

# Correlación entre una evaluación sumativa escrita y el promedio ponderado en estudiantes de medicina humana

Franco Ronald Romaní Romaní<sup>a,\*‡</sup>, César Gutiérrez<sup>b,§</sup>

Facultad de Medicina



## Resumen

**Introducción:** La evaluación sumativa es usualmente aplicada al final de un periodo académico para medir el aprendizaje de un estudiante, estas evaluaciones utilizan diversas representaciones de pacientes, desde pacientes reales hasta narraciones escritas basadas en casos clínicos. Existe limitada evidencia sobre la asociación entre los resultados de evaluaciones escritas basadas en casos clínicos como evaluación sumativa y el desempeño académico en estudiantes de medicina humana.

**Objetivo:** Evaluar la correlación entre el puntaje de un examen anual escrito basado en casos clínicos con preguntas de opción múltiple y el promedio ponderado –como indicador de desempeño académico– en estudiantes de una facultad de medicina humana del Perú.

**Método:** Estudio de fuentes secundarias, con diseño

longitudinal en panel para el periodo 2017 a 2020. El examen anual tiene 250 preguntas de opción múltiple basadas en diferentes contextos clínicos extraídos de reportes de casos publicados. Cada pregunta correctamente respondida brindó un punto, el puntaje fue transformado a escala vigesimal y se evaluó su correlación con el promedio ponderado en escala vigesimal. Se realizaron gráficos de dispersión y el cálculo del coeficiente de correlación de Spearman. El análisis fue realizado para cada año académico.

**Resultados:** Se analizaron los datos de 228 estudiantes, las medias de los puntajes del examen anual fluctuaron entre 9.61 (desviación estándar = 1.91) para el 2019 a 10.75 (desviación estándar = 2.34) para el 2018. En todos los años se encontró correlación lineal directa significativa entre el promedio ponderado anual y el examen anual.

<sup>a</sup> Unidad de Investigación, Facultad de Medicina Humana, Universidad de Piura, Lima, Perú

<sup>b</sup> Departamento de Ciencias de la Medicina, Facultad de Medicina Humana, Universidad de Piura, Lima, Perú  
ORCID ID:

<sup>‡</sup> <https://orcid.org/0000-0002-6471-5684>

<sup>§</sup> <https://orcid.org/0000-0002-0237-321X>

Recibido: 09-diciembre-2021. Aceptado: 14-febrero-2022.

\* Autor para correspondencia: Franco Ronald Romaní Romaní. Calle Mártir José Olaya 162, Miraflores, Lima, Perú. Teléfono: 051-990630678.

Correo electrónico: Franco.romani@udep.edu.pe

Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

La mayor correlación se dio en el 2019 (rho de Spearman [ $r_s = 0.812$ ]). En el análisis de correlación según sexo se mantuvieron correlaciones significativas.

**Conclusiones:** En los años analizados el puntaje obtenido en el examen anual se correlaciona de manera lineal directa con el desempeño académico, también se encontró correlación lineal directa entre los puntajes del examen anual.

**Palabras clave:** Rendimiento académico; evaluación educativa; educación médica; estudiantes de medicina; informes de casos.

Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Correlation between a written summative assessment and the weighted average in human medicine students

### Abstract

**Introduction:** Summative assessment is usually applied at the end of an academic period to measure a student's learning outcome, these assessments use various representations of patients, from real patients to written narratives based on clinical cases. There is scarce evidence about association between the results of written assessments based on clinical cases and academic performance in medical students.

**Objective:** To evaluate the correlation between the score of an annual written exam based on clinical cases with

multiple choice questions and the weighted average –as an indicator of academic performance– in students of human medicine in Peru.

**Method:** Secondary data analysis with longitudinal panel design for the period 2017 to 2020. The annual exam has 250 multiple-choice questions based on a clinical context drawn from published case reports. Each correctly answered question provided one point, the score was transformed to a vigesimal scale and its correlation with the weighted average was evaluated. Scatter plots and the calculation of the spearman correlation coefficient were performed. The analysis was performed for each academic year.

**Results:** Data from 228 students were analyzed, the means of annual exam scores varies between 9.61 (standard deviation = 1.91) for 2019 to 10.75 (standard deviation = 2.34) for 2018. In all years, a significant direct linear correlation was found between the annual weighted average and the annual exam score. The highest correlation occurred in 2019 (Spearman's rho [ $r_s = 0.812$ ]). In the correlation analysis according to sex, significant correlations were obtained.

**Conclusions:** In each of the years analyzed, the score obtained in the annual exam is directly linearly correlated with the academic performance, a direct linear correlation was also found between the annual exam scores.

**Keywords:** Academic performance; educational measurement; medical education; medicine students; case reports.

This is an Open Access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## INTRODUCCIÓN

La evaluación es fundamental en la enseñanza de la medicina humana, sus beneficios van más allá de medir el conocimiento y las competencias adquiridas por los estudiantes, también ayuda a direccionar y estimular el aprendizaje y en retroalimentar a los profesores y estudiantes<sup>1</sup>. Ninguna evaluación puede medir por sí sola todos los dominios teóricos y prácticos del aprendizaje. Se describen dos formas de evaluación, la formativa y la sumativa. La evaluación formativa es vital para la promoción

del aprendizaje, mientras que la sumativa es usualmente aplicada al final de un periodo académico para medir el resultado del aprendizaje de un estudiante<sup>1,2</sup>. Los resultados de la evaluación sumativa pueden aplicarse para mejorar la formulación de un curso, para evaluar la efectividad de la enseñanza, evaluar la calidad educativa de un programa; además, por sí sola, puede enriquecer el proceso de aprendizaje<sup>3</sup>.

Existen diversas modalidades de evaluación sumativa, todas ellas buscan determinar los niveles de

competencia de un estudiante usando diferentes representaciones de pacientes, que van desde un paciente real a formatos alternativos como casos clínicos, modelos anatómicos, maniqués, simuladores o pacientes simulados<sup>4</sup>. Algunos ejemplos son el examen clínico objetivo estructurado (OSCE por sus siglas en inglés), el examen práctico objetivo estructurado, la evaluación basada en casos o las evaluaciones escritas<sup>1,3,5</sup>.

Las evaluaciones con preguntas escritas pueden ser abiertas o de opción múltiple, además pueden clasificarse como aquellas con contexto de un caso clínico o sin él<sup>6</sup>. Las evaluaciones con contexto clínico generan un proceso cognitivo más complejo y próximo al de las situaciones clínicas<sup>1</sup>. Las evaluaciones escritas basadas en casos clínicos con opciones múltiples se han convertido en una submodalidad popular para evaluar el conocimiento y habilidad para resolver problemas, además ha mostrado ventajas sobre las preguntas de respuesta abierta<sup>7-9</sup>. Actualmente, una de las formas más conocidas de representación de pacientes reales son los casos narrativos que consisten en relatos que describen un paciente y su historial médico. Las narraciones suelen basarse en casos clínicos reales y son presentados al estudiante en texto o digitalmente, y con imágenes de apoyo<sup>4</sup>.

