías.

En cuanto a la correlación entre promedio de la carrera y desempeño en el examen, si bien

Tabla 6.
Promedio de carrera y desempeño en EU según tipo de universidad

	Aspirantes de universidades extranjeras	Aspirantes de universidades privadas	Aspirantes de universidades públicas
Promedio carrera de grado			
Media	7,12	7,50	7,02
Desvío estándar	1,31	0,93	0,99
Desempeño en el examen			
Media	48,03	59,31	63,35
Desvío estándar	10,53	11,24	10,23
Correlación Promedio/desempeño	0,44	0,62	0,60

instancia decisoria como lo es el acceso a la formación en residencias. En un trabajo comparativo de tres exámenes previos al EU en Argentina (EU, CABA y PBA) se observó una diferencia en la confiabilidad y el índice de dificultad: en CABA: 0.86 y

en las poblaciones de egresados y egresadas de universidades nacionales privadas y públicas es similar, en el caso de los promedios presentan una media, superior en las carreras privadas, mientras que en el caso de la población proveniente

de carreras públicas es mayor la media, del desempeño del examen. Este resultado, si bien aporta a la validez del examen, sugiere que la construcción de los promedios es diferente según la naturaleza de las carreras y su incorporación en la construcción de órdenes de mérito aportaría una varianza distorsiva, denominado en la literatura como “varianza irrelevante al constructo” (17) la cual no es irrelevante, valga la redundancia, desde una dimensión política. El hallazgo que el/la egresado/a de las carreras de medicina de gestión pública tiene mejor media de desempeño en el examen y una media de promedio de grado más bajo, es un fenómeno observado en otros países(20).

Estudios sobre el ENRAM, muestra que, el examen, de 450 preguntas de MO, tiene una media de desempeño similar a la hallada en nuestro trabajo en Argentina, de 62,4 y una moda de 60,9 (21). Sin embargo, en un trabajo publicado de Barajas Ochoa (14) sobre el ENRAM (México), plantea que el mismo presenta debilidades en los indicadores de calidad (26).

En conclusión, el instrumento tiene propiedades adecuadas para discriminar un orden de mérito. Alta confiabilidad, validez y capacidad de discriminación. La utilización del promedio para los concursos desconociendo los parámetros que los determinan, introduce un sesgo en los resultados del mismo a favor, en este caso, de los/las egresados/as de las universidades de gestión privada.

BIBLIOGRAFÍA

1. Dure I.; Daverio D. ; Dursi C. PLANIFICACIÓN Y GESTIÓN DEL SISTEMA NACIONAL DE RESIDENCIAS (SNR) EN ARGENTINA [Internet]. Buenos Aires; 2013. Available from: http://www.msal.gov.ar/residencias/images/stories/descargas/datos/DURE-DAVERIO-DURSI-ObservatRHUS_Planif_Gestion_SistResid-2013.pdf
2. Silberman P., Lopez E. MA. Innovación digital para el ingreso y adjudicación a residencias en salud en Artículo original. *Revista Argentina de Educación Médica* [Internet]. 2021;10:26–32. Available from: <https://raem.afacimera.org.ar/raem-marzo-2021/>
3. (Compiladoras) María Isabel Duré; María del Carmen Cadile. LA GESTIÓN DE RECURSOS HUMANOS EN SALUD EN ARGENTINA. Una estrategia de consensos 2007-2015. 2015. 378 p.
4. María Gabriela Serralunga, Marta del Valle, Horacio Yulita, Patricia Rodríguez, María Paz Molina, Federico Pikas, et al. Selección de residentes en Argentina: análisis y comparación de los exámenes de admisión de los tres sistemas públicos con mayor oferta de cargos. *Investigación en Educación Médica*. 2017 Apr;6(22):e28–9.
5. AERA. American educational research association. American Psychological Association and National Council on Measurement in Education, and Joint Committee on Standards for Educational and Psychological Testing. Standards for educational and psychological testing [Internet]. 2014. Available from: <https://www.aera.net/>
6. Instituto Nacional para la Evaluación de la Educación. Criterios Técnicos para el desarrollo y uso de Instrumentos de Evaluación Educativa. 2015.
7. Brennan RL. Educational measurement [Internet]. 4th ed. Publishers P, editor. Chicago: National Council on Measurement in Education. American Council on Education; 2006. 808 p. Available from: <http://www.loc.gov/catdir/toc/ecip0614/2006015706.html>
8. Márquez Jiménez A. Las pruebas estandarizadas en entredicho. *Perfiles Educativos*. 2014;36(144):3–9.
9. Nichols, Sharon L.; Berliner DC. Collateral Damage: How High-Stakes Testing Corrupts America's Schools. Harvard University GS of E, editor. Cambridge; 2007. 252 p.
10. Cizek GJ. More unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice*. 2001;20(4):19–27.
11. Flotts P, Rodríguez B. Definición del referente de la evaluación y desarrollo del marco de especificaciones [Internet]. Vol. 3, Cuadernillo Técnico de Evaluación Educativa. 2019. Available from: <https://www.inee.edu.mx/wp-content/uploads/2019/08/P2A353.pdf>
12. Guilbert JJ. Educational handbook for health personnel. Revised edition. 6th ed. OMS, editor. World Health Organization Offset Publication. Ginebra; 1987. 362 p.
13. Sánchez-Mendiola M, Delgado-Maldonado L. Exámenes de alto impacto: implicaciones educativas. *Investigación en Educación Médica*. 2017;6(21):52–62.
14. A. BO. Equidad, validez y confiabilidad del Examen Nacional para Aspirantes a Residencias Médicas (ENARM): oportunidades para mejorar Referencias Legislación y salud pública en Francia Referencias. *Salud Pública de México*. 2017;59(5):4–5.
15. Junta Nacional de Examinadores Médicos. Cómo elaborar preguntas para evaluaciones escritas en las áreas de ciencias básicas y clínicas. Vol. 3. 2016. 1–98 p.
16. 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. *Revista de la Fundación Educación Médica*. 2016;19(4):217.
17. Muñiz J. Teoría Clásica de los Tests. Piramide, editor. España; 2017. 392 p.
18. Ramos JM. ANÁLISIS MULTIRRASGO-MULTIMÉTODO EN LA VALIDACIÓN DE INSTRUMENTOS PARA LA EVALUACIÓN DE LA CALIDAD DOCENTE EN INSTITUCIONES UNIVERSITARIAS. *Revista española de pedagogía* [Internet]. 1999;214:417–44. Available from: <https://www.cin.edu.ar/instituciones-universitarias/>
19. Carrillo-Avalos BA, Sánchez Mendiola M, Leenen I. Amenazas a la validez en evaluación: implicaciones en educación médica. *Investigación en Educación Médica*. 2020;(34):100–7.
20. Lagunas-Flores A. El ENARM y las escuelas y facultades de medicina. Un análisis que no le va a gustar a nadie». *Revista Médica del Instituto Mexicano del Seguro Social*. 2021;58(3):498–511.
21. Gaxiola-García MÁ, de Jesús Villalpando-Casas J, García-Saisó S, García-Minjares M, Martínez-González A. National examination for medical residency admission: academic variables and performance among different schools. *Salud Pública de México*. 2020;63(1):60–7.