

Segunda rendición de la PAES: incremento de puntajes y brechas por dependencia

Un análisis longitudinal mediante panel intraindividual

Nicolás Pinto P. Observatorio PAES - Preuniversitario Gauss

Resumen ejecutivo. Este estudio analiza los cambios de puntaje entre la PAES Regular de 2025 y la PAES de Invierno de 2026 en estudiantes recientemente egresados que rindieron ambas aplicaciones. En el índice de Competencia Lectora y Matemática 1, el puntaje aumentó en promedio 42,6 puntos y el 79,2 % de los estudiantes mejoró. La magnitud del incremento difiere según la dependencia escolar. En el índice obligatorio, los estudiantes de establecimientos particulares pagados aumentaron más que los provenientes de SLEP, y esta diferencia se amplió al estandarizar por puntaje inicial. Además, la ventaja fue mayor entre estudiantes con puntajes iniciales bajos y medios. En Ciencias, el ordenamiento entre dependencias cambió al considerar el nivel inicial, en un patrón consistente con la paradoja de Lord. En conjunto, los resultados muestran una asociación entre segunda rendición y mayores puntajes, junto con un patrón de ventaja acumulativa entre dependencias, sin establecer mecanismos causales.

Palabras clave: PAES; admisión universitaria; repetición de pruebas; dependencia escolar; regresión a la media; efecto Mateo.

1 Introducción

La Prueba de Acceso a la Educación Superior (PAES) constituye uno de los principales instrumentos utilizados en el proceso de admisión a las universidades chilenas. Desde la implementación de una aplicación adicional de Invierno, los postulantes disponen de más de una oportunidad anual para rendir las pruebas y utilizar sus mejores puntajes en procesos posteriores de admisión. Esta posibilidad plantea una pregunta inmediata para quienes deciden rendir nuevamente la PAES: qué ocurre, en términos generales, con sus puntajes entre una aplicación y la siguiente.

La comparación entre ambas rendiciones no se reduce, sin embargo, a determinar si el puntaje medio aumenta o disminuye. Los estudiantes que repiten la prueba parten de niveles de desempeño diferentes y pertenecen a contextos educacionales heterogéneos. Por ello, resulta también relevante establecer si la magnitud de la variación entre ambas aplicaciones es similar entre distintos grupos de postulantes o si existen diferencias sistemáticas asociadas, por ejemplo, a la dependencia administrativa del establecimiento de egreso.

Este problema presenta además una dificultad estadística particular. La comparación directa de los incrementos puede estar condicionada por las diferencias existentes en el puntaje inicial de los grupos. Estudiantes que parten desde niveles distintos no disponen necesariamente del mismo margen de variación posterior, y fenómenos como la regresión a la media o las restricciones propias de la escala pueden modificar la interpretación de las diferencias observadas. En consecuencia, estudiar la evolución de los puntajes requiere considerar conjuntamente tanto el cambio entre rendiciones como el nivel de partida de los postulantes.

El presente estudio analiza los registros publicados por DEMRE correspondientes a la aplicación Regular de 2025, asociada al proceso de admisión 2026, y a la aplicación de Invierno de 2026, asociada al proceso de admisión 2027. El análisis se restringe a estudiantes recientemente egresados de enseñanza media que pueden ser identificados en ambas

aplicaciones y que, por tanto, rindieron nuevamente la prueba en la oportunidad inmediatamente posterior. A partir de este panel individual se estudia la diferencia entre el puntaje obtenido en ambas rendiciones para cada prueba y, adicionalmente, su asociación con la dependencia administrativa del establecimiento de egreso: Particular Pagado, Particular Subvencionado, Municipal y Servicio Local de Educación Pública (SLEP).

Los resultados muestran un incremento generalizado de los puntajes entre ambas aplicaciones. En el índice construido a partir de las dos pruebas obligatorias, Competencia Lectora y Matemática 1, el 79,2 % de los estudiantes obtiene un resultado superior en la segunda rendición. No obstante, la magnitud del incremento difiere entre dependencias: los estudiantes provenientes de establecimientos particulares pagados presentan, en promedio, mayores aumentos que los demás grupos, y esta diferencia persiste al comparar dependencias con una distribución común de puntaje inicial. El patrón puede describirse como una forma de ventaja acumulativa, en el sentido del denominado efecto Mateo [7], sin que ello implique atribuir a este estudio un mecanismo causal específico.

Un resultado particularmente ilustrativo aparece en Ciencias. La comparación cruda muestra un menor incremento entre los estudiantes de establecimientos particulares pagados que entre los particulares subvencionados; sin embargo, el ordenamiento se invierte al considerar las diferencias en el puntaje inicial. Este cambio constituye un caso consistente con la paradoja de Lord [6] y pone de manifiesto la importancia de distinguir entre la comparación directa de los cambios y la comparación entre grupos una vez considerado su nivel de partida.