Se ha estudiado la asociación entre diversos tipos de evaluaciones sumativas y el desempeño clínico de estudiantes de ciencias de la salud, por ejemplo, una revisión sistemática encontró que la mayor parte de los estudios ha evaluado la correlación del puntaje de un OSCE con el desempeño clínico, mientras que cuatro estudios evaluaron dicha correlación para evaluaciones escritas y solo uno, hasta el 2016, correlacionó la submodalidad de evaluación escrita basada en casos con preguntas de opciones múltiples y el desempeño clínico<sup>2</sup>. Por otro lado, algunos estudios han evaluado la correlación entre puntajes de OSCE con el promedio ponderado, este último como indicador de desempeño académico<sup>10-12</sup>; sin embargo, la correlación de evaluaciones escritas basadas en casos clínicos y el promedio ponderado no han sido igualmente estudiados en estudiantes de medicina humana.

Las evaluaciones escritas como herramienta de evaluación sumativa permiten evaluar el conoci-

miento y el saber qué hacer con dicho conocimiento, por ello, es necesario contar con nuevas metodologías de este tipo de evaluación; una primera aproximación es explorar si una evaluación escrita basada en casos clínicos con preguntas de opción múltiple se encuentra correlacionada con el desempeño académico.

## OBJETIVO

Evaluar la correlación entre el puntaje de un examen anual escrito basado en casos clínicos con preguntas de opción múltiple y el promedio ponderado –como indicador de desempeño académico– en estudiantes de medicina humana de una universidad en Perú.

## MÉTODO

### Ámbito y diseño de estudio

Se realizó un estudio de fuentes secundarias, la fuente de información fue una base de datos proporcionada por la facultad de medicina humana de la Universidad de Piura. Esta facultad se encuentra en la sede ubicada en la ciudad de Lima en Perú. Esta universidad privada fue seleccionada porque desde el inicio de su programa de medicina humana, el 2017, se implementó una evaluación denominada “Examen anual de medicina basado en casos clínicos” (EAM). El estudio tuvo un diseño longitudinal en panel para el periodo 2017 al 2020.

### La estrategia de evaluación

El EAM fue usado para fines de evaluación sumativa. Esta evaluación forma parte de la valoración integral del alumno y corresponde a un examen escrito objetivo basado en diferentes casos clínicos, a partir de los cuales los docentes de cada asignatura del plan de estudios formulan preguntas de su asignatura. La evaluación tiene 250 preguntas de opción múltiple con cinco alternativas, solo una de ellas es la respuesta correcta.

El examen se distribuye en tres cuadernillos, cada uno incluye preguntas en torno a cuatro o seis casos clínicos narrativos, extraídos de publicaciones del tipo reporte de caso en revistas médicas arbitradas. Este tipo de publicación tiene en su estructura una introducción, el reporte del caso, discusión y referencias bibliográficas; para la elaboración del EAM solo se extrae la sección del reporte de caso, inclu-

yendo las imágenes y tablas derivadas del estudio del paciente (**anexo 1**).

El EAM incluye todas las asignaturas que el alumno ha llevado, por ejemplo, un alumno de cuarto año será evaluado en las asignaturas desde el primer hasta el cuarto año de la carrera. La distribución del número de preguntas tiene dos criterios: a) El año de estudios que ha culminado el alumno (se rinde hasta el sexto año); y b) El número de créditos de la asignatura, en ambos casos el número de preguntas es directamente proporcional (**anexo 2**). El examen se rinde de manera no obligatoria al finalizar cada año académico en fecha única.

### Sujetos de estudio

La población de estudio estuvo conformada por los estudiantes de la facultad de medicina humana de una universidad, ingresantes entre el 2017 al 2020. Se incluyó a toda la población de estudio, se excluyeron a los estudiantes que no rindieron ningún EAM.

### Definición de variables

El puntaje del EAM varía de 0 a 250 puntos con base en el número de respuestas correctas, la respuesta correcta brinda un punto. Para el presente análisis convertimos dicho puntaje a la escala vigesimal. El índice acumulado es el promedio ponderado promocional registrado en el Sistema de Gestión Académica de la facultad para cada año académico, el puntaje varía de 0 a 20 puntos. Este promedio se obtiene al ponderar las calificaciones obtenidas en todas las asignaturas llevadas por el estudiante. Cada asignatura tiene un número de créditos (en función del número de horas académicas), la nota obtenida en la asignatura es multiplicado por el número de créditos, se realiza la suma total de estos elementos, y se divide entre el número total de créditos llevados. Otras variables fueron sexo (masculino/femenino), año de ingreso (2017, 2018, 2019, 2020) y número de créditos matriculados por semestre.

### Análisis estadístico

Se realizó un análisis descriptivo del sexo, año de estudio, número de EAM rendidos, puntaje del EAM y promedio ponderado para cada año analizado. Se evaluó el tipo de distribución mediante la prueba de Kolmogorov-Smirnov. Se aplicó la prueba U de

Mann-Whitney para determinar la asociación entre la mediana del número de créditos matriculados y haber rendido el EAM para cada año.

Para cada año, se evaluó la correlación entre el promedio ponderado y el puntaje del EAM mediante el coeficiente de correlación por rangos de Spearman, previamente se realizó un gráfico de dispersión. Este análisis fue realizado para toda la muestra y por sexo. Se evaluó la correlación entre los puntajes del EAM, por último, se elaboraron gráficos de Bland y Altman para explorar la concordancia entre el EAM y el promedio ponderado. La significancia estadística fue evaluada con pruebas de hipótesis a dos colas, se consideró como significativos valores de  $p < 0.05$ . El análisis fue realizado en SPSS versión 25.

### CONSIDERACIONES ÉTICAS

Se solicitó a la secretaría académica de la facultad los puntajes de los EAM y promedio ponderado promocional de los años 2017 al 2020. La base de datos contenía el código del alumno y las variables solicitadas. Dicha información fue recolectada para la gestión académica, y fue solicitada solo para fines del presente análisis. El código del alumno que es un identificador personal fue intercambiado por un identificador alfanumérico una vez que se realizó la limpieza de la base de datos. El estudio fue aprobado por el Comité Institucional de Ética en Investigación de la Universidad.

### RESULTADOS

Entre 2017 y 2020, hubo 234 estudiantes matriculados al final de año académico, seis no rindieron ningún EAM, por lo cual fueron excluidos del análisis. Tres excluidos fueron ingresantes del 2019 y otros tres del 2020. Ingresaron al análisis 228 estudiantes, 114 (50.0%) fueron mujeres. 51 estudiantes (22.4%) ingresaron en el año 2017; 57 (25.0%), en el 2018; 56 (24.6%), en el 2019; y 64 (28.1%), en el 2020.

