

Efecto de políticas informativas en los  
resultados del examen de egreso  
universitario: Caso colombiano

Universidad del Rosario

Facultad de Economía

Tesis de Maestría en Economía de la Políticas Públicas

Presentado por:

Daniela Gaitán Cotrino

Dirigido por:

Darwin Cortés Cortés

Bogotá, Colombia

2019

# Efecto de políticas informativas en los resultados de examen de egreso universitario: Caso colombiano

Daniela Gaitán Cotrino  
*Universidad del Rosario*  
Asesor: Darwin Cortés Cortés

20 de mayo de 2019

## Resumen

En este trabajo se estima el efecto que pueden tener políticas informativas referentes a los beneficios de un buen desempeño de la prueba Saber Pro sobre los resultados de la misma. Se utilizó como muestra a 537 estudiantes de la Facultad de Economía y la Escuela de Administración de la Universidad del Rosario, que presentaron el examen en los años 2017 y 2018. Los resultados arrojan que tanto ser elegido para recibir el tratamiento como leerlo no poseen ningún impacto a nivel global. Sin embargo, al analizar por componentes se halló que el efecto de ser elegido para recibir el tratamiento 1 genera un impacto negativo en la prueba de -0.152 desviaciones estándar y el efecto de leer el mensaje dado que fue elegido para recibir este tratamiento genera un efecto negativo de -0.181 desviaciones estándar.

**Palabras clave:** Economía del comportamiento, educación, rendimiento académico.

# 1. Introducción

El presente documento busca indagar cómo políticas informativas acerca de los posibles beneficios que se pueden derivar del buen desempeño en la prueba Saber Pro (Examen de egreso universitario colombiano); generan cambios en los resultados del mismo.

En Colombia, la evaluación de la educación superior inició en el año 2003 con la aplicación de Exámenes de Calidad de la Educación Superior (ECAES), que buscaba entregar resultados comparables entre carreras e instituciones, con la intención declarada de mejorar la calidad, la transparencia y la rendición de cuentas en el sector de la educación superior. En la prueba piloto del 2003, se aplicó el examen a 27 programas aumentando semestre a semestre el número de participantes hasta llegar a 55 programas en 2008. Teniendo en cuenta la curva de aprendizaje de este periodo y múltiples requerimientos legales recibidos por el Instituto Colombiano para el Fomento de la Educación Superior (ICFES), en 2009, se dio la reestructuración del examen, cambiando su nombre a Saber Pro e implementando su obligatoriedad para recibir el título universitario de pregrado.

Cada vez más, las universidades y los estudiantes usan los resultados del examen como una señal de habilidad y calidad. Por ejemplo, algunas universidades usan dichos resultados como requisito de admisión a programas de postgrado o para otorgar becas, y el ICETEX ofrece una línea de crédito para estudios de posgrado exclusiva para los mejores estudiantes a nivel nacional. Además, todos los años el presidente y el ministro de Educación reconocen públicamente a las personas con los diez puntajes más altos en cada campo. Como evidencia anecdótica, se encontró que los mejores estudiantes reciben ofertas de trabajo basadas en el conocimiento público de los puntajes de sus pruebas. Adicionalmente, MacLeod et al. (2015) proporcionan evidencia de que los exámenes de salida afectan los ingresos del mercado laboral de los graduados.

Sin embargo, al ser esta prueba exigida por las universidades solo como un requisito de grado, sin solicitar un desempeño mínimo, hace que el examen se profile como "Low Stakes tests"<sup>1</sup>, lo cual, implica que el rendimiento en esta termina afectando principalmente el posicionamiento de las instituciones. Además, Al Ahmad (2014) señala que si bien dichos exámenes de salida pueden mejorar la calidad y la eficiencia de los programas académicos, también pueden llegar a generar preocupación y ansiedad entre los estudiantes.

Basados en lo anterior, este estudio realizó un experimento aleatorio donde se encontró que tanto ser elegido para recibir el tratamiento como leerlo no poseen ningún impacto a nivel global. Sin embargo, al analizar por componentes se halló que el efecto de ser elegido para recibir el tratamiento 1 genera un impacto negativo en la prueba

---

<sup>1</sup>Los exámenes del tipo "low stakes" son aquellos que poseen bajo impacto en la vida académica del estudiante

de -0.152 desviaciones estándar y el efecto de leer el mensaje dado que fue elegido para recibir este tratamiento genera un efecto negativo de -0.181 desviaciones estándar.

En este orden de ideas, este trabajo es relevante pues existe limitada literatura colombiana relacionada con el impacto de incentivos sobre el desempeño en exámenes de egreso y espacialmente cuando son de tipo "low stakes", como el examen Saber Pro. Además, servirá de herramienta para la toma de decisiones institucionales y de política pública sobre incentivos en la educación superior.

La estructura de este documento es la siguiente: la sección 2 muestra la revisión de literatura relevante para este caso, la sección 3 describe los datos utilizados y sus fuentes, la sección 4 presenta la metodología utilizada y su aplicación empírica, la sección 5 muestra los resultados obtenidos en el modelo base y la implementación de variables instrumentales y finalmente, la sección 6, discusión y conclusiones.

## 2. Revisión de literatura

En esta sección se muestra la relevancia que, según la literatura, poseen los exámenes estandarizados, su impacto en la calidad educativa y en el proceso de aprendizaje de los estudiantes. Además, se analiza como se ha estudiado el uso de políticas informativas en la toma de decisiones de los estudiantes.

Uno de los principales retos que enfrentan las instituciones educativas es la identificación de adquisición de conocimientos y habilidades de sus estudiantes, para adaptar sus planes de estudio y así ajustarse a estándares de calidad internacional. Esto hace de la evaluación un aspecto importante en la educación, ya que, por un lado permite medir la calidad de esta y por otro lado, puede ser una herramienta de aprendizaje (Heitink. M, 2016).

Según Schlesselman y Coleman (2011), en la educación superior la medición es de vital importancia, ya que factores tradicionales, como el promedio de desempeño académico en los primeros semestres puede predecir el resultado final de los estudiantes en el programa. Adicionalmente, se debe procurar que se realicen a través de métodos estandarizados que permitan a los docentes y a las instituciones identificar las debilidades y las fortalezas de sus estudiantes (Crescentini. A, 2014). Dentro del concepto de estandarización debe comprenderse la similitud, la uniformidad y la continuidad de comportamientos y acciones (David. P.A, 1996), dicho de otra forma se debe disminuir la variación en los resultados (Santos. A, 2002), incrementar la productividad e incentivar la ventaja comparativa (P, 2013). Poseer buenos métodos de evaluación es un requisito indispensable para aquellas instituciones que buscan alcanzar altos estándares de calidad e influir de manera positiva en el aprendizaje de sus estudiantes (Al-Kadri, 2015; Askham, 1997).

La literatura muestra que las pruebas estandarizadas son predictores útiles del

desempeño subsecuente de los estudiantes, pronostican de forma más exacta, no muestran sesgos y no son influenciadas por los docentes. Según el trabajo de Kuncel y Hezlett (2007), que analiza múltiples estudios que toman exámenes estandarizados de egreso universitario en Estados Unidos, como el Graduate Record Examination (GRE), el Graduate Management Admissions Test (GMAT), entre otros; estas pruebas: (1) son predictores efectivos del desempeño en estudios de posgrado; (2) pronostican resultados académicos importantes más allá del desempeño académico obtenido en el pregrado; (3) predicen más medidas sobre el éxito futuro del estudiante al momento del egreso que el desempeño académico obtenido previamente y (4) si se combinan con el desempeño académico arroja predicciones de éxito más precisas.

Teniendo esto en cuenta y sumado a la relevancia que estas pruebas tienen a la hora de generar predicciones en el futuro académico de los estudiantes y la capacidad de maniobra que le otorgan a las instituciones a la hora de gestar planes de mejora, se abre la puerta para la innovación educativa y el fortalecimiento del rol del docente dentro del aula (Munoz-Comonfort. A, 2014; Herman. J.L, 1993). Además, se permite la evaluación y eventual reformulación de políticas públicas enfocadas a la calidad de la educación (Casas. F, 2001; Herman. J.L, 1993; Moses. M, 2007).