En conjunto, el estudio busca caracterizar empíricamente qué ocurre con los puntajes de quienes rinden nuevamente la PAES y determinar si dicha variación presenta patrones sistemáticos según la dependencia escolar de origen. Los resultados corresponden al censo de una cohorte específica de egreso y, por tanto, describen de manera exhaustiva a

la población observada; su eventual generalización debe entenderse referida a estudiantes comparables sometidos a condiciones de rendición y admisión similares.

2 Metodología

2.1 Diseño y población de estudio

Se realizó un estudio longitudinal de medidas repetidas, donde se construyó un panel intraindividual a partir de los registros de los resultados de las pruebas PAES declarados en [1] y [2], vinculando las rendiciones correspondientes a la PAES Regular de diciembre de 2025 con la PAES de Invierno de julio de 2026. El emparejamiento se realizó mediante el identificador individual ID_aux, conservando para cada estudiante los puntajes obtenidos en ambas aplicaciones.

La población de análisis se restringió a estudiantes clasificados en la PAES de Invierno con SITUACION_EGRESO = 5, correspondiente a promoción anterior nacional y alumno regular. Esta selección identifica a estudiantes que rindieron la PAES de Invierno inmediatamente después de egresar de enseñanza media. Al vincular estos registros con el proceso Regular inmediatamente anterior, el análisis se restringió a quienes habían rendido también la PAES al término de cuarto medio. Los puntajes fueron transformados a formato numérico y se consideraron válidos aquellos comprendidos entre 100 y 1000 puntos. Para cada prueba, el análisis se realizó sobre los estudiantes con puntaje válido en ambas rendiciones.

Se analizaron las pruebas de Competencia Lectora, Matemática 1, Matemática 2, Historia y Ciencias Sociales y Ciencias. Adicionalmente, se construyó un índice de pruebas obligatorias definido como el promedio entre Competencia Lectora y Matemática 1, únicamente para estudiantes con observaciones válidas en ambas pruebas y en ambas rendiciones.

En Ciencias se exigió, además, que el estudiante hubiese rendido el mismo módulo en ambas aplicaciones. Cuando el módulo difería entre la rendición Regular y la de Invierno, ambas observaciones fueron excluidas del análisis de esta prueba.

2.2 Cambio de puntaje entre rendiciones

Para evaluar el cambio de puntaje entre ambas aplicaciones se calculó, para cada estudiante, la diferencia entre sus puntajes obtenidos en la prueba de Invierno y la prueba Regular:

$$\Delta_i = Y_{i,\text{Invierno}} - Y_{i,\text{Regular}}. \quad (1)$$

El parámetro principal fue el incremento medio $\bar{\Delta}$. Su incertidumbre se estimó mediante un intervalo de confianza bootstrap BCa al 95 % [3], utilizando 1000 remuestreos con reemplazo y una semilla fija para asegurar la reproducibilidad de los resultados. Como referencia adicional se calculó el intervalo paramétrico basado en la distribución t .

La diferencia entre ambas rendiciones se contrastó mediante una prueba t para muestras pareadas. Como análisis complementario, se aplicó la prueba de rangos con signo de Wilcoxon para evaluar la localización de la distribución de las diferencias respecto de cero.

La magnitud del cambio se cuantificó mediante el tamaño de efecto d_z de Cohen para medidas pareadas [8].

$$d_z = \frac{\bar{\Delta}}{s_{\Delta}}, \quad (2)$$

donde s_{Δ} corresponde a la desviación estándar de las diferencias individuales. Como medida complementaria se calculó el incremento medio estandarizado por la desviación estándar del puntaje inicial.

La asociación entre los puntajes de ambas aplicaciones se describió mediante la correlación test-retest, calculada como la correlación de Pearson entre los puntajes Regular e Invierno, acompañada de un intervalo de confianza al 95 % obtenido mediante la transformación de Fisher. Finalmente, se estimó la proporción de estudiantes con $\Delta_i > 0$, junto con su intervalo de confianza de Wilson al 95 %.

2.3 Diferencias según dependencia del establecimiento

Los estudiantes fueron clasificados de acuerdo con la dependencia de su establecimiento de origen en cuatro grupos: particular pagado, particular subvencionado, municipal y Servicio Local de Educación Pública (SLEP).

En una primera etapa se estimó descriptivamente el incremento medio de puntaje en cada dependencia, acompañado de su intervalo de confianza al 95 %. Las diferencias globales entre grupos se evaluaron mediante ANOVA de Welch [10], debido a la posible heterogeneidad de varianzas entre dependencias. Posteriormente se realizaron contrastes unilaterales entre establecimientos particulares pagados y cada una de las restantes dependencias mediante pruebas de Welch, aplicando la corrección de Holm sobre la familia de contrastes [5]. Como análisis complementario de todas las comparaciones por pares se utilizó el procedimiento de Games–Howell [4].