De los ingresantes en el 2017, el 58.8% rindieron los cuatro EAM posibles; dicho porcentaje se incrementó a 82.5% en 2018; 85.7%, en 2019; y 100%, en el 2020 (**tabla 1**). El EAM del 2017 fue aplicado al 86.3% (44/51) de estudiantes; el del 2018, al 87.0% (94/108); el del 2019, al 86.0% (141/164); y el del 2020, al 96.5% (220/228). Aquellos que rindieron el EAM tenían un mayor número de créditos comparados con aquellos

**Tabla 1.** Número de examen anual de medicina basado en casos clínicos rendidos por los estudiantes según año de ingreso

Año		Número de exámenes anuales de medicina basado en casos clínicos				Total
		1	2	3	4	
2017	n	2	9	10	30	51
	%	3.9	17.6	19.6	58.8	
2018	n	0	10	47	0	57
	%	0.0	17.5	82.5	0.0	
2019	n	8	48	0	0	56
	%	14.3	85.7	0.0	0.0	
2020	n	64	0	0	0	64
	%	100.0	0.0	0.0	0.0	

**Tabla 2.** Análisis descriptivo de los puntajes del examen anual de medicina basado en casos clínicos y el promedio ponderado promocional según año

Variable	n	Media	DE	CV (%)	Mediana	Q1 - Q3
Puntaje del Examen Anual de Medicina Basado en Casos Clínicos (A)						
2017	44	9.68	2.15	22.25	9.28	8.40 - 10.86
2018	94	10.75	2.34	21.77	11.00	9.42 - 12.16
2019	141	9.61	1.91	19.88	9.68	8.40 - 10.92
2020	220	10.09	2.10	20.84	10.24	8.74 - 11.36
Promedio ponderado promocional (B)						
2017	51	13.37	1.63	12.16	13.14	12.28 - 14.44
2018	107	13.50	1.56	11.56	13.40	12.47 - 14.56
2019	162	13.42	1.51	11.28	13.35	12.23 - 14.56
2020	226	14.15	1.37	9.67	14.14	13.30 - 15.10
Diferencia entre puntajes (A - B)						
2017	44	-3.87	1.36	35.05	-3.76	-4.36 a -3.07
2018	94	-2.90	1.43	49.29	-2.70	-3.58 a -1.91
2019	141	-4.05	1.15	28.30	-4.10	-4.83 a -3.19
2020	220	-4.13	1.49	36.06	-4.02	-5.02 a -3.01

DE: desviación estándar. CV: coeficiente de variación. Q1: cuartil 1. Q3: cuartil 3.

que no lo rindieron (año 2017: comparación de medianas (rango intercuartílico): 25 (0) versus 15 (16), U de Mann-Whitney (U) = 94.0, valor de  $p = 0.001$ ; año 2018: 25 (0) versus 16 (20), U = 345.0, valor de  $p = 0.001$ ; año 2019: 25 (0) versus 17 (6), U = 554.0, valor de  $p < 0.001$ ; y año 2020: 25 (0) versus 25 (17), U = 472.0, valor de  $p = 0.021$ ).

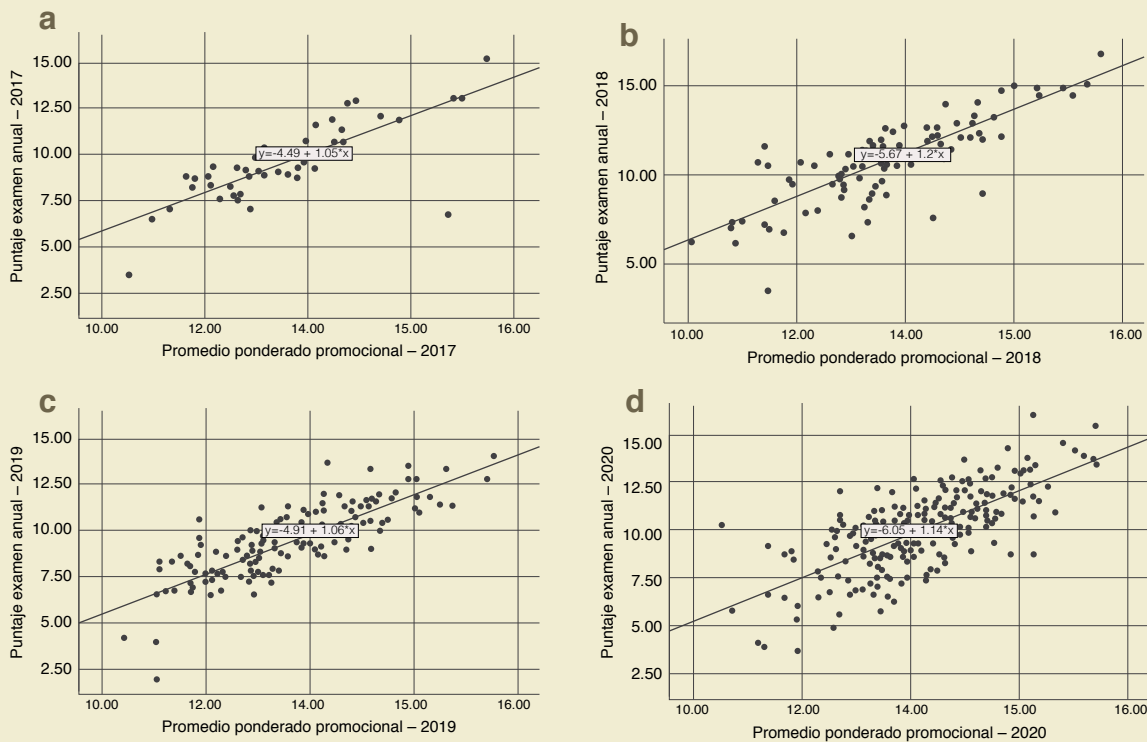
Las medias de los puntajes del EAM fluctuaron entre 9.61 para el 2019 a 10.75 para el 2018. El promedio ponderado tuvo un incremento de 0.78 puntos entre el 2017 y 2020, además el coeficiente de variación fue disminuyendo de 12.16% en 2017 a 9.67% en

2020. La variabilidad de los puntajes fue mayor para el EAM respecto al promedio ponderado (**tabla 2**).

En todos los años se encontró una correlación lineal directa significativa entre el promedio ponderado anual y el puntaje del EAM (**figura 1**). La mayor correlación se dio en el 2019 (rho de Spearman ( $r_s = 0.812$ )). En el análisis de correlación según sexo se mantuvieron correlaciones significativas. En dos años, 2017 y 2019, el coeficiente de correlación fue mayor entre mujeres (**tabla 3**).

