

Midiendo lo invisible: un análisis contrafactual del
abandono escolar y la desigualdad educativa en
Antioquia, Colombia

Marcela Galvis Restrepo

Octubre 2025

Índice

1. Introducción	4
2. Metodología: Modelo de Cohortes y Análisis Contrafactual del Aprendizaje Esperado	6
2.1. Justificación Metodológica	6
2.2. Diseño del modelo de cohortes	8
2.3. Medición del abandono y del aprendizaje contrafactual	9
2.4. Limitaciones y alcance del enfoque	10
3. Análisis de Brechas en el Abandono Escolar en Antioquia: Tendencias y Vulnerabilidades entre las Cohortes de Saber 5 y Saber 11 (2014-2023)	11
4. Análisis de Brechas en el Aprendizaje Escolar en Antioquia: Tendencias y Vulnerabilidades entre las Cohortes de Saber 5 y Saber 11 (2014-2023)	19
5. Amplificación de brechas durante la pandemia del Covid-19, 2019-2023	27
6. Conclusiones	35

Índice de figuras

1. Distribución del NSE individual en Saber 5 y Saber 11 para las cohortes que iniciaron Saber 5 entre 2014 y 2017.	12
2. Participación relativa de estudiantes con extraedad en Saber 5 y Saber 11 . . .	15
3. Participación relativa de estudiantes con internet en Saber 5 y Saber 11 . . .	16
4. Participación relativa de estudiantes por género en Saber 5 y Saber 11	17
5. Participación relativa de grupos seleccionados en Saber 5 y Saber 11	18
6. Resultados históricos de Saber 11 sin contrafactual	21

7.	Puntaje promedio por decil del INSE	23
8.	Puntaje promedio por decil del INSE, Matemáticas	26
9.	Puntaje promedio por decil del INSE, Lectura Crítica	26
10.	Puntaje promedio por decil del INSE, análisis de la pandemia	28
11.	Puntaje promedio por decil del INSE, análisis de la pandemia matemáticas .	30
12.	Puntaje promedio por decil del INSE, análisis de la pandemia lectura crítica	31
13.	Puntaje promedio por decil del INSE, por género año 2023	33
14.	Puntaje promedio por decil del INSE, por área año 2023	33
15.	Puntaje promedio por decil del INSE, por sector año 2023	34

Índice de cuadros

1.	Medias, medianas y prueba de Kolmogorov-Smirnov del NSE individual por cohorte y prueba.	12
2.	Estimación del modelo de datos de panel con efectos fijos por municipio y año. Variable dependiente: NSE promedio.	14

Nota: el analisis a continuacion se hizo para Antioquia, en este momento estoy trabajando para tener analisis para los datos agregados del pais y tambien hacer comparaciones a nivel departamental, este estudio se hizo inicialmente con financiacion de Proantioquia y los resultados son parte del informe 'El estado de la Educación en Antioquia. Yo soy la unica autora de la parte metodologica y empirica y mi plan es rehacer el analisis para Colombia, pero no he terminado los experimentos.

1. Introducción

La pandemia de COVID-19 profundizó desigualdades estructurales en los sistemas educativos latinoamericanos, afectando el aprendizaje y la permanencia escolar de millones de estudiantes. Según el Banco Mundial Banco Mundial (2021), el cierre prolongado de escuelas dejó a más de 120 millones de niños en riesgo de perder un año completo de escolaridad presencial y aumentó la “pobreza de aprendizaje” regional del 51 % al 62,5 %. En Colombia, estas pérdidas no se distribuyeron de manera uniforme: los estudiantes de menores ingresos, de zonas rurales y matriculados en instituciones oficiales fueron los más afectados Banco de la República (2021). Sin embargo, las métricas disponibles no permiten dimensionar con precisión la magnitud ni la distribución de estos efectos, debido a limitaciones en la cobertura y periodicidad de las evaluaciones nacionales.

La prueba Saber 11, actualmente la única de carácter censal en el país, ofrece información valiosa sobre el desempeño de los estudiantes que logran culminar la educación media (Instituto Colombiano para la Evaluación de la Educación – ICFES, 2020). No obstante, su cobertura restringida genera un sesgo estructural: mide el resultado final de un subconjunto de estudiantes, excluyendo a quienes abandonan el sistema o transitan con rezago. Este fenómeno, conocido como sesgo de supervivencia, implica que las desigualdades en el acceso y la permanencia pueden enmascarse detrás de aparentes mejoras en el desempeño promedio. Durante la pandemia, por ejemplo, instituciones con mayores pérdidas de matrícula muestra-

ron incrementos artificiales en sus puntajes en Saber 11, reflejando más bien un efecto de exclusión que de mejora real (Banco Interamericano de Desarrollo (BID), 2021).

Hasta 2017, las pruebas Saber 3°, 5° y 9° permitían seguir de manera censal los avances del aprendizaje en los ciclos iniciales y medios. Sin embargo, la decisión del ICFES de aplicar estos instrumentos de forma muestral desde 2018 redujo el acceso a datos desagregados y eliminó la posibilidad de construir trayectorias comparables a nivel territorial. Como consecuencia, los indicadores disponibles para medir la calidad educativa —centrados exclusivamente en Saber 11— no capturan la relación entre desempeño, abandono y desigualdad.

Frente a esta limitación, surge la necesidad de desarrollar metodologías que integren los datos históricos de las evaluaciones con información contextual para reconstruir trayectorias educativas y estimar desigualdades invisibles. Este estudio propone un enfoque de análisis contrafactual de cohortes esperadas, que conecta los resultados de Saber 5 (2014–2017) con los de Saber 11 (2020–2023), para estimar qué proporción de los estudiantes observados en primaria alcanzó efectivamente la educación media, y cómo la deserción y el rezago se relacionan con el desempeño académico y las condiciones socioeconómicas.

El análisis no solo cuantifica la magnitud del abandono, sino que examina cómo la desigualdad social se traduce en desigualdad educativa a lo largo del ciclo escolar. Al estimar los puntajes contrafactuales que habrían tenido los estudiantes desertores si hubieran permanecido en el sistema, se logra capturar una dimensión ausente en las métricas tradicionales: el aprendizaje perdido y su distribución desigual. Esta aproximación busca, por tanto, medir lo invisible —las trayectorias truncadas y las brechas acumuladas— y ofrecer un marco empírico para diseñar políticas que aborden simultáneamente la calidad y la equidad educativa.

El artículo se organiza de la siguiente manera: la sección 2 describe las fuentes de datos y la metodología de construcción de cohortes esperadas; la sección 3 presenta los resultados sobre abandono, rezago y desigualdad; y la sección 4 discute las implicaciones de los hallazgos para el monitoreo y la política educativa en contextos subnacionales.

2. Metodología: Modelo de Cohortes y Análisis Contrafactual del Aprendizaje Esperado

El estudio emplea un enfoque cuantitativo longitudinal para estimar la relación entre abandono escolar, desempeño académico y desigualdad educativa. La estrategia metodológica se basa en la construcción de cohortes esperadas de estudiantes desde la prueba Saber 5 hasta la prueba Saber 11, permitiendo comparar la composición sociodemográfica y el rendimiento académico potencial de quienes permanecen o abandonan el sistema.

La hipótesis central plantea que, aun antes de la pandemia, los estudiantes de nivel socioeconómico (NSE) bajo presentaban mayores tasas de deserción y rezago que sus pares de NSE alto, y que dicha brecha se amplió durante los años 2020–2021, cuando las cohortes en transición hacia la educación media enfrentaron una exposición desigual al cierre escolar.

2.1. Justificación Metodológica

En ausencia de datos longitudinales individuales, los *modelos de cohortes esperadas* (también denominados cohortes sintéticas o reconstruidas) ofrecen una alternativa sólida para analizar trayectorias educativas. Este enfoque, ampliamente utilizado en economía de la educación y demografía, permite simular el progreso de una cohorte hipotética de estudiantes a través del sistema escolar utilizando datos transversales repetidos en el tiempo. La idea se inspira en los trabajos de Deaton (1985) sobre pseudo-paneles, donde se agrupan observaciones repetidas de unidades similares (por ejemplo, por año de nacimiento o grado) para construir una serie temporal de cohortes. De manera análoga, la UNESCO ha distinguido desde hace décadas entre el análisis de cohortes verdaderas —que requieren seguimiento individual— y las *cohortes reconstruidas*, que infieren tasas de promoción, repetición y abandono a partir de registros agregados anuales (UNESCO, 1981).

El método de cohorte reconstruida ha sido adoptado por organismos internacionales como el Banco Mundial y el Instituto de Estadísticas de la UNESCO (UIS) para estimar indicadores

como la *tasa de supervivencia* y la *eficiencia interna* del sistema educativo. Estos indicadores se calculan aplicando sucesivamente las probabilidades de permanencia por grado a una cohorte ficticia de 100 estudiantes, asumiendo que los patrones recientes de transición se mantienen constantes (UNESCO Institute for Statistics, 2018). En el caso colombiano, este enfoque ha sido empleado para estimar la supervivencia de las cohortes mediante el descuento de desertores en cada grado según su último grado aprobado, constituyéndose en una práctica estadística reconocida para medir la retención y la progresión escolar.

Una vez establecida la cohorte sintética, es posible incorporar un *análisis contrafactual* que amplía su poder explicativo. Los economistas de la educación utilizan este tipo de análisis para estimar escenarios hipotéticos —por ejemplo, cómo cambiarían los resultados educativos si todos los estudiantes permanecieran en el sistema escolar o si la composición sociodemográfica fuera distinta. Pritchett and Sandefur (2017) aplicaron este tipo de enfoque al analizar perfiles de aprendizaje con encuestas demográficas y de salud, estimando que, bajo un escenario de escolarización universal, la tasa de alfabetización femenina en Etiopía aumentaría 57 puntos porcentuales, mientras que en Nigeria solo cinco puntos, dada la baja calidad de aprendizaje observada. De manera similar, Akmal and Pritchett (2019) mostraron que igualar los insumos educativos entre grupos ricos y pobres no bastaría para eliminar las brechas de aprendizaje, pues incluso entre los más aventajados el rendimiento promedio es insuficiente. Estos ejercicios contrafactuales ilustran cómo la combinación de cobertura, retención y calidad puede examinarse empíricamente sin requerir datos experimentales.

El uso combinado de cohortes esperadas y análisis contrafactual resulta particularmente pertinente en contextos con sistemas de información fragmentados o con evaluaciones censales históricas no vinculadas entre sí. Frente a métodos convencionales —como regresiones transversales o comparaciones de promedios entre grados— este enfoque aprovecha las evaluaciones disponibles en diferentes momentos para construir trayectorias hipotéticas comparables. Así, permite inferir cómo evolucionaría el desempeño promedio de una cohorte a lo largo del ciclo educativo, aun sin seguir a individuos específicos. Esta aproximación es especialmente útil pa-

ra corregir el sesgo de selección derivado de la deserción: al avanzar de grado, los estudiantes con menores logros tienden a abandonar el sistema, elevando artificialmente los promedios de quienes permanecen. Ignorar este fenómeno lleva a sobreestimar las ganancias de aprendizaje con la progresión escolar (Kaffenberger, 2019).

El análisis contrafactual corrige parcialmente este sesgo al estimar los puntajes que habrían obtenido los estudiantes desertores en caso de continuar su trayectoria educativa. De esta forma, se obtiene una medida más precisa del aprendizaje perdido y de la desigualdad acumulada a lo largo del ciclo escolar. En suma, la combinación de modelos de cohortes esperadas y análisis contrafactual constituye una estrategia metodológica robusta y adecuada para contextos donde no existen datos de panel estudiantiles, pero sí evaluaciones censales extensivas. Además de cuantificar el abandono escolar y el desempeño académico, esta metodología permite revelar dimensiones de la *desigualdad educativa invisible*, proporcionando evidencia empírica relevante para el diseño y evaluación de políticas públicas.

