

PREGUNTAS DE OPCIÓN MÚLTIPLE: SU REUTILIZACIÓN Y ESTABILIDAD PSICOMETRICA TEMPORAL

Di Bernardo, Juan José; Castillo Odena, Socorro; Semeza, Amada;
Palma Pérez, Ricardo; Di Bernardo; Gonzalo; Navarro; Viviana de los Ángeles

RESUMEN:

Introducción: La asignatura Medicina III, entre sus evaluaciones, incluye dos pruebas escritas con 80 preguntas de opción múltiple (POM) tipo A, que luego de aplicadas reciben análisis psicométrico, seleccionándose (según criterios establecidos) las POM que pasarán al banco de preguntas para ser reutilizadas, y aquellas con fallas que deben reformularse.

Objetivo: Analizar posibles diferencias en los indicadores psicométricos de las POM reutilizadas en diferentes exámenes.

Material y métodos: Estudio cuantitativo transversal donde se revisaron las pruebas escritas de Medicina III (2018-2023), y se seleccionaron POM que fueron reutilizadas, con idéntica redacción, en dos exámenes (E1 y E2) en diferentes cohortes. De las POM se registró su alcance cognitivo (taxonomía de Bloom) y de cada aplicación: el índice de dificultad (iP), el índice de discriminación (iD), y el funcionamiento de los distractores. Se compararon las medidas de resumen y diferencias en los índices; y se correlacionaron las variables entre E1 y E2.

Resultados: Se seleccionaron 50 POM reutilizadas, con diferentes alcances cognitivo. Comparando las medias de los índices registrados en E1 y E2, no hubo diferencias significativas en el total de preguntas: iP= 0.60 vs 0.56 e iD= 0.29 vs 0.27; ni en las diferentes categorías cognitivas. Los coeficientes de correlación (E1/E2) fueron significativos para: iP ($r=0,72$) e iD ($r=0,71$); así como la relación de dependencia: $R^2= 0,52$ y $0,49$ respectivamente. El iP fue muy estable en el 60% y el iD en el 54% de las POM; y una minoría tuvo diferencias amplias. Los distractores elegidos fueron similares en E1yE2.

Conclusiones: Las POM almacenadas en banco con criterios de calidad psicométrica, mostraron al ser reutilizadas, valores estables en los iP, iD y en el funcionamiento de los distractores, lo que indica que las preguntas de calidad comprobada tienen una adecuada estabilidad psicométrica temporal.

Palabras clave: educación médica – evaluación – cuestionarios - examen escrito - teoría clásica de los test - alcance cognitivo

INTRODUCCION:

La asignatura Medicina III, es una materia troncal de 5° año de la carrera de medicina de la UNNE que utiliza un sistema de evaluación longitudinal multimétodo (adaptado de Epstein, 2007)¹, con doble propósito, formativo y sumativo. El sistema incluye, entre otros instrumentos, dos pruebas escritas estructuradas, para evaluar principalmente conocimientos declarativos, razonamiento clínico y capacidad de resolución de problemas.

Cada prueba de conocimiento contiene 80 preguntas de opción múltiple (POM) de base clínica multi-items, formato A de tres opciones (una correcta y dos distractores)²⁻⁴ que, en su conjunto, demandan para las respuestas los primeros cinco niveles de procesamiento cognitivo de la taxonomía de Bloom⁵, en consonancia con los resultados de aprendizaje explorados.⁶

Luego de su aplicación a los estudiantes y su corrección con lector óptico, se realiza un análisis psicométrico del examen aplicando la teoría clásica de los test.⁷⁻⁹ A partir de estos resultados, se seleccionan (según los criterios consensuados por los profesores) las POM que pasarán al banco de preguntas (BP) de la materia, y aquellas con fallas o disfuncionales que deben ser reformuladas o descartadas.

De cada POM se analiza: el Índice de dificultad (iP), el Índice de Discriminación (iD) y el funcionamiento de los distractores.⁸⁻¹²

El iP indica el grado de dificultad (o facilidad) que presenta cada POM para decidir la respuesta correcta y se expresa con el valor de la proporción de estudiantes que respondió correctamente esa pregunta (un iP más bajo indica una mayor dificultad).

El iD indica en qué medida esa pregunta permite diferenciar a los estudiantes bien preparados (que saben más) de los menos preparados (que saben menos), y se obtiene correlacionando el rendimiento del estudiante en cada pregunta con su rendimiento global en el examen (calificación total). Para ello se calcula la diferencia en el número de aciertos a la pregunta, entre dos grupos de estudiantes, el 27% con las notas más altas vs el 27% con notas más bajas.

Para evaluar el funcionamiento de los distractores, se calcula el porcentaje de estudiantes que eligió cada opción (clave y distractores) de las preguntas. Cuando un distractor es elegido por menos del 5% de los estudiantes evaluados, se lo considera distractor “no funcionante”.