Se encontraron correlaciones lineales significativas entre los puntajes de los EAM, los mayores coeficientes

**Figura 1.** Gráficos de dispersión entre los puntajes del examen anual de medicina basado en casos clínicos y el promedio ponderado promocional



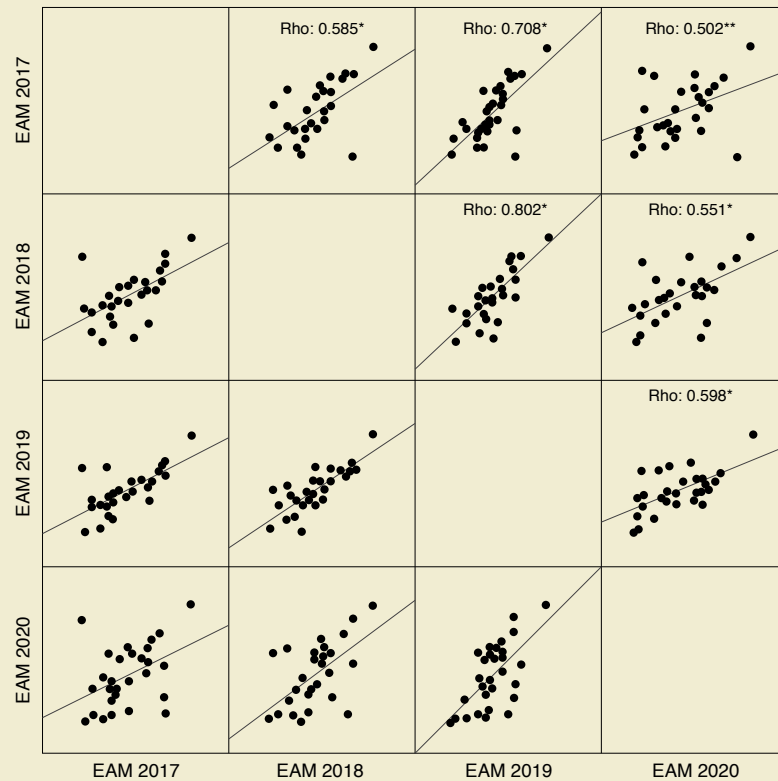
- a. Gráfico de dispersión para el 2017, ecuación de línea de tendencia (puntaje de examen anual =  $-4.49 + 1.05$  (promedio ponderado)).
- b. Gráfico de dispersión para el 2018, ecuación de línea de tendencia (puntaje de examen anual =  $-5.67 + 1.20$  (promedio ponderado)).
- c. Gráfico de dispersión para el 2019, ecuación de línea de tendencia (puntaje de examen anual =  $-4.91 + 1.06$  (promedio ponderado)).
- d. Gráfico de dispersión para el 2020, ecuación de línea de tendencia (puntaje de examen anual =  $-6.05 + 1.14$  (promedio ponderado)).

**Tabla 3.** Análisis de correlación lineal entre los puntajes del examen anual de medicina basado en casos clínicos y el promedio ponderado promocional por año y según sexo

Año	Grupo	Coefficiente de correlación por rangos de Spearman*	n
2017	Femenino	0.857	19
	Masculino	0.762	25
	Total	0.788	44
2018	Femenino	0.787	39
	Masculino	0.791	55
	Total	0.786	94
2019	Femenino	0.833	63
	Masculino	0.803	78
	Total	0.812	141
2020	Femenino	0.608	112
	Masculino	0.765	108
	Total	0.690	220

\*En todos los análisis de correlación lineal el valor de p fue <0.001.

**Figura 2.** Gráficos de dispersión y análisis de correlación entre los puntajes del examen anual de medicina basado en casos clínicos



EAM: examen anual de medicina basado en casos clínicos. \*Valor de  $p < 0.001$ . \*\*Valor de  $p = 0.001$ .

de correlación se identificaron entre los puntajes del 2018 y 2019 ( $r_s = 0.802$ , valor de  $p < 0.001$ ) y los puntajes del 2017 y 2019 ( $r_s = 0.708$ , valor de  $p < 0.001$ ) (figura 2). Los puntajes del EAM resultaron menores a los del promedio ponderado para todos los años de estudio. Se observó una tendencia que a mayor promedio de los puntajes menor diferencia entre los ellos. Solo un alumno en el EAM del 2018 logró un puntaje mayor que su respectivo promedio ponderado, no se identificaron valores concordantes (anexo 3).

## DISCUSIÓN

Se encontró que el puntaje de una evaluación sumativa escrita basada en casos clínicos está moderada o fuertemente correlacionado con el desempeño académico en términos de promedio ponderado. Algunos

estudios encontraron correlación entre la evaluación sumativa escrita basada en situaciones clínicas y el promedio ponderado, por ejemplo, en un estudio en 65 estudiantes de Corea del Sur se encontró un coeficiente de correlación de Pearson ( $r$ ) de 0.837 ( $p < 0.001$ ) entre el promedio ponderado y un examen de seis secciones y 400 preguntas de opción múltiple basado en el examen de licenciamiento médico de Corea<sup>5</sup>. Otro estudio en 119 estudiantes de los Estados Unidos de Norteamérica (EEUU), encontró una correlación significativa ( $r = 0.67$ ) entre el promedio ponderado y el resultado de un examen de certificación nacional para médicos que emplea preguntas cerradas de opción múltiple y única respuesta<sup>13</sup>. En 428 estudiantes en Australia se encontró que el 31.18% de la variabilidad ( $p < 0.01$ ) en el punta-

je de una evaluación basada en preguntas de opción múltiple diseñado para una evaluación sumativa fue atribuido al promedio ponderado<sup>12</sup>.

Nuestros resultados son consistentes con estudios que evaluaron la correlación entre el puntaje de una OSCE y el desempeño clínico. Si bien, la OSCE es otro tipo de evaluación sumativa, obtuvimos de manera análoga resultados que apoyan la existencia de correlación entre nuestra evaluación sumativa y el desempeño clínico medido a través del promedio ponderado. Una revisión sistemática encontró que 12 de 15 estudios mostraron una asociación significativa, y que el puntaje en la OSCE explicaba entre el 1.9% y 39.7% de la variabilidad en el desempeño clínico de los estudiantes de medicina<sup>2</sup>. Otra investigación en 302 estudiantes de Australia aplicó una herramienta de evaluación del desempeño clínico para médicos recién egresados, y encontró que el puntaje de dicha evaluación tuvo una correlación significativa ( $r = 0.257$ ,  $p < 0.001$ ) con el promedio ponderado y con el puntaje de un OSCE medido en el cuarto ( $r = 0.137$ ,  $p = 0.027$ ) y quinto año ( $r = 0.161$ ,  $p = 0.022$ ) de la carrera<sup>14</sup>.