2.2. Diseño del modelo de cohortes

Se construyeron cuatro cohortes educativas con base en los estudiantes que presentaron la prueba Saber 5 en los años 2014, 2015, 2016 y 2017. Cada cohorte fue emparejada con su generación esperada en Saber 11 —2020, 2021, 2022 y 2023, respectivamente—, asumiendo una progresión escolar regular de seis años entre ambos grados. Este emparejamiento permite evaluar cambios en la composición de las cohortes y estimar la proporción relativa de estudiantes que no alcanzan grado 11 dentro del periodo esperado.

El análisis considera cinco dimensiones estructurales: (i) nivel socioeconómico del estudiante, (ii) carácter del establecimiento (oficial o privado), (iii) ubicación geográfica (rural o urbana), (iv) educación parental, y (v) acceso a tecnología (computador e internet en el hogar). Estas variables permiten caracterizar la composición de las cohortes y analizar cómo evoluciona la representación de cada grupo a lo largo del ciclo educativo.

2.3. Medición del abandono y del aprendizaje contrafactual

La variable principal de interés es la *deserción relativa*, definida como la diferencia en la participación de un grupo determinado (por ejemplo, estudiantes de NSE bajo) entre Saber 5 y Saber 11 dentro de la misma cohorte esperada. Este indicador refleja la magnitud del abandono o rezago escolar en función de la pérdida proporcional de representación de cada grupo.

Para estimar el efecto de la deserción sobre la desigualdad del aprendizaje, se construyó un modelo predictivo de puntajes contrafactuales en Saber 11. Este modelo, basado en técnicas de aprendizaje automático, predice el desempeño hipotético de los estudiantes que realizaron Saber 5 pero no aparecen en las cohortes correspondientes de Saber 11, condicionando la predicción a las características observadas en el momento de la evaluación de primaria.

Formalmente, el modelo estima:

$$\hat{Y}_{i,11} = f(X_{i,5}) + \epsilon_i, \quad (1)$$

donde $\hat{Y}_{i,11}$ representa el puntaje esperado en Saber 11 del estudiante i dado su vector de características observadas en Saber 5 ($X_{i,5}$), y $f(\cdot)$ es una función no paramétrica ajustada mediante técnicas de *gradient boosting* (XGBoost). El entrenamiento se realiza sobre los estudiantes con observaciones en ambos grados, y las predicciones se aplican a las observaciones censales de Saber 5 para simular el aprendizaje perdido asociado al abandono.

El modelo fue validado mediante procedimientos de validación cruzada y comparaciones con especificaciones lineales y logarítmicas, garantizando la estabilidad de las estimaciones y su capacidad predictiva.

El diseño se aproxima a un esquema cuasi-experimental al aprovechar la exposición diferenciada de las cohortes al *shock* pandémico. La comparación entre cohortes previas y posteriores a 2020 permite inferir si la pandemia exacerbó el abandono entre los grupos más vulnerables. Adicionalmente, se estimaron regresiones con efectos fijos por año y estable-

cimiento educativo, controlando factores institucionales y tendencias generales del entorno socioeconómico.

2.4. Limitaciones y alcance del enfoque

Una restricción clave del análisis radica en la ausencia de identificadores únicos que permitan vincular a los mismos individuos entre Saber 5 y Saber 11. En consecuencia, el estudio opera a nivel agregado mediante cohortes, lo que implica un enfoque ecológico y no individual. Este diseño permite observar patrones poblacionales pero no atribuir causalidad a nivel micro.

Como toda inferencia basada en agregados, el modelo puede incurrir en la llamada “falacia ecológica”: asumir que relaciones observadas en promedio se replican a nivel individual. Por tanto, las conclusiones deben interpretarse en términos de tendencias estructurales, no de comportamientos individuales.

Pese a estas limitaciones, el uso de efectos fijos y pruebas de robustez aporta consistencia a los resultados, reduciendo la posibilidad de que los patrones observados reflejen simples variaciones institucionales o coyunturales. En conjunto, este enfoque ofrece una base empírica sólida para vincular el abandono escolar con la desigualdad del aprendizaje y estimar la magnitud del conocimiento perdido durante la pandemia.

3. Análisis de Brechas en el Abandono Escolar en Antioquia: Tendencias y Vulnerabilidades entre las Cohortes de Saber 5 y Saber 11 (2014-2023)

Este capítulo presenta un análisis empírico detallado sobre la composición sociodemográfica de las cohortes escolares que presentaron las pruebas Saber 5 y Saber 11 entre 2014 y 2023. El objetivo principal es identificar patrones sistemáticos de abandono escolar, destacando los grupos más vulnerables a la exclusión educativa a lo largo del ciclo escolar. Para ello, se comparan las distribuciones de variables clave entre grado 5° y grado 11° para las cohortes que finalizaron la educación primaria entre 2014 a 2017 y por tanto debieron finalizar su ciclo escolar entre 2020 y 2023, agrupadas por su año esperado de finalización del ciclo educativo.

Iniciamos este capítulo con un análisis empírico que compara la distribución del nivel socioeconómico individual (NSE) entre los estudiantes de cada cohorte en los grados 5° y 11°. El Nivel Socioeconómico (NSE) utilizado por el ICFES es una medida continua que refleja las condiciones socioeconómicas del entorno familiar de los estudiantes. Este índice se construye a partir de información recopilada durante la aplicación de las pruebas Saber, incluyendo variables como el nivel educativo y ocupacional de los padres, características del hogar, tipo de colegio y hábitos de lectura, entre otros Instituto Colombiano para la Evaluación de la Educación (ICFES) (2019). El resultado es un índice en una escala de 0 a 100, con una media de 50 y una desviación estándar de 10. A partir de este índice, los estudiantes se clasifican en cuatro categorías de nivel socioeconómico (NSE), donde el NSE 1 representa el nivel más bajo y el NSE 4 el más alto. Este índice es fundamental para caracterizar la población.

La figura 1 muestra las funciones de densidad estimadas para cada cohorte (2014 a 2017), contrastando los estudiantes que presentaron Saber 5 y los que efectivamente presentaron Saber 11 seis años después. En todos los casos, se observa un desplazamiento hacia la derecha en la distribución del NSE en Saber 11 en comparación con Saber 5. Esto indica que, en promedio, los estudiantes que llegaron a grado 11 pertenecen a niveles socioeconómicos más

altos que aquellos que estaban presentes en grado 5. Este patrón es consistente con la hipótesis de que los estudiantes de NSE bajo son más propensos a abandonar el sistema escolar antes de culminar la educación media.

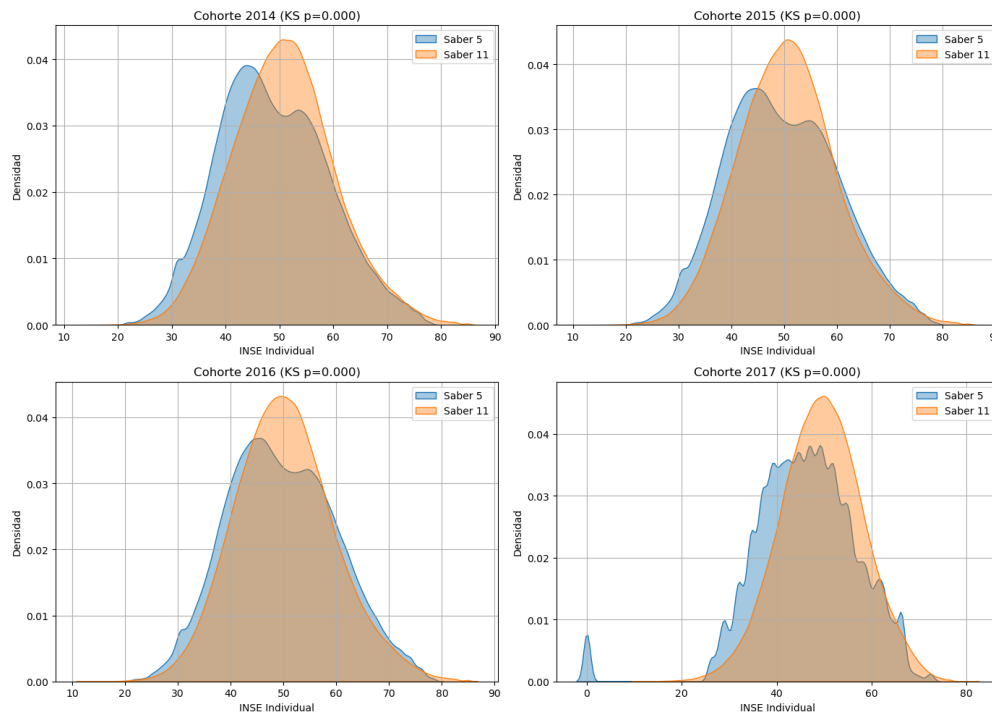


Figura 1: Distribución del NSE individual en Saber 5 y Saber 11 para las cohortes que iniciaron Saber 5 entre 2014 y 2017.

Fuente: Elaboración propia con datos del ICFES.

Cohorte	Media S5	Media S11	Mediana S5	Mediana S11	KS Stat	p-valor
2014	48.81	50.81	48.04	50.63	0.1161	0.000
2015	49.52	50.63	48.91	50.41	0.0861	0.000
2016	49.68	50.27	49.11	49.91	0.0603	0.000
2017	46.04	49.45	46.28	49.50	0.1588	0.000

Cuadro 1: Medias, medianas y prueba de Kolmogorov-Smirnov del NSE individual por cohorte y prueba.

Fuente: Elaboración propia con datos del ICFES.

Este patrón se confirma mediante la prueba de Kolmogorov-Smirnov (Tabla 1), que muestra diferencias estadísticamente significativas ($p < 0.001$) en todas las cohortes. La cohorte 2017, que enfrentó el mayor impacto de la pandemia, exhibe la mayor divergencia entre

ambas distribuciones.

Aunque las diferencias observadas entre los estudiantes que presentaron Saber 5 y los que llegaron a Saber 11 sugieren patrones claros de abandono escolar, es importante reconocer que este análisis no cuenta con un seguimiento individual (panel de estudiantes). En consecuencia, los cambios en la composición sociodemográfica pueden deberse tanto a procesos de deserción como a dinámicas de transición entre instituciones educativas, migración entre zonas, o cambios estructurales en el acceso a servicios.

Para abordar esta limitación, adoptamos un enfoque que combina análisis de diferencias agregadas, estimaciones de densidades y medias condicionadas por territorio, y un modelo de datos de panel con efectos fijos por municipio y año. Este enfoque permite identificar patrones sistemáticos que van más allá de fluctuaciones aleatorias o sesgos locales. En particular, los cambios consistentes entre cohortes, los desplazamientos hacia niveles más altos de NSE, el aumento sostenido de la proporción de estudiantes con acceso a internet y computador, y la caída en la representación del sector oficial sugieren la existencia de procesos de exclusión educativa que afectan de manera desproporcionada a los estudiantes más vulnerables.

Para fortalecer esta interpretación, estimamos un modelo de datos de panel con efectos fijos por municipio y año, utilizando información de todas las pruebas Saber 5 (2014-2017) y Saber 11 entre 2014 y 2023. Este modelo permite comparar el nivel socioeconómico promedio entre grados dentro del mismo municipio y año, controlando por variaciones estructurales a nivel territorial y temporal.

El modelo estimado tiene la siguiente forma:

$$\text{NSE}_{it} = \beta_0 + \beta_1 \cdot \text{GRADO11}_{it} + \alpha_i + \delta_t + \varepsilon_{it} \quad (2)$$

donde:

- NSE_{it} es el nivel socioeconómico promedio del municipio i en el año t ,
- GRADO11_{it} es una variable binaria que toma valor 1 si los datos corresponden a estu-

diantes de grado 11, y 0 para grado 5,

- α_i representa los efectos fijos por municipio (características inobservables constantes en el tiempo),
- δ_t representa los efectos fijos por año (tendencias macroeconómicas o políticas comunes a todos los municipios),
- ε_{it} es el término de error.