Para su inclusión al BP las POM deben cumplir tres criterios: iP entre 0,20 y 0,90; iD mayor a 0,10; y tener todos sus distractores funcionantes.

Las POM del BP, están alojadas en carpetas digitales, agrupadas por unidad temática y por su alcance cognitivo, para facilitar su identificación y reutilización en nuevos exámenes de acuerdo a los contenidos a evaluar y a la tabla de especificaciones elaboradas para cada prueba.¹³⁻¹⁵

Respecto al uso de las POM almacenadas en el BP, hay consenso entre los profesores, que las mismas pueden reutilizarse en años siguientes, sin superar el 15% del total de las preguntas de cada examen, para no poner en riesgo la confiabilidad del instrumento.

Con el propósito de valorar los grados de estabilidad psicométrica temporal de las POM que son reutilizadas del BP de la asignatura Medicina III, y en el marco del Proyecto de Investigación PI 221006: “La evaluación como estrategia de aprendizaje de las competencias clínicas en la carrera de medicina de la UNNE” acreditado por SGCT-UNNE, se planteó como objetivo: Analizar posibles diferencias en los indicadores psicométricos de las POM reutilizadas en diferentes exámenes.

MATERIAL Y MÉTODOS

En este estudio, cuantitativo transversal, se revisaron las pruebas de conocimientos aplicadas en la asignatura Medicina III entre 2018 y 2023, y se seleccionaron las POM que fueron utilizadas en dos exámenes (E1 y E2) aplicados en diferentes cohortes de estudiantes.

Se seleccionaron solo las preguntas que fueron aplicadas con idéntica redacción y contenido en ambos exámenes; y se excluyeron las POM que (entre E1 y E2) presentaban diferencias: en la extensión o en los datos de sus viñetas, en la construcción o alcance cognitivo de las preguntas, o en los contenidos planteados como distractores.

De cada aplicación de las POM seleccionadas se registró: el iP, el iD, y el funcionamiento de los distractores. Los índices habían sido calculados con las mismas ecuaciones en ambas oportunidades (E1 y E2). Dado que se analizaron preguntas aisladas de diferentes exámenes, no se consideraron los índices de confiabilidad (Alfa de Crombach) de las pruebas administradas.

Los datos obtenidos fueron analizados, en cada instancia de evaluación, en forma global y por categoría según el alcance cognitivo de las POM (recordar, comprender, aplicar, analizar y evaluar).

Tratamiento Estadístico:

Los datos se expresan en valores absolutos, porcentajes y proporciones según corresponda. Como medidas de resumen se calcularon los valores de tendencia central y de dispersión. Para la presentación de los datos resumidos se utilizan tablas de doble entrada. Para la representación gráfica de las variables se utilizan diagramas de caja (box plots), que permiten mostrar una serie de datos numéricos a través de sus cuartiles identificando claramente la media, la mediana, y los valores que se encuentran fuera de los límites aceptables; gráficos de puntos para mostrar la relación entre dos variables; gráficos de líneas apiladas para presentar la variación en el tiempo de los indicadores estudiados; y gráficos de columnas verticales para resumir las diferencias observadas.

Para los análisis estadísticos se utilizó el software XLSTAT (Addinsoft), se realizaron pruebas de “t” con los modelos y prueba “F” con las varianzas, tomando como nivel de significancia valores de $p < 0.05$. Para la correlación de las variables se utilizó el coeficiente de correlación (r) de Pearson y se realizó análisis de regresión calculando el coeficiente de regresión (r), de determinación (r^2) y curva de ajuste con los intervalos al 95% de confianza, asignándose $p < 0.05$ como nivel de significación para el ajuste.

RESULTADOS

Se identificaron 86 POM reutilizadas entre 2018 y 2023, de las cuales se seleccionaron 50 que fueron aplicadas en E1 y E2 con igual redacción y contenido en ambas oportunidades, que constituyen la muestra analizada para este estudio.

Estas 50 preguntas comparten la misma estructura, son POM tipo A ricas en contexto, con viñetas clínicas y tres opciones (una correcta y dos distractores); todas cumplieron en E1 los criterios para su inclusión al BP ($iP: 0,20 - 0,90$; $Id > 10$; y distractores todos funcionantes); y difieren en sus alcances cognitivos pues demandan diferentes niveles de la taxonomía de Bloom: el 26% (13) “recordar”; el 22% (11) “comprender”; el 24% (12) “aplicar”; el 14% (7) “analizar”; y el 14% (7) “evaluar”.

