

Comparación entre el rendimiento académico entre el modelo de exámenes escritos clásicos de la gestión 2022 y el modelo ajustado de la gestión 2023, obtenido por los estudiantes de 5° año de la cátedra de Medicina III

(Comparison of academic performance between the traditional written exam model of the 2022 term and the adjusted model of the 2023 term, achieved by 5th-year students in the Medicine III course)

Autores: Dr. Omar Rodas Funes^a, Dra. Mónica Guzmán Mendoza^b, Biotec. Melisza Valdez Gonzales^c

^a Docente Titular C – FMENT – UMSA.

^b Docente Emérito – FMENT – UMSA.

^c Biotecnóloga del Hospital de Clínicas de La Paz.

Investigación educativa original, Departamento de Medicina de la Carrera de Medicina, Facultad de Medicina, Enfermería, Nutrición, Tecnología Médica.

Autor principal: Omar Rodas, dirección: Unidad de Hematología, Hospital de Clínicas de La Paz, Av. Saavedra #2245, Miraflores. Correo electrónico: omarrodas20@gmail.com, omarillo_rodas@hotmail.com

RESUMEN

La evaluación es parte fundamental del proceso de enseñanza aprendizaje en la formación médica, dentro de los tipos de evaluación se encuentra la formativa, a lo largo del proceso y la sumativa, al final del proceso; esta última es la más utilizada en la Cátedra de Medicina III de la Facultad de Medicina, enfermería, Nutrición y Tecnología Médica de la Universidad Mayor de San Andrés, teniendo indicadores de aprobación muy bajos, debido a diversos factores, pero poco estudiados, entre ellos los exámenes con muchas preguntas para ser resueltas en 60 minutos.

El objetivo es comparar el rendimiento académico entre el modelo de exámenes escritos clásicos de la gestión 2022 y el modelo ajustado de la gestión 2023, obtenido por los estudiantes de 5° año de la cátedra de Medicina III, que permitió evaluar la reducción del tiempo de los exámenes parciales y finales, y de la cantidad de preguntas y reactivos por pregunta.

Es un estudio analítico, correlacional retrospectivo; evaluando exámenes entre parciales (4 parciales por cada estudiante) y finales (1 final por cada estudiante) de 405 estudiantes regulares (216 en la gestión 2022 y 189 en la gestión 2023), que rotaron por la Cátedra de Medicina III de ambas gestiones.

Los resultados muestran que reducir el tiempo de los exámenes parciales de 60 a 45 minutos y de 70 a 35 preguntas por examen, generan mayor rendimiento académico por parte de los estudiantes ($p > 0.05$), a diferencia de los exámenes finales donde no se halló esta correlación.

Se concluye que la reducción del tiempo de los exámenes parciales y de la cantidad de preguntas, logra la mayor probabilidad de aprobación, debiendo evaluarse el caso de los exámenes finales.

Palabras clave.

Exámenes, parciales, finales, preguntas, reactivos, tiempo.

SUMMARY:

Assessment is a fundamental part of the teaching-learning process in medical education. Among the types of assessment are formative evaluation, throughout the process, and summative evaluation, at the end of the process. This last way of evaluation is the most used in Medicine III Chair of the Faculty of Medicine, Nursing, Nutrition, and Medical Technology at the Universidad Mayor de San Andrés (UMSA), with very low approval rates, due to various factors that have been little studied, including exams with many questions to be completed in 60 minutes.

The main objective was to compare the academic performance between the traditional written exam model of the 2022 term and the adjusted model of the 2023 term, achieved by the 5th-year students of the Medicine III course, which allowed for the evaluation of the reduction in time for the midterm and final exams, as well as the number of questions and items per question.

It is a retrospective, correlational, analytical study; evaluating exams between midterms (4 midterms per student) and finals (1 final per student) of 405 regular students (216 in the 2022 term and 189 in the 2023 term), who rotated through the Chair of Medicine III during both terms.

The results show that reducing the time of midterm exams from 60 to 45 minutes and from 70 to 35

questions per exam generates greater success among students ($p > 0.05$), unlike the final exams where this correlation was not found.

It is concluded that reducing the time of midterm exams and the number of questions achieves a higher likelihood of passing, while the case of final exams should be further evaluated.

Keywords.

Exams, midterms, finals, questions, items, time.

Among the results, it was found that reducing the time of midterm exams from 60 to 45 minutes and reducing the number of questions from 70 to 35 per exam increases the probability of student passing ($p > 0.05$), unlike final exams where this correlation was not found.

It is concluded that reducing the time of midterm exams and the number of questions achieves a higher probability of passing, while the case of final exams should be further evaluated.

Keywords Exams, midterms, finals, questions, items, time.

A. INTRODUCCION.

Una de las metas para alcanzar la calidad educativa universitaria, y generar la mejora continua idóneos en las instituciones educativas superiores, está en la predisposición a querer mejorar los diferentes procesos que rigen el currículo, desde la planificación, hasta la evaluación (1).

La evaluación en la educación universitaria médica es un proceso integral que busca no sólo medir el conocimiento y la comprensión de los estudiantes, sino también desarrollar habilidades prácticas y competencias profesionales. Esta evaluación tiene al menos dos dimensiones de medición (exámenes o valoraciones cuantitativas, y en juicio de valor que es la interpretación del evaluador respecto a la o las mediciones). En la actualidad se centra en la retroalimentación continua (mediciones del proceso gradual y modificaciones paso a paso), la formación continua (educación y actualización permanente) y la adaptación a las necesidades del campo profesional. Debiendo ser inclusiva, equitativa y centrada en el estudiante, considerando su desempeño en múltiples contextos y situaciones (2).

En la práctica, la evaluación se resume en formativa y sumativa, siendo que esta última utiliza instrumentos que permiten medir el aprendizaje y el desempeño de los estudiantes al final de un periodo académico. Entre estos instrumentos que tienen el fin de verificar si el evaluado puede obtener el mejor puntaje, clasifica el rendimiento del estudiante como bueno, regular o malo; entre ellos se tiene ampliamente conocido al examen teórico de opción múltiple, y sus grados de dificultad (3).

La Facultad de Medicina, Enfermería, Nutrición y Tecnología Médica (FMENT), dependiente de la Universidad Mayor de San Andrés (UMSA), realiza una evaluación sumativa, mayormente, valorando las capacidades de cada estudiante al final de cada materia, cátedra y de cada año. Específicamente, en la cátedra de Medicina III de 5° año, dependiente de la Carrera de Medicina, la evaluación sumativa se basa en exámenes teóricos llamados parciales con 70 preguntas para ser resueltas en 60 minutos, y exámenes finales con 60 preguntas para ser resueltas en 60 minutos igualmente, ambas con 5 reactivos

en cada pregunta. Cada examen parcial tiene un valor de 35 puntos, teniendo que resolverse 4 que son promediados al final de la rotación, y cada examen final tiene un valor de 30 puntos que se resuelve 1 sola vez al final de la rotación o bloque que dura 13 semanas. Esto significa que entre parciales y el final, cada estudiante obtiene el 65% de la nota total de la rotación, restando 35 puntos promediados que se obtiene de las prácticas hospitalarias de cada materia que compone Medicina III.

Esta forma de evaluación teórica, usado en la cátedra de Medicina III, tenía una tasa alta de reprobación, que, junto al tipo de construcción dificultosa de las preguntas, condicionaron calificaciones bajas, no habiéndose investigado científicamente la correlación entre estas y la forma de evaluación.

Durante la Pandemia COVID-19 se tuvo que modificar el proceso educativo en las universidades, teniendo que recurrir al uso de la virtualidad exclusivamente como estrategia didáctica, resultando en desventajas educativas tanto para estudiantes, profesionales como para el mismo sistema que no estaba preparado para este cambio. Es por esto que en la Facultad de Medicina, la cátedra de Medicina III fue pionera en implementarlo, usando recomendaciones internacionales para la educación virtual, específicamente evaluación virtual, se redujo la cantidad de preguntas por examen parcial a 35 por el mismo valor y del examen final a 30 por el mismo valor, y se redujo la cantidad de respuestas (reactivos) a 4 por pregunta, obteniendo mejores calificaciones por parte de los estudiantes. En Postpandemia, se aceptó mantener dichas recomendaciones al modelo presencial de evaluación establecida en Medicina III.