Variable	Coefficiente	Error estándar	t-estadístico	p-valor	IC 95 %
GRADO11	-0.3364	0.0809	-4.16	0.000	[-0.4949, -0.1779]

Cuadro 2: Estimación del modelo de datos de panel con efectos fijos por municipio y año. Variable dependiente: NSE promedio.

Fuente: Elaboración propia con datos del ICFES.

El resultado muestra que, en promedio, el NSE en grado 11 es significativamente más alto que en grado 5, con una diferencia de 0.34 puntos ($p < 0,001$). Esta estimación confirma que las diferencias observadas no se deben únicamente a mejoras en las condiciones de vida, sino que reflejan un patrón sistemático de abandono escolar por parte de los estudiantes de menores recursos. Si bien no podemos afirmar con certeza que un estudiante en Saber 5 abandonó la escuela o migró a otra institución, el conjunto de evidencia presentada permite establecer inferencias robustas sobre la magnitud y perfil de la exclusión escolar.

Las gráficas 2 - 4 presentan la comparación de la proporción de estudiantes entre Saber 5 y Saber 11 según extraedad, acceso a internet, acceso a computador y género. Como se observa, en todos los casos analizados se evidencian cambios relevantes en la composición de las cohortes. En primer lugar, se registra un aumento en la proporción de estudiantes en situación de extraedad en Saber 11 en comparación con Saber 5. Este resultado es esperable, pues el paso del tiempo naturaliza que una parte de los estudiantes repita uno o más grados, pero también sugiere que muchos de quienes estaban en edad adecuada al momento de presentar Saber 5 no lograron avanzar con normalidad, acumulando rezago.

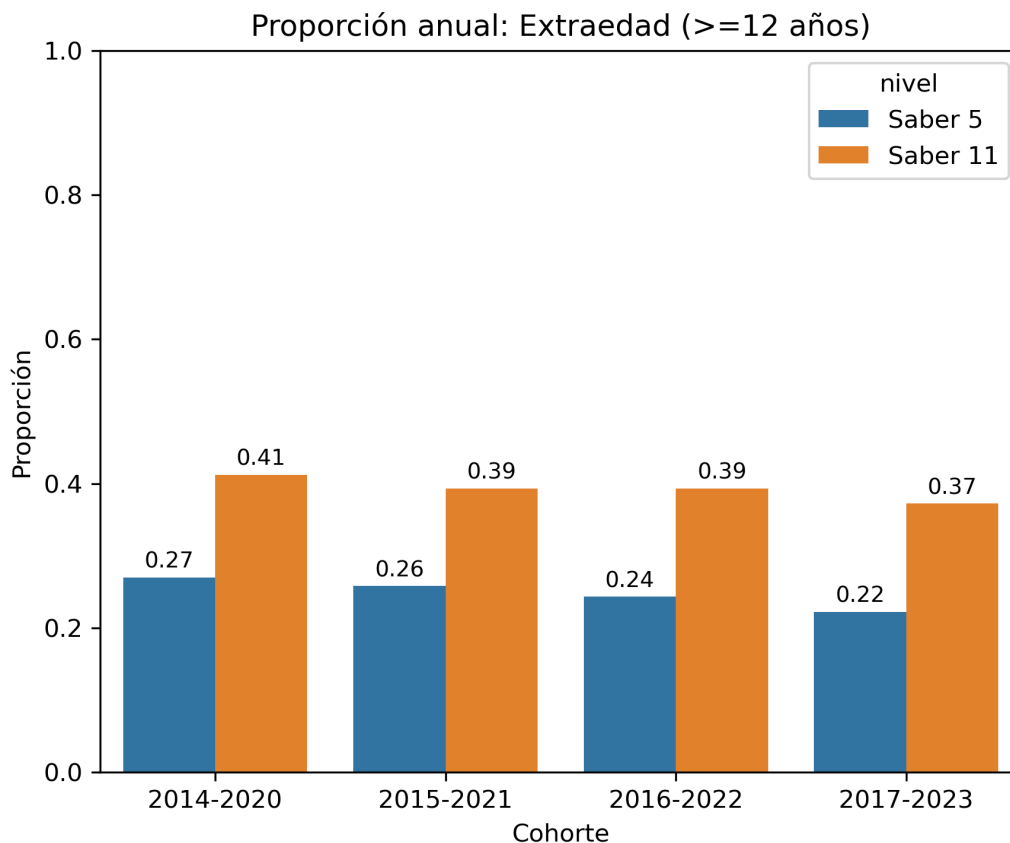


Figura 2: Participación relativa de estudiantes con extraedad en Saber 5 y Saber 11

Fuente: Elaboración propia con datos del ICFES.

Una tendencia importante es el incremento notable en la participación de estudiantes con acceso a internet y computador en el hogar. En Saber 11, estos grupos representan una proporción mayor de la cohorte, lo cual podría indicar que los estudiantes sin acceso a tecnología enfrentan mayores riesgos de exclusión del sistema educativo; sin embargo, dado el choque de la pandemia que implementó un modelo de alternancia con educación virtual, el patrón también podría indicar un incremento acelerado en el acceso a internet de toda la población.

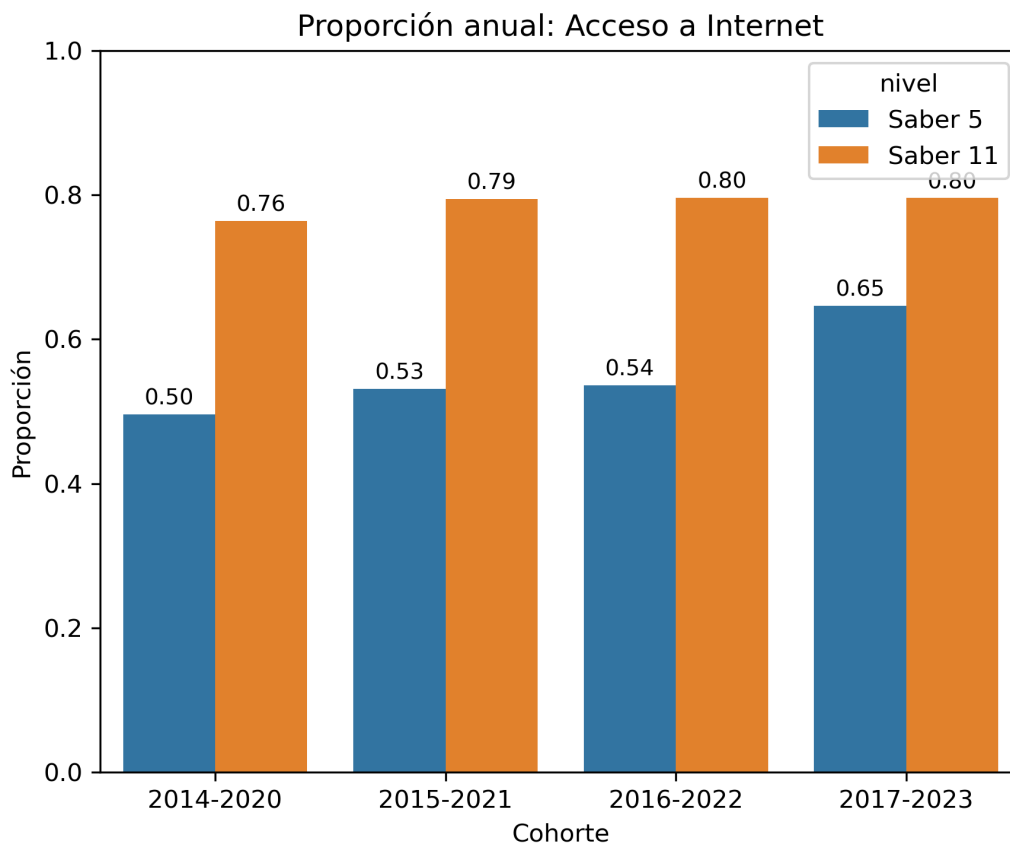


Figura 3: Participación relativa de estudiantes con internet en Saber 5 y Saber 11

Fuente: Elaboración propia con datos del ICFES.

En cuanto al género, se identifica una mayor participación de mujeres en Saber 11, lo que sugiere una mayor permanencia de las niñas en el sistema. Aunque la diferencia no es muy pronunciada, se alinea con evidencia previa que muestra una tasa de deserción ligeramente más alta entre varones en ciertos contextos. Uno de los hallazgos más consistentes es la reducción en la proporción de estudiantes de colegios oficiales entre grado quinto y grado once. Este resultado puede deberse a dos procesos simultáneos: por un lado, un mayor abandono escolar entre estudiantes del sector oficial, y por otro, una posible migración hacia instituciones privadas por parte de quienes logran permanecer en el sistema.

El gráfico 5 presenta la participación relativa de los estudiantes en Saber 5 y Saber 11 según el nivel educativo alcanzado por sus madres, considerando todas las cohortes disponibles

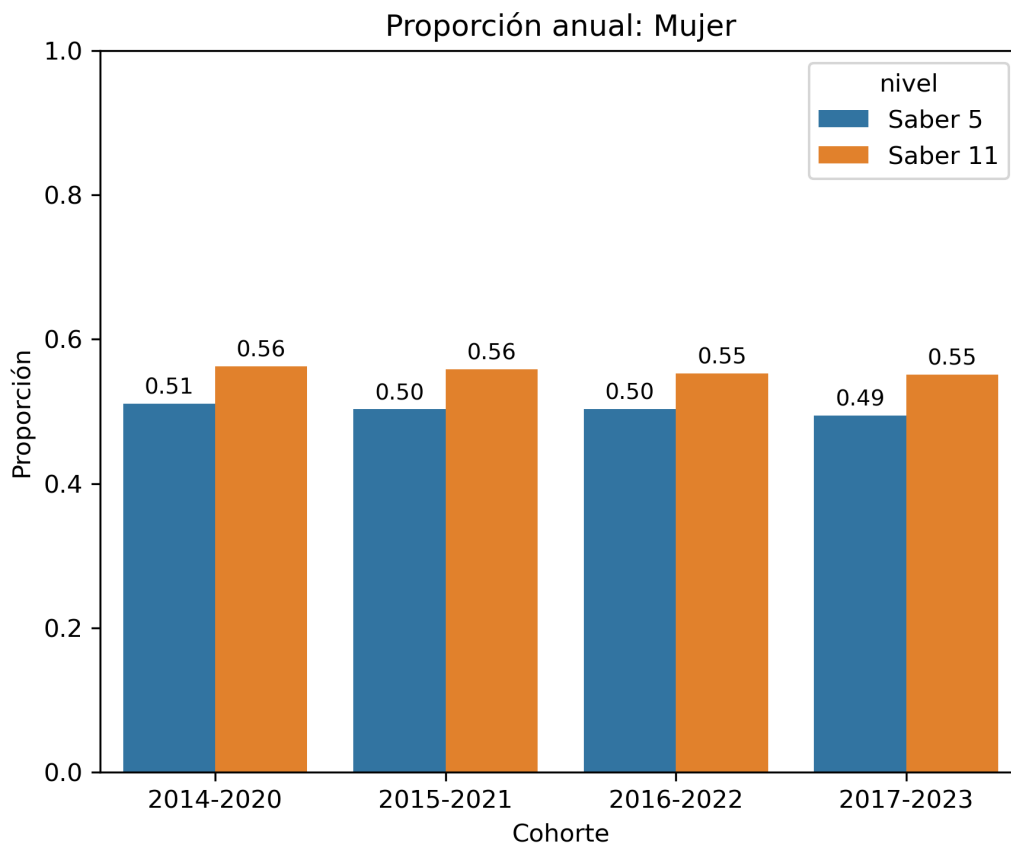


Figura 4: Participación relativa de estudiantes por género en Saber 5 y Saber 11

Fuente: Elaboración propia con datos del ICFES.

de manera agregada. Esta variable es clave para identificar factores estructurales que inciden en la permanencia o abandono escolar, ya que el capital educativo de las madres se relaciona fuertemente con el apoyo académico en el hogar y las trayectorias educativas de sus hijos.