2.4 Estandarización por puntaje inicial

Dado que las distribuciones de puntaje inicial difieren entre dependencias y que la magnitud del cambio puede depender del punto de partida, el análisis principal entre grupos se realizó mediante estandarización directa por nivel inicial.

Para cada prueba, la distribución conjunta del puntaje Regular se dividió en cinco quintiles. Dentro de cada combinación dependencia–quintil se calculó el incremento medio. Posteriormente, estos incrementos fueron ponderados utilizando una distribución común de quintiles, determinada a partir de la población total analizada. De este modo, el incremento estandarizado de cada dependencia representa el cambio esperado si todas presentaran la misma distribución de puntaje inicial.

Formalmente, para una dependencia g , el incremento estandarizado se definió como

$$\Delta_g^* = \sum_{q=1}^5 w_q \bar{\Delta}_{gq}, \quad (3)$$

donde w_q corresponde a la proporción común asignada al quintil q y $\bar{\Delta}_{gq}$ al incremento medio observado en la dependencia g dentro de dicho quintil.

La comparación principal se definió mediante los contrastes

$$C_g = \Delta_{\text{Pagado}}^* - \Delta_g^*, \quad (4)$$

para particular subvencionado, municipal y SLEP. La incertidumbre de estos contrastes se estimó mediante intervalos de confianza bootstrap percentil al 95 %. En cada réplica se remuestrearon estudiantes con reemplazo y se reconstruyeron los quintiles, sus pesos y los incrementos estandarizados, de manera que el procedimiento completo de estandarización quedara incorporado en la variabilidad bootstrap.

2.5 Variación de la brecha según el puntaje inicial

Como análisis complementario se estudió si la diferencia entre dependencias variaba según el puntaje inicial mediante un modelo lineal de mínimos cuadrados ordinarios con interacción entre el puntaje Regular y la dependencia del establecimiento. Se estimaron pendientes específicas para cada dependencia y se utilizaron errores estándar robustos HC3.

La homogeneidad de pendientes se evaluó mediante un test conjunto de Wald sobre los términos de interacción. Cuando esta hipótesis se rechaza, la diferencia entre establecimientos particulares pagados y las restantes dependencias se interpreta como dependiente del puntaje inicial. Para describir dicha variación se estimó la ventaja del grupo particular pagado en los percentiles 10, 50 y 90 de la distribución del puntaje Regular, acompañada en cada caso de su correspondiente intervalo de confianza robusto al 95 %.

Adicionalmente, se estimó una ANCOVA con pendiente común,

$$Y_{\text{Invierno}} = \beta_0 + \beta_1 Y_{\text{Regular}} + \beta_2 \text{Dependencia} + \varepsilon, \quad (5)$$

como análisis de referencia. Los contrastes ajustados entre particular pagado y las demás dependencias utilizaron covarianzas robustas HC3 y corrección de Holm.

2.6 Análisis de sensibilidad para Ciencias

Para evaluar la influencia de la composición de los módulos de Ciencias se realizó un análisis de sensibilidad mediante una segunda estandarización directa, controlando simultáneamente por quintil de puntaje inicial y módulo rendido.

Para garantizar un soporte común entre dependencias, la estandarización se restringió a combinaciones quintil-módulo representadas en los cuatro grupos, exigiéndose operacionalmente un mínimo de cinco observaciones por celda y dependencia. Esta condición excluyó únicamente el módulo TEC, que no presentó observaciones entre estudiantes de establecimientos particulares pagados. Los pesos de las celdas restantes se obtuvieron a partir de su frecuencia en la población conjunta y se aplicaron de forma idéntica a todas las dependencias.

Finalmente, se compararon los incrementos obtenidos mediante la estandarización exclusiva por puntaje inicial con aquellos obtenidos al controlar simultáneamente por puntaje inicial y módulo de Ciencias.

2.7 Reproducibilidad y código de análisis

Todos los análisis fueron implementados en Python utilizando las librerías numpy, pandas, scipy y matplotlib. Los procedimientos de remuestreo utilizaron 1000 réplicas bootstrap y una semilla fija (SEED = 2025), de modo que la ejecución completa del cuaderno sobre las mismas bases de datos reproduce las mismas estimaciones y figuras.

El código utilizado para reproducir los análisis y figuras se encuentra disponible públicamente en el sitio web del Observatorio PAES de Preuniversitario Gauss [9].

Para la elaboración, depuración y revisión de este código se utilizaron Claude (Anthropic) y ChatGPT (OpenAI) como herramientas de apoyo basadas en modelos de lenguaje de gran escala. El código producido con esta asistencia fue posteriormente revisado, modificado, ejecutado y verificado por el autor.