Por este motivo, se realizó esta investigación con los exámenes realizados por los estudiantes de 5° año de la cátedra de Medicina III de nuestra facultad, durante las gestiones 2022 y 2023, durante los cuales se aplicaron los ajustes a los exámenes parciales y finales, buscando la evidencia científica necesaria para contribuir a la mejora del proceso de calidad en formación de nuestros estudiantes.

Cabe resaltar que no existen estudios con este objetivo o similares, en nuestra población universitaria de salud, hallándose trabajos similares internacionales,

especialmente en aquellos que buscan el mejoramiento de la calidad formativa basados en un currículo de competencias o híbridos.

B. OBJETIVOS

1. Objetivo general:

Comparar el rendimiento académico entre el modelo de exámenes escritos clásicos de la gestión 2022 y el modelo ajustado de la gestión 2023, obtenido por los estudiantes de 5° año de la cátedra de Medicina III.

2. Objetivos específicos o resultados esperados:

a) Caracterizar los exámenes parciales escritos entre ambas gestiones, ajustando el tiempo de 60 minutos en la gestión 2022, y 45 minutos en 2023.

b) Caracterizar los exámenes finales entre ambas gestiones, ajustando el tiempo de 60 minutos en la gestión 2022, y 45 minutos en 2023.

c) Relacionar los tiempos alcanzados en los exámenes parciales y finales con la aprobación de éstos, en ambas gestiones.

d) Correlacionar los hallazgos de calificaciones, tiempo y aprobación logrados en los exámenes parciales y finales de las gestiones 2022 y 2023, en base a modelos estadísticos apropiados.

3. Materiales

- Listas oficiales de estudiantes de 5° año que cursan Medicina III, proporcionado por el Departamento de Medicina de nuestra facultad. Cada examen parcial y final realizado por cada estudiante incluido en el presente trabajo.

- Acceso a las calificaciones de los exámenes parciales y finales de las gestiones 2022 y 2023, proporcionado por la Jefatura de la Cátedra de Medicina III.

- Uso de programas estadísticos como: SPSS versión 21 y EXCEL versión 365, para tabular y analizar la información, así como generar gráficos, realizando además correlación de Spearman y Regresión lineal simple.

4. Métodos

a. Tipo de estudio

Estudio es analítico, correlacional, transversal, retrospectivo.

b. Hipótesis

El modelo ajustado de exámenes escritos (45 minutos) comparado al modelo clásico de exámenes escritos (60 minutos) mejora el rendimiento académico de los estudiantes de 5° año de la cátedra de Medicina III.

c. Metodología

Se utilizó como muestra, el total de los exámenes escritos realizados por estudiantes que cumplieron con los criterios de inclusión:

- Modelo clásico de exámenes escritos de la gestión 2022: de 216 estudiantes (4 exámenes parciales y 1 examen final de cada uno, total de 1080 exámenes): 2° gran rote de 109 estudiantes; 3° gran rote de 107 estudiantes (no fueron incluidos los es-

tudiantes del 1° gran rote debido a que no se pudo modificar el examen como instrumento de recolección de información).

- Modelo ajustado de exámenes escritos de la gestión 2023: de 189 estudiantes (4 exámenes parciales y 1 examen final de cada uno, total de 945 exámenes), como prueba piloto modificando ciertas características: 2° gran rote 94 estudiantes; 3° gran rote 95 estudiantes (no fueron incluidos los estudiantes del 1° gran rote debido a que no se pudieron modificar las hojas de respuestas a la prueba piloto.

- Se contó con la Aprobación de la FMENT, establecida a través de una resolución facultativa.

d. Criterios de Inclusión y Exclusión

I. Criterios de Inclusión.

- Todos los estudiantes que dieron sus exámenes parciales y finales en las gestiones 2022 y 2023.

II. Criterios de Exclusión.

- Todos los estudiantes que abandonaron la Cátedra de Medicina III desde el inicio o a lo largo de las gestiones 2022 y 2023.

- Todos los estudiantes que no dieron uno o más exámenes parciales o final, durante las gestiones 2022 y 2023.

- Todos los estudiantes a los cuales no se pudieron aplicar los instrumentos de evaluación de esta investigación.

5. ETICA

Protección de la privacidad y confidencialidad: El uso de exámenes ya elaborados garantiza que la información obtenida no afecte a los participantes, se consideró el manejo de los datos confidenciales de la privacidad de los estudiantes, cumpliéndose con los principios éticos y normativas locales e internacionales.

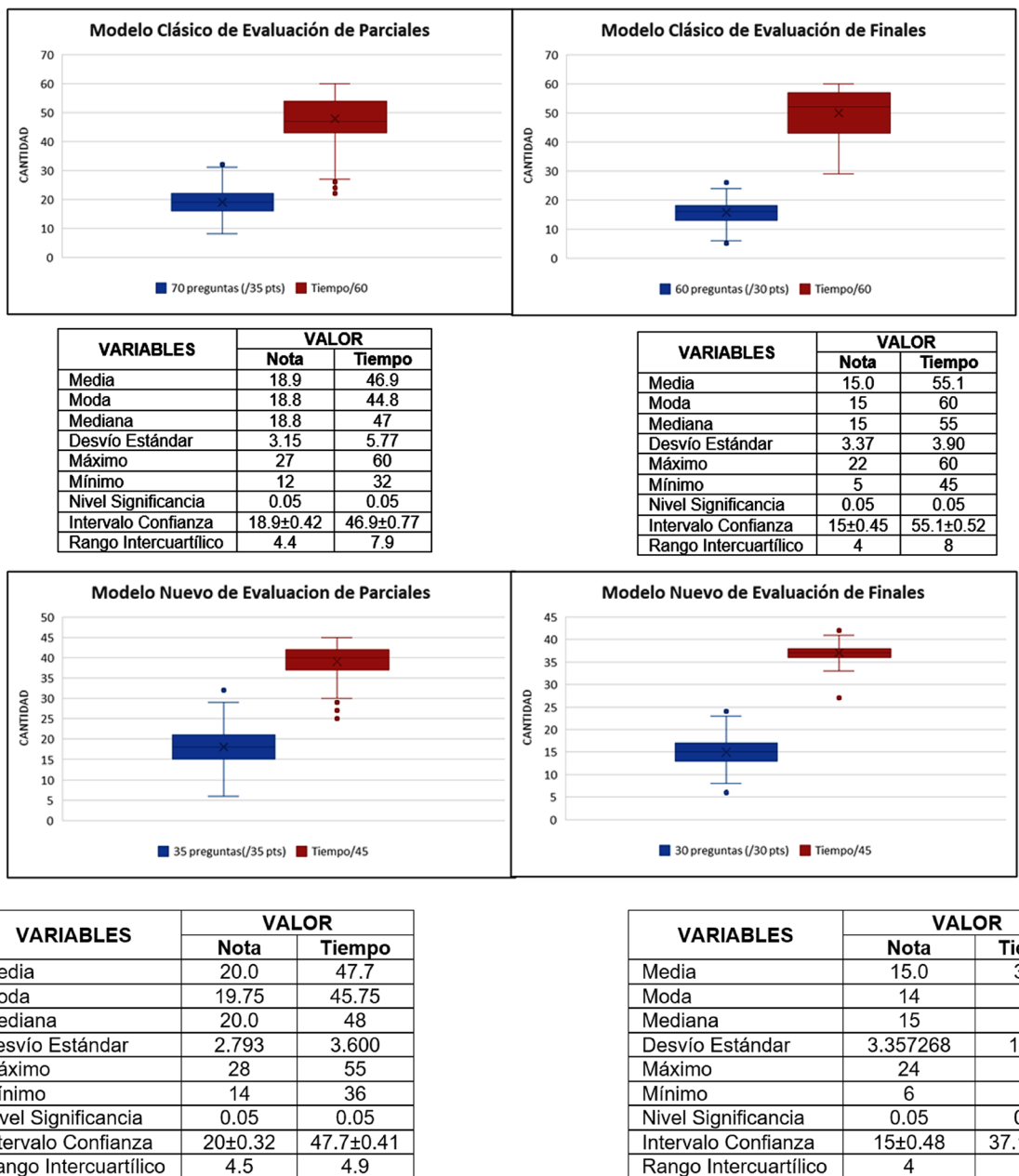
Consentimiento y uso de datos existentes: Dado que los datos provienen de calificaciones registradas en una fuente administrativa, se anuló el riesgo de invasión a la privacidad individual descartando alguna posibilidad de identificación personal.