En la Tabla 1 se presentan y comparan los valores (Media \pm DE) de los índices de dificultad y de discriminación, registrados en las dos evaluaciones (E1 y E2), agrupados según el nivel taxonómico de las POM. En este análisis comparativo no se observaron diferencias significativas en ninguna de las categorías cognitivas ni en el total de preguntas.

Tabla 1: Índices iP e iD en las dos evaluaciones según el alcance cognitivo de las POM

Para representar mejor la distribución y tendencia central de los valores iP e iD de las 50 POM, en los Gráficos 1 (A y B) se presentan los diagramas de caja y bigotes en forma comparativa entre las dos evaluaciones.

Gráficos 1 (A y B): Distribución y comparación del iP (1-A) e iD (1-B) en las dos evaluaciones.

El análisis estadístico comparativo de las Medias (DE) registradas en E1 y E2, no mostró diferencias significativas en ninguno de los índices estudiados.

Correlacionando los valores de cada indicador entre E1 y E2, se observaron coeficientes de correlación (r) de Pearson significativos en los dos índices evaluados: iP: $r=0,72$ (IC 95%:0.55-0.83) e iD: $r=0,71$ (IC 95%:0.54-0.82).

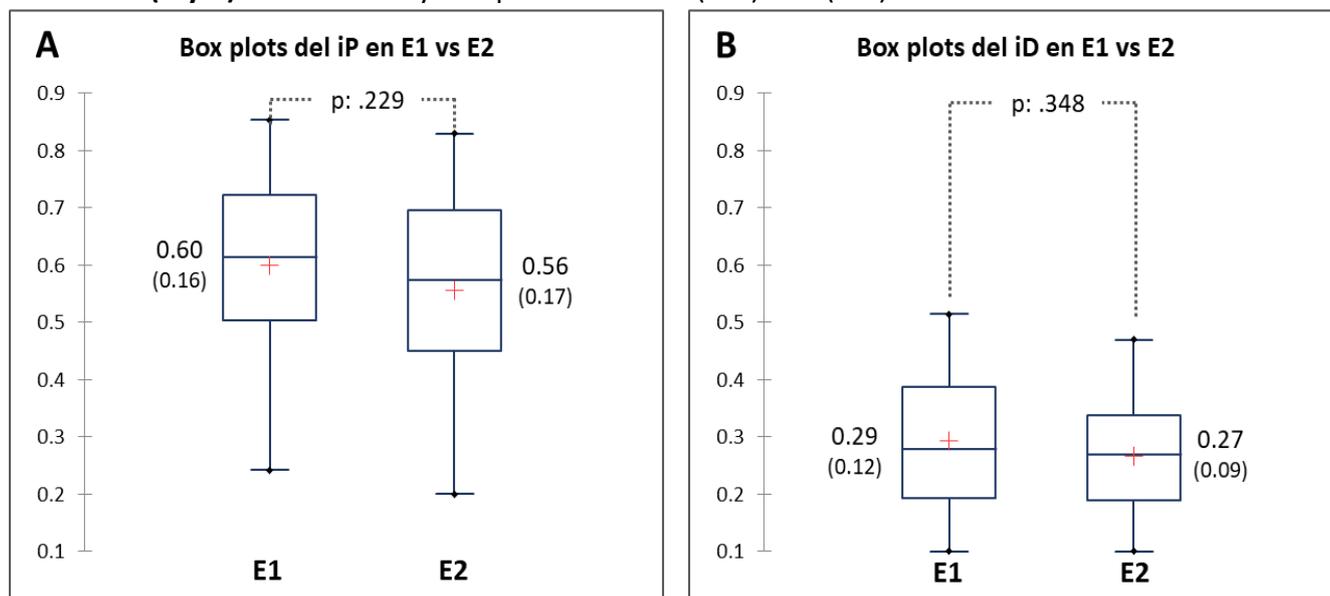
Para valorar la relación de dependencia entre las variables extraídas en E1 y E2, se obtuvieron las curvas de regresión lineal que se presentan en el Grafico 2 (A y B)

Tabla 1: Índices iP e iD en las dos evaluaciones según el alcance cognitivo de las POM

Alcance cognitivo*	n POM	Índices de Dificultad (iP)			Índices de Discriminación (iD)		
		E1	p	E2	E1	p	E2
Recordar	13	0.66 ±.14	.153	0.55 ±.23	0.33 ±.12	.635	0.31 ±.09
Comprender	11	0.63 ±.15	.512	0.59 ±.13	0.29 ±.11	.818	0.28 ±.09
Aplicar	12	0.60 ±.15	.868	0.61 ±.14	0.28 ±.12	.107	0.21 ±.08
Analizar	7	0.60 ±.12	.203	0.49 ±.18	0.28 ±.10	.840	0.27 ±.08
Evaluar	7	0.44 ±.17	.570	0.49 ±.15	0.29 ±.15	.677	0.26 ±.11
Total	50	0.60 ±.16	.229	0.56 ±.17	0.29 ±.12	.348	0.27 ±.09

* Según taxonomía de Bloom

Gráficos 1 (A y B): Distribución y comparación del iP (1-A) e iD (1-B) en las dos evaluaciones.