Si bien existe gran cantidad de autores que estudian sus ventajas, otros como Kuncel y Hezlett(2007) afirman que, aunque las pruebas estandarizadas ofrecen información útil para predecir el desempeño subsecuente de los estudiantes, la motivación y el interés a menudo pueden inferirse a través de pruebas no estandarizadas como cartas de recomendación, juicios personales y entrevistas. De acuerdo a los resultados de Kellermanns, Barnett, Pearson y Pearson (2016) en su trabajo sobre el papel mediador de la orientación hacia los objetivos en la relación entre las autoevaluaciones básicas y el rendimiento académico, los estudiantes que se encuentran más motivados poseen conductas que les permiten probar que son mejores que otros alcanzando sus metas y, como resultado, mejoran su desempeño académico. Otros señalan que el uso de estas pruebas tiene efectos negativos en la enseñanza y en el aprendizaje pues conducen a una concepción superficial del contenido programático e inducen al profesor a centrarse en habilidades específicas (R.L., 2000).

En este orden de ideas, Casas y Meaghan (2001) identifican las siguientes desventajas de los exámenes estandarizados: (1) desvían tiempo valioso del proceso de aprendizaje por tiempo invertido en preparar la prueba; (2) se diseñan con el fin de evaluar respuestas de temas específicos dejando fuera otros relevantes; (3) con el fin de obtener altas calificaciones, se descuidan otras habilidades cognitivas y de resolución de problemas; (4) el material evaluado puede ser de relevancia limitada; (5) no son adecuados para todos los estudiantes; (6) el formato y los procesos de evaluación pueden contraponerse a la forma en que los estudiantes interactúan de forma colaborativa en el salón de clases; (7) la confiabilidad es limitada, carece de validez y de justicia respecto a factores como la raza, el género y el ingreso familiar; (8) la creatividad y la flexibilidad del profesor puede ser reducida para enfocarse en determinados aspectos; y (9) los resultados pueden

ser exagerados o dar una falsa impresión del logro de los estudiantes.

Con base en las ventajas y las desventajas de las pruebas estandarizadas, Linn (2000) propone combinarlas con pruebas no estandarizadas para incrementar la validez de inferencias sobre los logros. Tal como señalan, Herman y Golan (1993) si los resultados del examen representan aprendizaje significativo<sup>2</sup>, entonces las pruebas pueden contribuir a la construcción de aprendizaje. De otra forma, dicha práctica puede ser considerada como una herramienta equivocada. Por lo cual, deben diseñarse de tal forma que contribuyan al proceso formativo y a la evaluación del estudiante.

Considerando que para las instituciones educativas el examen de egreso de la educación superior representa un examen de alto valor<sup>3</sup> pero que para los individuos puede o no serlo dado que este no es requerido para obtener empleos ni para acceder a niveles de educación superiores, solo para acceder a becas o como predictor del desempeño futuro, es necesario proporcionar a los estudiantes incentivos. En este caso, información sobre los beneficios que obtener un buen resultado otorga, para que apliquen un grado mayor de esfuerzo.

Al analizar el uso de políticas informativas, una parte de esta literatura explora el ego, valoración excesiva de uno mismo, y como cada individuo reacciona a múltiples escenarios basados en sus antecedentes (Grether.D, 1980; El-Gamal, 1995). Y otra, estudia las reglas del procesamiento de la información en entornos más realistas donde las creencias son de importancia directa, como la capacidad, el rendimiento, la evaluación de riesgos, entre otros (Viscusi and O'Connor, 1984; Delavande.A, 2008; Eil.D, 2011; Grossman.Z, 2012) para entender como los individuos, en este caso estudiantes, toman decisiones basados en información adicional.

Es importante tener en cuenta que si bien este estudio analiza si la información afecta las decisiones de los estudiantes y por ende su rendimiento en el examen, no puede arrojar información sobre los mecanismos subyacentes que conducen a las revisiones y el proceso de formación de expectativas, en gran parte porque se necesitan datos más detallados para hacerlo.

---

<sup>2</sup>Debe entenderse el aprendizaje significativo, como un tipo de aprendizaje en que un estudiante relaciona la información nueva con la que ya posee; reajustando y reconstruyendo ambas informaciones en este proceso.

<sup>3</sup>Una prueba de alto valor es cualquier prueba que se usa para tomar decisiones importantes sobre estudiantes, educadores, escuelas o distritos, generalmente con el propósito de rendir cuentas. En general, "alto valor" significa que los puntajes de las pruebas se usan para determinar los castigos (como sanciones, penalizaciones, reducciones de fondos, publicidad negativa), reconocimientos (premios, celebración pública, publicidad positiva), avance (promoción de grado o graduación para estudiantes), o compensación (aumentos salariales o bonificaciones para administradores y profesores).

### 3. Tratamiento

En esta sección se describe el contexto del experimento, cada una de las piezas de información que se entregó a los estudiantes y la selección de los grupos de tratamiento.

La Universidad del Rosario es una de las instituciones de educación superior más antiguas e influyentes en la historia de Colombia. Es reconocida por los altos estándares de calidad que posee y por el acompañamiento que otorga a sus estudiantes, lo cual le ha permitido destacarse desde 2014 como una de las mejores en el examen de egreso universitario Saber Pro, con resultados por encima de la media nacional en todos los grupos de referencia en los que participa tal como se señala en el documento de acreditación institucional.

Si bien la evolución del desempeño en exámenes de egreso universitario ha crecido en los últimos años, para la institución es de gran importancia generar reflexiones y acciones de mejora como resultado del análisis del rendimiento en la prueba. Por ello, desde 2016, con la creación de la Dirección de Estudiantes y de la Dirección de Planeación y Efectividad Institucional<sup>4</sup> se establecieron responsabilidades centrales de generación de análisis y evaluación de los resultados de la prueba, proveyendo a las unidades académicas de mayor cantidad de herramientas para la toma de decisiones e implementación de estrategias encaminadas al mejoramiento en este ámbito.

Basados en dichos informes, facultades como la de Economía decidieron abrir cursos de libre entrada para la preparación del Saber pro, donde profesores expertos en cada área (Pensamiento económico, microeconomía, macroeconomía, econometría, constitución, lecto -escritura y matemáticas) realizan repaos cortos y concisos para enfrentar dicha prueba de la mejor forma. Sin embargo, estos espacios no son aprovechados por los estudiantes presentando asistencias cada vez más deficientes, tal como se observa en el siguiente cuadro.

Cuadro 1: Asistencia Sesiones de repaso 2017 - 2018

Área	Sesión 1	Sesión 2	Promedio
Lectura Crítica	1.8 %	10 %	5.9 %
Comunicación escrita	8.18 %	9.1 %	8.6 %
Competencias ciudadanas	2.7 %	4.5 %	7.3 %

Fuente: Dirección de pregrados, Facultad de Economía

Teniendo en cuenta el objetivo y los esfuerzos institucionales por mejorar el desempeño en la prueba y la baja respuesta recibida por los estudiantes, el presente trabajo

<sup>4</sup>Áreas de la universidad encargadas del acompañamiento y la permanencias estudiantil, y del aseguramiento de la calidad.

pretende indagar que tan sensibles son estos ante información adicional referente a los posible beneficios de un buen desempeño en el examen.

Para evaluar dicha sensibilidad se envió dos piezas informativas, de forma aleatoria; a los estudiantes de la Facultad de Economía y la Escuela de Administración, a través de las secretarías académicas<sup>5</sup> de cada unidad, el día previo al examen. La primera pieza consistía en alertar a los estudiantes sobre la posibilidad de obtener beneficios académicos derivados de su buen desempeño, como becas de hasta el 80 % en estudios de posgrado y educación continua dentro de la universidad<sup>6</sup>; así como, acceso a becas en el exterior por parte del ICETEX. La segunda pieza indicaba efectos salariales y económicos, mostrando que como resultado del buen rendimiento en la prueba, la calificación y reputación de la Universidad aumentarían y de esta forma, incrementarían sus oportunidades y salarios de enganche, tal como lo señala (MacLeod. W, 2015) y adicionalmente, aquellos que poseyeran algún crédito con el ICETEX podrían solicitar su condonación.

Dado que la Facultad de Economía presenta una cantidad menor de estudiantes, su población se dividió en 2 grupos, uno de control y otro de tratamiento que recibió ambas piezas, mientras que la Escuela de Administración se dividió en 3, un grupo de control, un grupo de tratamiento 1 que recibió ambas piezas y otro, de tratamiento 2, que recibió la pieza que contenía información académica; buscando identificar el impacto de cada una de las piezas

Es importante aclarar que el anterior ejercicio se pudo aplicar para ambos periodos de estudio en la Facultad de Economía, mientras que en la Escuela de Administración solo se aplicó en 2017, pues en el segundo periodo dicha facultad se encontraba en periodo de transición administrativa.

