INFORMES

Medición de los efectos de la educación superior en Colombia sobre el aprendizaje estudiantil

Informe técnico

MINEDUCACIÓN

icfes mejor saber

TODOS POR UN NUEVO PAÍS
PAZ EQUIDAD EDUCACIÓN

Presidente de la República
Juan Manuel Santos Calderón

Ministra de Educación Nacional
Gina María Parody d'Echeona

Viceministra de Educación Superior
Natalia Ariza Ramírez

Directora General
Ximena Dueñas Herrera

Secretaria General
María Sofía Arango Arango

Director de Evaluación
Hugo Andrés Gutierrez

Director de Producción y Operaciones
Francia María del Pilar Jiménez

Directora de Tecnología e Información
Ingrid Picón Carrascal

Jefe Oficina Asesora de Comunicaciones y Mercadeo
Diego Fernando Suárez Manzur

Jefe Oficina Gestión de Proyectos de Investigación
Luisa Fernanda Bernat Díaz

Subdirectora de Producción de Instrumentos
Claudia Lucía Saenz Blanco

Subdirectora de Diseño de Instrumentos
Flor Patricia Pedraza Daza

Subdirectora de Análisis y Divulgación
Silvana Godoy Mateus

Adaptación y elaboración del documento
Benjamin Domingue
Julián Mariño von Hildebrand
Adriana Molina Mantilla
Alvaro Uzaheta
Carolina Lopera
Edwin Cuéllar

Revisión de estilo
Fernando Carretero Socha

Diseño y diagramación
Jessica Pacateque Flórez

ISBN de la versión digital: 978-958-11-0672-1

Bogotá D.C., mayo de 2015

MINEDUCACIÓN

icfes mejor saber

TODOS POR UN NUEVO PAÍS
PAZ EQUIDAD EDUCACIÓN

ADVERTENCIA

Con el fin de evitar la sobrecarga gráfica que supondría utilizar en español "o/a" para denotar uno u otro género, el ICFES opta por emplear el masculino genérico en el que todas las menciones de este se refieren siempre a hombres y mujeres.

Todo el contenido es propiedad exclusiva y reservada del ICFES y es el resultado de investigaciones y obras protegidas por la legislación nacional e internacional. No se autoriza su reproducción, utilización ni explotación a ningún tercero. Solo se autoriza su uso para fines exclusivamente académicos. Esta información no podrá ser alterada, modificada o enmendada.

ICFES. 2015. Todos los derechos de autor reservados ©.

TÉRMINOS Y CONDICIONES DE USO PARA PUBLICACIONES Y OBRAS DE PROPIEDAD DEL ICFES

El Instituto Colombiano para la Evaluación de la Educación (ICFES) pone a la disposición de la comunidad educativa y del público en general, **DE FORMA GRATUITA Y LIBRE DE CUALQUIER CARGO**, un conjunto publicaciones a través de su portal www.icfes.gov.co. Dichos materiales y documentos están normados por la presente política y están protegidos por derechos de propiedad intelectual y derechos de autor a favor del ICFES. Si tiene conocimiento de alguna utilización contraria a lo establecido en estas condiciones de uso, por favor infórmenos al correo prensaicfes@icfes.gov.co.

Queda prohibido el uso o publicación total o parcial de este material con fines de lucro. **Únicamente está autorizado su uso para fines académicos e investigativos**. Ninguna persona, natural o jurídica, nacional o internacional, podrá vender, distribuir, alquilar, reproducir, transformar ([1]), promocionar o realizar acción alguna de la cual se lucre directa o indirectamente con este material. Esta publicación cuenta con el registro ISBN (International Standard Book Number, o Número Normalizado Internacional para Libros) que facilita la identificación no sólo de cada título, sino de la autoría, la edición, el editor y el país en donde se edita.

En todo caso, cuando se haga uso parcial o total de los contenidos de esta publicación del ICFES, el usuario deberá consignar o hacer referencia a los créditos institucionales del ICFES respetando los derechos de cita; es decir, se podrán utilizar con los fines aquí previstos transcribiendo los pasajes necesarios, citando siempre la fuente de autor) lo anterior siempre que estos no sean tantos y seguidos que razonadamente puedan considerarse como una reproducción simulada y sustancial, que redunde en perjuicio del ICFES.