Gráficos 2 (A y B): Curvas de regresión lineal del iP (2-A) e iD (2-B) en las dos evaluaciones. Los valores R^2 (en ambas curvas) indican que los índices correspondientes al E1 explican el 52% (Gráfico 2-A) y el 49% (Gráfico 2-B), de la variabilidad de los índices del E2; y los valores de $p < 0.0001$ le otorgan significación estadística e indican que la información aportada por estas curvas es significativamente mejor que las que podrían aportar únicamente las medias.

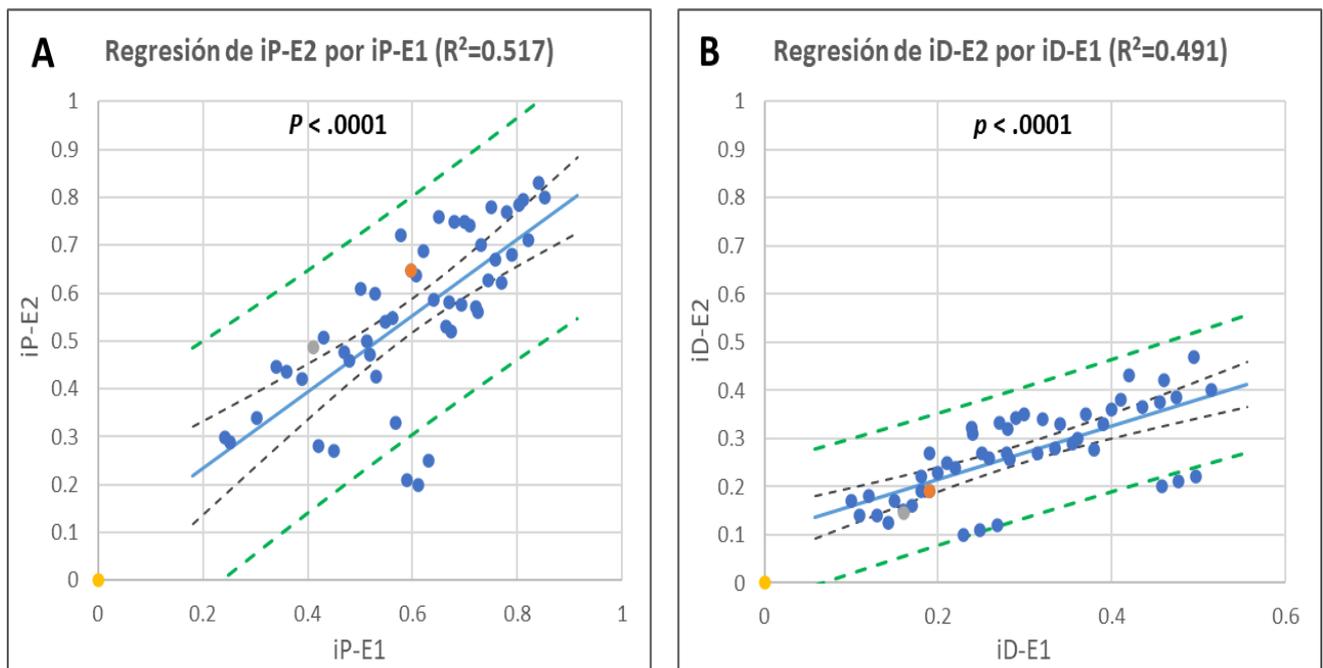
Las tendencias y magnitud de los cambios registrados en los indicadores iP e iD, en los dos exámenes aplicados en tiempos diferentes (E1 y E2), se muestran en los Gráficos 3 y 4 con líneas apiladas y mediante barras verticales que agrupan las POM en tres categorías según la magnitud de las diferencias observadas en los indicadores.

Gráfico 3: Variaciones de los índices de dificultad de cada POM en las dos evaluaciones.

Gráfico 4: Variaciones de los índices de discriminación de cada POM en las dos evaluaciones.

Analizando el funcionamiento de los distractores en las dos evaluaciones, no se observaron diferencias llamativas en ninguna de las POM, las opciones elegidas por los estudiantes mostraron proporciones similares en ambas aplicaciones, y no se detectaron distractores no funcionantes.

Gráficos 2 (A y B): Curvas de regresión lineal del iP (2-A) e iD (2-B) en las dos evaluaciones.



----- Intervalo de confianza (media 95%)

----- Intervalo de confianza (observaciones 95%)

Gráfico 3: Variaciones de los índices de dificultad de cada POM en las dos evaluaciones.