Finalmente, se realizaron encuestas telefónicas<sup>7</sup> en las que se indagó sobre la lectura de la pieza y sobre la intención de llevar a cabo estudios de posgrado en el corto, mediano y largo plazo.

## 4. Datos

Esta sección inicia con un análisis de las bases de datos utilizadas en el trabajo, posteriormente se realiza una caracterización descriptiva de los estudiantes que presentaron su examen de egreso universitario Saber Pro en los periodos 2017 y 2018, la cual, incluye características académicas y socio económicas; luego, se analizan los resultados obtenidos en dicho examen y por último, se explica cómo se construyó un instrumento

---

<sup>5</sup>La secretaria académica es el área encargada de los servicios estudiantiles, generando comunicaciones oficiales, certificaciones, validaciones, procesos disciplinarios, entre otros.

<sup>6</sup>Decreto Rectoral 1508 de 2017, Normativa de becas, Universidad del Rosario

<sup>7</sup>Ver anexos

de encuesta en donde, entre varios aspectos, se preguntó qué tan pronto quieren los estudiantes acceder a estudios de posgrado.

### **Saber Pro**

El Saber Pro es un examen de egreso universitario colombiano que busca medir la calidad de la educación superior en el país. Este se realiza en dos sesiones, la primera es obligatoria para todos los inscritos, ya que está compuesta por 5 módulos que se consideran genéricos para cualquier programa de formación de nivel profesional que permite realizar comparación de todos los estudiantes universitarios del país, mientras que, a la segunda sesión solo asisten quienes hayan sido inscritos por su institución de educación superior (IES) para presentar entre 1 y 3 módulos específicos relativos a su área de formación.

Tal como se señaló anteriormente, las pruebas genéricas permiten hacer comparaciones entre estudiantes de diferentes programas, evaluando habilidades de comunicación escrita, razonamiento cuantitativo, lectura crítica, competencias ciudadanas e inglés, que se miden a través de puntajes que reportan una escala de 0 a 300 y niveles de desempeño que van de 1 a 4, siendo 4 el de mejor rendimiento. Adicionalmente, el puntaje global del examen se obtiene a partir del promedio simple de los puntajes obtenidos en los módulos mencionados.

### **Saber 11**

El examen de egreso de la educación media en Colombia, Saber 11, lo deben presentar los estudiantes que estén finalizando el último grado, con el fin de obtener resultados oficiales que les permitan ingresar a la educación superior. También pueden presentarlo quienes hayan obtenido el título de bachiller o superado el examen de validación del bachillerato. Esta prueba habilidades en matemáticas, lectura crítica, sociales y competencias ciudadanas, ciencias naturales e inglés, que se miden a través de puntajes que reportan una escala de 0 a 100. Por su parte, el puntaje global del examen se obtiene a partir del promedio ponderado de los resultados de las cinco pruebas, siendo inglés el que menor peso posee y su escala es de 0 a 500.

Además, en esta prueba se aplica un cuestionario socio económico que contiene 24 preguntas, que buscan caracterizar la población joven escolarizada de Colombia. A partir de esta, se analizarán algunos aspectos socio económicos como el Índice Dotación del Hogar, que se calculo por el método de componentes principales, en el que participan 7 componentes: internet, servicio de televisión, computador, lavadora, microondas, automóvil y motocicleta, y como resultado al obtener un índice alto, el nivel de vida es mayor. El estrato de la vivienda<sup>8</sup> y la educación de los padres, que es categorizada en

---

<sup>8</sup>La estratificación socio económica es la clasificación de los inmuebles residenciales de un municipio, que se hace en atención al régimen de servicios públicos domiciliarios en Colombia (Ley 142 de 1994). Se realiza principalmente para cobrar de manera diferencial, permitiendo asignar subsidios y cobrar contribuciones en esta área. De esta manera, quienes tienen más capacidad económica pagan más

9 niveles que parten de la educación primaria incompleta hasta estudios de posgrado.

### **Historial académico**

La tercera y última base de datos que dispone este trabajo es el historial académico de los estudiantes en el momento que presentaron su examen. Esta contiene su promedio acumulado que reporta una escala de 0 a 5, el porcentaje de créditos aprobados<sup>9</sup>; y los resultados obtenidos en cada una de las asignaturas cursadas. De igual forma, presenta información referente a doble programa, paso por el programa de fortalecimiento e información financiera.

A partir de estos datos se pudo diseñar un modelo que incluyó a 537 estudiantes de la Facultad de Economía y la Escuela de Administración en los periodos 2017 y 2018. En el Cuadro 2, se encuentran las estadísticas descriptivas para los componentes académicos y socio económicos de los mismos. Los estudiantes en promedio poseen 22 años, su promedio acumulado es de 3.97, la media del porcentaje de créditos aprobados es del 88 %, el 12 % está cursando un segundo programa. Ese mismo porcentaje ha pasado por Fortalecimiento. Al indagar sobre su rendimiento en pasadas pruebas de Estado como el examen Saber 11, en promedio los estudiantes obtuvieron 310 puntos de un máximo de 500, la prueba de ingreso a la educación superior con mayor puntaje fue de 483 y la de menor, de 210.

Por otro lado, al analizar las características socio económicas encontramos que el 16 % de los estudiantes cuenta con un crédito con el ICETEX, en promedio residen en hogares de estrato 4 con una desviación de 0.04, según el Índice de Dotación del Hogar en promedio cuentan con una dotación de 0.84 en una escala de 0 a 1 y en la media, los padres poseen estudios profesionales.

---

y contribuyen para que los estratos bajos puedan pagar sus facturas. Actualmente, se cuenta con 6 categorías.

<sup>9</sup>El crédito académico se refiere a la valoración del tiempo que un estudiante dedica a la asignatura, incluidas las horas presenciales de clase y el tiempo de trabajo independiente. Este es equivalente a 48 horas. Cada programa académico posee un número diferente de créditos totales entre obligatorias, electivas y opción de grado para obtener el diploma de pregrado, en el caso de los programas de la Facultad de Economía los planes de estudio exigen 140 créditos mientras que en Escuela de Administración son 170 créditos

Cuadro 2: Estadísticas descriptivas

	Muestra Completa			Control	Tratamiento 1			Tratamiento 2		
	Media Err. Estd	Min	Max	Media Err. Estd	Media Err. Estd	<i>p</i> -value Medias	<i>p</i> -value Distribución	Media Err. Estd	Medias	<i>p</i> -value Distribución
<b>Saber Pro</b>										
Puntaje Global	175.66 (0.92)	67	256	177.58 (1.46)	175.53 (1.38)	0.91	0.97	171.96 (2.28)	0.07*	0.24
Razonamiento Cuanti	178.01 (1.16)	81	300	179.82 (1.80)	179.77 (1.82)	0.21	0.21	170.65 (2.58)	0.00***	0.07*
Comp. ciudadanas	164.82 (1.32)	59	233	169.35 (1.97)	163.59 (2.08)	0.45	0.33	158.02 (3.20)	0.02**	0.03**
Comunicación escrita	168.99 (1.67)	72	300	166.96 (2.72)	168.07 (2.49)	0.64	0.99	175.24 (3.92)	0.08*	0.12
Lectura crítica	172.90 (1.28)	71	230	176.93 (1.90)	170.97 (2.01)	0.22	0.17	168.55 (3.17)	0.12	0.29
Inglés	196.89 (0.83)	109	300	196.70 (1.24)	198.51 (1.31)	0.11	0.28	193.92 (2.06)	0.10*	0.01***
Sexo	0.54 (0.02)	0	1	0.56 (0.03)	0.53 (0.03)	0.72	1.00	0.53 (0.05)	0.85	1.00
Edad	22.35 (0.02)	19	31	22.22 (0.03)	22.43 (0.03)	0.39	0.96	22.48 (0.05)	0.36	0.29
<b>Académico</b>										
Promedio Acumulado	3.97 (0.01)	3.3	7.75	3.97 (0.01)	3.97 (0.01)	0.83	0.88	3.97 (0.02)	0.95	0.88
Créditos aprobados	0.88 (0.01)	0	1.25	0.86 (0.01)	0.87 (0.01)	0.16**	0.15**	0.96 (0.01)	0.00***	0.00***
Doble Programa	0.12 (0.01)	0	1	0.15 (0.02)	0.13 (0.02)	0.96	1.00	0.50 (0.02)	0.19	0.99
Fortalecimiento	0.12 (0.03)	0	1	0.14 (0.02)	0.13 (0.02)	0.62	1.00	0.06 (0.02)	0.05**	0.66
P. Global Saber 11	310.00 (0.05)	210	483	315.70 (0.07)	310.14 (0.07)	0.95	0.96	297.99 (0.07)	0.01***	0.01***
<b>Expectativa posgrado</b>										
Corto plazo (Menor 1 año)	0.09 (0.01)	0	1	0.10 (0.01)	0.09 (0.02)	0.97	1.00	0.09 (0.03)	0.77	1.00
Mediano plazo (Entre 1 y 5 años)	0.44 (0.05)	0	1	0.43 (0.03)	0.43 (0.03)	0.79	1.00	0.43 (0.04)	0.66	1.00
Largo plazo (Mayor a 5 años)	0.37 (0.05)	0	1	0.35 (0.07)	0.39 (0.07)	0.41	0.99	0.38 (0.07)	0.84	1.00
No desea estudiar	0.05 (0.05)	0	1	0.06 (0.07)	0.04 (0.07)	0.44	1.00	0.06 (0.07)	0.73	1.00
<b>Socio económicos</b>										
Crédito ICETEX	0.16 (0.02)	0	1	0.18 (0.02)	0.14 (0.02)	0.33	1.00	0.15 (0.03)	0.76	1.00
Estrato	4.11 (0.04)	0	6	4.10 (0.05)	4.20 (0.07)	0.08	0.78	4.13 (0.09)	0.80	1.00
Índice Dotación hogar	0.84 (0.01)	0.43	1	0.83 (0.01)	0.84 (0.01)	0.13	0.97	0.84 (0.01)	0.81	1.00
Educación padres	0.71 (0.02)	0	1	0.68 (0.03)	0.74 (0.02)	0.06*	0.30	0.67 (0.04)	0.30	0.76
Educación padre	0.70 (0.02)	0	1	0.68 (0.03)	0.74 (0.03)	0.08**	0.54	0.67 (0.05)	0.37	0.99
Educación madre	0.71 (0.02)	0	1	0.69 (0.03)	0.74 (0.03)	0.13	0.74	0.68 (0.05)	0.42	1.00
N	537	537	537	216	216	216	2016	105	105	105