RESULTADOS

Se muestran los resultados obtenidos de los exámenes parciales escritos y posteriormente de los exámenes finales escritos de ambos modelos evaluativos.

En el caso del Modelo clásico de exámenes parciales escritos con 70 preguntas para resolverse en 60 minutos, y exámenes finales con 60 preguntas para resolverse en 60 minutos, de la gestión 2022; comparados a aquellas obtenidas de los exámenes parciales con 35 preguntas para resolverse en 45 minutos, y exámenes finales con 30 preguntas para resolverse en 45 minutos, de la gestión 2023. (Figura 1)

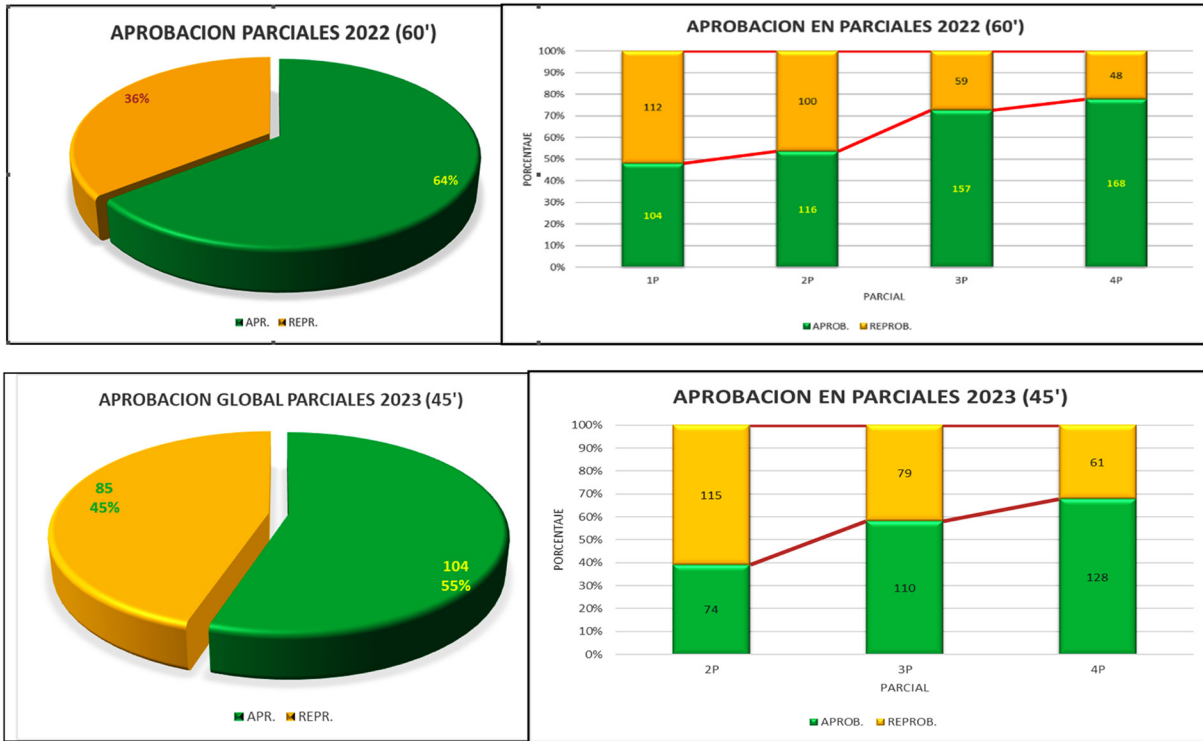
Figura 1. Comparación global entre el modelo clásico de exámenes (gestión 2022) y el modelo ajustado de exámenes (gestión 2023).



Fuente propia

- a. Datos de los exámenes parciales completados a los 60 minutos de la gestión 2022.
- El promedio de calificaciones fue de 18.9 ± 0.42 (IC 95%) sobre 35 puntos, con extremos que oscilaron entre 8 y 32 puntos.
 - El tiempo para completar dichos exámenes fue de 46.9 ± 0.77 (IC 95%) minutos, con extremos que oscilaron entre 22 y 60 minutos.
 - La tasa de aprobación promedio fue de 64% (138 de 216 estudiantes) del total de estudiantes evaluados. (Figura 2)
 - En la evaluación del rango intercuartílico, se observó que en el caso de la calificación se obtuvo un valor de 4.4 que se interpreta como el rango de calificaciones dentro del cual se encuentra el 50% central de los datos; en el caso del tiempo se obtuvo un valor de 7.9 que se interpreta como el rango de tiempo dentro del cual se encuentra el 50% central de los datos. (Figura 1)
- Figura 2. Tasa de aprobación en el modelo clásico de exámenes parciales (gestión 2022) y el modelo ajustado de exámenes (gestión 2023).

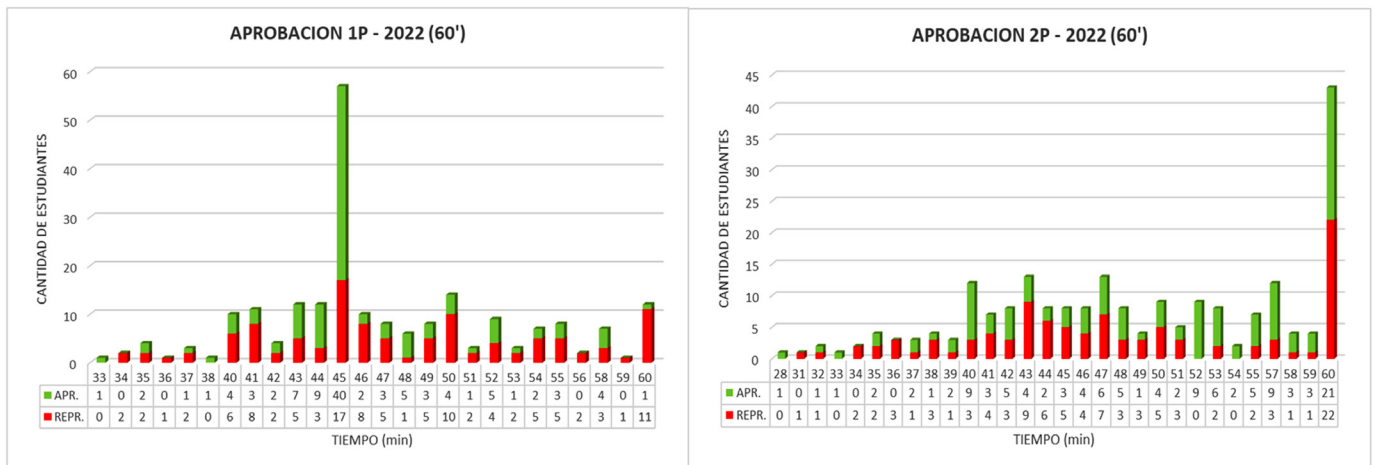
Figura 2. Tasa de aprobación en el modelo clásico de exámenes parciales (gestión 2022) y el modelo ajustado de exámenes (gestión 2023).