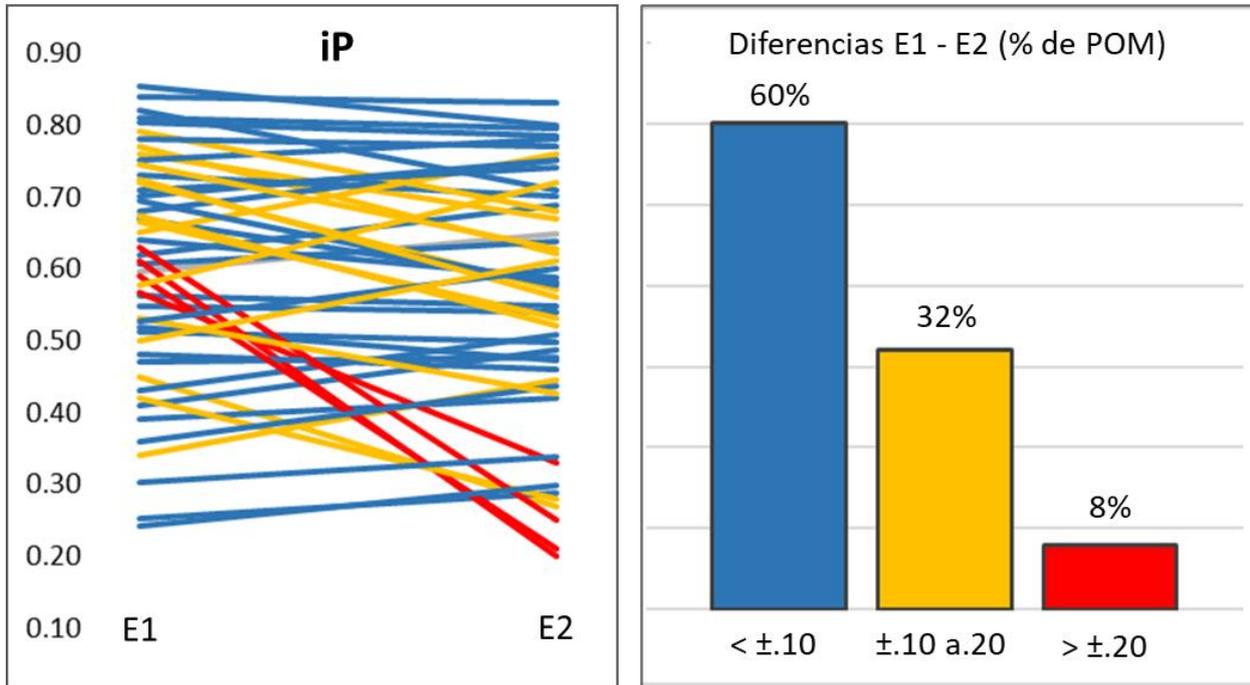
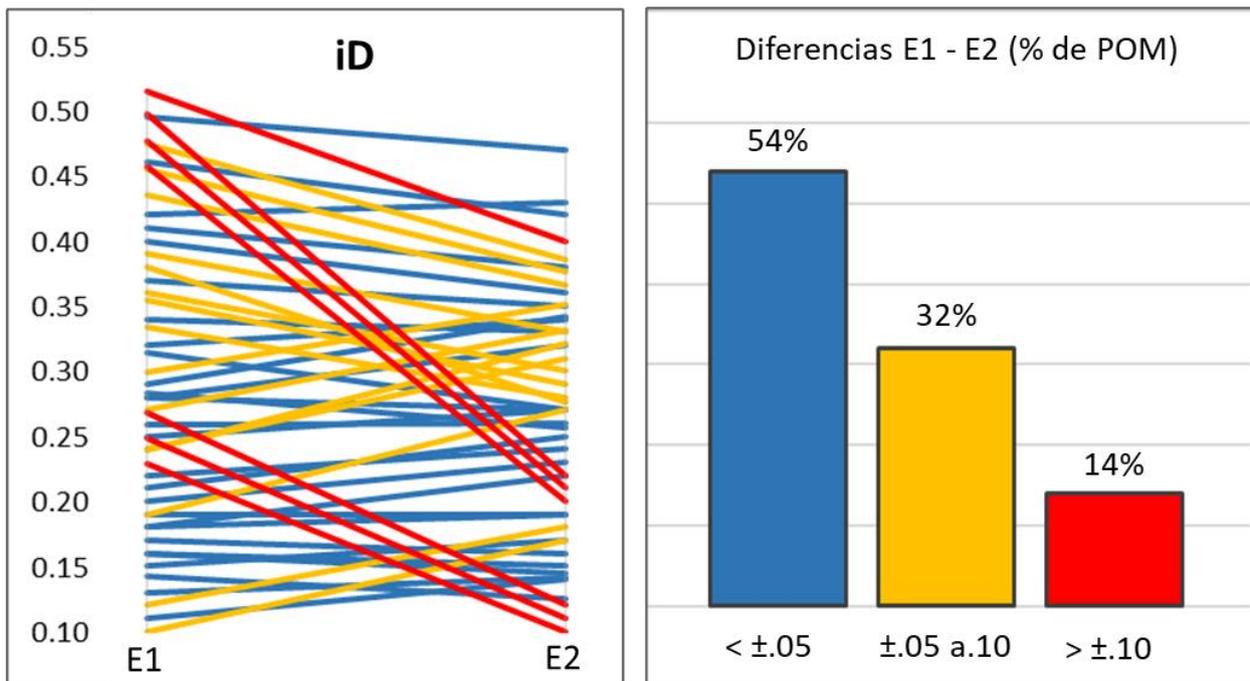