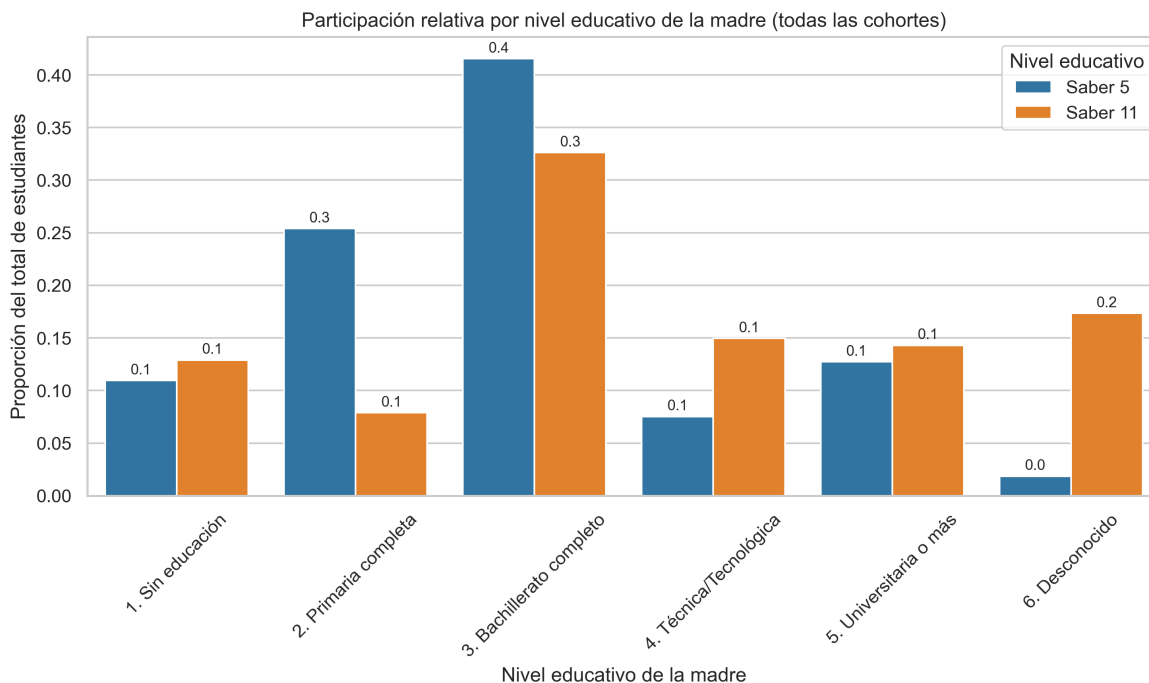


Figura 5: Participación relativa de grupos seleccionados en Saber 5 y Saber 11

Fuente: Elaboración propia con datos del ICFES.

Los resultados muestran un patrón claro: los estudiantes cuyas madres tienen un nivel educativo medio o alto (especialmente técnica, tecnológica o universitaria) tienden a aumentar su representación en grado 11. En particular, el grupo con madres con educación técnica o tecnológica prácticamente *duplica su participación relativa* entre grado 5 y grado 11, mientras que el grupo con madres con formación universitaria también gana peso. Esto sugiere que el capital educativo de las madres actúa como un factor protector frente al abandono.

Por otro lado, los estudiantes cuyas madres solo completaron la primaria pierden participación de forma importante en Saber 11. Este grupo pasa de representar una proporción significativa de los estudiantes en Saber 5 a menos de la mitad de su valor inicial en grado

11. Esta caída indica un mayor riesgo de abandono y retraso para este perfil familiar, lo que podría estar relacionado con una menor capacidad de acompañamiento académico, mayores barreras socioeconómicas, o condiciones estructurales de exclusión educativa.

También se observa una leve disminución en la participación de estudiantes con madres que completaron el bachillerato, aunque sigue siendo el grupo mayoritario en ambos grados. Finalmente, la participación del grupo de madres sin ningún nivel educativo formal se mantiene relativamente estable, aunque con un leve aumento. Esto sugiere que este grupo puede estar más heterogéneamente compuesto de lo esperado, posiblemente incluyendo estudiantes de contextos rurales o con registros incompletos. En conjunto, los resultados refuerzan la idea de que la educación materna tiene un papel central en la permanencia escolar, y que las políticas públicas orientadas a reducir el abandono deberían considerar este indicador para diseñar intervenciones más focalizadas y efectivas.

Estos resultados permiten identificar algunos de los factores estructurales que explican la exclusión educativa. El abandono no ocurre de manera aleatoria ni homogénea, sino que tiende a afectar de forma desproporcionada a los estudiantes con mayores niveles de vulnerabilidad, en particular aquellos con menos recursos tecnológicos y provenientes de colegios del sector oficial. Lo anterior refuerza la importancia del entorno familiar en la trayectoria escolar y sugieren que los esfuerzos de política pública deben considerar dichas variables como factores clave para diseñar intervenciones focalizadas.

4. Análisis de Brechas en el Aprendizaje Escolar en Antioquia: Tendencias y Vulnerabilidades entre las Cohortes de Saber 5 y Saber 11 (2014-2023)

Saber 11 tiene como objetivo medir las habilidades y competencias adquiridas por los estudiantes en su trayectoria a lo largo del sistema educativo de Colombia. Esta prueba se

realiza dos veces al año, una vez por semestre, y está dirigida a todos los estudiantes de undécimo grado. Además, puede ser presentada por personas que están finalizando su educación para adultos o por aquellos que desean validar su bachillerato. Aquí nos concentramos en los estudiantes de grado undécimo y en la aplicación a la segunda mitad del año, pues más del 90 % de los estudiantes de Antioquia presentan el examen en esta aplicación, mientras en la prueba de principio de año participan principalmente colegios privados de calendario B. En este capítulo nos concentramos en el puntaje promedio global y por áreas. El puntaje promedio global es calculado usando las pruebas de Lectura Crítica, Matemáticas, Ciencias Naturales, Sociales y Ciudadanas e inglés (ICFES, 2018). Para cada prueba, el puntaje promedio se mide en una escala de 0 a 100 puntos. Luego, a partir de un promedio ponderado de la calificación en las cinco pruebas, se deduce un puntaje global del examen, dado en una escala entre 0 y 500 puntos.

Al analizar la evolución del puntaje global en Antioquia entre 2014 y 2023, se observa una caída sostenida hasta 2021 y una posterior recuperación parcial. En particular, el puntaje pasó de 266 puntos en 2016 a 250 en 2021 (una caída de 16 puntos), desde 2022, ambos sectores muestran una recuperación de aproximadamente 4 puntos. Una mirada ingenua de estas cifras sugeriría una cierta estabilidad relativa del sistema y podría llevar a la conclusión de que, tras un breve periodo de afectación, la educación se encuentra nuevamente en una senda de mejora.

Sin embargo, estas observaciones son incompletas. Se basan exclusivamente en los estudiantes que lograron culminar su educación media y presentar la prueba, lo que introduce un sesgo estructural importante: el universo de estudiantes evaluado no incluye a quienes abandonaron el sistema educativo antes de llegar a grado 11, ni a quienes, por diversos motivos, no lograron presentar el examen. Este sesgo de selección puede distorsionar la interpretación de los resultados, especialmente en contextos como el de la pandemia, donde la deserción escolar aumentó y afectó desproporcionadamente a estudiantes en condiciones de mayor vulnerabilidad.

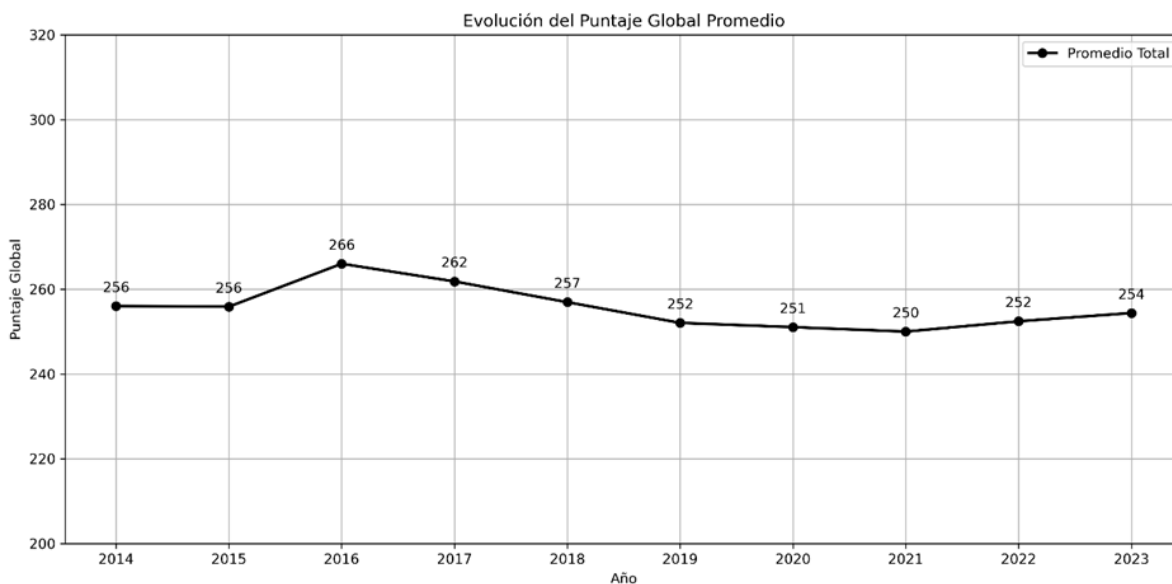


Figura 6: Resultados históricos de saber 11, observados sin contrafactual.

Fuente: Elaboración propia con datos del ICFES.

La estabilidad aparente en los puntajes promedio puede ocultar un fenómeno más preocupante: una recomposición de la población evaluada, en la que los estudiantes con mayores dificultades de aprendizaje o en condiciones más precarias simplemente ya no están presentes. En este escenario, los promedios globales no reflejan mejoras reales del sistema educativo, sino una muestra cada vez más selectiva de quienes logran permanecer.

Para superar estas limitaciones, este estudio propone una metodología basada en el análisis contrafactual. El objetivo es estimar cómo habrían sido los puntajes en Saber 11 si todos los estudiantes que comenzaron el ciclo educativo hubieran llegado a grado 11, simulando un escenario sin deserción o retraso. Esto permite no solo dimensionar el impacto oculto de la deserción sobre los resultados agregados, sino también identificar cómo este impacto varía según nivel socioeconómico, género, zona, tipo de institución y territorio.

Puesto que solo tenemos datos completos de Saber 5 de los años 2014-2017, el seguimiento hasta Saber 11 se hace entre 2020 y 2023. El punto de partida consiste en reconstruir las cohortes educativas utilizando los resultados de las pruebas Saber 5 aplicadas entre 2014-2017, y vincularlas con los resultados de Saber 11 observados entre 2020-2023. De esta manera, iden-

tificamos las trayectorias escolares esperadas y estimamos qué estudiantes deberían haber presentado la prueba Saber 11, 6 años después de culminar la primaria. Para cada cohorte, desarrollamos una estrategia de emparejamiento basada en *propensity score matching*, que nos permite identificar pares comparables entre los estudiantes que efectivamente presentaron Saber 11 y aquellos que deberían estar allí de acuerdo con los presentes en Saber 5, 6 años antes, condicionando por características observables como nivel socioeconómico, tipo de colegio, ubicación geográfica, edad y otros factores contextuales disponibles en los microdatos del ICFES.

Este enfoque metodológico permite revelar una dimensión oculta del sistema educativo: no solo cuántos estudiantes logran llegar al final del ciclo escolar, sino qué tan representativos son esos estudiantes de la cohorte original. Al hacerlo, visibiliza el verdadero alcance de las brechas educativas acumuladas —especialmente las que se intensificaron durante la pandemia— y permite un análisis más riguroso sobre la equidad y eficacia del sistema educativo a lo largo del tiempo.