3 Resultados

3.1 El puntaje sube al repetir

Al rendir la prueba por segunda vez tras egresar, el estudiante promedio eleva el puntaje en las cinco pruebas (y por consecuencia, en el índice), con efectos de magnitud variable entre pruebas (véase Tabla 1, Figura 1). Las mayores alzas se observan en Matemática 1 (54,9 puntos; $d_z = 0,76$) e Historia (50,6). Por otra parte, este incremento es menor en Ciencias (13,8; $d_z = 0,21$). En el índice de pruebas obligatorias, construido para los estudiantes con observaciones válidas en ambas pruebas y rendiciones, el incremento medio es de 42,6 puntos y el 79,2 % de los estudiantes mejora respecto de su primera rendición.

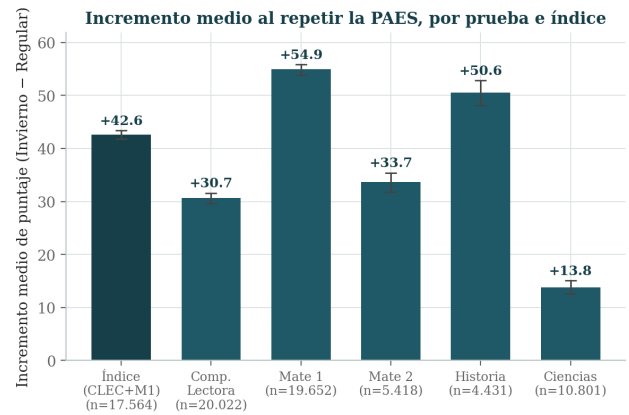


Figura 1: Incremento medio de puntaje al repetir, con intervalo de confianza al 95 %, para cada prueba y para el índice obligatorio. Los intervalos son estrechos debido a la gran cantidad de observaciones.

La correlación test-retest es alta en todas las pruebas (entre 0,77 y 0,91), lo que indica una fuerte asociación entre los puntajes obtenidos en ambas rendiciones. Así, aunque los puntajes presentan un desplazamiento general hacia valores superiores en la aplicación de Invierno, el desempeño inicial continúa siendo un predictor importante del desempeño posterior: quienes obtienen puntajes relativamente altos en la primera rendición tienden también a obtenerlos en la segunda. Esta asociación no implica que las posiciones relativas permanezcan inalteradas, pues existe variación individual entre ambas mediciones. La Figura 2 lo muestra

Cuadro 1: Hipótesis 1. Incremento al repetir por prueba e índice. Δ es el incremento medio (Invierno menos Regular) con su intervalo de confianza al 95 % por *bootstrap*; d_z y d son el tamaño de efecto sobre la desviación de las diferencias y del puntaje inicial; r es la correlación test–retest.

Prueba / índice	<i>n</i>	Δ (IC 95 %)	d_z	d	r	% mejora
Índice obligatorias (CLEC + M1)	17.564	+42,6 [41,9; 43,4]	0,80	0,37	0,904	79,2
Competencia Lectora	20.022	+30,7 [29,6; 31,6]	0,42	0,27	0,794	66,6
Matemática 1	19.652	+54,9 [53,8; 55,9]	0,76	0,39	0,886	78,6
Matemática 2	5.418	+33,7 [31,8; 35,4]	0,48	0,39	0,770	70,0
Historia y Cs. Sociales	4.431	+50,6 [48,1; 52,9]	0,64	0,40	0,844	73,3
Ciencias	10.801	+13,8 [12,6; 15,1]	0,21	0,10	0,877	59,3

con claridad para Matemática 1: la nube se sitúa por encima de la diagonal que representa ausencia de variación en el puntaje. Además, la recta de ajuste de mínimos cuadrados ordinarios (MCO) tiene pendiente inferior a uno, lo que constituye un patrón compatible con el fenómeno de regresión a la media.

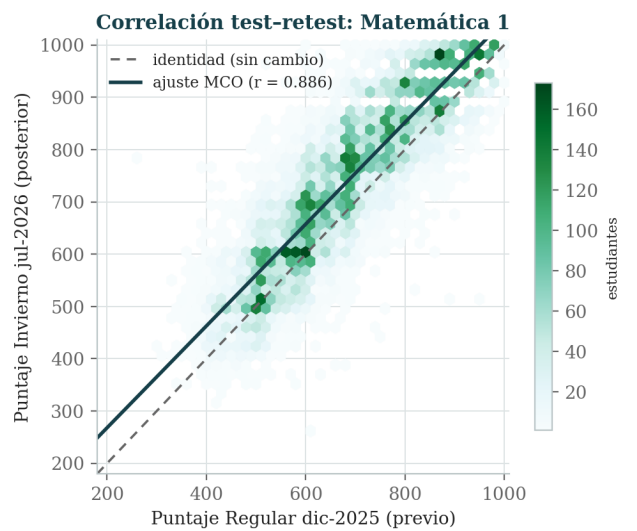


Figura 2: Puntaje posterior frente al previo en Matemática 1.