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1 significativamente diferente

El tamaño de la muestra en comunicación escrita es de 527

*P*-value de medias corresponde al t-test, *p*-value de distribución corresponde al test Kolmogorov-Smirnov, respecto cada grupo de tratamiento con el grupo de control

Los resultados Saber Pro poseen una escala de 0 a 300, mientras que los resultados Saber 11 poseen una escala de 0 a 500

El porcentaje de créditos aprobados puede ser mayor a 1, pues los estudiantes sobrepasan los créditos requeridos por el programa.

Las expectativas de posgrado, son resultado de la encuesta telefónica realizada posterior al examen Saber Pro. Esta se divide en cuatro categorías.

La educación familiar esta categorizada en 9 niveles, los cuales fueron agrupados de 1 a 7 y de 8 a 9, es decir, un nivel educativo mayor a pregrado completo.

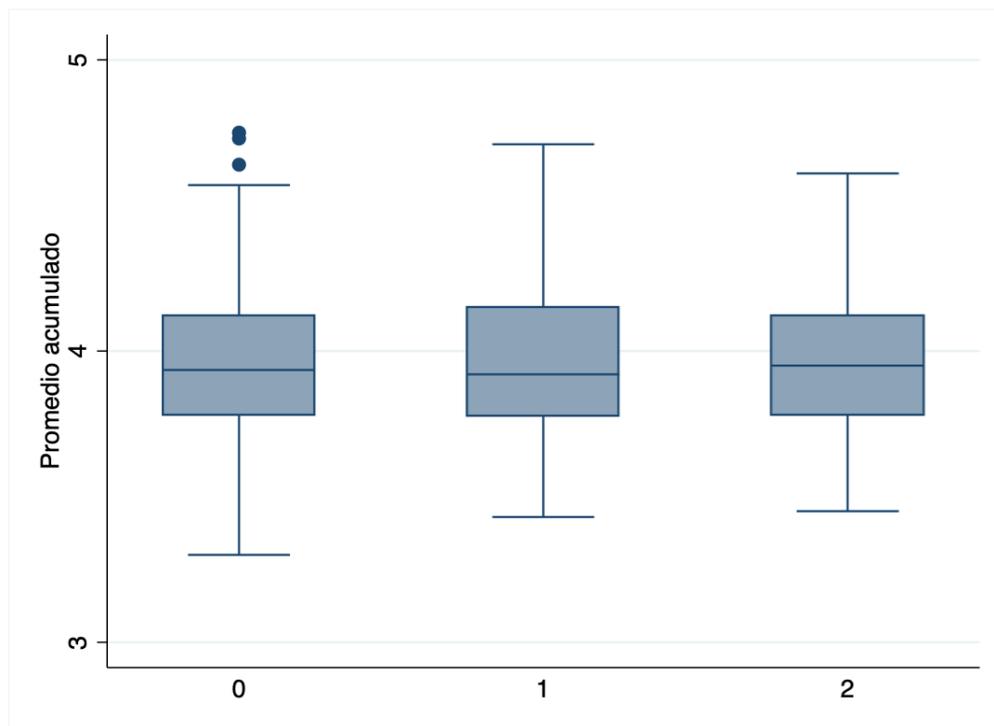


Figura 1: Promedio acumulado

Dentro del modelo, se recategorizaron el estrato y la educación de los padres. En el primer caso se unieron los primeros 3 estratos en una categoría y los estratos 5 y 6 en otra. En el segundo caso, se unieron las primeras 7 categorías, dejando las categorías de profesional completo y posgrado independientes.

Cuando se analizan los resultados obtenidos en la prueba de acuerdo al puntaje global, se observa que en promedio los estudiantes obtuvieron un puntaje de 175.66 de un máximo de 300, con desviación estándar de 21.44. Las pruebas en las que obtuvieron los mejores promedios fueron inglés y razonamiento cuantitativo. Mientras que en comunicación escrita y competencias ciudadanas, tuvieron los peores resultados.

Finalmente, para instrumentar los tratamientos aplicados se construyó una encuesta que permitiera indagar sobre la recepción y lectura de las piezas enviadas. Adicionalmente, se preguntó a los estudiantes acerca de sus planes académicos futuros. En dicha sección se acotó las respuestas a corto plazo, en un tiempo menor a un año; mediano plazo, entre uno y cinco años; y largo plazo, en más de cinco años. Se obtuvieron los siguientes resultados: el 9% quiere empezar dicho proceso en el corto plazo, el 44% lo tiene contemplado en el mediano plazo, el 37% en el largo plazo, 5% no lo contempla y el otro 5% no la contestó.

## 5. Metodología

Los experimentos aleatorios han sido ampliamente utilizados en evaluaciones de impacto de diversos temas (Glewwe y Kremer 2005) para intentar responder preguntas sobre política o probar teorías de forma precisa. En estos la muestra de estudio se divide en un grupo que recibirá la intervención (el grupo de tratamiento) y otro grupo que no (el grupo de control). Esto se hace con el fin de encontrar el efecto causal de la intervención comparando ambos estados.

Sin embargo, estimar un efecto causal es esencialmente un problema de datos faltantes, ya que no observamos las unidades tratadas en el caso hipotético de no haber sido tratadas, así como no observamos las unidades tratadas en el caso hipotético de haber sido tratadas. En otras palabras, no observamos los contrafactuales de las unidades tratadas y no tratadas, por lo cual se deben establecer condiciones bajo las cuales lo que observamos para el caso del grupo de control se puede utilizar para abordar el contrafactual faltante para el caso del grupo de tratamiento. Por tal motivo, se debe realizar una aleatorización que, al comparar el test de diferencia de medias y el de distribución, permita inferir sobre qué tan parecidos son ambos grupos.

### 5.1. Aleatorización

La aleatorización es un método de control experimental que permite por una lado, obtener grupos comparables y por el otro, reducir el sesgo de selección.

En este caso, la aleatorización fue del tipo simple controlando por: el sexo, el programa al que el estudiante pertenece, el desempeño académico acumulado, doble programa <sup>10</sup>, el paso por el programa institucional de fortalecimiento <sup>11</sup>, el acceso al crédito a través del ICETEX, los resultados en la prueba Saber 11 y algunas características socio económicas como: el estrato, el índice de Dotación del Hogar, la educación de los padres, etc.