Para cada cohorte, se implementa una estrategia de emparejamiento basada en propensity score weighting utilizando modelos de bosques aleatorios, lo que permite estimar la probabilidad de que un estudiante pertenezca a la cohorte original (Saber 5 o Saber 9) en función de características observables como el nivel socioeconómico, el tipo de establecimiento, la ubicación geográfica y la edad. Estas probabilidades se utilizan para ponderar los resultados de los estudiantes que sí presentaron Saber 11, simulando una distribución de características similar a la de las cohortes originales. De este modo, se construye un escenario contrafactual que permite estimar cuál habría sido el puntaje global promedio en Saber 11 si no hubiera ocurrido deserción, haciendo visible el impacto oculto de las desigualdades educativas y los efectos acumulados de la pandemia.

Una vez construidas las cohortes educativas y estimados los puntajes contrafactuales, es posible analizar la magnitud de las desigualdades educativas desde una perspectiva más completa. El gráfico 7 presenta el puntaje global promedio en Saber 11 por decil del índice

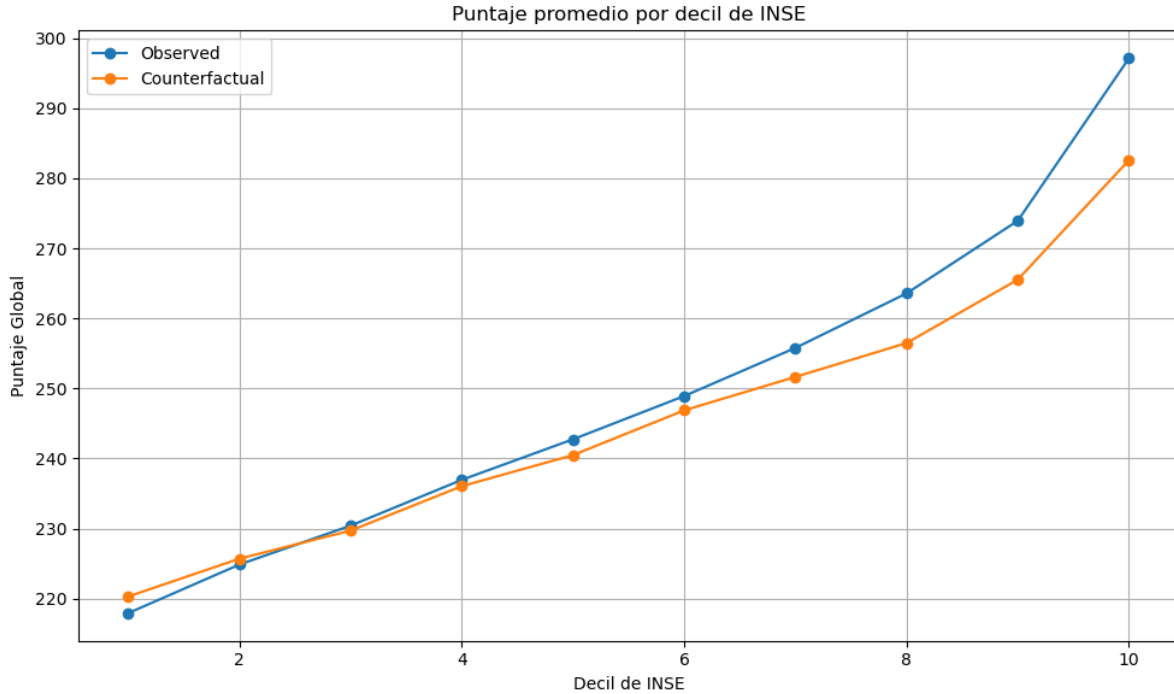


Figura 7: Puntaje promedio por decil del INSE

Fuente: Elaboración propia con datos del ICFES.

socioeconómico (INSE), diferenciando entre dos grupos: los estudiantes que efectivamente presentaron la prueba (observados) y aquellos que, según nuestro modelo contrafactual, deberían haber permanecido en el sistema (contrafactual).

La curva ascendente en ambos grupos confirma la fuerte asociación entre nivel socioeconómico y rendimiento académico. Sin embargo, lo que resulta especialmente relevante es la divergencia entre ambas líneas, que se acentúa hacia los extremos de la distribución. Mientras que en los deciles bajos (1 al 4) las diferencias entre los estudiantes observados y los contrafactuales son menores, en los deciles superiores se observa una creciente brecha a favor de los estudiantes que permanecieron en el sistema educativo. Por ejemplo, el decil más alto (10) presenta una diferencia de más de 10 puntos entre el puntaje promedio observado y el puntaje contrafactual.

Este patrón sugiere que el abandono escolar tiende a ser más común entre estudiantes de niveles socioeconómicos bajos, pero además que los estudiantes con mayor INSE que per-

manecen en el sistema tienen un rendimiento excepcionalmente alto, lo que magnifica la desigualdad cuando se compara con un escenario contrafactual en el que todos los estudiantes hubieran completado su trayectoria educativa a tiempo. De hecho, al calcular la brecha entre los extremos de la distribución (decil 10 menos decil 1), se observa una diferencia de aproximadamente 76 puntos en el grupo observado, frente a 63 puntos en el contrafactual, lo que implica que la desigualdad en los resultados educativos es mayor cuando se excluye a los desertores del análisis.

La relación entre nivel socioeconómico (INSE) y puntaje promedio se mantiene en todas las áreas evaluadas, lo que reafirma el fuerte vínculo estructural entre origen social y logros educativos. Sin embargo, la comparación entre las curvas observadas y contrafactuales permite identificar matices importantes sobre cómo opera la desigualdad y la selección en cada dimensión de aprendizaje.

En Lectura Crítica, la curva contrafactual es más cercana a la observada en los deciles bajos e intermedios, pero se abre progresivamente en los deciles altos. La diferencia entre los grupos se amplía a partir del decil 7, y en el decil 10 alcanza cerca de 3 puntos de diferencia. Aunque la magnitud no es tan dramática como en el puntaje global, este patrón sugiere que los estudiantes con mayor INSE que permanecieron en el sistema no solo tenían mejores condiciones estructurales, sino que también desarrollaron habilidades de comprensión lectora por encima de lo que se habría esperado incluso para su perfil socioeconómico. Es probable que el capital cultural, el acceso a lectura en el hogar y entornos enriquecidos expliquen parte de esta brecha.

En Matemáticas, en cambio, se observa un patrón más complejo. En los deciles bajos, el puntaje contrafactual es ligeramente superior al observado en el decil 1, y muy cercano en los deciles 2 a 6. Esto podría sugerir que, en esta área, la pérdida de aprendizajes entre los estudiantes que permanecieron fue mayor, o que los desertores tenían un desempeño potencialmente más alto. A partir del decil 7, las curvas comienzan a separarse más claramente, y en el decil 10 la brecha alcanza cerca de 3 puntos, como en Lectura Crítica, pero partiendo

de un patrón distinto: menor rendimiento en los deciles bajos y mayor diferenciación en los altos.

Este contraste revela que la selección del sistema no opera de manera uniforme entre áreas: mientras que en Lectura Crítica el patrón parece estar más asociado al capital cultural acumulado en hogares de alto INSE, en Matemáticas se evidencian pérdidas de aprendizaje más pronunciadas en los sectores bajos, posiblemente ligadas a las dificultades de mantener continuidad pedagógica en ausencia de clases presenciales. La comparación entre Lectura Crítica y Matemáticas revela diferencias importantes no solo en el nivel de puntaje alcanzado, sino también en la forma en que opera la desigualdad a lo largo de la distribución socioeconómica. En términos absolutos, los puntajes promedio en Lectura Crítica son consistentemente más altos que en Matemáticas en todos los deciles de INSE, tanto en el grupo observado como en el contrafactual. Esta diferencia sugiere que, a pesar de las dificultades impuestas por la pandemia, los estudiantes lograron sostener mejor las competencias asociadas a comprensión lectora que aquellas relacionadas con el pensamiento matemático

Estos hallazgos refuerzan la necesidad de incorporar a los desertores en la evaluación de la calidad educativa. Ignorar este grupo puede llevar a una subestimación sistemática de las desigualdades reales del sistema. Al mismo tiempo, la diferencia entre los dos escenarios refleja la capacidad del sistema educativo de ampliar las brechas cuando los mecanismos de exclusión afectan de manera desproporcionada a los estudiantes más vulnerables.

En conclusión, el análisis contrafactual no solo permite estimar el impacto agregado del abandono escolar en el rendimiento escolar en Antioquia, sino que también ofrece una mirada más precisa sobre las dinámicas de desigualdad estructural. El contraste entre los puntajes observados y contrafactuales revela que la deserción escolar actúa como un filtro que elimina del sistema a los estudiantes con menor rendimiento esperado, elevando artificialmente el promedio del examen pero a costa de profundizar las brechas sociales.

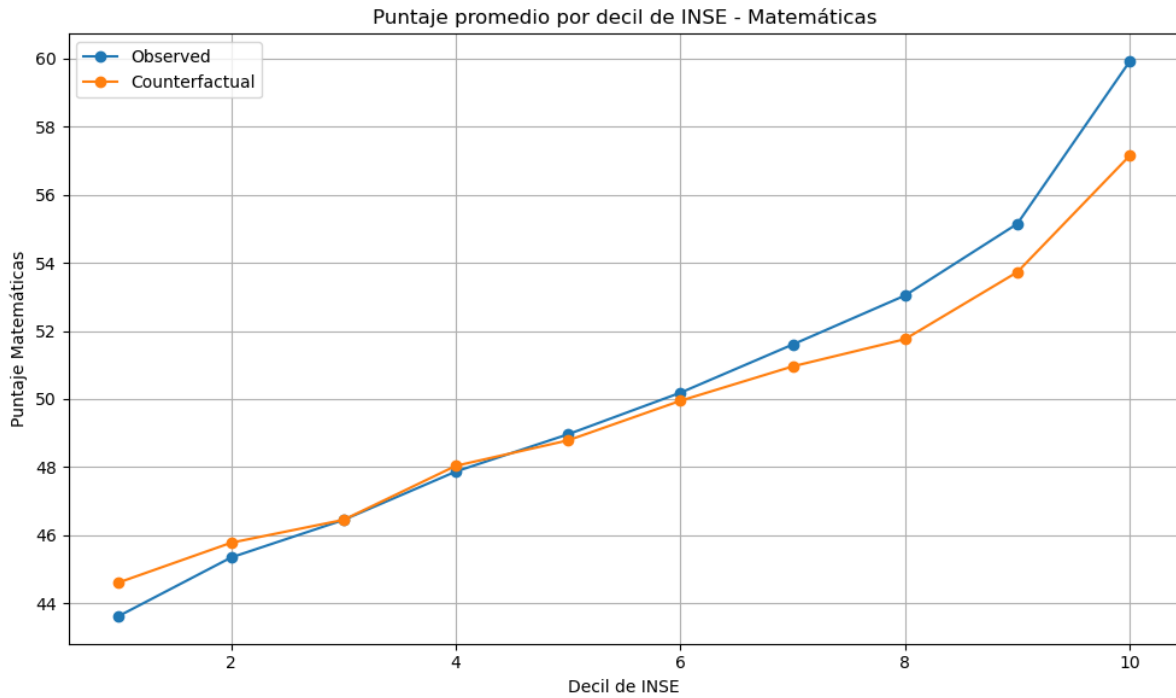


Figura 8: Puntaje promedio por decil del INSE, Matemáticas

Fuente: Elaboración propia con datos del ICFES.