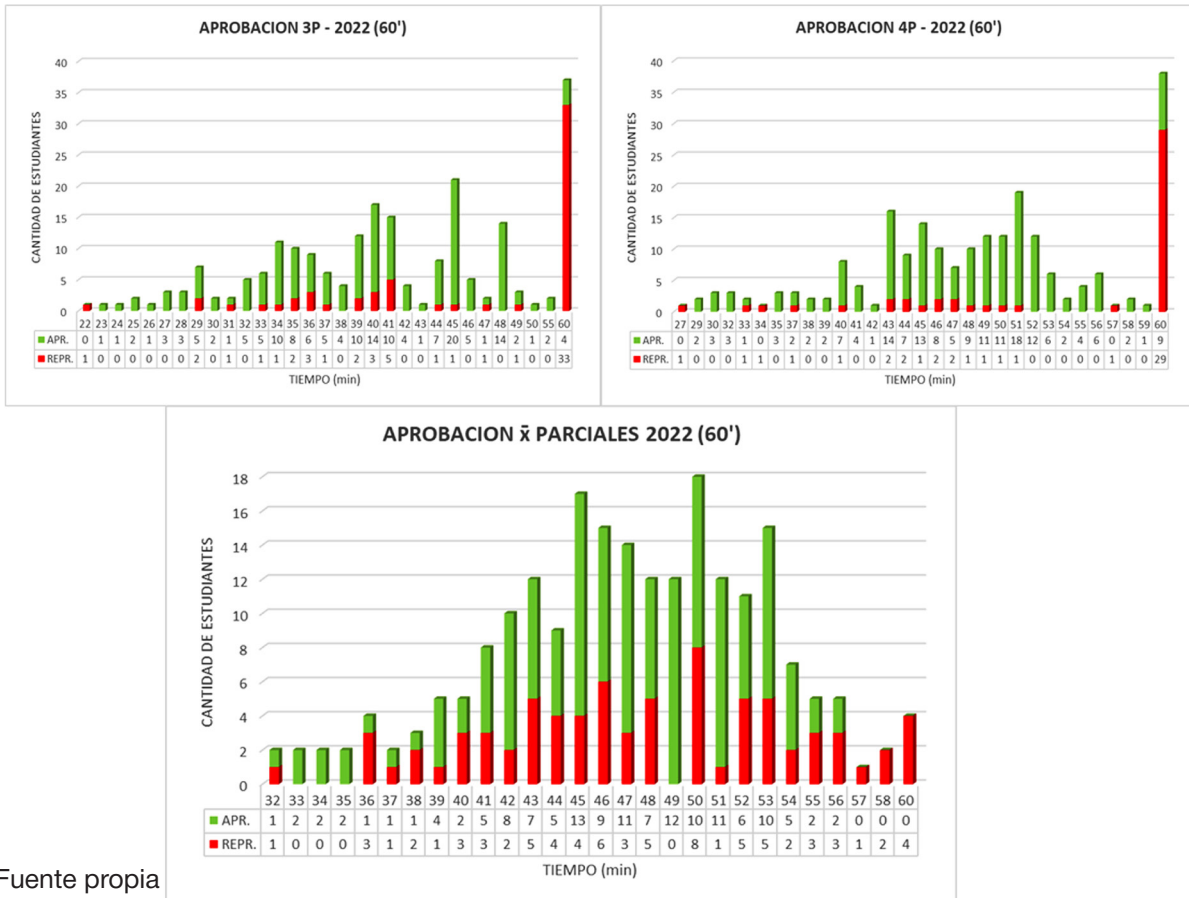
Fuente propia

- En la evaluación cruzada de Frecuencias entre calificación obtenida de cada estudiante y tiempo requerido para resolver cada parcial, se encontró: para el primer parcial un incremento gradual de aprobación entre el minuto 40 y el minuto 52, con un pico significativo de aprobación al minuto 45; para el segundo parcial un incremento gradual de aprobación entre el minuto 40 y el minuto 57, con un pico significativo de aprobación al minuto 52; para el tercer parcial un incremento gradual de aprobación entre el minuto 27 y el minuto 48, con un pico significativo de aprobación al minuto 45; y para el cuarto parcial un incremento gradual de aprobación entre el minuto 40 y el minuto 52, con dos picos significativos de aprobación al minuto 43 y 51. (Figura 3)

Figura 3. Relación entre calificaciones obtenidas en el tiempo de duración cada examen parcial/global y el rendimiento en el modelo clásico de exámenes parciales de la gestión 2022.



Fuente propia



Fuente propia

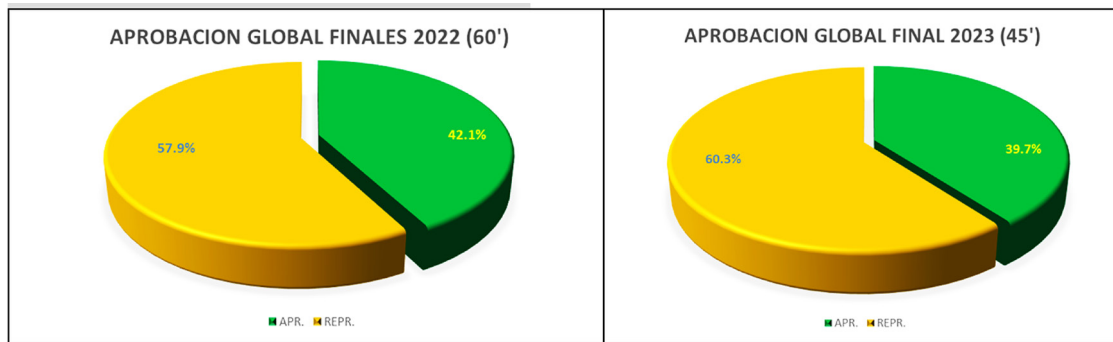
- En la evaluación cruzada de Frecuencias entre los promedios de la calificación obtenida de cada estudiante y el promedio del tiempo requerido para resolver los parciales, hubo un incremento gradual de aprobación entre el minuto 40 y el minuto 56, con dos picos significativos de aprobación al minuto 45 y 51. (Figura 3)
- En la Correlación de Spearman (por tratarse de distribuciones asimétricas) de cada parcial por separado, se observó una asociación significativa e inversa de ambas variables (calificación de parcial y tiempo de parcial), confirmándolo en el primer, tercer y cuarto parciales ($p < 0.05$), siendo excepcional en el segundo parcial no siendo significativo ($p = 0.13$). (Tabla 1)

			Correlations							
			Calificaciones del Primer Parcial	Tiempo del Primer Parcial	Calificaciones del Segundo Parcial	Tiempo del Segundo Parcial	Calificaciones del Tercer Parcial	Tiempo del Tercer Parcial	Calificaciones del Cuarto Parcial	Tiempo del Cuarto Parcial
Spearman's rho	Calificaciones del Primer Parcial	Correlation Coefficient	1.000	-.158 [*]	.596 ^{**}	.243 ^{**}	.340 ^{**}	.152 [*]	.373 ^{**}	.093
		Sig. (2-tailed)	.	.020	<.001	<.001	<.001	.026	<.001	.171
		N	216	216	216	216	216	216	216	216
	Tiempo del Primer Parcial	Correlation Coefficient	-.158 [*]	1.000	-.069	.174 [*]	-.060	.178 ^{**}	-.152 [*]	.158 [*]
		Sig. (2-tailed)	.020	.	.314	.011	.380	.009	.026	.021
		N	216	216	216	216	216	216	216	216
	Calificaciones del Segundo Parcial	Correlation Coefficient	.596 ^{**}	-.069	1.000	.105	.468 ^{**}	.083	.420 ^{**}	.067
		Sig. (2-tailed)	<.001	.314	.	.126	<.001	.223	<.001	.327
		N	216	216	216	216	216	216	216	216
	Tiempo del Segundo Parcial	Correlation Coefficient	.243 ^{**}	.174 [*]	.105	1.000	-.082	.424 ^{**}	-.068	.465 ^{**}
		Sig. (2-tailed)	<.001	.011	.126	.	.232	<.001	.317	<.001
		N	216	216	216	216	216	216	216	216
	Calificaciones del Tercer Parcial	Correlation Coefficient	.340 ^{**}	-.060	.468 ^{**}	-.082	1.000	-.412 ^{**}	.541 ^{**}	-.270 ^{**}
		Sig. (2-tailed)	<.001	.380	<.001	.232	.	<.001	<.001	<.001
		N	216	216	216	216	216	216	216	216
	Tiempo del Tercer Parcial	Correlation Coefficient	.152 [*]	.178 ^{**}	.083	.424 ^{**}	-.412 ^{**}	1.000	-.179 ^{**}	.562 ^{**}
		Sig. (2-tailed)	.026	.009	.223	<.001	<.001	.	.008	<.001
		N	216	216	216	216	216	216	216	216
	Calificaciones del Cuarto Parcial	Correlation Coefficient	.373 ^{**}	-.152 [*]	.420 ^{**}	-.068	.541 ^{**}	-.179 ^{**}	1.000	-.388 ^{**}
		Sig. (2-tailed)	<.001	.026	<.001	.317	<.001	.008	.	<.001
		N	216	216	216	216	216	216	216	216
	Tiempo del Cuarto Parcial	Correlation Coefficient	.093	.158 [*]	.067	.465 ^{**}	-.270 ^{**}	.562 ^{**}	-.388 ^{**}	1.000
		Sig. (2-tailed)	.171	.021	.327	<.001	<.001	<.001	<.001	.
		N	216	216	216	216	216	216	216	216