Asimismo, los logotipos institucionales son marcas registradas y de propiedad exclusiva del Instituto Colombiano para la Evaluación de la Educación (ICFES). Por tanto, los terceros no podrán usar las marcas de propiedad del ICFES con signos idénticos o similares respecto de cualesquiera productos o servicios prestados por esta entidad, cuando su uso pueda causar confusión. En todo caso queda prohibido su uso sin previa autorización expresa del ICFES. La infracción de estos derechos se perseguirá civil y, en su caso, penalmente, de acuerdo con las leyes nacionales y tratados internacionales aplicables.

El ICFES realizará cambios o revisiones periódicas a los presentes términos de uso, y los actualizará en esta publicación.

El ICFES adelantará las acciones legales pertinentes por cualquier violación a estas políticas y condiciones de uso.

* La transformación es la modificación de la obra a través de la creación de adaptaciones, traducciones, compilaciones, actualizaciones, revisiones, y, en general, cualquier modificación que de la obra se pueda realizar, generando que la nueva obra resultante se constituya en una obra derivada protegida por el derecho de autor, con la única diferencia respecto de las obras originales que aquellas requieren para su realización de la autorización expresa del autor o propietario para adaptar, traducir, compilar, etcétera. En este caso, el ICFES prohíbe la transformación de esta publicación.

Contenido

Introducción **5**

1. Aspectos conceptuales: inferencia causal **7**

1.1. Supuestos subyacentes a las estimaciones de medidas de valor agregado (VA) 8
1.2. Decisiones operativas y complicaciones en los análisis de VA 10
1.2.1. Tratamientos 10
1.2.2. Unidades 12
1.2.3. Resultados 13
1.3. Desafíos Técnicos: sesgo 15

2. Aspectos metodológicos **17**

2.1. Datos 17
2.1.1. Relación entre las pruebas 17
2.1.2. Propiedades psicométricas de las medidas de SABER 11° 18
2.1.3. Índice de nivel socioeconómico (INSE) 19
2.1.4. Selección de los estudiantes 19
2.2. Métodos 22

3. Resultados **24**

3.1. Modelos estimados: efectos fijos y componentes de varianza 24
3.2. Relaciones entre modelos 25

4. Análisis de sensibilididad **28**

4.1. Efectos heterogéneos del tratamiento 28
4.2. Términos de orden superior 30
4.3. Diferentes efectos contextuales 30

5. Consideraciones finales **31**

Referencias **32**

Anexo **35**

Introducción

Los modelos de valor agregado *(VA)* son modelos estadísticos que intentan evaluar la efectividad educativa haciendo énfasis en el progreso de los estudiantes en el tiempo. En este sentido, el valor agregado de una institución se entiende como la contribución al progreso neto de los estudiantes, hacia objetivos de aprendizaje establecidos, una vez eliminada la influencia de otros factores ajenos a la institución que pueden contribuir a ese progreso (Matinez R., Gaviria J., Castro M., 2009). En Estados Unidos, los modelos de *VA* en educación son un componente que genera controversia en muchos sistemas de evaluación docente. Investigaciones han sugerido que tienen propiedades deseables (Chetty, Friedman, Rockoff, 2011; KaneyStaiger, 2008), aunque otros autores no están de acuerdo (Newton, Darling-Hammond, Haertely Thomas, 2010)[1].

Colombia se encuentra en una posición única para reportarle el valor agregado a su sistema de educación superior. Esto porque la mayoría de los países no requieren que los estudiantes presenten un examen estandarizado después de su educación secundaria y, por consiguiente, este tipo de estudios son, en gran parte, imposibles en instituciones de educación post-secundaria. Existen excepciones importantes, como el trabajo que implica la Evaluación del Aprendizaje en la Educación Superior (*Collegiate Learning Assessment*, CLA, Steedle, 2012). Sin embargo, los datos obtenidos a través de CLA son transversales y no representativos. Por tanto, su carácter se diferencia muy poco de muchos de los estudios de *VA* (en parte debido a que también se utilizan en evaluaciones de desempeño).

En el estudio sobre valor agregado para la educación superior en Colombia se utilizan los puntajes actuales de SABER PRO en comparación con los puntajes predichos a partir de los resultados observados en SABER 11° (y otras características del estudiante o de la institución educativa) para evaluar el desempeño de las instituciones. Dado que no se pretende evaluar a los docentes a nivel individual, esta tarea se asemeja a los estudios sobre los efectos del colegio (Raudenbush y Willms, 1995).

Este documento se compone de cinco secciones. En la primera se analizan los aspectos conceptuales del marco tradicional de inferencia causal para los modelos de valor agregado (*VA*), los desafíos ocasionados las diferencias estructurales entre las instituciones de educación superior y las de primaria y secundaria y los aspectos relacionados con los sesgos los cuales han motivado las decisiones tomadas en esta investigación.