La utilidad de las evaluaciones sumativas escritas para predecir el desempeño clínico es aún discutible con la evidencia disponible. Las evaluaciones con preguntas de opción múltiple y otras modalidades como preguntas de emparejamiento ampliado explicarían solo el 29.2% de la variación de un puntaje de desempeño clínico<sup>16</sup>, otro estudio encontró que un examen escrito tomado en el quinto año de la carrera no estuvo correlacionado con el desempeño clínico medido con la herramienta denominada *Junior Doctor Assessment Tool* (JDAT) ( $r = 0.076$ , valor de  $p = 0.148$ ), mientras que para el examen en el sexto año si hubo correlación ( $r = 0.178$ , valor de  $p = 0.014$ )<sup>14</sup>. La evidencia, incluyendo nuestros hallazgos, muestra correlaciones lineales directas significativas entre los resultados de evaluaciones sumativas escritas y el promedio ponderado<sup>5,12,13</sup>, más que con las evaluaciones de desempeño clínico.

En el presente análisis, el desempeño académico del estudiante fue medido a través del promedio ponderado, este puntaje ha sido empleado como criterio de selección de graduados para programas de posgrado o para fines de empleabilidad<sup>15</sup>. Además, es una medida objetiva y estable en el tiempo, por su

naturaleza acumulativa y ha sido empleada como un aproximado al desempeño académico en estudiantes de medicina<sup>16</sup>. En el Perú, entre el 2008 y 2010 se encontró que el promedio ponderado tuvo correlación ( $r_s = 0.582$ ,  $p < 0.001$ ) con el puntaje de un examen nacional de medicina, el cual busca medir los conocimientos de los estudiantes en su último año de formación profesional<sup>17</sup>. Para el 2013, dicha correlación también estuvo presente ( $r_s = 0.566$ ,  $p < 0.001$ )<sup>18</sup>.

Se empleó un examen escrito con preguntas de opción múltiple para fines de evaluación sumativa. Entre las limitaciones de esta modalidad esta la falta de aplicación de habilidades clínicas y su enfoque en el recuerdo de información<sup>19,20</sup>. También genera el efecto de señalización, situación en la que los alumnos pueden responder mediante el reconocimiento de la opción correcta, y que en ausencia de opciones no podrían responder<sup>1</sup>. Sin embargo, la novedad del examen descrito radica en su formulación, la cual emplea casos clínicos publicados como reportes de casos para brindarle contexto a las preguntas; de esta manera fue posible evaluar conocimientos, entendimiento, aplicación y análisis, es decir un mayor nivel de habilidades del pensamiento según la taxonomía de Bloom. Así, hemos podido evaluar la capacidad del estudiante para sintetizar información, así como su razonamiento clínico<sup>21</sup>.

La forma más frecuente de preguntas de opción múltiple es la de única mejor respuesta<sup>22</sup>, y fue la empleada en nuestra evaluación. Estas preguntas fueron formuladas agrupadas en función de un escenario clínico, cada pregunta aborda conocimientos aprendidos en asignaturas que el estudiante ha completado hasta el momento de rendir el EAM. De esta manera los alumnos rendirán progresivamente seis EAM durante la carrera universitaria, y en cada EAM se van incluyendo preguntas de una mayor variedad de asignaturas.

El EAM no es obligatorio, por ello, no todos los estudiantes realizaron la evaluación. Sin embargo, el porcentaje de estudiantes que dieron todos los exámenes posibles fue incrementándose. El principal determinante para realizar el examen fue haberse matriculado en el mayor número de créditos posibles. Los estudiantes que han completado todas las asignaturas tendrían un mayor nivel de motivación para evaluar su nivel de aprendizaje, así como para



identificar las brechas en su conocimiento y esperar una retroalimentación luego de la evaluación<sup>20,21</sup>; por el contrario, aquellos estudiantes que no se matricularon en la totalidad de créditos o llevaron solo algunos cursos evitaron rendir la evaluación.

Se encontraron resultados consistentes en la correlación entre el promedio ponderado y el puntaje del EAM en los cuatro años analizados, lo cual brinda evidencia robusta sobre una real correlación lineal directa entre dichos puntajes. Si se considera un coeficiente de correlación de 0.8 o más como fuertemente positivo y de 0.5 o más como moderadamente positiva<sup>23</sup>, en todos los años se encontró por lo menos una moderada correlación, que se mantuvo en ambos sexos. Otra característica relevante fue que los puntajes del EAM son menores a los promedios ponderados, este hallazgo se encontró en todos los años analizados. Este hallazgo podría explicarse por la naturaleza acumulativa del EAM y por la exigencia de evaluar la capacidad de síntesis y aplicación del conocimiento de parte del estudiante.

El estudio tiene limitaciones, en vista que el EAM no es obligatorio, no todos los alumnos rindieron la evaluación. El diseño del estudio no permitió determinar si el puntaje del promedio ponderado predice el puntaje en la evaluación sumativa escrita, o viceversa. Debido a que es un estudio de fuentes secundarias, se utilizó el dato disponible de promedio ponderado como un aproximado al desempeño académico. El estudio fue realizado en una sola universidad, los hallazgos no necesariamente serán generalizables a otros contextos.

A pesar de lo mencionado, el presente estudio presenta una novedosa experiencia de evaluación sumativa basada en casos clínicos descritos en reportes de casos de revistas científicas. Las preguntas son de opción múltiple, y formuladas en torno a una hipotética situación clínica. A nuestro conocimiento, esta submodalidad de evaluación sumativa escrita es la primera descrita en Perú y puede resultar novedosa para el contexto latinoamericano.

## CONCLUSIONES

Se ha demostrado en cada uno de los años analizados que el puntaje obtenido en el EAM se correlaciona de manera lineal directa con el promedio ponderado, también se ha encontrado que existe co-

relación lineal directa en los puntajes de EAM. Se recomienda continuar la investigación en el desarrollo y validación de nuevas estrategias escritas de evaluación sumativa basadas en casos clínicos que puedan complementar al examen clínico estructurado objetivo, la evaluación directa de las habilidades clínicas, entre otras.

## CONTRIBUCIÓN INDIVIDUAL

- FRRR: Concepción y diseño del estudio, limpieza de base de datos, análisis de datos, redacción del primer borrador, revisiones de las versiones del manuscrito, revisión y aprobación de la versión final. Asume responsabilidad de lo publicado.
- CAGV: Concepción y diseño del estudio, análisis de datos, revisiones de las versiones del manuscrito, revisión y aprobación de la versión final. Asume responsabilidad de lo publicado.

## AGRADECIMIENTOS

Ninguno.

## PRESENTACIONES PREVIAS

Datos preliminares del presente análisis fueron presentados como una experiencia innovadora en el VI Congreso Internacional de Educación Médica organizado por la Asociación Peruana de Facultades de Medicina del 25 al 29 de mayo del 2021.

## FINANCIAMIENTO

El estudio es financiado por la Universidad de Piura.