### 5.2. Modelo base

El modelo inicial busca identificar el efecto que puede tener ser elegido para recibir el mensaje sobre los resultados de la prueba Saber Pro.

$$Y_i = \beta_0 + \beta_1 Z_i + \beta C_i + \varepsilon_i$$

Este modelo se denomina Intención de tratamiento (ITT), pues incluye a cada sujeto que se asigno al azar de acuerdo con la aleatorización del tratamiento, ignorando el incumplimiento, las desviaciones del protocolo, el retiro y todo lo que ocurra después

---

<sup>10</sup>Estudiantes que cursan dos programas académicos dentro de la universidad.

<sup>11</sup>Programa diseñado para estudiantes que han cursado menos del 80 % de los créditos de su programa y que por algún motivo han perdido el cupo y desean reingresar.

de la aleatorización. En el análisis de ITT, la estimación del efecto del tratamiento es generalmente conservadora. En este caso las personas que leyeron las piezas son un subconjunto de quienes las recibieron, por cuanto el análisis se hace sobre el último grupo.

### 5.3. Modelo IV

Tal como se ilustra en la sección de tratamiento, la aleatorización de los grupos basada en los conceptos antes mencionados se alcanzó teniendo en cuenta las diferencias de medias y distribución y el poder de las muestras. Sin embargo, el modelo básico por OLS muestra el efecto de ser elegido para recibir el tratamiento sobre los resultados de la prueba, por lo cual es necesario instrumentar la encuesta de lectura con el modelo base para indagar el efecto de leer el mensaje sobre los resultados y de esta forma corregir el problema de endogeneidad, que se puede dar porque los estudiantes dadas sus características son quienes deciden leer o no el correo.

La endogeneidad, entendida como la correlación entre la variable independiente del modelo y el término del error, originada a través de variables omitidas o errores de medida puede hacer que los estimadores de mínimos cuadrados ordinarios sean sesgados e inconsistentes. Para solucionar este problema se usa el estimador de variables instrumentales, conocido como estimador de mínimos cuadrados en dos etapas, donde la primera etapa corresponde a la regresión entre el instrumento y la variable endógena y la segunda a la regresión entre la variable endógena estimada y la variable de resultado.

Dada la presencia de endogeneidad en el modelo, es necesario incluir un instrumento para la variable independiente. En este caso, la variable a instrumentar es la lectura. El instrumento aquí adoptado es el tratamiento inicial. Como consecuencia, en la primera etapa se estima la relación entre el tratamiento y la lectura de las piezas, y en la segunda etapa se estima el modelo con los coeficientes estimados en la primera. Es decir, el efecto que haber leído las piezas dado que recibió el tratamiento posee sobre los resultados en la prueba Saber Pro. A continuación, se presenta el modelo de estimación inicial y modelo con estimación de variables instrumentales.

$$Y_i = \beta_0 + \beta_1 Z_i + \beta_2 C_i + \varepsilon_i$$

Primera etapa:

$$X_i = \gamma_0 + \gamma_1 Z_i + \gamma_2 C_i + \mu_i$$

Segunda etapa:

$$Y_i = \beta_0 + \beta_1 \hat{X}_i + \beta_2 C_i + \varepsilon_i$$

Donde  $Y_i$  es el puntaje en la prueba Saber Pro.  $C_i$  son los controles a nivel individual tenidos en cuenta en el modelo, tanto académicos como socio económicos (programa, promedio acumulado, créditos aprobados, resultados Saber 11, estrato, dotación del hogar, entre otros).  $X_i$ , lectura de cada una de las piezas informativas obtenido a través de las encuestas. Finalmente,  $Z_i$  es el instrumento, tratamiento enviado a cada uno de los grupos aleatoriamente seleccionados a través del correo institucional de la Facultad e ilustrado con piezas de información.

Cabe resaltar que el estimador de variables instrumentales debe cumplir las siguientes condiciones para ser insesgado y consistente:

*Condición de relevancia*

$$Cov(Z_i, X_i) \neq 0$$

Para corroborar la relevancia del instrumento se analiza la relación de este con la variable independiente haciendo la regresión de primera etapa. Si su covarianza es muy pequeña, el instrumento es débil y conllevará a sobrestimar el estimador por variables instrumentales. Para este fin, se utiliza el test de Wald. Este señala que, cuando la F-wald es mayor a 10, el instrumento es relevante. En este caso, la covarianza es alta, lo que indica que la lectura de la pieza es un buen instrumento del tratamiento.

*Restricción de exclusión o exogeneidad del instrumento*

$$Cov(Z_i, \mu) = 0$$

La restricción de exclusión o exogeneidad del instrumento señala que no debe existir correlación entre el instrumento y el error para poder inferir causalidad, lo cual implica que la lectura de las piezas solo puede estar correlacionada con el tratamiento y bajo ningún motivo puede correlacionarse con otras variables. En este caso, el tratamiento práctico (instrumento) solo se correlaciona con el tratamiento teórico, por lo cual se podría afirmar que se está trabajando con un instrumento adecuado. En otras palabras, cambiar el instrumento sin cambiar la variable endógena, no debe cambiar la variable resultado.

Sin embargo, es importante tener en cuenta que este resultado se denomina, Efecto de Tratamiento Promedio Local (LATE). Es decir, la asignación aleatoria al tratamiento se relaciona imperfectamente con el estado real del tratamiento, pues la relación causal es obtenida de una submuestra que en este caso es quienes leyeron el mensaje.

Adicionalmente, se debe realizar otras pruebas para comprobar el poder de la muestra y su robustez.

El poder de una prueba es la probabilidad de rechazar correctamente la hipótesis nula cuando esta es falsa. Es decir,  $\theta = 1 - \beta = \Pr(\text{rechazar } H_0 \mid H_0 \text{ es falso})$ . Aumentar la potencia o poder de una prueba disminuye la probabilidad de un error de tipo II <sup>12</sup>, por lo que se espera que esta sea alta, normalmente se busca que se encuentre entre el 80 % y 90 %. Especialmente, cuando se manejan experimentos aleatorios es muy importante determinar el tamaño de la muestra a partir del poder que se pretende manejar, pues esta debe ser lo suficientemente grande para detectar una diferencia práctica cuando exista.

Debido a que se incorporó variables instrumentales para mitigar el problema de endogeneidad, es importante seguir los pasos señalados por Murray (2006) para alcanzar una buena práctica en el uso de este método: (i) chequear la significancia y el signo estimado del instrumento, (ii) hacer un testeo de instrumentos débiles, (iii) obtener las estimaciones usando Fuller<sup>13</sup> en vez de mínimos cuadrados en dos etapas e (iv) interpretar con cuidado los resultados de variables instrumentales.

Además de la prueba de poder realizada previamente y tal como lo señala Murray (2006), es importante realizar un chequeo de robustez mediante el "Limited Information Maximum Likelihood"(LIML). Este se basa en el supuesto de normalidad de los errores en la ecuación estructural y en la ecuación de la primera etapa, demostrando tener un sesgo menor que el estimador de mínimos cuadrados en dos etapas. Como complemento, se analiza el estimador Jackknife de variables instrumentales para establecer la robustez del modelo. Según Angrist (1999), este es independiente a las perturbaciones que pueden generar las muestras pequeñas, al intentar eliminar el sesgo de las muestras finitas.

## 6. Resultados

En esta sección se muestran los resultados obtenidos en la prueba de egreso de educación superior Saber Pro y el efecto que los tratamientos tuvieron en la misma. Inicialmente, se realizará el análisis de los resultados de la regresión de mínimos cuadrados ordinarios (ITT) como línea base y, luego, se estudiará los resultados por el método de variables instrumentales (LATE). Finalmente, se realizará un análisis de la relevancia y la robustez de las muestras.

### 6.1. Resultados OLS (ITT)

En el Cuadro 3, se analizan los resultados bajo la especificación OLS . En la primera columna de cada prueba, se analiza el efecto de tratamiento sin ninguna variable de

---

<sup>12</sup>Cuando la hipótesis nula es falsa y no se rechaza, existe un error de tipo II. La probabilidad de cometer un error de este tipo es  $\beta$ , que depende de la potencia de la prueba. Puede reducirse el riesgo de cometer un error de tipo II al asegurarse de que la prueba tiene suficiente poder.

<sup>13</sup>Variaciones de Información Limitada de Máxima Probabilidad (LIML) que llevó a cabo Fuller (1977).

control, mientras que en la segunda se tienen en cuenta los controles que se encuentran desbalanceados en el cuadro 2. El tratamiento 1 en el puntaje global no presenta resultados significativos en presencia de controles, entretanto, el tratamiento 1 muestra un impacto negativo de 0.152 desviaciones estándar y significativo al 90% cuando se introducen controles para la prueba de Lectura Crítica.