Gráfico 4: Variaciones de los índices de discriminación de cada POM en las dos evaluaciones.



DISCUSION

En el período 2018 a 2023, la asignatura Medicina III aplicó en distintas cohortes de estudiantes, diez (10) “pruebas de conocimiento” en modalidad presencial, utilizando cuestionarios con 80 POM tipo A de base clínica multi-items, con tres opciones (una correcta y dos distractores). De estas 800 preguntas, 86 (11%) fueron reutilizadas en el mismo período, lo que se ajusta al criterio adoptado por la materia de no superar el 15% del total de las preguntas.

Treinta y seis de las POM reutilizadas no se incluyeron en este estudio por correcciones en sus viñetas o cambios de distractores antes del E2, las otras 50 POM (sin modificaciones), constituyen la muestra. La mayoría de estas preguntas (72%) abarcan niveles cognitivos de Orden Inferior según la Taxonomía de Bloom⁵ (recordar, comprender, aplicar) y las restantes (28%) tienen alcances cognitivos de Orden Superior (analizar y evaluar). Esta proporción (0.28) de preguntas de Orden Superior reutilizadas en pruebas dirigidas a evaluar conocimientos que en su mayoría requieren habilidades de Orden Inferior, es muy beneficiosa para los estudiantes pues les permite una comprensión conceptual más profunda de los contenidos evaluados.¹⁶

Analizando y comparando los indicadores psicométricos (iP e iD) registrados en los dos momentos (E1 y E2), no se observaron diferencias significativas entre las medias del total de POM (iP= 0,60 vs 0,56; iD= 0,29 vs 0,27), ni en las diferentes categorías cognitivas de las preguntas; lo que permite inferir, desde este primer análisis, que las preguntas reutilizadas tienen estabilidad psicométrica en el tiempo. Dicha inferencia se fortalece con la fuerte correlación de Pearson (iP: $r=0,72$ e iD: $r=0,71$) observada en las dos aplicaciones, y con los altos coeficientes de determinación entre las variables (iP: $r^2=0,52$ e iD: $r^2=0,49$) que se obtuvieron en las curvas de regresión lineal.

Esta “estabilidad psicométrica temporal” (como la denominamos) no fue una característica uniforme de las POM reutilizadas, pues como se ve en los Gráficos 3 y 4, la mayoría de las preguntas mostraron intensa estabilidad de sus iP (60%) e iD (54%); y una minoría (8% y 14 % respectivamente), tuvieron diferencias amplias de los indicadores. Además, la estabilidad psicométrica no estuvo influida por el alcance cognitivo de las preguntas (**Tabla 1**); ni por los valores absolutos de los indicadores como se puede ver en los **Gráficos 3 y 4** de líneas apiladas.

Esta propiedad de estabilidad en el tiempo de los parámetros psicométricos de las POM reutilizadas, ha sido comprobada por varios autores¹⁷⁻¹⁹ en diferentes contextos educativos; pero al mismo tiempo cuestionada por otros cuando las POM: son aplicadas en cohortes pequeñas de estudiantes²⁰; se reutilizan tres veces o más en un corto período de tiempo²¹; son divulgadas entre los estudiantes, oficialmente por los docentes con fines de retroalimentación²², u obtenidas por los mismos estudiantes a través de conductas irregulares^{21, 23}.

Young et al (2017)²⁰ observaron una gran variabilidad de los iP e iD, reutilizando 32 POM (combinadas con preguntas nuevas) en 22 cohortes pequeñas (10 a 15 estudiantes) secuenciales de pasantías, en la Universidad de McGill (Quebec-Canadá). Esta evidencia no es válida para el contexto de Medicina III donde el número de alumnos evaluados oscila entre 180 y 300 en cada examen.