1 Los modelos de VA se describen ampliamente en otro lugar de este documento. Un punto de partida potencial

La segunda parte analiza las cuestiones metodológicos de esta investigación, haciendo énfasis en los aspectos relacionados con los datos y los modelos de estimación utilizados en este estudio. En la tercera parte se presentan los principales resultados; en la cuarta se muestra el análisis de sensibilidad de las estimaciones de valor agregado para algunos supuestos importantes. Por último, la quinta parte contiene algunas consideraciones finales del estudio de valor agregado para la educación superior en Colombia, haciendo énfasis en aspectos importantes para los reportes de resultados institucionales.

1. Aspectos conceptuales: inferencia causal[2]

Para estimar el valor agregado de una institución sobre el aprendizaje de los estudiantes, ya sea implícita o explícitamente, se requiere la inferencia causal. En el marco de resultados potenciales, Holland (1986) aclara los desafíos enfrentados al hacer inferencias causales. El marco de los resultados potenciales se entiende más fácilmente en términos de una prueba aleatoria. Un estudiante es expuesto a un tratamiento (T) o a una condición de control (C). Después de la exposición, se observa cierto resultado. Para el estudiante i, esto sería $Y_i(T)$ o $Y_i(C)$. El *problema fundamental de la inferencia causal* (Holland, 1986) es que solo se observa uno de estos resultados para cada estudiante. Si los estudiantes se han asignado aleatoriamente a una condición de tratamiento (T) o a una condición de control (C), entonces hacer una inferencia causal es relativamente sencillo: la diferencia promedio entre el resultado de los individuos del grupo T y los del C proporciona una estimación insesgada del efecto causal promedio.

En escenarios donde no es posible hacer una asignación aleatoria (por ejemplo, datos observacionales, asignación no aleatoria de estudiantes a las instituciones, etcétera), la inferencia es más difícil dado que no se puede asumir que los estudiantes asignados a T y a C son equivalentes al principio del tratamiento. Las diferencias previas existentes en los dos grupos constituyen el sesgo de selección: las diferencias al principio de la educación superior pueden explicar las diferencias al final del tratamiento, en lugar de pertenecer al grupo T por sí mismo.

En consecuencia, comparar los promedios en el resultado medido para el grupo de estudiantes en T y C, sin el ajuste correspondiente, podría llevar a conclusiones sesgadas. Por ejemplo, si los estudiantes con altas capacidades van a la institución A en lugar de la B, las diferencias en los resultados podrían reflejar parcialmente diferencias en la capacidad del estudiante que son anteriores a la formación impartida en cada institución. Con datos observacionales, el objetivo es estimar el efecto causal ajustando las diferencias preexistentes relacionadas con la asignación al grupo de tratamiento y los resultados.

2 Algunas partes de esta sección se basan en el manuscrito presentado en la Conferencia de *American Educational Research Association* (AERA), 2014. Retos para el estudio de los efectos de las instituciones en la educación superior.

1.1. Supuestos subyacentes a las estimaciones de medidas de valor agregado (VA)

En contextos observacionales, las inferencias causales válidas requieren diversos supuestos. Reardon y Raudenbush (2009) identifican los siguientes supuestos para una inferencia causal válida con modelos de *VA*:

a. **Manipulabilidad:** teóricamente, los estudiantes podrían estar expuestos a cualquier tratamiento.
b. **No interferencia entre unidades:** el resultado de un estudiante depende solamente de su asignación a un determinado tratamiento.
c. **El supuesto métrico:** los resultados del puntaje de la prueba están en una escala de intervalo.
d. **Homogeneidad:** el efecto causal no varía como función de una característica del estudiante.
e. **Tratamiento altamente ignorable:** la asignación al tratamiento es esencialmente aleatoria después de condicionarse por las variables de control.
f. **Forma funcional:** la forma funcional empleada para el control de las características del estudiante es la correcta.

La lógica con los supuestos *(c)* y *(d)* es similar tanto en la educación primaria como en la superior. El supuesto *(c)*, establece que los puntajes de la prueba están en una escala de intervalo. Mientras que Ballou (2009) plantea interrogantes acerca de la validez de este supuesto, la cuestión es psicométrica y no es exclusiva para la educación superior. En el supuesto *(d)* establece que los efectos de la escuela son una constante independiente de los antecedentes del estudiante. Este supuesto puede ser potencialmente erróneo respeto a la educación primaria y a la educación superior. En Raudenbush y Willms (1995) se discute que este supuesto debe ser objeto de una cuidadosa investigación en los estudios del *VA*.