Fuente propia

- En la Correlación de Spearman del promedio de ambas variables (calificaciones y tiempos de parciales), se observó una asociación no significativa e inversa de ambas variables ($p=0.98$).
- En la Correlación de Spearman entre el promedio del tiempo de resolución y la Aprobación, se observó una asociación no significativa y directa de ambas variables ($p=0.16$).
- En la Correlación de Spearman de cada parcial, entre el tiempo y la Aprobación, se observó una asociación significativa y directa, en el primer y cuarto parcial ($p<0.05$), hallando una asociación no significativa en el segundo y tercer parcial ($p=0.55$ y 0.29 respectivamente).
- b. Datos de los exámenes finales completados a los 60 minutos de la gestión 2022.
- El promedio de calificación fue de 15 ± 0.45 (IC 95%) sobre 35 puntos; con extremos que oscilaron entre 6 y 29 puntos.
- El tiempo para completar dichos exámenes fue de 55.1 ± 0.52 (IC 95%) minutos, con extremos que oscilaron entre 45 y 60 minutos.
- El porcentaje de aprobación global de exámenes finales fue de 42.1% (91 de 216 estudiantes) del total de estudiantes evaluados. (Figura 4)
- En la evaluación del rango intercuartílico, se observó que en el caso de la calificación se obtuvo un valor de 4 que se interpreta como el rango de calificaciones dentro del cual se encuentra el 50% central de los datos; en el caso del tiempo se obtuvo un valor de 8 que se interpreta como el rango de tiempo dentro del cual se encuentra el 50% central de los datos. (Figura 1)

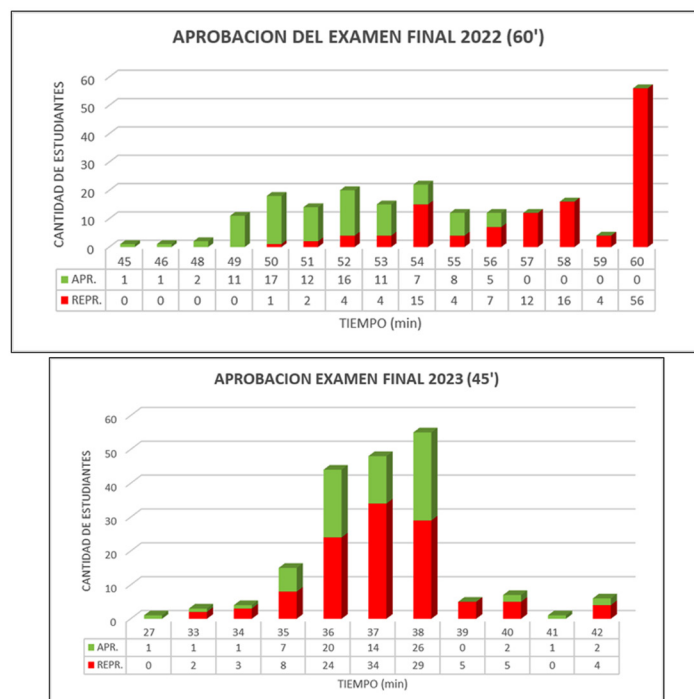
Figura 4. Tasa de aprobación en el modelo clásico de exámenes finales (gestión 2022) y el modelo ajustado de exámenes finales (gestión 2023).



Fuente propia

- En la evaluación cruzada de Frecuencias entre calificación obtenida de cada estudiante y tiempo requerido para resolver el examen final, se observó un incremento gradual de aprobación entre el minuto 49 y el minuto 53, con un pico significativo de aprobación al minuto 50. (Figura 5)

Figura 5. Relación entre calificaciones obtenidas en el tiempo de duración de cada examen final y el rendimiento en el modelo clásico de exámenes finales de la gestión 2022 y el modelo ajustado de exámenes finales de la gestión 2023.



Fuente propia

- En la Correlación de Spearman (por tratarse de distribuciones asimétricas) de los exámenes finales, se observó una asociación fuertemente significativa e inversa de ambas variables (calificación y tiempo de los exámenes finales) ($p < 0.01$). Cuando se sometió a una regresión lineal para establecer el nivel de correlación, ésta era fuerte y cierta en el modelo ($R = 0.861$), con una correlación significativa en ella ($p < 0.01$). (Tabla 2)

Tabla 2. Correlación de Spearman y Regresión lineal simple entre calificaciones y tiempo de los exámenes finales al minuto 60, gestión 2022.

Correlations				
		Calificación final		Tiempo de examen
Spearman's rho	Calificación final	Correlation Coefficient	1.000	-.903**
		Sig. (2-tailed)	.	<.001
		N	216	216
	Tiempo de examen	Correlation Coefficient	-.903**	1.000
		Sig. (2-tailed)	<.001	.
		N	216	216

** . Correlation is significant at the 0.01 level (2-tailed).

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.861 ^a	.742	.741	1.718

a. Predictors: (Constant), Tiempo de examen

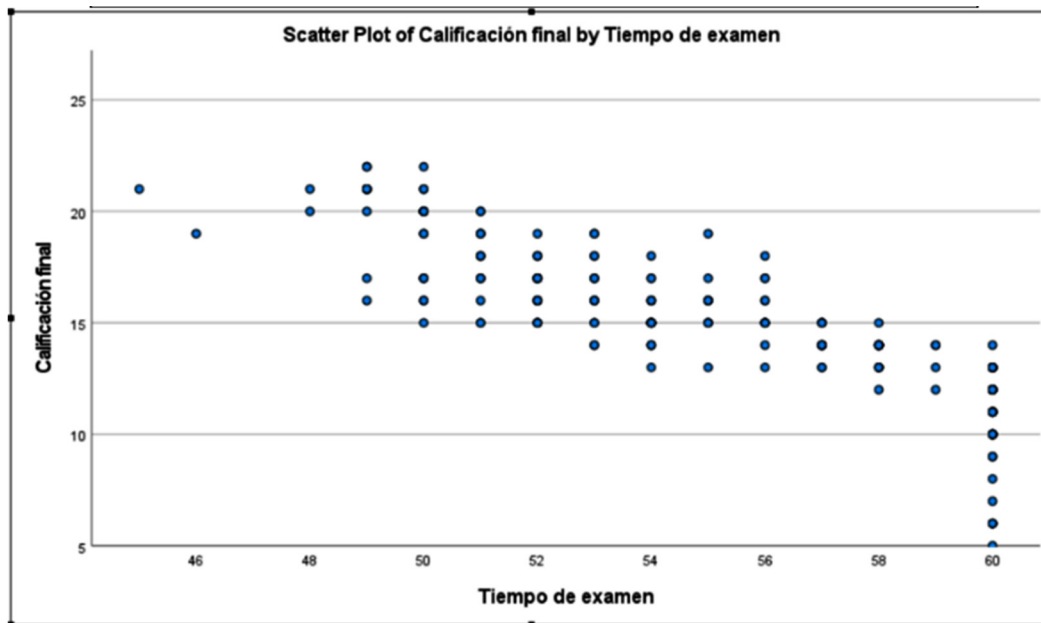
ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1814.276	1	1814.276	614.884	<.001 ^b
	Residual	631.428	214	2.951		
	Total	2445.704	215			

a. Dependent Variable: Calificación final
b. Predictors: (Constant), Tiempo de examen

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	55.969	1.658		33.761	<.001
	Tiempo de examen	-.744	.030	-.861	-24.797	<.001

a. Dependent Variable: Calificación final

Fuente propia



- En la Correlación de Spearman entre el tiempo de resolución y la posibilidad de Aprobación, se observó una asociación significativa y directa de ambas variables ($p < 0.01$). Cuando se sometió a una regresión lineal para establecer el nivel de correlación, ésta era fuerte y cierta en el modelo ($R = 0.755$), con una correlación significativa en ella ($p < 0.01$). (Tabla 3)

Tabla 3. Correlación de Spearman y Regresión lineal simple entre tiempo y aprobación de los exámenes al minuto 60, gestión 2022.