Joncas et al (2018)²¹ analizaron los cambios en los indicadores psicométricos de 1.629 POM que fueron reutilizadas hasta cuatro veces en un período de cinco años en la Universidad de Sherbrooke (Quebec, Canadá), y observaron que, desde la primera a la cuarta aplicación, la media del iP aumentó de 0,79 a 0,85 respectivamente (fueron más fáciles), y el iD disminuyó de 0,17 a 0,14 respectivamente. Esta observación excede a nuestro trabajo pues las POM solo se reutilizaron una vez.

Appelhaus et al (2023)²² en la Universidad de Berlín revisaron 199 exámenes aplicados en cuatro trimestres, analizaron los índices iP e iD de 10.148 POM (25% nuevas, 50% reutilizadas no divulgadas, y 25% reutilizadas y divulgadas) y observaron diferencias entre los tres grupos, en las medias de los iP= 0,66; 0,71 y 0,83; y en los iD= 0,20; 0,21 y 0,25 respectivamente. Cuando analizaron los efectos de la reutilización, observaron que los iP disminuyeron en las POM no divulgadas y aumentaron en las POM divulgadas (fueron más fáciles). Mientras que los iD aumentaron en ambos grupos (no hubo efectos por la divulgación). Las POM aplicadas en nuestro trabajo no fueron divulgadas entre los estudiantes antes de su reutilización.

La divulgación de las POM y su posterior reutilización, es una cuestión muy controvertida en la educación médica. Los argumentos en contra se basan principalmente en los riesgos que conlleva la posible disminución de la confiabilidad y validez de las POM reutilizadas, lo que es particularmente importante en las evaluaciones sumativas de riesgo (promoción, selección, graduación). Pero también hay argumentaciones a favor de la divulgación como, los beneficios de una retroalimentación más precisa cuando la evaluación es formativa; la mayor transparencia que da al proceso de administración de las pruebas; las oportunidades que tienen los estudiantes de prepararse mejor; la reducción de la ansiedad en los exámenes; y el menor esfuerzo de los docentes para mantener la confidencialidad del banco de preguntas.^{17-19, 21, 22, 24-26}

Otro beneficio que puede ofrecer la divulgación y posterior reutilización de las POM es su utilidad para el “aprendizaje mejorado mediante pruebas” (del inglés “Test-enhanced learning”) ²⁷, fenómeno que tiene su explicación en que la recuperación activa de información en una evaluación (o a través de la exposición repetida a preguntas), contribuye a la creación de redes organizadas de conocimiento que favorecen la retención a largo plazo y promueven la recuperación de información en exámenes posteriores o en contextos donde se necesita.^{27, 28}

CONCLUSIONES

Los resultados de este estudio revelaron que, las POM almacenadas en un banco de preguntas en base a criterios de calidad psicométrica, mostraron al ser reutilizadas, valores estables en los índices de dificultad y discriminación, y en el funcionamiento de los distractores, lo que indica que las preguntas de calidad comprobada tienen una adecuada estabilidad psicométrica temporal.

Esta propiedad de las POM, permite a los docentes que administran las evaluaciones, seleccionar y reutilizar las preguntas necesarias para cubrir todos los contenidos a evaluar y en los niveles cognitivos requeridos, manteniendo una adecuada calidad psicométrica con el propósito de lograr una evaluación confiable, de dificultad global intermedia y con suficiente poder de discriminación.

BIBLIOGRAFIA

1. Epstein RM. Assessment in Medical Education. *N Engl J Med*. 2007; 356:387-396.
2. Schuwirth LWT, van der Vleuten CPM. Different written assessment methods: what can be said about their strengths and weaknesses? *Med Educ* 2004; 38: 974–979
3. Rodriguez MC. Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational measurement: issues and practice* 2005; 24(2): 3-13.
4. Paniagua, MA. Swygert, KA. (Editores). *Cómo elaborar preguntas para evaluaciones escritas en las áreas de ciencias básicas y clínicas*. Philadelphia. National Board of Medical Examiners (NBME). Cuarta Edición (2016). En https://www.nbme.org/sites/default/files/2020-01/DownloadingtheGoldBook_ES.pdf.
5. Krathwohl DR. A Revision of Bloom's Taxonomy: An Overview. *Theory Pract* 2002; 41(4): 212-218.
6. Di Bernardo JJ, Navarro V. La evaluación como aprendizaje de habilidades cognitivas y metacognitivas. Efectos de un sistema de evaluación multimétodo. *FEM* 2023; 26 (4):167-173. doi:10.33588/fem.264.1294