Los otros supuestos tienen implicaciones únicas para los estudios del *VA* en la educación superior. El supuesto *(a)* requiere que sea teóricamente factible que un estudiante tenga un resultado potencial para todas las posibles instituciones de educación superior *(IES)*. En este sentido, los estudiantes son "manipulables" (los estudiantes podrían asistir a todas las *IES* disponibles)[3]. Reardon y Raudenbush (2009) argumentan que esto es poco plausible para algunos sistemas de escuela primaria, dadas las enormes diferencias en los antecedentes de los estudiantes de las diferentes escuelas. Este problema aumenta en la educación superior dado que gran parte de las universidades solamente matriculan a los estudiantes seleccionados. ¿Si una universidad le niega la admisión a un estudiante, se puede afirmar razonablemente que este estudiante tiene un resultado potencial en esta institución?

3 Problemas con este supuesto son una de las causas que llevan a la definición de la vecindad de comparación en los reportes institucionales.

En situaciones en las que las tasas de aceptación son comparativamente bajas, el enfoque debe cambiar del efecto del tratamiento promedio hasta el efecto del tratamiento sobre el tratado[4] dado que el interés podría estar sobre los efectos de las instituciones élite en los estudiantes con mayor probabilidad de asistir a estas. Como consecuencia, las estimaciones solo serían relevantes para subconjuntos de estudiantes. Esto significaría un reto para la elaboración de informes de resultados sobre medidas de valor agregado a nivel de las instituciones.

En el supuesto *(b)* se requiere que los resultados potenciales de un estudiante i no dependan de la escuela asignada a otros estudiantes i'. Este supuesto, también conocido como el supuesto de estabilidad del valor de tratamiento unitario (SUTVA, por su sigla en inglés), no toma en cuenta los efectos del grupo par. El efecto de la escuela j sobre el estudiante i es independiente de si los compañeros de i son estudiantes fuertes o débiles, ricos o pobres, etcétera. La investigación de los efectos del grupo par en la educación superior (Winston y Zimmerman, 2004; Sacerdote, 2011) sugiere que existen efectos moderados, de manera que el supuesto *(b)* no puede ser sostenible.

Una posible manera de evitar estos dilemas es considerar el tratamiento como si se aplicara a todos los grupos de estudiantes, en lugar de aplicarlo a estudiantes individuales como sugieren Rubin, Stuart y Zanutto (2004). Esto lleva a cambios sustanciales en la interpretación de los resultados (los efectos solo se calculan en relación con la forma como están constituidos los grupos de estudiantes, en lugar de agrupaciones alternativas). En particular, las estimaciones solo serían relevantes para la asistencia a una institución junto a la cohorte previa de estudiantes empleados en el análisis, disminuyendo claramente la relevancia para los estudiantes potenciales.

En el supuesto *(e)* se requiere que la asignación a las instituciones pueda tratarse aleatoriamente después de controlar por las covariables. Esto pone un peso sustancial sobre cómo elegir las covariables correctas. Por ejemplo, la inclusión de una variable como el nivel socioeconómico *(NSE)* podría tener sentido dada la asociación entre la riqueza y la asistencia a ciertos tipos de instituciones. Sin embargo, la inclusión de tal covariable puede ser poco atractiva en un sistema de rendición de cuentas formal, puesto que equivale a establecer expectativas diferentes para los estudiantes en función de su NSE.

Esta tensión entre la inclusión o exclusión de ciertas covariables puede corregirse con la inclusión de muchas covariables a nivel del estudiante, ya que las estimaciones son normalmente poco sensibles a su inclusión (Ballou, Sanders y Wright, 2004). Es decir, las estimaciones de los efectos de la institución no se afectan dramáticamente por la inclusión de covariables adicionales, después de incluir un control apropiado de los logros previos. Sin embargo, cuando se consideran covariables a nivel de institución, las estimaciones pueden cambiar dramáticamente, situación que se ilustra empíricamente más adelante.

4 Ver Imai, King y Stuart, 2008 quienes realizan una discusión sobre estimaciones causales alternativas

El supuesto *(f)* afirma que la forma funcional de la ecuación de regresión es la correcta. Por ejemplo, los puntajes previos a menudo se incluyen como predictores lineales y, por tanto, se asume que la verdadera relación entre los resultados y los puntajes previos es lineal[5]. Lo más importante es quizá que hay suficiente "soporte común", es decir, que a cada universidad asiste un número suficiente de estudiantes a través de las distribuciones de los controles relevantes (NSE, aptitud académica, etcétera). Este supuesto se relaciona con el supuesto *(a)*. Si muy pocos estudiantes de un determinado tipo (por ejemplo, NSE alto) asisten a una escuela determinada, entonces el efecto estimado, para esos estudiantes de asistir a esa institución, dependerá en gran medida de la especificación del modelo de regresión más que de comparaciones basadas en datos entre unidades comparables en diferentes instituciones. Los problemas con tales supuestos subyacentes en el caso de datos escasos, son bien conocidos en la literatura sobre inferencia causal (Ho & Rubin, 2011).