Correlations				
			Tiempo de examen	Aprobación de examen
Spearman's rho	Tiempo de examen	Correlation Coefficient	1.000	.753**
		Sig. (2-tailed)	.	<.001
		N	216	216
	Aprobación de examen	Correlation Coefficient	.753**	1.000
		Sig. (2-tailed)	<.001	.
		N	216	216

**. Correlation is significant at the 0.01 level (2-tailed).

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.755 ^a	.570	.568	.325

a. Predictors: (Constant), Tiempo de examen

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	30.032	1	30.032	283.996	<.001 ^b
	Residual	22.630	214	.106		
	Total	52.662	215			

a. Dependent Variable: Aprobación de examen
b. Predictors: (Constant), Tiempo de examen

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-3.697	.314		-11.780	<.001
	Tiempo de examen	.096	.006	.755	16.852	<.001

a. Dependent Variable: Aprobación de examen

Fuente propia

En el caso del Modelo ajustado de exámenes parciales y finales escritos con 35 preguntas, cada una con 4 reactivos, para resolverse en 45 minutos, de la gestión 2023; comparados a aquellas obtenidas en el modelo clásico. (Figura 1)

c. Datos de los exámenes parciales completados a los 45 minutos de la gestión 2023.

- Los exámenes parciales tuvieron una duración ajustada de 45 minutos, cuyos valores se midieron hasta ese rango. En el caso del primer gran rote de 98 estudiantes y del primer parcial del segundo gran rote de dicha gestión, no se ajustó el tiempo a 45 minutos por motivos administrativos, permaneciendo el tiempo en 60 minutos, por lo que se tuvo que excluir a todos estos estudiantes con considerarse variable confusional.

d. Datos de los exámenes parciales completados a los 45 minutos de la gestión 2023.

- El promedio de calificaciones fue de 20 ± 0.32 (IC 95%) sobre 35 puntos, con extremos que oscilaron entre 12 y 30 puntos.

- El tiempo para completar dichos exámenes fue de 47.7 ± 0.41 (IC 95%) minutos en promedio, con extremos que oscilaron entre 31 y 45 min.

- La tasa de aprobación global de exámenes parciales fue de 55% (104 de 189 estudiantes) del total de estudiantes evaluados. (Figura 2)

- La tasa de aprobación por cada parcial fue: del segundo parcial de 39.2% (74 estudiantes); del

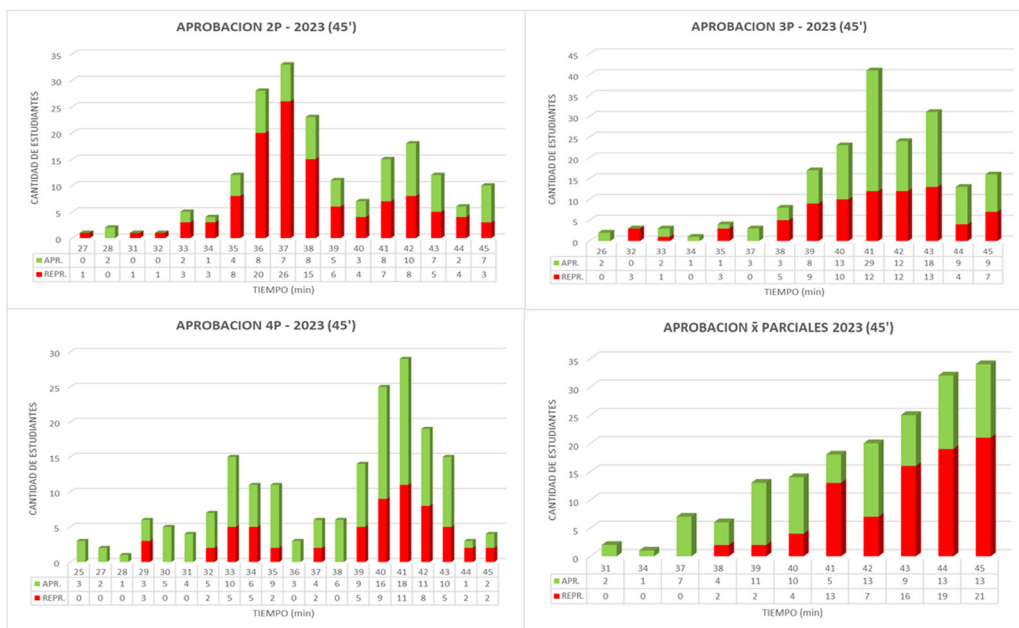
tercer parcial de 58.2% (110 estudiantes); y del cuarto parcial de 67.7% (128 estudian.). (Figura 2).

- En la evaluación del rango intercuartílico, se observó que en el caso de la calificación se obtuvo un valor de 4.5 que se interpreta como el rango de calificaciones dentro del cual se encuentra el 50% central de los datos; en el caso del tiempo se obtuvo un valor de 4.9 que se interpreta como el rango de tiempo dentro del cual se encuentra el 50% central de los datos. (Figura 1)

- En la evaluación cruzada de Frecuencias entre calificación obtenida de cada estudiante y tiempo requerido para resolver cada parcial, se obtuvo: en el segundo parcial, un incremento gradual de aprobación entre el minuto 35 y el minuto 45, con pico significativo de aprobación al minuto 42; para el tercer parcial, un incremento gradual de aprobación entre el minuto 39 y el minuto 45, con pico significativo de aprobación al minuto 41; para el cuarto parcial, un incremento gradual de aprobación entre el minuto 30 y el minuto 43, con dos picos significativos de aprobación al minuto 33 y 41. (Figura 6)

- En la evaluación cruzada de Frecuencias entre los promedios de la calificación obtenida de cada estudiante y el tiempo requerido para resolver los parciales, se observó un incremento gradual de aprobación entre el minuto 37 y el minuto 45, con dos picos significativos de aprobación al minuto 39 y 42. (Figura 6)

Figura 6. Relación entre calificaciones obtenidas en el tiempo de duración de cada examen parcial y el rendimiento en el modelo clásico de exámenes finales de la gestión 2022 y el modelo ajustado de exámenes finales de la gestión 2023.



Fuente propia

- En la Correlación de Spearman (por tratarse de distribuciones asimétricas) de cada parcial por separado, se observó una asociación significativa e inversa de ambas variables (calificación de parcial y tiempo de parcial), confirmándolo en el segundo y cuarto parciales ($p < 0.05$), no siendo significativo en el tercer parcial ($p = 0.91$). (Tabla 4)

Tabla 4. Correlación de Spearman entre calificaciones y tiempos de cada examen parcial en el modelo ajustado de la gestión 2022.