7. De Champlain AF. A primer on classical test theory and item response theory for assessments in medical education. *Med Educ.* 2010;44(1):109-117. doi:10.1111/j.1365-2923.2009.03425.x
8. Morales P. Análisis de ítems en las pruebas objetivas. Madrid: Universidad Pontificia Comillas (2012). En <http://educreea.ci/wp-content/uploads/2014/11/19-nov-analisis-de-ite-mos-en-las-pruebas-objetivas.pdf>.
9. Bonillo, A. Análisis de los Ítems. Universitat Oberta de Catalunya. PID: 00198631. *Psicometría.* 2013. En http://openaccess.uoc.edu/webapps/o2/bitstream/10609/69325/7/Psicometr%C3%ADa_M%C3%B3dulo%205_An%C3%A1lisis%20de%20los%20%C3%ADtems.pdf.
10. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. *J Pak Med Assoc.* 2012;62(2):142-147.
11. Chauhan GR, Chauhan BR, Vaza JV, Chauhan PR. Relations of the Number of Functioning Distractors With the Item Difficulty Index and the Item Discrimination Power in the Multiple Choice Questions. *Cureus.* 2023;15(7):e42492. doi:10.7759/cureus.42492
12. Tarrant M, Ware J, Mohammed AM. An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis. *BMC Med Educ* 2009; 9: 40.
13. Vale CD. Computerized item banking. In Downing SM & Haladyna TM (Eds.), *Handbook of test development.* Mahwah NJ: Erlbaum 2006. p. 261-285
14. Sadaf S, Khan S, Ali SK. Tips for developing a valid and reliable bank of multiple choice questions (MCQs). *Educ Health (Abingdon).* 2012;25(3):195-197. doi:10.4103/1357-6283.109786
15. Bhat SK, Prasad KHL. Item analysis and optimizing multiple-choice questions for a viable question bank in ophthalmology: A cross-sectional study. *Indian J Ophthalmol.* 2021;69(2):343-346. doi:10.4103/ijo.IJO_1610_20
16. Jensen JL, McDaniel MA, Woodard SM, Kummer TA. Teaching to the test...or testing to teach: Exams requiring higher order thinking skills encourage greater conceptual understanding. *Educ Psychol Rev.* 2014; 26:307-329.
17. Herskovic P. Reutilization of multiple-choice questions. *Med Teach.* 1999;21(4):430-431.
18. Wood TJ. The effect of reused questions on repeat examinees. *Adv Heal Sci Educ.* 2009;14(4):465-473.
19. Yang EB, Lee MA, Park YS. Effects of test item disclosure on medical licensing examination. *Adv Heal Sci Educ.* 2018;23(2):265-274.
20. Young M, Cummings BA, St-Onge C. Ensuring the quality of multiple-choice exams administered to small cohorts: A cautionary tale. *Perspect Med Educ.* 2017;6(1):21-28. doi:10.1007/s40037-016-0322-0
21. Joncas SX, St-Onge C, Bourque S, et al. Re-using questions in classroom-based assessment: an exploratory study at the undergraduate medical education level. *Perspect Med Educ.* 2018;7(6):373-378.
22. Appelhaus S, Werner S, Grosse P, Kämmer JE. Feedback, fairness, and validity: effects of disclosing and reusing multiple-choice questions in medical schools. *Med Educ Online.* 2023;28(1):2143298. doi:10.1080/10872981.2022.2143298
23. Henning MA, Ram S, Malpas P, Shulruf B, Kelly F, Hawken SJ. Academic dishonesty and ethical reasoning: pharmacy and medical school students in New Zealand. *Med Teach.* 2013;35:e1211-e7
24. Wagner-Menghin M, Preusche I, Schmidts M. The effects of reusing written test items: a study using the Rasch model. *ISRN Educ.* 2013; <https://doi.org/10.1155/2013/585420>.
25. Park YS, Yang EB. Three controversies over item disclosure in medical licensure examinations. *Med Educ Online.* 2015;20(1):28821.
26. O'Neill TR, Sun L, Peabody MR, Royal KD. The Impact of Repeated Exposure to Items. *Teach Learn Med.* 2015;27(4):404-409. doi:10.1080/10401334.2015.1077131
27. Larsen DP, Butler AC, Roediger III HL. Test-enhanced learning in medical education. *Med Educ* 2008; 2(10): 959-966.
28. Green ML, Moeller JJ, Spak JM. Test-enhanced learning in health professions education: a systematic review: BEME Guide No. 48. *Med Teach* 2018; 40(4), 337-350.

DATOS DE AUTOR

Título:

Preguntas de opción múltiple: su reutilización y estabilidad psicométrica temporal

Autores:

Di Bernardo Juan José *

Castillo Odena Socorro *

Semeza Amada *

Palma Pérez Ricardo *

Di Bernardo Gonzalo *

Navarro Viviana de los Ángeles *

*Facultad de Medicina. Universidad Nacional del Nordeste. Corrientes. Argentina

Correo electrónico de contacto: jjdibernardo@med.unne.edu.ar