Cuadro 3: Resultados OLS (ITT)

VARIABLES	Puntaje Global		Lectura Crítica		Razonamiento Cuantitativo	
Tratamiento 1	-0.0959 (0.0937)	-0.0313 (0.0776)	-0.201** (0.0934)	-0.152* (0.0824)	-0.00190 (0.0957)	0.0530 (0.0829)
Tratamiento 2	-0.262** (0.126)	-0.0461 (0.108)	-0.282** (0.124)	-0.110 (0.110)	-0.342*** (0.117)	-0.137 (0.110)
T1-T2	0,1663 (0.124)	0,018 (0.104)	0.081 (0.126)	-0.043 (0.107)	0,1663 (0.124)	0,018 (0.104)
R2	0.009	0.322	0.013	0.261	0.018	0.235
Observaciones	537	537	537	537	537	537
Controles		✓		✓		✓

Errores estándar robustos en paréntesis. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Controles: Edad, créditos aprobados, puntaje saber 11, Índice de Dotación del Hogar, educación de los padres

Tratamiento 1: Pieza información académica + pieza información económica

Tratamiento 2: Pieza información académica

Adicionalmente, como el segundo tratamiento se encuentra contenido en el primero, es posible inferir el sentido del impacto de la piezas a partir de la diferencia en sus coeficientes. Sin embargo, ni dichas diferencias ni los tratamientos sobre la prueba de Razonamiento Cuantitativo son significativos.

## 6.2. Resultados IV (LATE)

Por otro lado, al analizar los resultados de variables instrumentales, se encuentra que el instrumento es significativo para ambos tratamientos, como para cada uno de los escenarios, es decir, con y sin controles. Y al analizar los resultados ninguna de las variables es significativa, ni las piezas arrojan significancia en el sentido de sus efectos.

Teniendo en cuenta que los anteriores resultados no son concluyentes, es necesario analizar cada uno de los componentes de la prueba. A continuación, se muestran las dos áreas más relevantes, dejando los otros tres componentes en la sección de anexos.

Cuadro 4: Resultados IV (LATE)

VARIABLES		Puntaje Global
Tratamiento 1	-0.139 (0.104)	-0.0722 (0.0842)
Tratamiento 2	-0.271** (0.133)	-0.0644 (0.114)
T1-T2	0.132 (0.128)	-0.0054 (0.108)
R2	0.009	0.321
Observaciones	537	537

VARIABLES		Lectura Crítica
Tratamiento 1	-0.235** (0.104)	-0.181** (0.0901)
Tratamiento 2	-0.306** (0.132)	-0.142 (0.116)
T1-T2	0.0714 (0.129)	-0.0397 (0.1096)
R2	0.013	0.260
Observaciones	537	537

VARIABLES		Razonamiento Cuantitativo
Tratamiento 1	-0.0248 (0.106)	0.0289 (0.0919)
Tratamiento 2	-0.365*** (0.124)	-0.173 (0.117)
T1- T2	0.132 (0.129)	-0.0054 (0.1096)
R2	0.018	0.235
Observaciones	537	537

Instrumento 1	0.909*** (0.009)	0.906*** (0.011)
Instrumento 2	0.979*** (0.009)	0.975*** (0.011)
F-Wald Instrumento 1	24.05	10770.85
F-Wald Instrumento 2	11502.18	107565.44

Controles ✓

Errores estándar robustos en paréntesis. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1  
 Controles: Edad, créditos aprobados, puntaje saber 11,  
 Índice de Dotación del Hogar, educación de los padres  
 Tratamiento 1: Pieza información académica + pieza información económica  
 Tratamiento 2: Pieza información académica

En el componente de lectura crítica, cuando el modelo se controla con las variables desbalanceadas, el coeficiente estimado por variables instrumentales es negativo y significativo a un nivel del 95 % para el tratamiento 1. Es decir, este tiene un impacto de -0.181 desviaciones estándar sobre los resultados en la prueba del componente, y tal como se observó en los resultados globales, el tratamiento 2 también posee un impacto negativo pero no significativo.

Al revisar los resultados de la prueba de razonamiento cuantitativo, el efecto de los

tratamientos no es significativo.

Por otra parte, al ver los resultados de ambas pruebas en cuadro 10<sup>14</sup> y tal como se menciona en la sección metodológica al comprobar los resultados por los métodos LIML y Jackknife para el método de variables instrumentales se puede observar que los resultados son consistentes con los encontrados por variables instrumentales. La única variación hallada fue el cambio del nivel de significancia de la prueba de lectura crítica que pasó de 95 % al 90 %.

## 7. Discusión

En este trabajo se encontraron efectos diversos sobre el impacto que políticas informativas pueden tener en los resultados de la prueba Saber Pro. A nivel global no se encontró ningún resultado significativo, por esta razón, se analizó cada componente de forma individual. Se encontró que el efecto de haber sido elegido para recibir el tratamiento 1 (ambas piezas) y controlando por las variables desbalanceadas, presentó un efecto negativo y significativo al 90 % de -0.152 desviaciones estándar, mientras que al analizar el componente de razonamiento cuantitativo no se evidencia ningún impacto significativo. Adicionalmente, se validó su robustez con las estimaciones por LIML y Jackknife, que son pruebas más precisas.

Estos resultados pueden deberse a la caracterización de cada componente, la prueba de lectura crítica, se evalúa su comprensión de texto, lo cual requiere del estudiante un mayor grado de atención. Mientras que la prueba de razonamiento cuantitativo evalúa lógica matemática y abstracta, lo cual requiere procesos cognitivos diferentes.

Ahora cabe analizar que aspectos se podrían mejorar del estudio.

### **Focalización**

La muestra objeto de estudio fueron estudiantes de la Facultad de Economía y de la Escuela de Administración de la Universidad del Rosario, que dentro de sus planes de estudio poseen un alto porcentaje de materias cuantitativas y que por el diseño del mismo cuentan con pocas materias que se centren en su comprensión de lectura. Adicionalmente, poseen un nivel socioeconómico alto, lo cual hace que la proporción de estudiantes que cuenta con crédito ICETEX se baja por lo que no tenían interés en condonaciones de la deuda. Por lo cual se debería incluir estudiantes de programas multidisciplinarios y de varias instituciones.

### **Información**

Al ser un examen del tipo "Low Stakes Test", requiere mayor creatividad del diseñador. Por una parte, el mensaje pudo haber sido más motivador dando recomendaciones

---

<sup>14</sup>Ver anexos

de relajación y manejo del estrés, y por otra, la información que se suministro estaba orientada a las personas que desean continuar con sus estudios de posgrado.

### **Tiempo**

Si bien el tratamiento se envió el día antes del examen para reducir la probabilidad de "spillovers"<sup>15</sup>, este puede generar externalidades negativas como aumento en el nivel de ansiedad, remordiando, angustia por el futuro, entre otros aspectos psicológicos. Por lo cual, se podría replantear el diseño y extender su aplicación a otras instituciones o unidades académicas para mitigar la difusión de información, y así reducir el impacto de la externalidad. Además, se podría aplicar el tratamiento antes de la realización del curso de preparación, de esta forma los estudiantes podrían tomar una decisión más consiente.

Teniendo en cuenta los resultados y los planes de mejora anteriormente mencionados, ampliar el estudio para observar la efectividad de recibir mensajes con incentivos en este caso informativos respecto al buen desempeño en la prueba Saber Pro en distintas disciplinas, sería una estrategia útil para generar políticas que pretendan modificar la percepción de examen tipificándolo como un examen "high stakes." de alta prioridad.

---

<sup>15</sup>Spillover es el impacto que tienen fenómenos, eventos o políticas de un sector en otros grupos que no fueron los que indujeron o participaron en dicho evento. Uno de los casos más comunes es la difusión del conocimiento.