1.2 Decisiones operativas y complicaciones en los análisis de VA

Existen varias implicaciones sobre los análisis de valor agregado derivados de las decisiones operativas –identificación de unidades, resultados y tratamiento– sobre los supuestos expuestos anteriormente. En esta sentido, se presenta un conjunto de consideraciones técnicas que deben ser tenidas en cuenta en el análisis de valor agregado. Como lo señalan Rubin *et al*. (2004), estas son las primeras tareas de la inferencia causal y son fundamentales.

1.2.1 Tratamientos

La especificación de una condición de tratamiento requiere que se seleccione una condición de control que la acompañe. ¿Debe compararse a la institución educativa j con una institución "promedio", con otra institución a la que el estudiante i podría razonablemente asistir, o con otra institución a la que definitivamente no podría asistir? ¿Deben compararse carreras académicas dentro de la misma institución? La opción estándar en el trabajo de *VA* es permitir que con la condición de control, o el contrafactual, se obtenga un grado de la institución "promedio". Esta elección tiene implicaciones para la selección de las covariables, debido a que estas se utilizan para definir la institución "promedio" (Manzi, San Martin y Van Bellegem, 2014).

Una preocupación en la educación superior es que naturalmente hay más variabilidad (con relación a la escuela primaria y secundaria) en la duración de la exposición al tratamiento. En promedio, completar la educación universitaria puede tomar de 3 a 6 años (o incluso más). Esta variabilidad es inevitable, pero puede plantear desafíos. Por ejemplo, los estudiantes que se demoran mucho más tiempo para obtener su título, pueden haber estado trabajando más horas o tener empleos más exigentes que los que se graduaron en menos tiempo.

5 Ver Lockwoody McCaffrey (*s.f.*)) para explorar la inclusión de términos de orden superior.

Como consecuencia de sus experiencias laborales, es posible que estos estudiantes hayan aprendido mucho más fuera de la institución educativa, lo cual podría atribuirse inadecuadamente al tratamiento.

Otra consideración respecto a la definición del tratamiento es la especificación de los aspectos de asistir a la institución j que se intentan evaluar. Raudenbush (2004) separa los efectos causales que las instituciones (o los docentes) pueden tener en los estudiantes en efectos de práctica y contexto. Los primeros se relacionan con el aprendizaje debido a prácticas escolares intencionales, mientras que los últimos son efectos debido al contexto en el cual funciona la institución. Es decir, distingue las prácticas escolares de contextos escolares, como el NSE del lugar donde se encuentra ubicada la institución.

Raudenbush y Willms (1995) describen dos estrategias (denominadas modelos tipo A y B) que ayudan a analizar estos efectos causales. Los modelos tipo A incluyen tanto el efecto del contexto como el de la práctica, pero no pretenden analizar la proporción en que cada uno de esos componentes contribuyen al aprendizaje de los estudiantes. En la educación primaria, estos modelos deberían ser de interés para padres y estudiantes, debido a que describen el efecto sobre el aprendizaje que tiene lugar dentro de la escuela, independientemente de lo que causó el efecto (por ejemplo, un profesor o los efectos asociados con los compañeros del salón de clases).

Los modelos tipo B solo miden el efecto de la práctica. Estos podrían ser de interés para los encargados de las políticas públicas en educación primaria, debido a que describen los efectos sobre el aprendizaje derivados de los elementos que una escuela puede controlar, como la calidad de la enseñanza y altas expectativas para el éxito de los estudiantes. En estos modelos, por ejemplo, no se responsabiliza a las escuelas por los efectos de pares.

En contraste, en la educación superior es posible que los modelos tipo A sean de interés tanto para los estudiantes como para los responsables de las políticas, ya que las IES controlan el ingreso de sus estudiantes y, por tanto, se podría afirmar que causan el aprendizaje contextual. En un sentido técnico, Raudenbush y Willms (1995) señalan que es difícil obtener estimaciones tipo B de manera adecuada debido a que tipicamente la práctica escolar no es observable.