Al considerar el índice construido a partir de los incrementos del promedio de ambas pruebas obligatorias, la distribución del incremento se encuentra desplazada hacia valores positivos, como se ilustra en la Figura 3. Se observa que una minoría de los postulantes disminuye su puntaje, mientras que una amplia mayoría (79 %), observa mejoría en sus resultados.

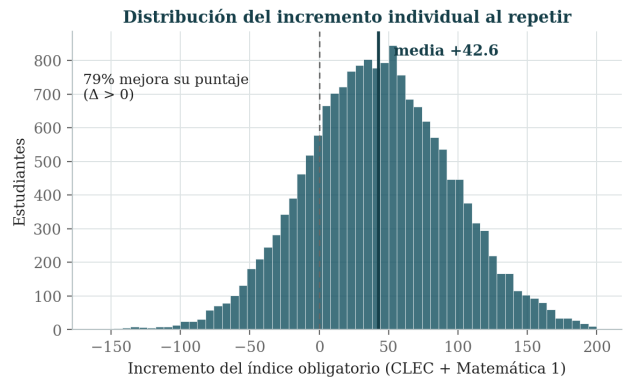


Figura 3: Distribución del incremento individual en el índice de pruebas obligatorias.

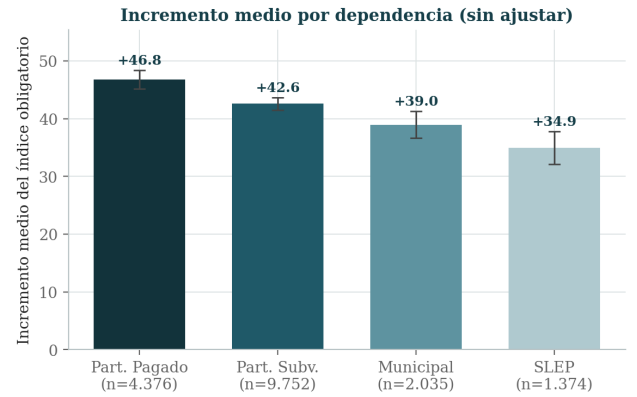


Figura 4: Gradiente del incremento según dependencia administrativa en el índice obligatorio (sin ajustar)

3.2 El gradiente por dependencia

El incremento de puntajes difiere entre dependencias. En el índice de pruebas obligatorias crece de forma monótona con la dependencia: 34,9 puntos en SLEP, 39,0 en Municipal, 42,6 en Subvencionado y 46,8 en Particular Pagado (Figura 4). El contraste de acuerdo a ANOVA de Welch rechaza la igualdad de incrementos ($F = 21,4$ aproximado; $p < 10^{-13}$) y los contrastes planificados por metodología confirman que el Particular Pagado supera a cada una de las otras tres dependencias. El mismo ordenamiento aparece en Competencia Lectora, Matemática 1, Matemática 2 e Historia.

La única excepción en la lectura cruda es Ciencias, donde el Particular Pagado exhibe el menor incremento aparente (8,5 frente a 16,4 del Subvencionado). Sin embargo, este patrón se invierte al estandarizar por puntaje inicial, lo que indica que la comparación cruda está fuertemente condicionada por las diferencias en el nivel de partida entre dependencias.

3.3 Regresión a la media y moderación por el nivel inicial

Dado que las dependencias presentan distribuciones distintas de puntaje inicial, resulta necesario considerar la posibilidad de que parte del gradiente observado esté asociada a regresión a la media. Los análisis ajustados muestran que el gradiente persiste al considerar el nivel de partida y, adicionalmente, evidencian que la magnitud de la ventaja del postulante perteneciente a dependencias de tipo Particular Pagado varía según el puntaje inicial.

El análisis principal corresponde a la estandarización directa por puntaje inicial. Tras igualar la distribución de puntajes iniciales entre las dependencias, el ordenamiento observado se mantiene y la magnitud de la brecha aumenta. En el ín-