## CONFLICTOS DE INTERESES

Los autores declaran que tienen vínculo laboral con la Universidad de Piura. 🔍

## REFERENCIAS

1. Epstein RM. Assessment in Medical Education. *N Engl J Med.* 2007;356(4):387-96. doi: 10.1056/NEJMra054784
2. Terry R, Hing W, Orr R, Milne N. Do coursework summative assessments predict clinical performance? A systematic review. *BMC Med Educ.* 2017;17(1):40. doi:10.1186/s12909-017-0878-3
3. Kibble JD. Best practices in summative assessment. *Adv Physiol Educ.* 2017;41(1):110-9. doi:10.1152/advan.00116.2016
4. Bauer D, Lahner F-M, Schmitz FM, Guttormsen S, Huwendiek S. An overview of and approach to selecting appropriate patient representations in teaching and summative assessment in medical education. *Swiss Med Wkly.* 2020;150:w20382. doi: 10.4414/smw.2020.20382

5. Park WB, Kang SH, Myung SJ, Lee Y-S. Does Objective Structured Clinical Examinations Score Reflect the Clinical Reasoning Ability of Medical Students? *Am J Med Sci*. 2015;350(1):64-7. doi:10.1097/MAJ.0000000000000420
6. Schuwirth LWT, van der Vleuten CPM. Different written assessment methods: what can be said about their strengths and weaknesses? *Med Educ*. 2004;38(9):974-9. doi:10.1111/j.1365-2929.2004.01916.x
7. Hift RJ. Should essays and other “open-ended”-type questions retain a place in written summative assessment in clinical medicine? *BMC Med Educ*. 2014;14(1):249. doi:10.1186/s12909-014-0249-2
8. Pepple DJ, Young LE, Carroll RG. A comparison of student performance in multiple-choice and long essay questions in the MBBS stage I physiology examination at the University of the West Indies (Mona Campus). *Adv Physiol Educ*. 2010;34(2):86-9. doi:10.1152/advan.00087.2009
9. 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):49. doi:10.1186/1472-6920-7-49
10. Eftekhar H, Labaf A, Anvari P, Jamali A, Sheybaee-Moghaddam F. Association of the pre-internship objective structured clinical examination in final year medical students with comprehensive written examinations. *Med Educ Online*. 2012;17(1):15958. doi:10.3402/meo.v17i0.15958
11. Dong T, Saguil A, Artino AR, Gilliland WR, Waechter DM, Lopreaito J, et al. Relationship Between OSCE Scores and Other Typical Medical School Performance Indicators: A 5-Year Cohort Study. *Mil. Med*. 2012;177(9S):44-6. doi:10.7205/MILMED-D-12-00237
12. David M, Eley D, Schafer J, Davies L. Risk assessment of student performance in the International Foundations of Medicine Clinical Science Examination by the use of statistical modeling. *Adv Med Educ Pract*. 2016;7:653-660. doi: 10.2147/AMEP.S122841
13. Brown G, Imel B, Nelson A, Hale LS, Jansen N. Correlations Between PANCE Performance, Physician Assistant Program Grade Point Average, and Selection Criteria. *J Physician Assist Educ*. 2013;24(1):42-4. doi:10.1097/01367895-201324010-00006
14. Carr SE, Celenza A, Puddey IB, Lake F. Relationships between academic performance of medical students and their workplace performance as junior doctors. *BMC Med Educ*. 2014;14(1):157. doi:10.1186/1472-6920-14-157
15. Richardson M, Abraham C, Bond R. Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychol. Bull*. 2012;138(2):353-87. doi:10.1037/a0026838
16. Chisholm-Burns MA, Berg-Poppe P, Spivey CA, Karges-Brown J, Pithan A. Systematic review of noncognitive factors influence on health professions students' academic performance. *Adv in Health Sci Educ*. 2021;26(4):1373-445. doi:10.1007/s10459-021-10042-1
17. Huamani C, Gutiérrez C, Mezones-Holguín E. Correlación y concordancia entre el examen nacional de medicina y el promedio ponderado universitario: análisis de la experiencia peruana en el periodo 2007 - 2009. *Rev Peru Med Exp Salud Pública*. 2011;28(1):62-71.
18. Moreno-Loaiza M, Mamani-Quispe PV, Moreno-Loaiza O. Diferencias entre el promedio ponderado promocional y la nota del Examen Nacional de Medicina en el proceso de adjudicación de plazas para el Servicio Rural y Urbano Marginal de Salud. *Rev Peru Med Exp Salud Pública*. 2013;30(2): 360-362.
19. Khan MU, Aljarallah BM. Evaluation of Modified Essay Questions (MEQ) and Multiple Choice Questions (MCQ) as a tool for Assessing the Cognitive Skills of Undergraduate Medical Students. *Int J Health Sci (Qassim)*. 2011;5(1):39-43.
20. Douthit NT, Norcini J, Mazuz K, Alkan M, Feuerstein M-T, Clarfield AM, et al. Assessment of Global Health Education: The Role of Multiple-Choice Questions. *Front Public Health*. 2021;9:640204. doi:10.3389/fpubh.2021.640204
21. Vuma S, Sa B. A comparison of clinical-scenario (case cluster) versus stand-alone multiple choice questions in a problem-based learning environment in undergraduate medicine. *J Taibah Univ Med Sci*. 2016;12(1):14-26. doi: 10.1016/j.jtumed.2016.08.014
22. Campbell DE. How to write good multiple-choice questions: Annotation. *J. Paediatr. Child Health*. 2011;47(6):322-5. doi:10.1111/j.1440-1754.2011.02115.x
23. Zou KH, Tuncali K, Silverman SG. Correlation and Simple Linear Regression. *Radiology*. 2003;2.

## ANEXO 1

Ejemplo de preguntas de opción múltiple basadas en casos clínicos extraídos de reportes de casos publicados. El ejemplo fue extraído en el 2019, de cuadernillo aplicado a alumnos del tercer año de la carrera de medicina humana. Se observa el enunciado extraído de la sección de descripción clínica del caso, no se incluye el texto de

la introducción o discusión del artículo. No incluimos las figuras que pueden visualizarse en la fuente.

Se presentan 14 preguntas que abordan conocimientos y aplicación del conocimiento en distintas asignaturas y relativas al contexto clínico.

### CASO 05

**Fuente:** Aldama O, Domínguez L, Rivelli V, Gorostiaga G, Mendoza G, Aldama A. Síndrome LEOPARD (lentiginosis progresiva cardiomiopática). *Dermatología CMQ*. 2009; 7(2):113-116.

Paciente de sexo masculino, de 14 años, remitido por el servicio de cardiología infantil al servicio de dermatología de un hospital nacional para evaluación de manchas oscuras, asintomáticas, presentes desde el nacimiento a nivel de rostro, cuello y tronco, y que paulatinamente fueron aumentando en número y tamaño. El paciente es portador de cardiopatía congénita compensada, medicado con un antihipertensivo (enalapril) como remodelador miocárdico. El paciente tiene una hermana gemela no idéntica, sin lesiones, y niega otros familiares con lesiones similares.