1.2.2 Unidades

La elección de la unidad de análisis se relaciona con el debate sobre el supuesto *(b)*. Específicamente, responde a si el tratamiento debe aplicarse a estudiantes individuales o a grupos de estudiantes. Un enfoque útil es considerar cuál sería la forma ideal de realizar un experimento aleatorio. Es decir, El tratamiento se aplicaría a grupos de estudiantes conformados aleatoriamente o a grupos de estudiantes que existen dentro de los datos

observacionales. La respuesta a esta pregunta depende claramente del enfoque sobre los efectos de los pares o compañeros. Si pensamos que estos son un componente inevitable (tal vez deseable) del aprendizaje de los estudiantes, entonces solamente la segunda opción es sostenible.

El supuesto de estabilidad del valor del tratamiento unitario (SUTVA, por su sigla en inglés) tiene implicaciones en la selección de las unidades de análisis. Rubin *et al*. (2004) señalan que el SUTVA requiere que haya solamente una versión del tratamiento. En universidades donde los estudiantes tienen muchas opciones diferentes para el enfoque académico, pueden surgir interrogantes acerca de si se está aplicando un único tratamiento para todos los estudiantes. Si el tratamiento se define como la obtención de un grado de la institución j, la implicación es que el tratamiento no varía por estudiante i en la institución j.

Se podría argumentar que obtener en una universidad un título de ingeniería en lugar de uno en historia del arte son tratamientos muy diferentes. Por supuesto, se podría argumentar que incluso enfocar el tratamiento dentro de "grupos de referencia"[6] podría ser una categorización demasiado subjetiva y, en su lugar, proponer que se modelen los resultados dentro de carreras académicas (por ejemplo, ingeniería eléctrica versus ingeniería mecánica).

Pero reducir la definición de unidades de tratamiento complica el análisis de VA. En primer lugar, un análisis minucioso lleva a una reducción del tamaño de la muestra y de la precisión de las estimaciones. En segundo lugar, si una universidad está dividida en múltiples unidades, la atribución causal se vuelve más complicada para los estudiantes que cambian de programa académico o que toman clases en múltiples facultades. Considere esta simple ilustración. Imagínese un estudiante de literatura en una Facultad de Artes Liberales. Él cursa dos clases de matemáticas, ambas en la Facultad de Ciencias Naturales, mientras asiste a la universidad. Si la medición de los conocimientos matemáticos es un resultado relevante, ¿cuál debería ser la facultad responsable del aprendizaje de este estudiante? El efecto del cambio entre universidades puede ser mínimo si la mayoría de los cambios de los estudiantes se producen dentro de un grupo de referencia (por ejemplo, de ingeniería eléctrica a ingeniería mecánica), asumiendo que esa es la unidad de análisis.

1.2.3 Resultados

Establecer las medidas de resultados pertinentes en la educación superior es complejo, debido a la diversidad de aprendizajes que se llevan a cabo. Una posibilidad son las competencias genéricas – como las que se miden, por ejemplo, a través de la Evaluación del Aprendizaje en la Educación Superior (Collegiate Learning Assessment) (Shavelson, 2010), o una parte

6 En Colombia corresponden a categorías de carreras académicas análogas a las de los colegios universitarios (college) en Estados Unidos (por ejemplo, Escuela de Artes Liberales) dentro de una universidad.

de lo desarrollado por la OECD con la Evaluación de los Resultados de la Educación Superior AHELO (Ewell, 2012), y el perfil de competencias de *ETS* (http://www.ets.org/proficiencyprofile/about/content/).

Estas evaluaciones se enfocan en competencias, como cuantitativas básicas, lenguaje, comunicación y pensamiento crítico, en las que todos los estudiantes matriculados en la educación superior deberían demostrar un crecimiento. El uso de tales medidas hace que las comparaciones entre campos de estudio sean posibles. Sin embargo, podría ocurrir que tales pruebas pasen por alto el aprendizaje específico, el cual es un componente clave del tratamiento. Desarrollar las evaluaciones a través de un rango de campos podría ser costoso, pero benéfico cuando se considere crucial el énfasis en las competencias específicas dentro de un campo.