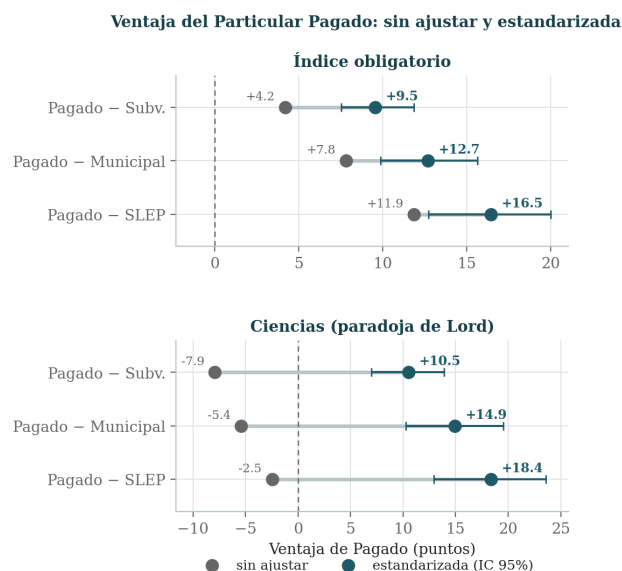


Figura 5: Ventaja del Particular Pagado frente a cada dependencia, sin ajustar y estandarizada a una misma distribución de puntaje inicial.

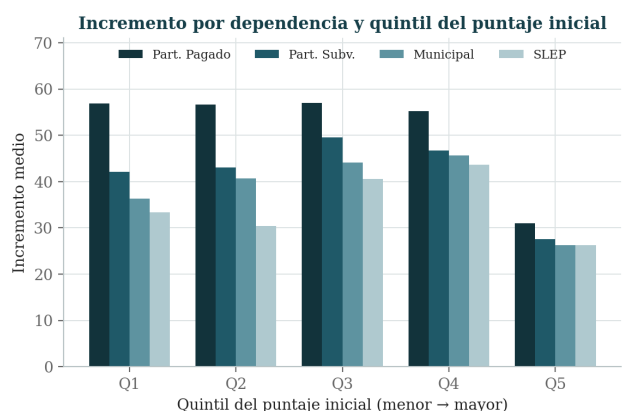


Figura 6: Incremento por dependencia dentro de cada quintil del puntaje inicial.

dice obligatorio, el incremento estandarizado alcanza 34,9 puntos en SLEP y 51,3 puntos en establecimientos Particulares Pagados. El contraste Pagado-SLEP es de 16,5 puntos (IC 95 % bootstrap: [12,7; 20,0]), superior a la diferencia no ajustada de 11,9 puntos. La diferencia entre los contrastes crudo y estandarizado refleja la distinta distribución de puntajes iniciales entre dependencias y la relación existente entre el nivel inicial y el incremento posterior (Figura 5). La estratificación por quintiles (Figura 6) muestra que el Pagado lidera dentro de cada nivel de partida, aunque esta ventaja es menos pronunciada en el quintil superior.

La magnitud de la diferencia entre dependencias varía según el puntaje inicial. En el índice obligatorio, las pendientes estimadas del puntaje posterior sobre el previo difieren entre grupos, desde 0,89 en Particular Pagado hasta 1,00 en SLEP. Un test conjunto de Wald con covarianza robusta HC3 rechaza la hipótesis de homogeneidad de pendientes ($p < 10^{-23}$), lo que indica que la brecha entre dependencias no es constante a lo largo de la distribución del puntaje inicial.

En particular, la ventaja estimada del Particular Pagado sobre SLEP es de 37,1 puntos en el percentil 10 del puntaje inicial (IC 95 % [31,6; 42,6]), de 18,9 puntos en la mediana y

cercana a cero en el percentil 90 (-0,0; IC 95 % [-6,0; 5,9]). Esta reducción de la brecha en los puntajes iniciales altos debe interpretarse con cautela, debido a la restricción impuesta por el límite superior de 1.000 puntos, consistente con un posible efecto techo.

En Ciencias, las pendientes estimadas entre el puntaje previo y el posterior presentan diferencias de magnitud relativamente pequeña entre dependencias (0,78-0,84). No obstante, la comparación entre grupos cambia de manera sustantiva al considerar el puntaje inicial: la diferencia cruda entre Particular Pagado y Particular Subvencionado es de -7,9 puntos, mientras que el contraste obtenido mediante estandarización por puntaje inicial alcanza 10,5 puntos. Este cambio de signo es consistente con la paradoja de Lord [6].

Dado que la composición por módulo de Ciencias difiere entre dependencias, se realizó adicionalmente un análisis de sensibilidad mediante una estandarización que controla simultáneamente por puntaje inicial y módulo, restringida a celdas con soporte común. Bajo esta especificación, los incrementos estandarizados fueron de 22,0 puntos en Particular Pagado y 12,9 puntos en Particular Subvencionado, manteniéndose positivo el contraste entre ambos grupos (9,1 puntos). Por tanto, el cambio de signo observado al considerar el puntaje inicial persiste al controlar adicionalmente por el módulo de Ciencias rendido.