Al examen físico se observaron máculas hipercrómicas color café con bordes definidos e irregulares, las mayores son de hasta 1 cm de diámetro. Así también otras cuatro manchas mayores de 1.5 cm y con una tonalidad más clara. Se tomó muestra de ambas lesiones para estudio histológico.

Al examen general presentó facies característica con hipertelorismo, raíz nasal ancha e implantación baja de orejas (**figura 1**); además de baja estatura, ausencia de desarrollo genital y puberal, lordosis y escápulas aladas (**figura 3**). El estudio oftalmológico (a excepción del hipertelorismo) y el otorrinolaringológico fueron normales. Los estudios complementarios informaron cardiomegalia observada en la radiografía de tórax. En el electrocardiograma se encontró ondas T negativas de V1 a V3 y el ecocardiograma informó cardiopatía hipertrófica obstructiva con estenosis subaórtica severa.

La histopatología de las manchas de ambas tonalidades fue semejante, mostrando acantosis de las crestas

interpapilares, hiperplasia de los melanocitos a nivel de la capa basal, aumento del pigmento melánico basal y en estrato de Malpighi hasta la capa córnea. Se observan melanófagos en la dermis, así como un discreto infiltrado linfocitario perivascular.

Por las características del caso, el diagnóstico fue el de síndrome LEOPARD, llamado también lentiginosis progresiva cardiomiopática o síndrome de los léntigos múltiples. El acrónimo LEOPARD significa: Lentiginosis, Electrocardiograma con trastornos de la conducción, hipertelorismo Ocular, estenosis Pulmonar, Anormalidades genitales, Retardo del crecimiento, sordera neurosensorial (del inglés Deafness).

### CUESTIONARIO DEL CASO 05

1. El hipertelorismo que presenta el paciente significa que hay un aumento en la distancia que separa las paredes internas de la órbita derecha e izquierda. De las siguientes estructuras, señale cuál corresponde a la pared interna de la cavidad o fosa orbitaria:
  - a) Hueso unguis.
  - b) Cara orbitaria del hueso frontal.
  - c) Ala mayor del esfenoides.
  - d) Hueso nasal.
  - e) Ala menor del esfenoides.
2. Muchos síndromes congénitos se asocian a malformaciones cardíacas, como es el caso de la paciente. La tetralogía de Fallot es una cardiopatía caracterizada por 4 malformaciones: Estenosis pulmonar, comunicación interventricular, cabalgamiento de aorta e Hipertrofia ventricular. ¿Cuál de las siguientes premisas es fundamental en el desarrollo del corazón?

- a) La cascada de marcación se inicia con la arginina y la serotonina.
- b) En la formación del corazón debe haber una rotación de 90 grados del bulbo cordis.
- c) Vía de Señalización y proceso de lateralidad.
- d) Las aurículas deben formarse del campo cardiogénico secundario.
- e) Las células de la cresta neural no intervienen en ningún proceso en la formación del corazón.

3. El ecocardiograma del paciente informo cardiopatía hipertrófica obstructiva con estenosis subaórtica severa. Una válvula aórtica normal se compone de tres velos finos. En la estenosis aórtica congénita, estos velos están fusionados o son demasiado gruesos. ¿Qué tejido recubre la porción fibrosa y esponjosa de las válvulas?

- a) Epitelio plano simple.
- b) Epitelio cúbico simple.
- c) Conjuntivo laxo.
- d) Conjuntivo denso.
- e) Adiposo unilocular.

4. En este caso se observa hiperplasia de los melanocitos que son células productoras de pigmentos, ubicados en la capa basal del epitelio y en determinadas regiones de la dermis. Este pigmento (melanina) se produce por

- a) Secreción citocrina.
- b) Actividad lisosómica.
- c) Síntesis de eumelanina.
- d) Oxidación de la tirosina.
- e) Producción de premelanosomas.

5. La embriología humana define como organogénesis el período comprendido entre la tercera a la octava semana de desarrollo. Primero se produce el paso de embrión bilaminar a trilaminar (gastrulación). El sistema nervioso proviene del

- a) Ectodermo.
- b) Mesodermo.
- c) Endodermo.
- d) Amnios.
- e) Somite.

6. Sobre sordera neurosensorial. Marcar verdadero (V) o falso (F) según corresponda:
- ( ) Causadas por enfermedades u obstrucciones en el oído exterior o medio.
  - ( ) Las pérdidas auditivas afectan a todas las frecuencias del oído de manera uniforme.
  - ( ) A menudo afectan a la habilidad de la persona para escuchar ciertas frecuencias más que otras.
  - ( ) Ocurre por daños en las células ciliadas del oído interno o los nervios que lo abastecen.
  - ( ) El origen real del nervio vestibulococlear se encuentra en ganglios periféricos (ganglio de Corti y Scarpa).

- a) V V F F F
- b) F F F V V
- c) V V V F F
- d) F F V V V
- e) F V V V F

7. El paciente se incorporó a una investigación que tuvo por objetivo describir las características clínicas y moleculares en pacientes con síndrome LEOPARD y compararlos con pacientes con otros síndromes neurocardiofaciocutáneos. Para la redacción de esta investigación usted sugiere se sigan las pautas de la guía.

- a) CONSORT
- b) PRISMA
- c) STARD
- d) STROBE
- e) CARE

8. Parte de los resultados de la investigación en la cual fue enrolado el paciente se muestran a continuación:

Manifestación clínica	Síndrome LEOPARD (n = 20)	Otros síndromes neurocardiofaciocutáneos (n = 80)
Manchas café con leche	80%	90%
Pectus excavatum	25%	20%

¿Qué análisis usted sugeriría realizar con estos datos?

- a) Prueba exacta de Fisher.
- b) Prueba chi cuadrado de independencia.
- c) Prueba de bondad de ajuste.