## Referencias

- [1] A., P. (2014). Other-rated personality and academic performance: Evidence and implications. *Learning and Individual Differences*, 34:24–32.
- [2] Al Ahmad. M, Al Marzouqi. A, H. M. (2014). Exit exam as academic performance indicator. *TOJET: The Turkish Online Journal of Educational Technology*, 13.
- [3] Al-Kadri (2015). From assessment cocktail to assessment symphony: The development of best assessment practices. *Health Professions Education*, 1:58–64.
- [4] Ansong, D. Ansong, E. A. A. A. S. (2015). A spatio-temporal analysis of academic performance at the basic education certificate examination in ghana. *Applied Geography*, 65:1–12.
- [5] Askham (1997). An instrumental response to the instrumental student: Assessment for learning. *Studies in Educational Evaluation*, 23(4):299–317.
- [6] Astin.A, O. (2012). *Pre-college and institutional influences on degree attainment*. Westport.
- [7] Blanco. M, Engelmannb. D, K. A. N. (2014). Preferences and beliefs in a sequential social dilemma: A within-subjects analysis. *Düsseldorf university press*, 45.
- [8] Casas. F, Meaghan, D. (2001). Renewing the debate over the use of standardized testing in the evaluation of learning and teaching. *Interchange*, 32(2):147–181.
- [9] Chamorro-Premuzic. T, Furnham. A, D. G. H. P. (2005). Personality and preference for academic assessment: A study with australian university students. *Personality and Individual Differences*, 15:247–256.
- [10] Chamorro-Premuzic. T, F. A. (2008). Personality, intelligence and approaches to learning as predictors of academic performance. *Personality and Individual Differences*, 44:1596–1603.
- [11] Chen. T.C, H. M. (2013). Influence of course design on learning approaches and academic performance in physical therapy students. *Procedia Social and Behavioral Sciences*, 93:97–101.
- [12] Ciorbea. I, Pasarica, F. (2013). The study of the relationship between personality and academic performance. *Procedia Social and Behavioral Sciences*, 78:400–404.
- [13] Crescentini. A, Z. G. (2014). The evaluation of mathematical competency: elaboration of a standardized test in ticino (southern switzerland). *Procedia Social and Behavioral Sciences*, 112:180–189.
- [14] D, L. (2016). Correlation between facebook use for educational purposes and academic performance of students. *Computers in Human Behavior*, 61:313–320.

- [15] Darling-Hammond. L, S. J. (2000). Authentic assessment of teaching in context. *Teaching and Teacher Education*, 16:523–545.
- [16] David. P.A, R. G. (1996). Standardization, diversity and learning: strategies for the coevolution of technology and industrial capacity. *International Journal of Industrial Organization*, 14(2):181–201.
- [17] Debicki. B.J, Kellermanns. F.W, B. T. P. A. P. R. (2016). Beyond the big five: The mediating role of goal orientation in the relationship between core self-evaluations and academic performance. *The International Journal of Management Education*, 14:273–285.
- [18] Delavande.A (2008). Measuring revisions to subjective expectations. *Journal of Risk and Uncertainty*.
- [19] E, P. (2006). How college affects students: Ten directions for future research. *Journal of College Student Development*, 47.
- [20] Eil.D, R. (2011). The good news-bad news effect: Asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*.
- [21] El-Gamal (1995). Are people bayesian? uncovering behavioral strategies. *Journal of the American Statistical Association*.
- [22] English. D, Lambert. S.F, L. N. (2016). Adding to the education debt: Depressive symptoms mediate the association between racial discrimination and academic performance in african americans. *Journal of School Psychology*, 57:29–40.
- [23] F, H. (2015). Do girl peers improve your academic performance? *Economics Letters*, 137:54–58.
- [24] Grether.D (1980). Bayes rule as a descriptive model: The representativeness heuristic. *The Quarterly Journal of Economics*.
- [25] Grossman.Z, O. (2012). An unlucky feeling: Overconfidence and noisy feedback. *Journal of Economic Behavior and Organization*.
- [26] Grunschel. C, Schwinger. M, S. R. F. S. (2016). Effects of using motivational regulation strategies on students academic procrastination, academic performance, and well-being. *Learning and Individual Differences*, 49:162–170.
- [27] Heissel. J, Adam.E, D. J. F. D. M. J. (2018). Testing, stress and performance: How students respond physiologically to high stakes testing. *National Bureau of Economic Research*, 25305.
- [28] Heissel. J. A., Levy. D. J., A. E. K. (2007). Stress, sleep, and performance on standardized tests: Understudied pathways to the achievement gap. *AERA Open*, 3(3).

- [29] Heitink. M, Van der Kleij. F.M, V. B. S. K.-K. W. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational Research Review*, 17:50–62.
- [30] Herman. J.L, G. S. (1993). The effects of standardized testing on teaching and schools. *Educational Measurement, Issues and Practice*, 12:20–25.
- [31] J, D. (2009). Exhibitions: Connecting classroom assessment with culminating demonstrations of mastery. *Theory Into Practice*, 48:36–43.
- [32] Kuncel. N.R, H. S. (2007). Standardized tests predict graduate students success. *Science*, 315:1080–1081.
- [33] Lozano. J.H, Gordillo. F, P. M. (2014). Impulsivity, intelligence, and academic performance: Testing the interaction hypothesis. *Personality and Individual Differences*, 61:62–68.
- [34] Lu. Y, Chou. S, L. E. (2014). Gender differences in the impact of weight status on academic performance: Evidence from adolescents in taiwan. *Children and Youth Services Review*, 46:300–314.
- [35] M, M. (2006). The bad, the weak, and the ugly: Avoiding the pitfalls of instrumental variables estimation. *Bates College*.
- [36] MacLeod. W, Riehl. E, S. J. U. M. (2015). How college affects students: Ten directions for future research. *National Bureau of Economic Research*, 21230.
- [37] Malarkey. W. B., Pearl. D. K., D. L. M. K.-G. J. K. G. R. (1995). Influence of academic stress and season on 24-hour mean concentrations of acth cortisol, and endorphin. *Psychoneuroendocrinology*, 20(5):499–508.
- [38] Masrek. M, Z. N. (2015). The relationship between knowledge conversion abilities and academic performance. *Procedia Social and Behavioral Sciences*, 174:3603–3610.
- [39] Mitrofan. N, I. A. (2013). Predictors of academic performance. the relation between the big five factors and academic performance. *Procedia Social and Behavioral Sciences*, 78:125–129.
- [40] Moses. M, N. M. (2007). The testing culture and the persistence of high stakes testing reforms. *Education and Culture*, 23:55–72.
- [41] Munoz-Comonfort. A, Leenen. I, F.-v. d. G. T. (2014). Correlación entre la evaluación diagnóstica y el rendimiento académico de los estudiantes de medicina. *Investigación en Educación Médica*, 3(10):85–91.
- [42] Mushtaq. I, K. S. (2012). Factors affecting students' academic performance. *Global Journal of Management and Business Research*, 12(9):17–22.

- [43] P, P. . (2013). The view of freedom and standardisation among managers in swedish construction contractor projects. *International Journal of Project Management*, 31:299–306.
- [44] Rashid. T, A. H. (2016). Technology use, self-directed learning, student engagement and academic performance: Examining the interrelations. *Computers in Human Behavior*, 63:604–612.
- [45] Riehl. E, Saavedra. J, U. (2016). Learning and earning: An approximation to college value added in two dimensions. Technical report.
- [46] R.L., L. (2000). Assessment and accountability. *Educational Researcher*, 29(2):4–16.
- [47] Santos. A, Torres Formoso. C, T. J. (2002). Expanding the meaning of standardisation within construction processes. *The TQM Magazine*, 14:25–33.
- [48] Schlesselman. L, C. C. (2011). Predictors of poor student performance at a single, accreditation council for pharmacy education - accredited school of pharmacy. *Currents in Pharmacy Teaching and Learning*, 3:101–105.
- [49] Struyven. K, Dochy. F, J. (2005). Students’ perceptions about evaluation and assessment in higher education: A review. *Assessment and Evaluation in Higher Education*, 30(4):325–341.
- [50] Viscusi, W. K. and O’Connor, C. (1984). Adaptive responses to chemical labeling: Are workers bayesian decision makers? *American Economic Review*.
- [51] Weekes. N., Lewis. R., P. F. G.-J. J. B. D. E. L. S. J. (2006). Examination stress as an ecological inducer of cortisol and psychological responses to stress in undergraduate students. *Stress*, 9(4):199–206.

## 8. Anexos

### 8.1. Tratamientos

Figura 2: Pieza incentivos académicos



## 8.2. Encuestas telefónicas: Metodología y dinámica

### Justificación

Teniendo en cuenta la naturaleza del medio por el cual se enviaron los tratamientos, es decir, correo electrónico, en el cual más allá de la confirmación de recepción y apertura de la pieza, no es posible confirmar la lectura de la información. Se hizo necesario emplear un método de recolección que más allá de rescatar dichos datos permitiera obtener información referente a las expectativas que poseían los estudiantes respecto a futuros estudios de posgrado.

Con el fin de conocer dicha información, se empleó la metodología de encuestas telefónicas. La cual, es uno de los métodos más efectivos para recopilar información rápidamente, proporcionando mayor flexibilidad que otro tipo de cuestionarios, pues se conforma por preguntas cerradas con un máximo de respuestas.