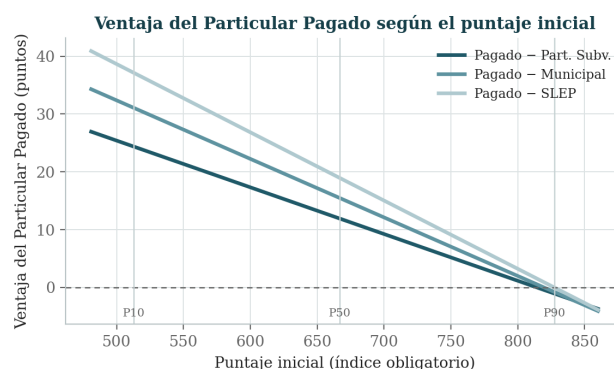


Figura 7: Ventaja del Particular Pagado en el incremento del índice obligatorio, según el puntaje inicial (modelo con pendientes por dependencia).

4 Discusión

Los resultados muestran un patrón consistente de incremento entre ambas rendiciones. En todas las pruebas analizadas, el puntaje medio de la aplicación de Invierno supera al de la aplicación Regular, aunque la magnitud del incremento varía considerablemente entre pruebas. El índice de pruebas obligatorias presenta un aumento medio de 42,6 puntos y cerca de cuatro de cada cinco estudiantes obtienen un puntaje superior en la segunda rendición. Estos resultados describen una asociación clara entre la repetición de la PAES y un mayor puntaje posterior en la cohorte estudiada, pero el diseño observacional no permite atribuir ese incremento exclusivamente al hecho de volver a rendir la prueba ni identificar los mecanismos que lo producen.

Las correlaciones test-retest, altas en todas las pruebas, indican además una considerable estabilidad relativa de los puntajes entre ambas aplicaciones. Esto significa que el

incremento medio coexiste con una asociación fuerte entre el nivel de desempeño inicial y el posterior: los estudiantes que obtienen puntajes relativamente altos en la primera rendición tienden también a obtenerlos en la segunda. Esta estabilidad no implica, sin embargo, que las posiciones relativas permanezcan inalteradas, pues existe variación individual entre ambas mediciones.

La segunda regularidad relevante corresponde a las diferencias según dependencia administrativa. En el índice de pruebas obligatorias, los estudiantes provenientes de establecimientos particulares pagados presentan un incremento medio mayor que los demás grupos, y esta diferencia persiste al estandarizar las dependencias a una misma distribución de puntaje inicial. El patrón puede describirse en términos de un efecto Mateo [7]: una diferencia inicialmente favorable entre grupos aparece acompañada por una diferencia posterior también favorable en la magnitud del incremento. En este estudio, sin embargo, el efecto Mateo se utiliza únicamente como descripción de un patrón de ventaja acumulativa. Los datos no permiten establecer que este resultado sea producido por diferencias en recursos económicos, acceso a preparación, características de los establecimientos u otros mecanismos asociados a la dependencia.

La magnitud de esta diferencia tampoco es constante a lo largo de la distribución inicial. En el índice obligatorio, la ventaja estimada del Particular Pagado es mayor entre quienes presentan puntajes iniciales bajos y medios y se atenúa hacia la parte superior de la distribución. Esta reducción debe interpretarse considerando la restricción impuesta por el máximo de 1.000 puntos, consistente con un posible efecto techo: quienes parten cerca del límite superior disponen mecánicamente de un menor margen para incrementar su puntaje. Por esta razón, la ausencia de una diferencia estadísticamente detectable en los niveles iniciales más altos no permite concluir que las dependencias tengan necesariamente el mismo comportamiento en ese segmento.

El caso de Ciencias ilustra además la importancia de considerar el puntaje inicial al comparar incrementos entre grupos. En la comparación cruda, los estudiantes de establecimientos particulares pagados presentan un incremento menor que los particulares subvencionados. Sin embargo, ambos grupos parten de distribuciones de puntaje inicial diferentes y, al estandarizarlos a una distribución común, el contraste cambia de signo y pasa a favorecer al Particular Pagado. Este fenómeno es consistente con la paradoja de Lord [6]: una comparación basada directamente en los cambios puede conducir a un ordenamiento distinto del obtenido al comparar grupos después de considerar sus diferencias iniciales. No se trata de que una de las dos comparaciones sea necesariamente incorrecta, sino de que responden a preguntas estadísticas distintas. En este caso, el cambio de signo se mantiene al controlar adicionalmente por el módulo rendido en Ciencias, lo que indica que el resultado no se explica por la diferente composición de módulos entre dependencias.

Estos resultados deben interpretarse dentro de los límites del diseño. En primer lugar, la dependencia administrativa constituye un indicador agregado y, por sí solo, imperfecto de las condiciones socioeconómicas de los estudiantes; no puede interpretarse como una medición directa de recursos individuales o familiares. En segundo lugar, la rendición de Matemática 2, Historia y Ciencias depende de las elecciones

de los propios postulantes, por lo que las poblaciones observadas en estas pruebas son selectivas y no representan necesariamente al conjunto de quienes repiten la PAES. Asimismo, aunque el análisis utiliza observaciones emparejadas de los mismos estudiantes, sigue siendo observacional y no permite identificar causalmente por qué algunos grupos incrementan más sus puntajes que otros.