			Correlations					
			Calificación 2P	Tiempo 2P	Calificación 3P	Tiempo 3P	Calificación 4P	Tiempo 4P
Spearman's rho	Calificación 2P	Correlation Coefficient	1.000	.219**	.338**	.165*	.525**	-.399**
		Sig. (2-tailed)	.	.002	<.001	.023	<.001	<.001
		N	189	189	189	189	189	189
Tiempo 2P	Correlation Coefficient	.219**	1.000	-.103	.509**	.149*	.027	
	Sig. (2-tailed)	.002	.	.158	<.001	.041	.712	
	N	189	189	189	189	189	189	
Calificación 3P	Correlation Coefficient	.338**	-.103	1.000	.009	.276**	-.098	
	Sig. (2-tailed)	<.001	.158	.	.907	<.001	.180	
	N	189	189	189	189	189	189	
Tiempo 3P	Correlation Coefficient	.165*	.509**	.009	1.000	.038	.023	
	Sig. (2-tailed)	.023	<.001	.907	.	.605	.754	
	N	189	189	189	189	189	189	
Calificación 4P	Correlation Coefficient	.525**	.149*	.276**	.038	1.000	-.276**	
	Sig. (2-tailed)	<.001	.041	<.001	.605	.	<.001	
	N	189	189	189	189	189	189	
Tiempo 4P	Correlation Coefficient	-.399**	.027	-.098	.023	-.276**	1.000	
	Sig. (2-tailed)	<.001	.712	.180	.754	<.001	.	
	N	189	189	189	189	189	189	

** . Correlation is significant at the 0.01 level (2-tailed).
* . Correlation is significant at the 0.05 level (2-tailed).

Fuente propia

- En la Correlación de Spearman del promedio de ambas variables (calificaciones y tiempos de parciales), se observó una asociación significativa e inversa de ambas variables ($p < 0.001$). Cuando se sometió a una regresión lineal para establecer el nivel de correlación, ésta era intermedia y cierta en el modelo ($R = 0.373$), con una correlación significativa en ella ($p < 0.001$), traduciéndose que a menor tiempo mayor calificación. (Tabla 5)

Tabla 5. Correlación de Spearman y Regresión lineal entre calificaciones y tiempos de cada examen parcial en el modelo ajustado de la gestión 2022.

		Promedio Calificaciones		Promedio Tiempo	
Spearman's rho	Promedio Calificaciones	Correlation Coefficient	1.000		-.346**
		Sig. (2-tailed)		.	<.001
		N	189	189	
	Promedio Tiempo	Correlation Coefficient	-.346**	1.000	
		Sig. (2-tailed)	<.001		.
		N	189	189	

** . Correlation is significant at the 0.01 level (2-tailed).

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	233.113	1	233.113	30.151	<.001 ^b
	Residual	1445.779	187	7.731		
	Total	1678.892	188			

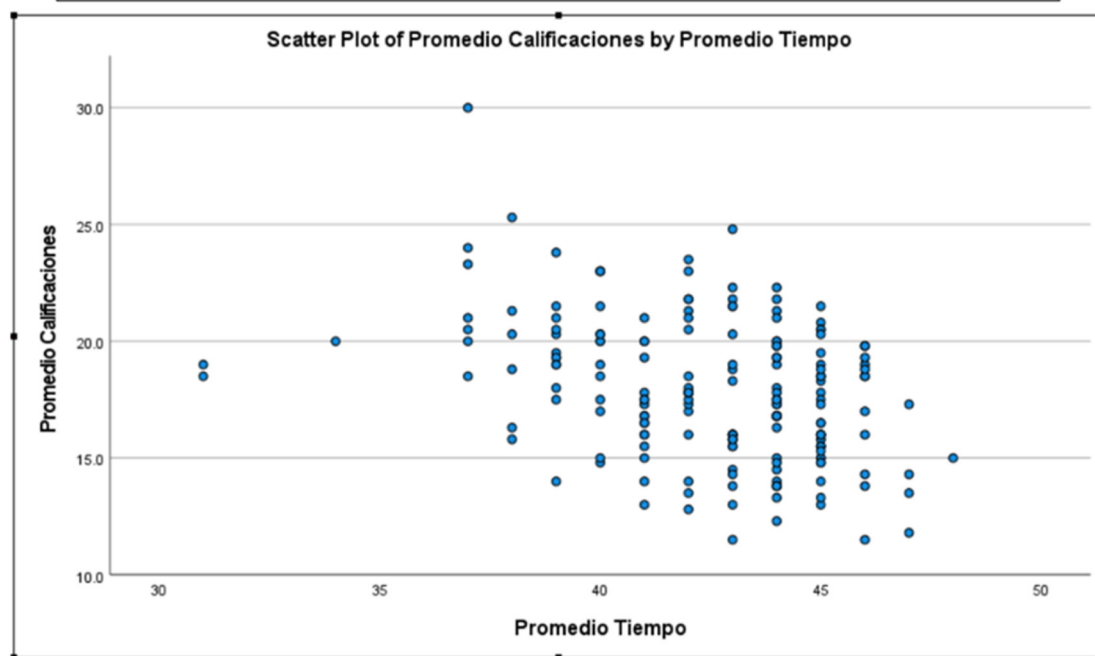
a. Dependent Variable: Promedio Calificaciones
b. Predictors: (Constant), Promedio Tiempo

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.373 ^a	.139	.134	2.7805

a. Predictors: (Constant), Promedio Tiempo

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	34.534	3.045		11.340	<.001
	Promedio Tiempo	-.393	.072	-.373	-5.491	<.001

a. Dependent Variable: Promedio Calificaciones



Fuente propia

- En la Correlación de Spearman entre el Promedio del tiempo de resolución y la posibilidad de Aprobación, se observó una asociación significativa y directa de ambas variables ($p < 0.001$). Cuando se sometió a una regresión lineal para establecer el nivel de correlación, ésta era intermedia y cierta en el modelo ($R = 0.310$), con una correlación significativa en ella ($p < 0.001$), en cuya grafica de correlación por puntos suplementaria se observa una correlación significativa homogénea y lineal, traduciéndose que los aprobados se encuentran entre los de menor tiempo. (Tabla 6)

Tabla 6. Correlación de Spearman y Regresión lineal entre Promedio de Tiempo y Aprobación de cada examen parcial al minuto 60, gestión 2022.

Correlations			Promedio de Tiempo	Promedio de Aprobacion
Spearman's rho	Promedio de Tiempo	Correlation Coefficient	1.000	.273**
		Sig. (2-tailed)	.	<.001
		N	189	189
	Promedio de Aprobacion	Correlation Coefficient	.273**	1.000
		Sig. (2-tailed)	<.001	.
		N	189	189

** . Correlation is significant at the 0.01 level (2-tailed).

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.310 ^a	.096	.091	.478

a. Predictors: (Constant), Promedio de Tiempo

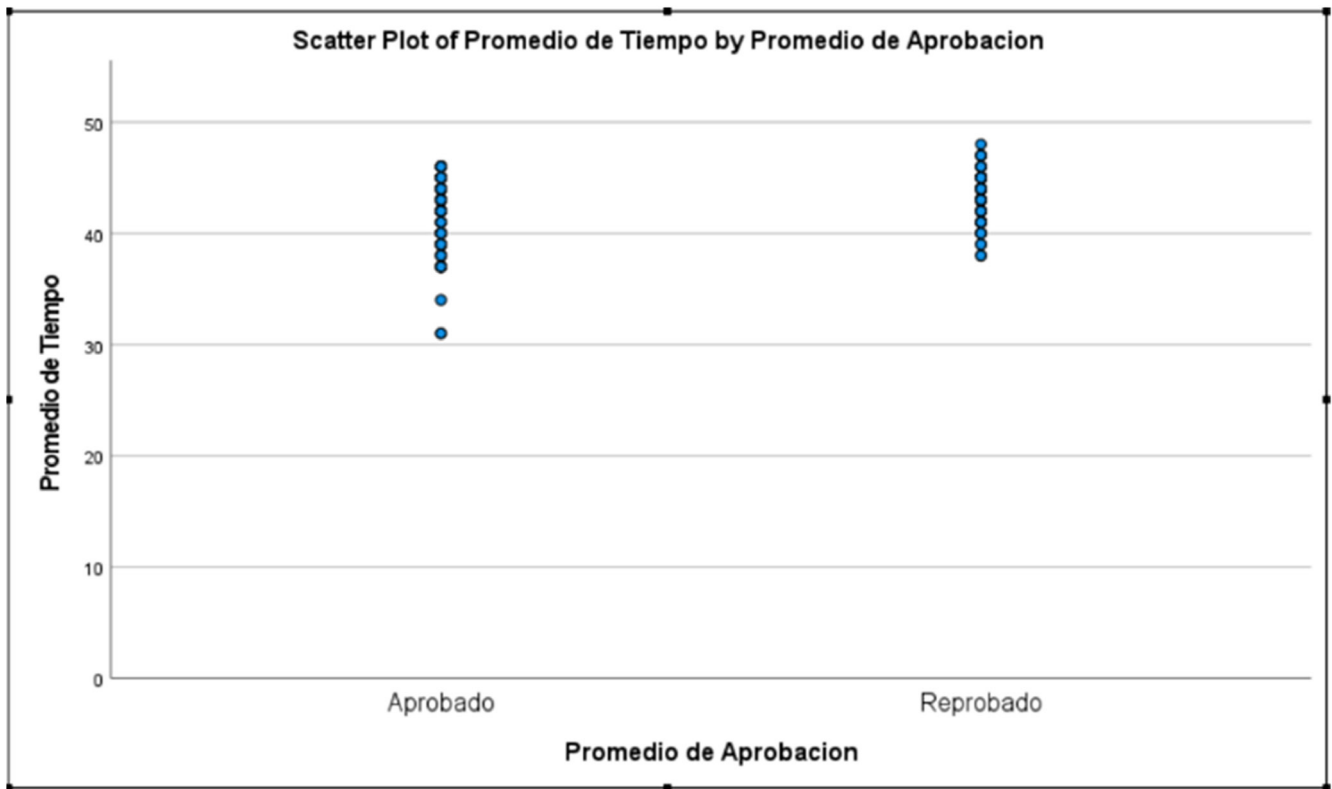
ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4.530	1	4.530	19.831	<.001 ^b
	Residual	42.718	187	.228		
	Total	47.249	188			