Además de los desafíos relacionados con los tratamientos, las unidades y los resultados, se identifican tres aspectos adicionales que merecen más atención cuando se pretende estimar la efectividad de las *IES*. Estos son la elección de covariables, datos faltantes y la selección de estudiantes. Estos aspectos están estrechamente relacionados con los supuestos y con las decisiones operativas que se mencionaron anteriormente y constituyen la base para las preguntas de investigación que subyacen el análisis empírico.

a. Selección de covariables

La elección de las covariables en un análisis de *VA* es una de las primeras preguntas en las investigaciones de este tipo (Ballou *et al.,* 2004). La principal consideración es cómo dar cuenta del logro previo a nivel del estudiante. En los estudios de *VA* en educación primaria, los investigadores utilizan normalmente el puntaje previo de una única prueba para la misma materia. Pero, incluso en la configuración principal, hay buenas razones para considerar la inclusión de puntajes previos adicionales, pues reducen el efecto del error de medición (Lockwood y McCaffrey, 2014).

En educación superior, un mismo puntaje de una prueba previa puede no estar disponible para cada resultado, especialmente cuando las pruebas específicas se usan como medidas de resultados. Incluso si se tiene disponible una prueba directamente comparable, el lapso entre la prueba previa y el resultado de las IES[7] puede ocasionar la reducción de las correlaciones entre los puntajes previos y los resultados (un mayor lapso entre las pruebas también reduce la precisión de la estimación de los efectos; Leckie y Goldstein, 2009). En general, se considera que la inclusión o exclusión de covariables adicionales a nivel del estudiantes (como el *NSE*) tiene efectos mínimos en las estimaciones de *VA*, una vez que los logros previos se han incluido de forma adecuada (Ballou *et al.,* 2004).

7 En educación superior, el lapso entre la matrícula y la graduación es sustancialmente mayor, comparada con el de la educación básica, en el que las medidas están normalmente disponibles para cada año

Una segunda consideración es si se desean incluir covariables contextuales o agregadas (por ejemplo, el promedio de la institución de la habilidad previa de los estudiantes que ingresan). Incluir tales covariables, se cree, es una forma de producir estimaciones tipo *B,* aunque no necesariamente abordan las dificultades observadas por Raudenbush y Willms (1995), pues en la práctica todavía no se han observado. En cambio, sí se sabe que la inclusión de tales covariables agregadas puede cambiar sustancialmente el rango de las estimaciones resultantes del *VA*, en comparación con los modelos sin covariables contextuales. Además, la inclusión de tales covariables tiene el efecto de cambiar la interpretación de las estimaciones del *VA*, cambiando sutilmente la definición del contrafactual (Manzi *et al.,* 2014).

b. Datos faltantes

Los problemas relacionados con datos faltantes se agravan en la educación superior debido a las mayores tasas de deserción. Es posible que los estudiantes de primaria y secundaria no presenten exámenes en caso de ausencia o reubicación, pero normalmente no se retiran completamente del sistema educativo (por lo menos en Estados Unidos, aunque la situación puede ser diferente en algunas zonas de Colombia).

Por otro lado, los estudiantes universitarios abandonan sus carreras en un gran porcentaje. En Colombia, la tasa promedio de deserción anual es de más del 10%. Con supuestos razonables (por ejemplo, que los estudiantes no abandonan sus estudios aleatoriamente), esta pérdida de estudiantes puede sesgar las estimaciones. Haciendo caso omiso al sesgo, todavía podría ser significativo el hecho de que algunos programas cuentan con tasas de deserción más altas que otros, de este modo, obtener el título pueden ser tan significativo como el aprendizaje que obtiene un estudiante durante el programa (Spence, 1973).

c. Selección de estudiantes

La literatura de *VA* en educación primaria y secundaria se ha enfocado en la selección de los estudiantes como una amenaza significativa para la inferencia causal (Rothstein, 2010; Briggs y Domingue, 2011; Chetty *et al*., 2011; Koedel y Betts, 2011; Goldhaber y Chaplin, 2012). Aunque el supuesto de grupos de estudiantes formados al azar dentro de las universidades es claramente insostenible, sigue siendo importante examinar la distribución del logro previo a través de las unidades de análisis (*IES*, grupos de referencia, etc.) para comprender el nivel de la selección no aleatoria. Si la distribución de habilidades de los estudiante (capturadas por puntajes de pruebas previas) a través de las unidades de análisis es relativamente equitativa, entonces el análisis de inferencia causal se facilita, porque se asigna menos peso a la forma funcional de la ecuación de regresión si se ha cumplido el aspecto del supuesto *(f)* de “soporte común”. Para ilustrarlo, imagínese un escenario en el que a cada institución asiste un grupo de estudiantes con idénticas distribuciones de habilidades previas. En tal

caso, no es necesario ningún ajuste de regresión y bastaría una comparación de los resultados promedio.