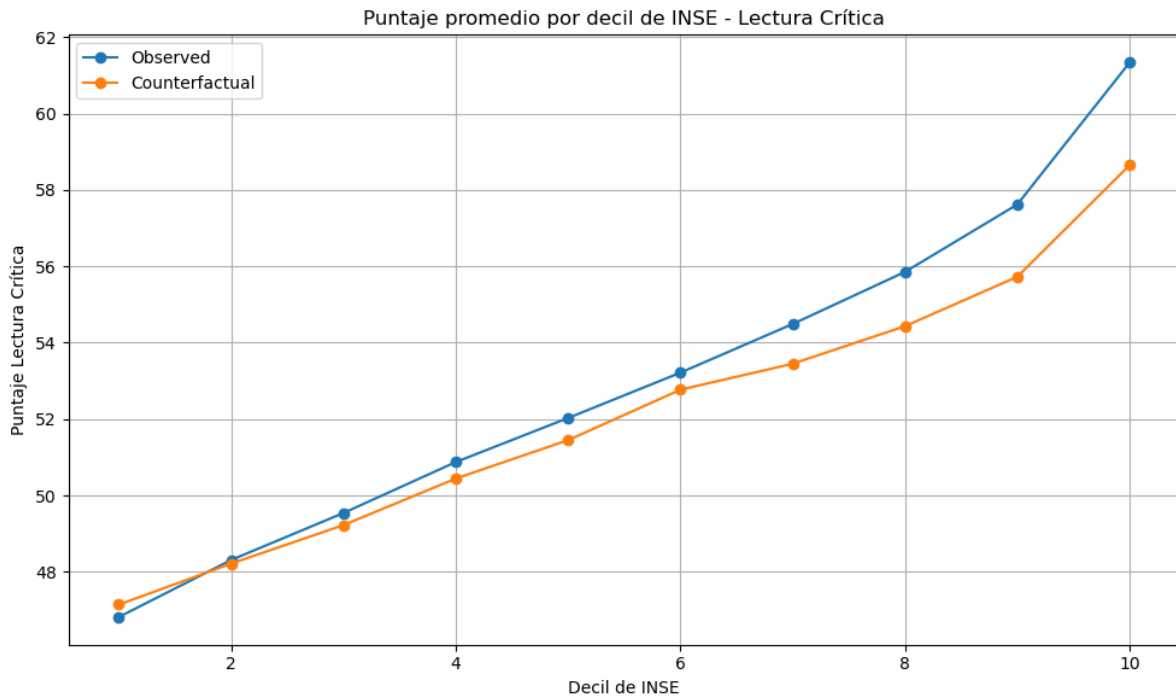


Figura 9: Puntaje promedio por decil del INSE, Lectura Crítica

Fuente: Elaboración propia con datos del ICFES.

5. Amplificación de brechas durante la pandemia del Covid-19, 2019-2023

La pandemia de COVID-19 representó el mayor shock educativo global en la historia reciente. En Colombia, las escuelas cerraron de forma generalizada entre marzo de 2020 y mediados de 2021, y en muchas regiones rurales y vulnerables el retorno a la presencialidad fue parcial y lento hasta bien entrado 2022. Esto implicó que durante casi dos años el proceso educativo dependiera de modalidades remotas y autogestionadas, con fuertes diferencias en el acceso a tecnología, conectividad y acompañamiento familiar. Las investigaciones nacionales e internacionales han documentado importantes retrocesos en los aprendizajes durante este periodo, pero también señalan que las consecuencias fueron desiguales, ampliando las brechas entre estudiantes según su nivel socioeconómico, género, ubicación geográfica y tipo de institución educativa (Abadía et al., 2021; Banco Interamericano de Desarrollo (BID), 2021).

En este estudio, nos centramos en analizar la cohorte de estudiantes que presentó la prueba Saber 11 entre 2020 y 2023, es decir, aquellos que debían ingresar a grado 11 seis años después de haber cursado quinto grado entre 2014 y 2017. La última cohorte es particularmente relevante por haber atravesado los grados más críticos de su trayectoria educativa (sexto a décimo) durante el periodo pandémico, en condiciones de alta incertidumbre, baja presencialidad y fuerte disparidad en el acceso a medios educativos. Analizar sus resultados permite estimar los efectos acumulados de la pandemia sobre los aprendizajes y, sobre todo, sobre la configuración de las desigualdades en el sistema escolar. Los comparamos con la cohorte de estudiantes que debieron finalizar su ciclo entre 2020 y 2022, por tanto su afectación fue más baja respecto al otro grupo, al desarrollar su educación secundaria en pandemia.

Para este análisis, desarrollamos una estrategia contrafactual basada en la comparación entre dos grupos: por un lado, los estudiantes que lograron permanecer en el sistema y presentaron la prueba Saber 11 entre 2020 y 2023 (grupo observado); y por otro, los estudiantes que presentaron la prueba Saber 5 entre 2014 y 2017. Mediante un modelo de emparejamiento

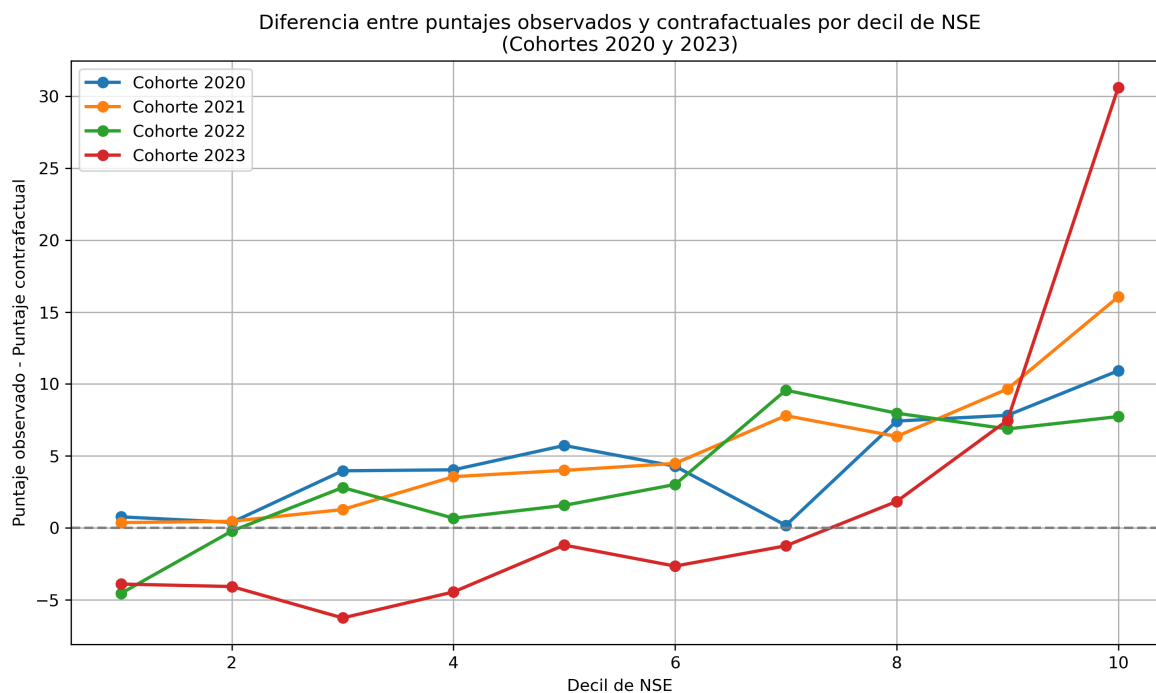


Figura 10: Puntaje promedio por decil del INSE, análisis de la pandemia

Fuente: Elaboración propia con datos del ICFES.

basado en puntajes de propensión (propensity scores), estimamos los puntajes contrafactuales que estos estudiantes de grado 5 habrían obtenido de haber llegado a grado 11 en condiciones similares a sus pares emparejados, manteniendo constantes sus características sociodemográficas iniciales.

La Figura 10 muestra de forma contundente que la pandemia de COVID-19 no solo afectó los niveles generales de aprendizaje, sino que exacerbó de manera sustantiva las desigualdades educativas existentes. Para ello, se comparan dos cohortes de estudiantes: la cohorte 2020, cuya trayectoria escolar se desarrolló en condiciones prepandemia, y la cohorte 2023, que estuvo expuesta al cierre de escuelas y modalidades remotas durante buena parte de su educación secundaria.

El eje vertical del gráfico representa la diferencia entre el puntaje observado en Saber 11 y el puntaje contrafactual estimado para estudiantes que no aparecen seis años después de haber presentado la prueba Saber 5, y que por tanto pueden considerarse desertores o reza-

gados. Este diferencial puede interpretarse como un indicador de selección dentro del sistema educativo: valores positivos indican que los estudiantes que permanecieron y presentaron la prueba obtuvieron, en promedio, mejores resultados que sus pares contrafactuales; valores negativos, por el contrario, sugieren que quienes permanecieron tienen un rendimiento inferior al que se habría esperado para sus cohortes con características similares en grado 5.

La evolución entre cohortes permite observar con nitidez cómo la pandemia de COVID-19 alteró estos patrones. En la cohorte de 2020 —última antes de la pandemia— las diferencias son relativamente estables y positivas a lo largo de los deciles, con una pendiente leve que sugiere un sistema moderadamente selectivo pero no profundamente desigual. A partir de 2021, sin embargo, se observa un cambio gradual: la curva se vuelve más inclinada, las diferencias en los deciles bajos se reducen o vuelven negativas, y las brechas en los deciles altos se amplifican. Esta transformación se profundiza en las cohortes de 2022 y 2023, donde el patrón se torna más extremo. En particular, en la cohorte de 2023 se registra una brecha negativa en los deciles 1 a 6, lo que implica que los estudiantes que permanecieron en el sistema en esos grupos de ingresos obtuvieron puntajes más bajos que los estimados para sus contrafactuales. En contraste, en el decil 10 la diferencia se dispara a más de 30 puntos, reflejando una fuerte sobre-representación de estudiantes con alto rendimiento entre quienes finalizaron el ciclo educativo en los estratos más altos.

En Lectura Crítica, la brecha negativa en los deciles 1 a 6 se hace evidente solo a partir de la cohorte 2023, lo que sugiere que los impactos sobre la comprensión lectora acumulados durante la pandemia tardaron más tiempo en manifestarse plenamente. Hasta 2022, los estudiantes que permanecieron en el sistema aún lograban rendimientos iguales o superiores a los contrafactuales, pero a partir de 2023 esta ventaja desaparece en los sectores bajos. La creciente diferencia positiva en el decil 10 a lo largo de los años —que alcanza casi 6 puntos en 2023— muestra que los estudiantes de mayor INSE no solo resistieron los efectos de la pandemia, sino que incluso profundizaron su ventaja relativa.

En Matemáticas, el patrón de deterioro es más rápido y más profundo. A partir de la

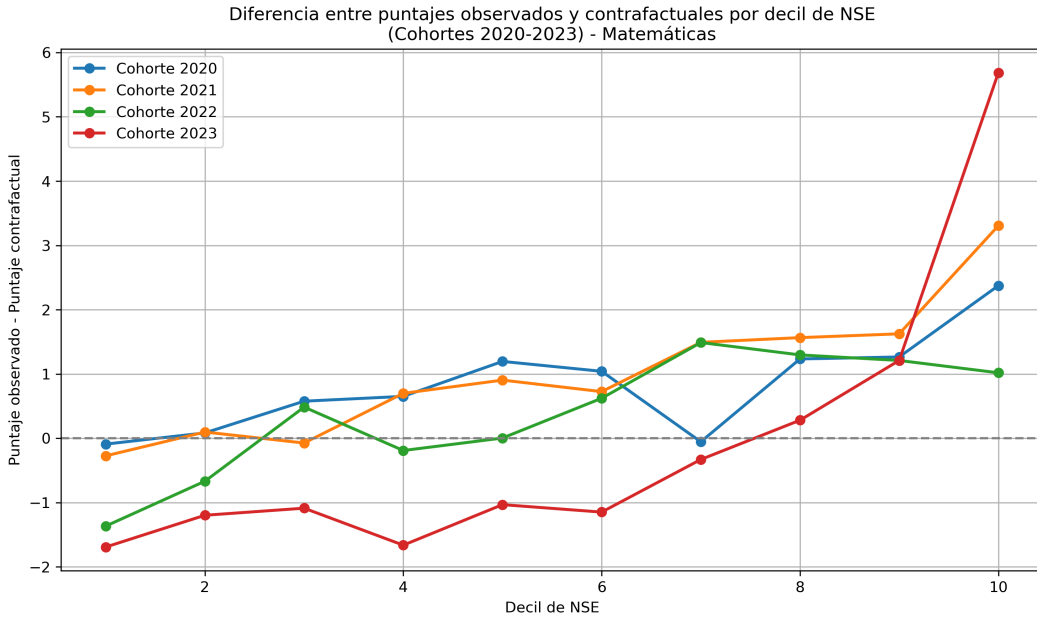


Figura 11: Puntaje promedio por decil del INSE, análisis de la pandemia matemáticas

Fuente: Elaboración propia con datos del ICFES.

cohorte 2021 ya se observan brechas negativas en los deciles bajos, que se acentúan con el tiempo. La cohorte de 2023 presenta diferencias negativas más marcadas que en Lectura Crítica, y una curva más empinada, con una diferencia de casi 6 puntos en el decil 10. Esto indica que los aprendizajes en Matemáticas fueron más sensibles a las condiciones adversas de la pandemia: menor interacción docente-estudiante, dificultad para seguir contenidos de forma autónoma y menor disponibilidad de apoyo en el hogar podrían haber contribuido a un mayor rezago en esta área.