Finalmente, los datos corresponden al censo de los estudiantes de una única cohorte de egreso que satisfacen los criterios de inclusión y pueden ser observados en ambas rendiciones. En consecuencia, las estimaciones describen exhaustivamente a esta población observada, pero su extrapolación requiere una interpretación de superpoblación. Las inferencias realizadas deben entenderse como referidas a estudiantes comparables a los de esta cohorte, sometidos a condiciones de rendición y admisión similares. Cambios futuros en la estructura de la PAES, en las cohortes de postulantes, en las condiciones de preparación o en el sistema de admisión podrían modificar tanto la magnitud de los incrementos como las diferencias observadas entre dependencias. Por ello, la replicación del análisis en próximas cohortes constituye una extensión natural del estudio.

5 Conclusiones

En la cohorte analizada, los estudiantes que rindieron nuevamente la PAES en la aplicación de Invierno obtuvieron, en promedio, puntajes superiores a los de su rendición Regular. El incremento se observa en todas las pruebas estudiadas y alcanza 42,6 puntos en el índice de pruebas obligatorias, donde el 79,2 % de los estudiantes mejora su resultado.

La magnitud de este incremento difiere según la dependencia administrativa del establecimiento de origen. En el índice obligatorio se observa un gradiente que favorece a los estudiantes provenientes de establecimientos particulares pagados. Esta diferencia no desaparece al considerar las distintas distribuciones de puntaje inicial entre dependencias; por el contrario, el contraste estandarizado entre Particular Pagado y SLEP alcanza 16,5 puntos (IC 95 % [12,7; 20,0]), frente a 11,9 puntos en la comparación cruda.

La ventaja asociada a la dependencia tampoco es uniforme a lo largo de la distribución inicial. En el índice obligatorio es mayor entre estudiantes con puntajes iniciales bajos y medios y se atenúa hacia los niveles superiores, donde la interpretación debe considerar la restricción impuesta por el máximo de 1.000 puntos. En este sentido, las diferencias entre dependencias deben entenderse como dependientes del nivel de partida y no como una brecha constante.

El caso de Ciencias muestra especialmente la importancia de considerar dicho nivel inicial. La comparación cruda favorece al Particular Subvencionado frente al Particular Pagado, mientras que la estandarización por puntaje inicial invierte el signo del contraste. Este resultado, consistente con la paradoja de Lord, se mantiene al incorporar además el módulo rendido en el análisis de sensibilidad.

En conjunto, los resultados son compatibles con un patrón de ventaja acumulativa entre dependencias, que puede describirse en términos de efecto Mateo, sin que el diseño permita identificar los mecanismos responsables de dicha asociación. La dependencia administrativa no constituye una medida directa del nivel socioeconómico y los análisis realizados son

observacionales, por lo que no permiten atribuir causalmente las diferencias observadas a recursos, preparación u otras características de los estudiantes o establecimientos.

Finalmente, estas conclusiones describen exhaustivamente a la cohorte observada, pero su generalización requiere una interpretación de superpoblación. Los resultados deben entenderse como aplicables a estudiantes comparables a los de esta cohorte bajo condiciones similares de rendición y admisión, y no como parámetros necesariamente estables para futuras cohortes o futuras versiones de la PAES. La replicación del análisis en sucesivos procesos de admisión permitirá evaluar la estabilidad temporal de los patrones encontrados.

Referencias

- [1] Datos Abiertos DEMRE, Archivo C, Datos de Rendición de la PAES del Proceso de Admisión 2026, <https://portal-transparencia.demre.cl/portal-base-datos>
- [2] Datos Abiertos DEMRE, archivo C, Datos de Rendición de las PAES de Invierno del Proceso de Admisión 2027, <https://portal-transparencia.demre.cl/portal-base-datos>
- [3] Efron, B. (1987). Better Bootstrap Confidence Intervals. *Journal of the American Statistical Association*, 82(397), 171–185.
- [4] Games, P. A., & Howell, J. F. (1976). Pairwise multiple comparison procedures with unequal n 's and/or variances: A Monte Carlo study. *Journal of Educational Statistics*, 1(2), 113–125.
- [5] Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70.
- [6] Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin*, 68(5), 304–305.
- [7] Merton, R. K. (1968). The Matthew Effect in Science. *Science*, 159(3810), 56–63.
- [8] Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, 7(1), 105–125.
- [9] **Dummy Reference para el código Python en la página del observatorio**
- [10] Welch, B. L. (1951). On the comparison of several mean values: An alternative approach. *Biometrika*, 38(3/4), 330–336.