a. Dependent Variable: Promedio de Aprobacion
b. Predictors: (Constant), Promedio de Tiempo

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-.829	.523		-1.583	.115
	Promedio de Tiempo	.055	.012	.310	4.453	<.001

a. Dependent Variable: Promedio de Aprobacion

Fuente propia



Fuente propia

e. **Datos de los exámenes finales completados a los 45 minutos de la gestión 2023.**

- El **promedio de calificación fue de 15.2 ± 0.48 (IC 95%)** sobre 35 puntos, con extremos que **oscilaron entre 9 y 24 puntos**.
- El **tiempo para completar dichos exámenes fue de 37.1 ± 0.25 minutos**, con extremos que **oscilaron entre 27 y 38 min.**
- La **tasa de aprobación global** de exámenes finales fue de **39.7%** (75 de 189 estudiantes) del total de estudiantes evaluados. (Figura 4)
- En la evaluación del **rango intercuartílico**, se observó que en el caso de la **calificación se obtuvo un valor de 4** que se interpreta como el rango de calificaciones dentro del cual se encuentra el 50% central de los datos; en el caso del **tiempo se obtuvo un valor de 2** que se interpreta como el rango de tiempo dentro del cual se encuentra el 50% central de los datos. (Figura 1)
- En la evaluación cruzada de Frecuencias entre calificación obtenida de cada estudiante y tiempo requerido para resolver el examen final, **se observó un incremento gradual de aprobación entre el minuto 35 y el minuto 38, con un pico significativo de aprobación al minuto 36.**
- En la **Correlación de Spearman (por tratarse de distribuciones asimétricas) de los exámenes finales, no se observó una asociación significativa** de ambas variables (calificación y tiempo de los exámenes finales) ($p=0.976$).
- En la Correlación de Spearman entre el tiempo de resolución y la posibilidad de Aprobación, no se observó una asociación significativa de ambas variables ($p=0.879$).

6. DISCUSION

En este trabajo, se evaluó el rendimiento estudiantil en base a los exámenes parciales y finales de tipo opción múltiple realizados, modificando el modelo clásico de evaluación acortando la cantidad de preguntas a la mitad, acortando la cantidad de reactivos por cada pregunta, y acortando el tiempo total de 60 a 45 minutos, utilizando las observaciones preliminares del rendimiento de los últimos 5 años en dicha cátedra.

Este tipo de trabajo, nunca se hizo en la Cátedra de Medicina III, o en alguna otra que se haya registrado en nuestra facultad, de modo que se constituye en un estudio pionero, con el propósito de mejorar el proceso educativo a nivel de evaluación.

Fue observado en diversos estudios pioneros norteamericanos de los años 90 que el tiempo necesario para rendir los exámenes de tipo opción múltiple, depende de varios factores estableciendo desde hace varios años recomendaciones por la GENEVAL (Centro Nacional de Evaluación para la Educación Superior en México) donde se asoció un minuto para responder a una pregunta de opción múltiple con un grado intermedio de dificultad, teniendo que extenderse a 1 minuto con 30 segundos por pregunta de opción múltiple de tipo caso problémico corto (grado alta de dificultad) (GENEVAL, 2013).

Los exámenes parciales hasta ese momento en la gado L y Sánchez M, 2012), usaron la Teoría de Respuesta al Ítem (TRI) para evaluar la calidad de los reactivos y concluyeron que es posible reducir el número de preguntas manteniendo una alta confiabilidad, lo que permite ajustar la duración del examen sin comprometer su validez.

Todos los estudios sugieren que, en medicina, la calidad y representatividad de las preguntas son más importantes que la cantidad. Ajustar el número de ítems según el tiempo disponible puede ser viable sin afectar la evaluación, siempre que se mantenga la cobertura adecuada de contenidos y se utilicen herramientas psicométricas para validar el examen. La evidencia desarrollada en este trabajo, comparable a otros estudios similares internacionales, establece que la reducción del total de preguntas e incorporando casos clínicos concretos en los exámenes escritos parciales en la cátedra de Medicina III, mejora el rendimiento académico, cuando se comparan los dos modelos del presente trabajo con un nivel de evidencia suficiente para recomendar mantener el nuevo modelo.

Por tanto, este cambio de modelo evaluativo ha mostrado ser eficaz garantizando el rendimiento de los estudiantes al momento de realizar los exámenes parciales, no así en el caso de los exámenes finales, por lo que se recomienda mantener este modelo para ser evaluado a futuro.

7. CONFLICTO DE INTERESES

Los tres autores no tienen conflicto de interés para realizar esta investigación, no se percibió apoyo económico de empresas terciarias ni universitarias.

8. REFERENCIAS BIBLIOGRÁFICAS

Ramos L Mayra. La importancia de la evaluación y la acreditación como procesos de mejora de la calidad educativa en las Instituciones de Educación Superior. Revista Conexión. [Internet]. 2024 [cited 2022 Feb 6] año 5. Número 14.

Flores Hernández Fernando, Contreras Michel Nancy, Martínez González Adrián. Evaluación del aprendizaje en la educación médica. Rev. Fac. Med. (Méx.) [revista en la Internet]. 2012 Jun [citado 2022 Feb 22]; 55(3):42-48. Disponible en:
http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0026-17422012000300008&lng=es.

Díaz Díaz Alejandro Antuan. Modelo de evaluación de los procesos formativos de los residentes en Medicina General Integral. Educ Med Super [Internet]. 2014 Mar [citado 2022 Feb 23]; 28(1):127-144. Disponible en:
http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412014000100014&lng=es.

Sarrias-Ramis R., Mateu L., Baillès E., Pérez J. Longitud (número de preguntas) y resultado de un examen. Educ. méd. [Internet]. 2010 Sep [citado 2022 Feb 23]; 13(3): 187-192. Disponible en:
http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1575-18132010000300010&lng=es.

CENEVAL. Lineamientos para la construcción de reactivos de opción múltiple. Comité Técnico de Normas y Estándares, 2ª Edición, Febrero de 2013. Cap. VII, Pg. 46.

Delgado-Maldonado Laura, Sanchez-Mendiola Melchor. Análisis del examen profesional de la Facultad de Medicina de la UNAM: Una experiencia de evaluación objetiva del aprendizaje con la teoría de respuesta al ítem. Inv Ed Med [Internet]. 2012 Feb [citado 2022 Feb 25] ;1(3):130-139. Disponible en:
 Análisis del examen profesional de la Facultad de Medicina de la UNAM: Una experiencia de evaluación objetiva del aprendizaje con la teoría de respuesta al ítem | Investigación en Educación Médica