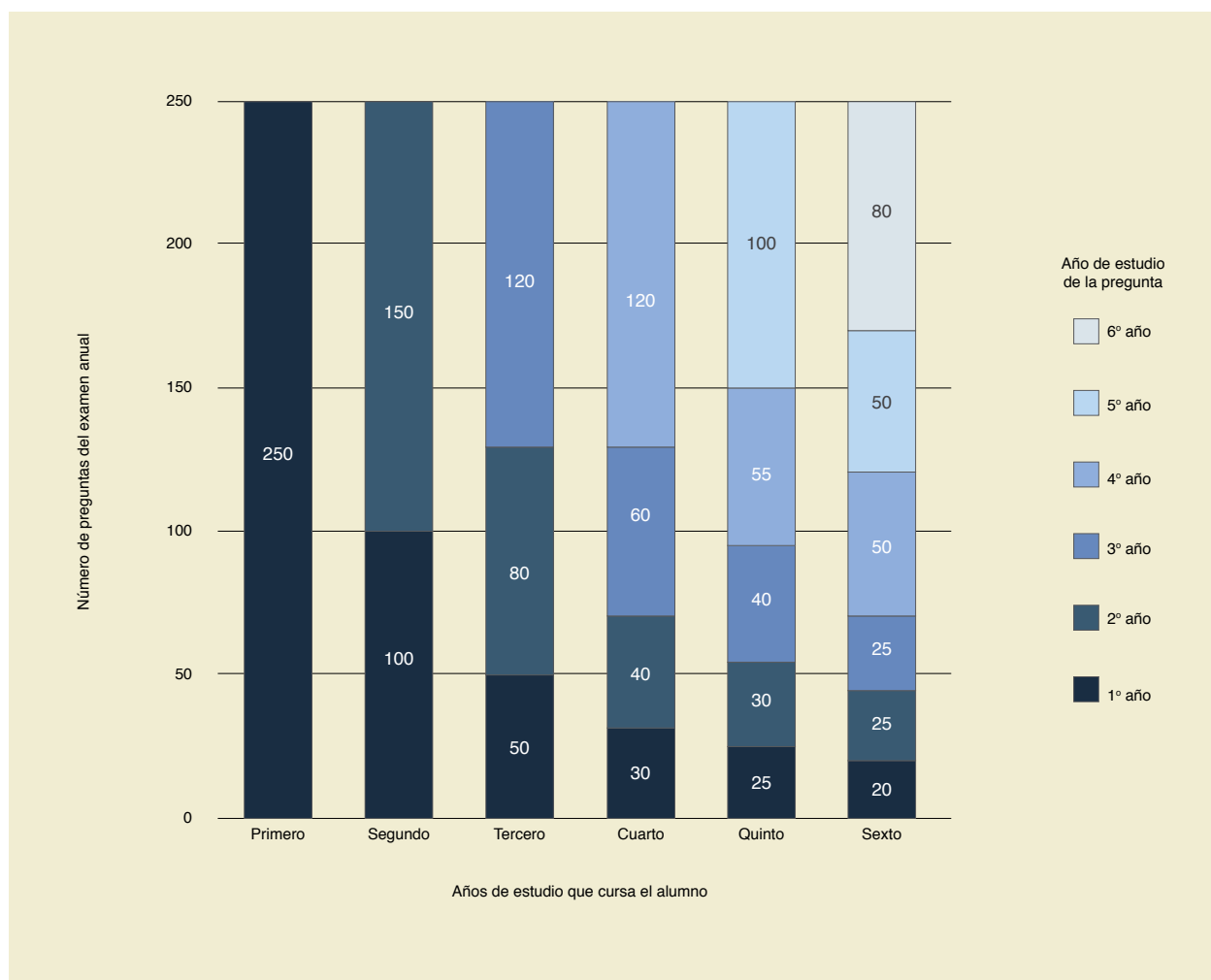
- d) Prueba de McNemar.  
e) Prueba de Cochran-Mantel-Haenszel.
9. En el caso de un léntigo, usted lo puede diferenciar de un nevo porque:
- a) En el nevo los queratinocitos son basales.  
b) En el nevo hay proliferación de melanocitos basales.  
c) En el nevo hay nidos o tecas.  
d) En el nevo hay disminución de los nevos.  
e) En el nevo siempre hay pigmento en dermis.
10. En el síndrome de LEOPARD puede haber una miocardiopatía hipertrofica. La hipertrofia se caracteriza por:
- a) Aumento del tamaño de las organelas.  
b) Cambio del fenotipo de las células.  
c) Alteraciones cromosómicas.  
d) Double hit.  
e) Eosinofilia del citoplasma.
11. En el caso se describe que el paciente posee clínica de una estenosis subaórtica. Dentro de los hallazgos del examen clínico, se espera encontrar:
- a) Soplo sistólico de eyección que puede o no acompañarse de frémito.  
b) Presión de pulso aumentada.  
c) Pulso saltón o hiperdinámico.  
d) Cianosis.  
e) Soplo diastólico de eyección sin frémito.
12. En caso de que el paciente hubiese presentado estenosis pulmonar como parte de la cardiopatía congénita, esta presentación clínica no se caracterizaría por:
- a) Hipertrofia del ventrículo izquierdo.  
b) Cianosis central sobre todo al momento de actividad física.  
c) Hipertrofia del ventrículo derecho.  
d) Dificultad respiratoria progresiva.  
e) Soplo sistólico a nivel de 2do espacio intercostal con línea paraesternal izquierda.
13. Se relata en el caso que el paciente posee máculas hiperpigmentadas de color café en la superficie del cuerpo. ¿Cuál de las siguientes entidades no formaría parte del diagnóstico diferencial?
- a) Neurofibromatosis.  
b) Pitiriasis versicolor.  
c) Síndrome de Marfan.  
d) Lentigo solar.  
e) Queratosis seborreica macular.
14. Algunos pacientes pueden desarrollar sordera neurosensorial como parte de las manifestaciones clínicas. ¿Cuáles son las pruebas que corroboran este diagnóstico si el paciente tuviera hipoacusia neurosensorial del lado derecho?
- a) Test de Weber que lateraliza al lado izquierdo, Rinne (+) lado derecho.  
b) Test de Weber indiferente, Rinne (+) lado derecho.  
c) Test de Weber que lateraliza al lado derecho, Rinne (+) lado derecho.  
d) Test de Weber indiferente, Rinne (+) lado izquierdo.  
e) Test de Weber que lateraliza al lado derecho, Rinne (-) lado derecho.

## ANEXO 2

### Distribución de preguntas por año de estudio en el examen anual de medicina basado en casos clínicos

Los alumnos que culminan el primer año son evaluados con preguntas de asignaturas correspondientes a ese año. Los alumnos de segundo año son evaluados en las asignaturas del segundo año (150 preguntas) y del primer año (100 preguntas) y así sucesivamente. El número de preguntas es mayor para las asignaturas del último año que ha cursado el alumno. El número de preguntas por asignatura es proporcional al número de créditos. Las preguntas que son tomadas a los alumnos del segundo

año en adelante son un subgrupo de aquellas tomadas en los años anteriores. Por ejemplo: las 20 preguntas de asignaturas del primer año tomadas a los alumnos de sexto, son tomadas del grupo de 25 preguntas de asignaturas del primer año, tomadas a los alumnos de quinto año; estas 25 preguntas a su vez son tomadas del grupo de 30 preguntas de asignaturas del primer año, tomadas a los alumnos de cuarto año; y así sucesivamente.



### ANEXO 3

Gráficos de Bland y Altman para evaluar la concordancia entre el promedio del examen anual de medicina basado en casos clínicos, y el promedio ponderado versus la diferencia entre dichos puntajes, para cada año de estudio (2017-2020).