El contraste entre ambas áreas revela un patrón claro: Matemáticas muestra pérdidas más tempranas y pronunciadas entre los sectores vulnerables, mientras que en Lectura Crítica las brechas se amplifican más en los sectores altos. Esto sugiere que la desigualdad generada por la pandemia no solo operó sobre el acceso o la permanencia, sino también sobre la profundidad y naturaleza de los aprendizajes adquiridos, diferenciando las trayectorias educativas por área de conocimiento.

Estos resultados permiten explorar al menos dos mecanismos complementarios. Primero, es probable que los estudiantes de sectores más desfavorecidos que permanecieron en el siste-

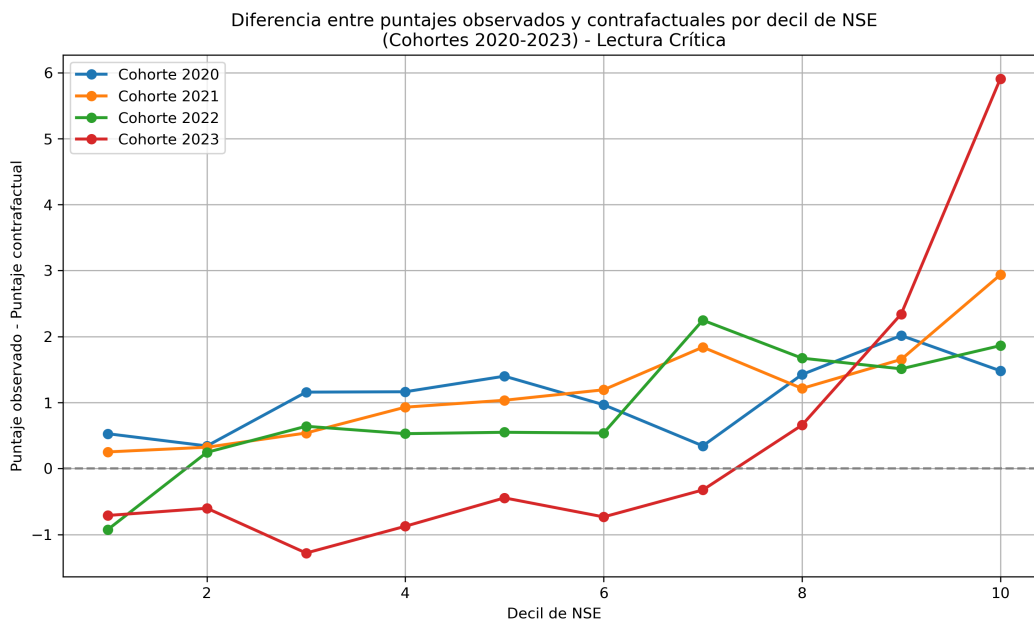


Figura 12: Puntaje promedio por decil del INSE, análisis de la pandemia lectura crítica

Fuente: Elaboración propia con datos del ICFES.

ma lo hayan hecho en condiciones precarias para el aprendizaje: sin conectividad estable, con espacios domésticos poco propicios, y con escaso acompañamiento pedagógico. La educación remota profundizó estas desigualdades, y eso podría explicar por qué sus puntajes fueron inferiores a los esperados. Segundo, aunque no puede afirmarse con certeza, es plausible que la deserción escolar durante la pandemia haya operado de forma selectiva, excluyendo a estudiantes con trayectorias más prometedoras dentro de los sectores populares. No obstante, esta interpretación debe asumirse con cautela, ya que el análisis se basa en una estrategia contrafactual con comparación ecológica y no en el seguimiento longitudinal de estudiantes individuales.

En conjunto, los datos muestran que los estudiantes más pobres no solo enfrentaron mayor riesgo de abandono escolar, sino que también experimentaron una pérdida significativa de aprendizajes, incluso si lograron mantenerse en el sistema. A la inversa, los estudiantes de mayores ingresos no solo resistieron los efectos de la pandemia, sino que ampliaron su ventaja relativa, probablemente gracias a mejores condiciones para la educación en el hogar, acceso a tecnología y apoyo familiar. Así, la pandemia actuó como un acelerador de desigualdades

preexistentes, profundizando las brechas tanto en términos de permanencia como de logros académicos.

El gráfico 13 muestra los resultados para la cohorte de 2023 por género, que en todos los deciles de NSE, las mujeres presentan brechas contrafactuales más negativas que los hombres, especialmente en los niveles socioeconómicos más bajos (deciles 1 a 6). En otras palabras, las mujeres que permanecieron en el sistema educativo tuvieron un desempeño inferior al que se habría esperado según sus características iniciales, más marcado que el de sus pares hombres. Aunque ambos géneros convergen en los deciles más altos, donde las brechas se tornan muy positivas, la diferencia en los sectores bajos sugiere que las condiciones estructurales que afectaron el aprendizaje durante la pandemia impactaron más fuertemente a las mujeres.

Esto podría estar relacionado con cargas de trabajo doméstico desiguales, menores condiciones de estudio en el hogar, o mayores riesgos de abandono escolar ligados a embarazo adolescente o tareas de cuidado. También cabe considerar que las políticas de respuesta educativa pudieron no haber considerado adecuadamente las barreras específicas que enfrentan las niñas y adolescentes en contextos de crisis.

El gráfico 14 muestra que las brechas observadas son más negativas en las zonas rurales a lo largo de casi todos los deciles, pero especialmente marcadas en los niveles más bajos de NSE. De hecho, en los deciles 1 a 6, los estudiantes rurales que permanecieron en el sistema obtuvieron puntajes significativamente más bajos que sus contrafactuales, en contraste con el patrón urbano, donde las brechas, aunque también negativas, son menos pronunciadas.

Esto es coherente con la evidencia sobre la infraestructura limitada en zonas rurales: menor acceso a internet, menor cobertura de plataformas digitales, dificultades para mantener contacto con docentes durante el cierre escolar, y en general, menores capacidades de resiliencia institucional en contextos rurales. La similitud de la curva en los deciles altos indica que el segmento más privilegiado logra superar estas diferencias, probablemente por acceso a instituciones privadas o urbanas incluso si el hogar está en zona rural. El gráfico 15 muestra diferencias dramáticas entre estudiantes del sector oficial y del no oficial. En particular:

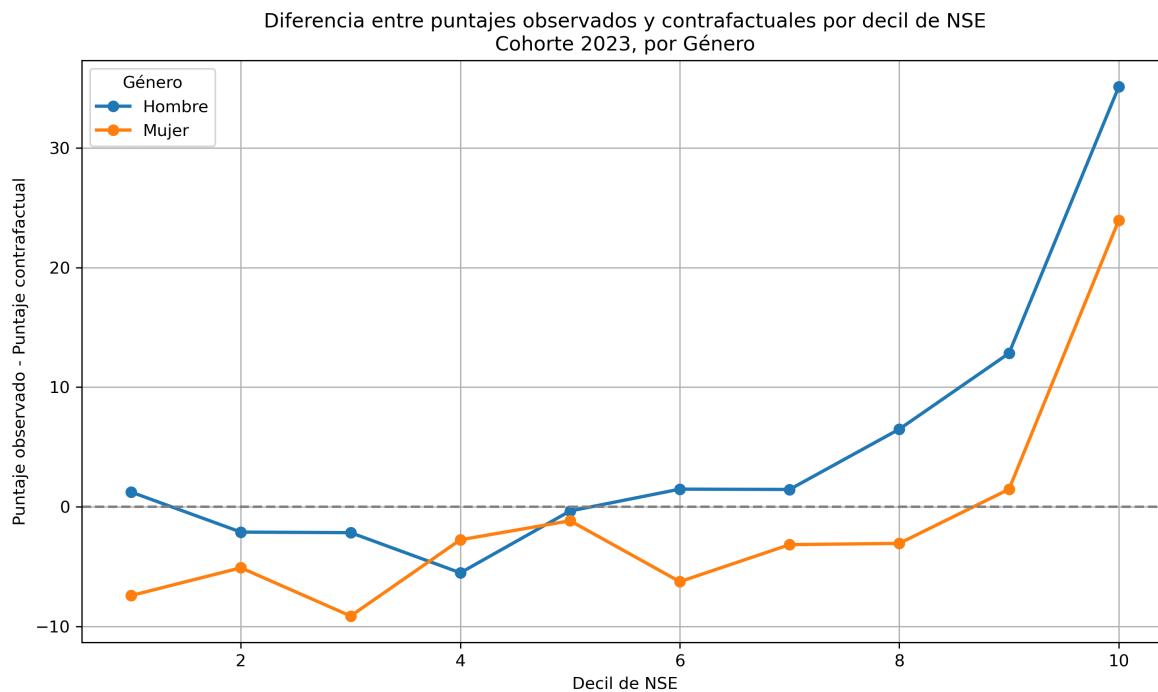


Figura 13: Puntaje promedio por decil del INSE, por género año 2023

Fuente: Elaboración propia con datos del ICFES.

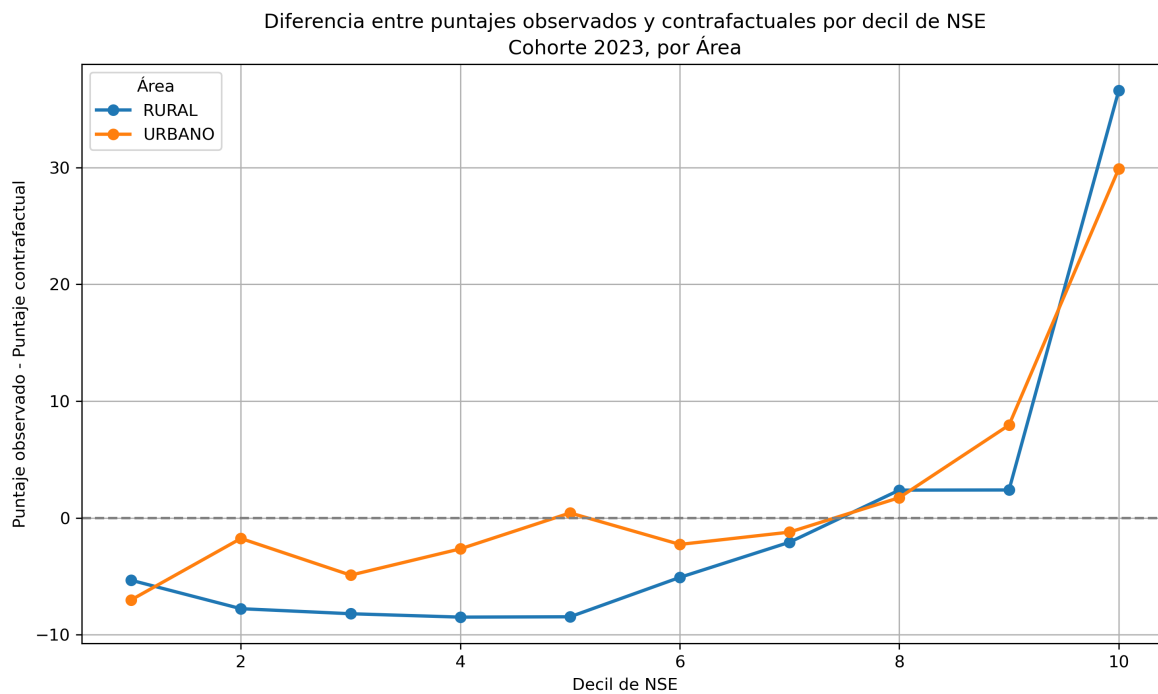


Figura 14: Puntaje promedio por decil del INSE, por área año 2023

Fuente: Elaboración propia con datos del ICFES.