### Objetivos

- Rastrear la lectura de la pieza con información académica o económica por parte de los estudiantes pertenecientes a alguno de los tratamientos.

Figura 3: Pieza incentivos económicos

Universidad del Rosario

Tus resultados en el **Saber Pro** tendrán efectos salariales y económicos

Al obtener mejores resultados la reputación y calificación de la UR crece\*

Aumentando oportunidades y salarios de enganche.

Si tienes créditos con el ICETEX podrás solicitar condonaciones de la deuda (decreto 2636/12)

**UN BUEN DESEMPEÑO EN EL EXÁMEN PAGA**

\*Macloed et al. 2015.

- Indagar acerca de las expectativas de estudio de posgrado en el corto, mediano y largo plazo.

### Estructura y desarrollo de la entrevista

- Fechas
  - Cohorte 2017, 19 de febrero al 2 de marzo
  - Cohorte 2018, 12 al 23 de noviembre
- Duración: 5 a 10 minutos. El número de preguntas se debe ajustar a este tiempo, de tal forma que el encuestado si lo requiere pueda transmitir sus ideas.
- Participantes:
  - Encuestados: 537 estudiantes. 315 estudiantes Escuela de Administración, 222 estudiantes Facultad de Economía.
  - Encuestadores: Dirección de cohorte Facultad de Economía.
- Características de los encuestados: Estudiantes de últimos semestres de la Facultad de Economía y de la Escuela de Administración que presentaron la prueba Saber Pro en 2017 o 2018.

## Estructura de la llamada

- Saludo

Buenos días / Buenas tardes.

Mi nombre es xxxxx de la Dirección de Cohorte de la Facultad de Economía. Me comunico con xxxxx estudiante de la Facultad de Economía/ Escuela de Administración de la Universidad del Rosario.

Una vez se recibe la afirmación, se procede.

- Habeas data

Antes de iniciar la llamada, quisiera recordarle que esta puede ser grabada y monitorizada, todo lo que esta contenga servirá para investigación y será completamente anonimizada.

¿Desea continuar?

- Preguntas

¿El pasado 28 de octubre de 2017/ 06 de octubre de 2018 día anterior a la presentación de la prueba Saber Pro recibió información referente a este por parte de su secretaria académica?

¿Leyó la información contenida en las piezas del correo electrónico?

Ahora le realizaré algunas preguntas respecto a sus expectativas académicas

¿Piensa iniciar estudios de posgrado, es decir, especialización, maestría o doctorado en el corto plazo? Entendido este periodo como un espacio de tiempo menor a un año.

¿Piensa iniciar estudios de posgrado, es decir, especialización, maestría o doctorado en el mediano plazo? Entendido este periodo como un espacio de tiempo mayor a un año y menor a 5 años.

¿Piensa iniciar estudios de posgrado, es decir, especialización, maestría o doctorado en el largo plazo? Entendido este periodo como un espacio de tiempo mayor a 5 años.

¿No piensa iniciar estudios de posgrado, es decir, especialización, maestría o doctorado?

- Despedida

Muchas gracias por su colaboración.

Si desea conocer alguna información adicional, puede contactarse a la extensión 4119 o al correo electrónico daniela.gaitanc@urosario.edu.co.

A medida que los estudiantes reportan sus respuestas, estas son consignadas a un excel que registra hora y fecha de la llamada, obteniendo los siguientes resultados.

Cuadro 5: Expectativas de Posgrado

	Muestra Completa			Control		Tratamiento 1			Tratamiento 2		
	Media Err. Estd	Min	Max	Media Err. Estd		Media Err. Estd	Medias	<i>p</i> -value Distribución	Media Err. Estd	Medias	<i>p</i> -value Distribución
Corto plazo (Menor 1 año)	0.09 (0.01)	0	1	0.10 (0.01)		0.09 (0.02)	0.97	1.00	0.09 (0.03)	0.77	1.00
Mediano plazo (Entre 1 y 5 años)	0.44 (0.05)	0	1	0.43 (0.03)		0.43 (0.03)	0.79	1.00	0.4 (0.04)	0.66	1.00
Largo plazo (Mayor a 5 años)	0.37 (0.05)	0	1	0.35 (0.07)		0.39 (0.07)	0.41	0.99	0.38 (0.07)	0.84	1.00
No desea estudiar	0.05 (0.05)	0	1	0.06 (0.07)		0.04 (0.07)	0.44	1.00	0.06 (0.07)	0.73	1.00

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$  significativamente diferente

*P*-value de medias corresponde al t-test, *p*-value de distribución corresponde al test Kolmogorov-Smirnov; respecto cada grupo de tratamiento con el grupo de control Las expectativas de posgrado, son resultado de la encuesta telefónica realizada posterior al examen Saber Pro. Esta se divide en cuatro categorías.

### 8.3. Poder y Robustez del modelo

El poder de una prueba es la probabilidad de rechazar correctamente la hipótesis nula cuando esta es falsa. Es decir,  $\theta = 1 - \beta = \Pr(\text{rechazar } H_0 \mid H_0 \text{ es falso})$ . Aumentar la potencia o poder de una prueba disminuye la probabilidad de un error de tipo II <sup>16</sup>, por lo que se espera que esta sea alta, normalmente se busca que se encuentre entre el 80 % y 90 %. Especialmente, cuando se manejan experimentos aleatorios es muy importante determinar el tamaño de la muestra a partir del poder que se pretende manejar, pues esta debe ser lo suficientemente grande para detectar una diferencia práctica cuando exista.

Teniendo en cuenta la importancia de esta prueba, antes de iniciar el experimento aleatorio, se evaluó el rendimiento histórico de los estudiantes en la prueba y en cada uno de sus componentes, para así determinar la línea base para indagar cuál sería el tamaño de muestra ideal. Se obtuvieron los siguientes resultados.

Cuadro 6: Tamaño de muestra por componente

Área	Poder		
	0.8	0.85	0.9
Puntaje Global	99	112	131
Lectura Crítica	266	304	356
Comunicación escrita	91	108	136
Competencias ciudadanas	39	44	51
Razonamiento cuantitativo	85	97	113
Inglés	180	205	240

Prueba de poder por grupo

Tal como se observa en el cuadro anterior, bajo un nivel de poder del 80 %, el presente experimento posee el tamaño de muestra por grupo para mitigar el error tipo II, en cada uno de los componentes. Por lo tanto, es posible analizar el impacto de los tratamientos en cada prueba.

Por otro lado, teniendo en cuenta el tamaño de las muestras y su poder. Tal como lo señala Murray (2006), es importante realizar un chequeo de robustez mediante el "Limited Information Maximum Likelihood"(LIML). Este se basa en el supuesto de normalidad de los errores en la ecuación estructural y en la ecuación de la primera etapa, demostrando tener un sesgo menor que el estimador de mínimos cuadrados en dos

<sup>16</sup>Cuando la hipótesis nula es falsa y no se rechaza, existe un error de tipo II. La probabilidad de cometer un error de este tipo es  $\beta$ , que depende de la potencia de la prueba. Puede reducirse el riesgo de cometer un error de tipo II al asegurarse de que la prueba tiene suficiente poder.

etapas. Como complemento, se analiza el estimador Jackknife de variables instrumentales para establecer la robustez del modelo. Según Angrist (1999), este es independiente a las perturbaciones que pueden generar las muestras pequeñas, al intentar eliminar el sesgo de las muestras finitas de 2SLS.

Cuadro 7: Pruebas de Robustez

VARIABLES	P.Global		P.Lectura		P.Escrita		P.Ciudadanas	
	LIML	Jackknife	LIML	Jackknife	LIML	Jackknife	LIML	Jackknife
Tratamiento_1	-0.0722 (0.0888)	-0.0729 (0.0897)	-0.181* (0.0927)	-0.181* (0.0937)	0.0320 (0.103)	0.0316 (0.105)	-0.134 (0.0955)	-0.134 (0.0965)
Tratamiento_2	-0.0644 (0.107)	-0.0646 (0.108)	-0.142 (0.112)	-0.143 (0.113)	0.351*** (0.126)	0.352*** (0.128)	-0.218* (0.115)	-0.218* (0.116)
Observaciones	537	537	537	537	527	527	537	537
R-cuadrado	0.321	0.321	0.260	0.260	0.085	0.085	0.215	0.215

Controles ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓

Errores estándar robustos en paréntesis

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Controles: Edad, créditos aprobados, puntaje saber 11, Índice de Dotación del Hogar, educación de los padres

Tratamiento 1: Pieza información académica + pieza información económica

Tratamiento 2: Pieza información académica