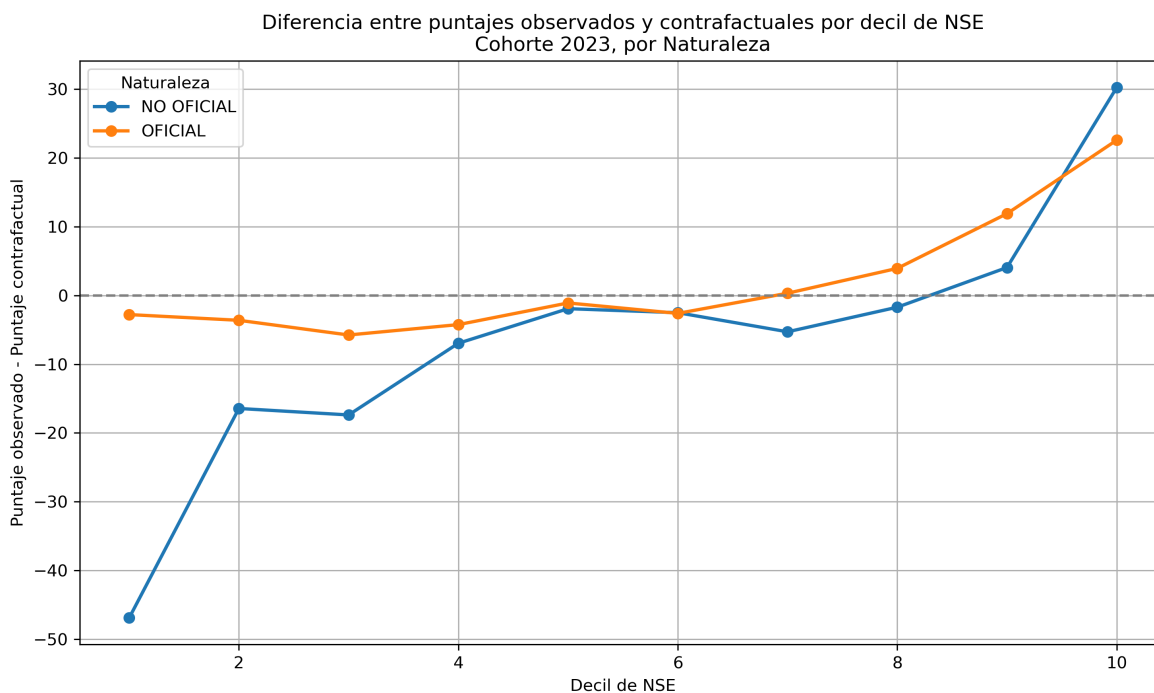


Figura 15: Puntaje promedio por decil del INSE, por sector año 2023

Fuente: Elaboración propia con datos del ICFES.

en los primeros cuatro deciles de NSE, los estudiantes de establecimientos no oficiales que permanecieron en el sistema muestran brechas contrafactuales mucho más negativas (hasta -45 puntos), en comparación con los del sector oficial, cuya brecha es más contenida (aunque igualmente negativa). La curva para el sector no oficial muestra una forma más extrema: una caída profunda en los deciles bajos y una subida abrupta en los deciles altos.

Esto sugiere una mayor segregación por nivel socioeconómico dentro del sector no oficial: estudiantes vulnerables que lograron mantenerse en el sistema privado lo hicieron en condiciones críticas y con bajo rendimiento, mientras que los más ricos —en los deciles 9 y 10— experimentaron rendimientos extraordinariamente superiores a lo esperado. El sector oficial, si bien con brechas relevantes, parece haber contenido más estas desigualdades, lo que sugiere que la cobertura pública jugó un rol de amortiguación social frente a la desigualdad educativa, aunque de forma insuficiente.

Desde la perspectiva de política pública, estos hallazgos plantean un desafío urgente: los

indicadores convencionales de calidad educativa, que se basan únicamente en los estudiantes que logran culminar el ciclo escolar, tienden a invisibilizar el impacto real de la crisis. La aparente “recuperación” pospandemia que muestran algunas métricas oficiales puede ser en realidad una ilusión estadística generada por una muestra cada vez más sesgada hacia los estudiantes con mayores ventajas estructurales. En este contexto, el uso de análisis contrafactuales como el aquí presentado se vuelve crucial para revelar las brechas ocultas por la deserción escolar y para orientar intervenciones educativas que realmente respondan a las pérdidas acumuladas en los sectores más vulnerables.

6. Conclusiones

El análisis de las trayectorias educativas en Antioquia, desde grado 5 hasta grado 11, revela transformaciones sistemáticas en la composición de las cohortes que permiten comprender mejor los mecanismos del abandono escolar. Se observa un incremento generalizado de estudiantes en extraedad, reflejo de rezagos acumulados en la progresión escolar; al mismo tiempo, aumentan las proporciones de estudiantes con acceso a internet y computador en el hogar, lo que indica tanto una expansión en el acceso tecnológico como un posible sesgo en la permanencia escolar a favor de quienes contaban con mejores condiciones para la educación remota. Las mujeres presentan una mayor permanencia relativa frente a los varones, y se reduce notablemente la representación de estudiantes de colegios oficiales, lo que sugiere procesos de abandono o migración hacia el sector privado. De forma especialmente destacada, el nivel educativo de las madres emerge como un factor protector clave: los estudiantes con madres que tienen educación técnica, tecnológica o universitaria incrementan su participación relativa, mientras que los de madres con baja escolaridad experimentan caídas importantes. Estos hallazgos no deben interpretarse como determinismos, sino como puntos de partida para diseñar políticas públicas focalizadas que reconozcan el peso de las condiciones familiares y tecnológicas en la trayectoria escolar y actúen decididamente para ampliar las oportunidades

de permanencia.

En cuanto al aprendizaje, el análisis contrafactual nos permite mirar más allá de los promedios y visibilizar un fenómeno crítico: la aparente recuperación en los puntajes globales en Saber 11 no representa necesariamente una mejora del sistema, sino que puede estar influida por la exclusión de estudiantes con mayores vulnerabilidades. Al contrastar el desempeño observado con el de estudiantes contrafactuales (es decir, aquellos con características similares que no llegaron a grado 11), se evidencia una brecha aún mayor por nivel socioeconómico, lo cual sugiere que el sistema está dejando atrás precisamente a quienes más podrían beneficiarse de un apoyo adicional. Esta herramienta metodológica, lejos de ser solo una crítica, se convierte en una oportunidad: nos permite construir indicadores más precisos para orientar decisiones, identificar zonas ciegas en las métricas tradicionales y proponer intervenciones más inclusivas. No basta con celebrar mejoras agregadas; necesitamos mirar con mayor profundidad quiénes están aprendiendo, quiénes están siendo excluidos y por qué.

La pandemia de COVID-19 amplificó estos patrones de desigualdad. El análisis de las cohortes que cursaron su secundaria durante los años de emergencia sanitaria revela dos procesos complementarios: por un lado, una exclusión selectiva de estudiantes con trayectorias educativas más frágiles; por otro, una pérdida real de aprendizajes incluso entre quienes lograron mantenerse en el sistema. Este doble impacto fue especialmente marcado en Matemáticas y se manifestó con más fuerza entre los estudiantes de menores ingresos, las mujeres, las zonas rurales y los colegios del sector privado con menor capacidad institucional. Lejos de ser un fenómeno inevitable, estas desigualdades nos muestran con claridad cuáles son los nudos críticos del sistema educativo frente a situaciones de crisis y nos dan pistas sobre qué factores deben fortalecerse para prevenir que se repitan.

Finalmente, el análisis territorial confirma que las brechas no se distribuyen de manera uniforme en Antioquia. La pandemia no fue un shock homogéneo: en algunos territorios —como los clasificados como Territorios de Vida o Rurales en Desarrollo— se profundizó la brecha entre estudiantes de distinto nivel socioeconómico, reflejando un sistema que refuerza

las ventajas acumuladas. En otros, como los Territorios Urbanos, la pérdida de aprendizajes fue más transversal, afectando incluso a quienes lograron permanecer. Y en los Territorios con Grandes Retos, la baja diferenciación en los resultados educativos apunta a limitaciones estructurales tan profundas que requieren más que ajustes marginales: se necesitan estrategias integrales que conecten políticas educativas con salud, bienestar, conectividad y protección social.

Estos hallazgos, en lugar de ser leídos con pesimismo, deben asumirse como un insumo valioso para la transformación educativa. Contamos ahora con evidencia robusta para definir prioridades: dónde intervenir, a quién priorizar, qué capacidades fortalecer y cómo monitorear avances con mayor precisión. En las recomendaciones que siguen, proponemos una ruta basada en este diagnóstico: intervenciones territoriales diferenciadas, estrategias de recuperación de aprendizajes por grupo poblacional, y un uso más sistemático de herramientas como el análisis contrafactual para informar la política pública. En suma, este estudio no solo documenta los desafíos; también ilumina caminos posibles para construir un sistema educativo más justo, resiliente y capaz de sostener el aprendizaje de todos los estudiantes, incluso en los contextos más difíciles.

Referencias

- Abadía, L. K., Gómez, S., and Cifuentes, J. (2021). Gone with the pandemic: effects of covid-19 on academic performance in colombia. Technical Report 019339, Universidad de los Andes, Facultad de Economía.
- Akmal, M. and Pritchett, L. (2019). Learning equity requires more than equality: Learning goals and achievement gaps. Working Paper 19/028, Research on Improving Systems of Education (RISE), Oxford.
- Banco de la República (2021). Impacto de la pandemia sobre el desempeño educativo en colombia.
- Banco Interamericano de Desarrollo (BID) (2021). El poder transformador de las evaluaciones de aprendizaje censales: 3 lecciones desde brasil.
- Banco Mundial (2021). The covid-19 pandemic: Shocks to education and policy responses.
- Deaton, A. (1985). Panel data from time series of cross sections. *Journal of Econometrics*, 30(1–2):109–126.
- Instituto Colombiano para la Evaluación de la Educación – ICFES (2020). Marcos de referencia de las pruebas saber 3, 5 y 9 por área del conocimiento.
- Instituto Colombiano para la Evaluación de la Educación (ICFES) (2019). Boletín saber al detalle: ¿cómo se construye el Índice de nivel socioeconómico (inse) en el contexto de las pruebas saber?
- Kaffenberger, M. (2019). A typology of learning profiles. Insight Note 8, Research on Improving Systems of Education (RISE), Oxford.
- Pritchett, L. and Sandefur, J. (2017). Girls’ schooling and women’s literacy: Schooling targets alone won’t reach learning goals. CGD Working Paper 440, Center for Global Development, Washington, D.C.

UNESCO (1981). *Metodología de cálculo de indicadores de eficiencia interna en educación*.

UNESCO, París. Documento técnico de la Oficina Internacional de Educación.

UNESCO Institute for Statistics (2018). Reconstructed cohort method for education indicators. Technical report, UNESCO Institute for Statistics, Montreal.