1.3 Desafíos técnicos: sesgo

El sesgo es un aspecto clave al hacer una investigación de valor agregado. En particular, este se puede observar en las estimaciones de los parámetros debido a diversos factores. Esta sección se enfoca en tres tipos de sesgo particulares:

- Sesgo en los coeficientes de los puntajes previos, debidos al error de medición.
- Sesgo en las estimaciones del efecto escolar para IES con características contextuales favorables, efectos tipo B.
- Sesgo en las estimaciones del efecto escolar para IES con características contextuales desfavorables.

Las implicaciones de estos asuntos son importantes y merecen comprenderse cuando se trata de usar modelos de VA en la rendición de cuentas.

Un error de medición en una variable predictora es un problema bien conocido en la modelación estadística. En el contexto de los modelos de VA, Lockwood y McCaffrey (2014) analizan los problemas relevantes. En general, la preocupación es que el error de medición en la prueba previa generará una estimación atenuada, porque esa prueba puede producir un sesgo en las estimaciones del efecto escolar. En particular, podría preocupar que se sobreestimen los efectos escolares y que estos resulten altamente correlacionados, de manera artificial, con el resultado. Es posible realizar ajustes técnicos como SIMEX (Cook y Stefanski, 1994), pero su implementación en la educación superior podría ser desafiante, porque los puntajes de las pruebas previas pueden provenir de varias aplicaciones diferentes (las cuales presumiblemente tienen diferentes niveles de errores de medición).

En general, si hay una correlación positiva entre el efecto escolar y la capacidad promedio en la prueba previa, entonces las estimaciones tipo B estarán sesgadas negativamente para las instituciones con un promedio alto de habilidades al entrar (Raudenbush y Willms, 1995). Una correlación positiva podría ser razonable en este caso. Por ejemplo, estudiantes cuyo promedio de habilidades fuera alto, podrían estar dispuestos a pagar más por su matrícula, lo cual podría hacer que las instituciones invirtieran en mejores profesores, instalaciones, computadores, etcétera.

Es importante distinguir los efectos de la práctica escolar de los generados por los pares (compañeros), es decir, el aumento potencial del aprendizaje que los estudiantes pueden experimentar debido a sus compañeros. Esta diferencia la exploran Castellano, Rabe-Hesketh y Skrondal (*s.f.*). En síntesis, los autores demuestran que las estimaciones del efecto aleatorio de la institución estarán sesgadas si existe una correlación, pero no hay un efecto contextual. Por otra parte, los estimadores (*HT*) de Hausman-Taylor (Hausmany Taylor, 1981) son consistentes. No obstante, si hay un efecto contextual, entonces las estimaciones *HT* estarán sesgadas aunque sirven como límites para el verdadero efecto.

Finalmente, vale la pena hacer énfasis en que la mayor fuente potencial de problemas por sesgo proviene de subestimar sistemáticamente las mediciones del valor agregado para las instituciones o profesores que atienden a ciertas poblaciones, como los estudiantes de bajos ingresos. Esto podría llevar a una alta correlación entre las covariables y las medidas de *VA*, como se discutió en el contexto de errores de medición. La evidencia de que existe tal sesgo ha sido documentada en EE.UU. (Newton *et al.,* 2010; Ehlert, Koedel, Parsons y Podgursky, 2012). Existe la preocupación de que tal sesgo podría existir incluso si se han incluido controles para las características de los antecedentes de los estudiantes.

2. Aspectos metodológicos

2.1 Datos

Este informe se basa en los resultados de SABER PRO en 2012 y 2013. La tabla 1, muestra el número de estudiantes que presentaron las pruebas SABER 11° y SABER PRO en cada periodo de aplicación. Es importante observar que la información de estudiantes con puntajes en las pruebas SABER 11° solamente se utilizó entre 4 y 6 años antes del examen de SABER PRO. Para los estudiantes que presentaron SABER PRO en 2012 se tomó la información de SABER 11° entre 2006 y 2008. Para los que presentaron las pruebas SABER PRO en 2013, se tomó la información de SABER 11° entre 2007 y 2009. Esta restricción se sustenta, en parte, debido a cambios en la estructura de los exámenes SABER 11°, lo cual no posibilita la comparabilidad y podría tener consecuencias determinantes sobre los resultados. Cerca del 20% de quienes presentaron las pruebas SABER PRO en el segundo semestre de 2013, tenían puntajes de SABER 11° desde 2006. Por tanto, es importante tener en cuenta que, dado que no se consideraron los puntajes de SABER 11° antes de 2006, al menos un cuarto de quienes presentaron las pruebas SABER PRO en este periodo no están incluidos en el análisis.

Tabla 1. Estudiantes en cada combinación de las pruebas SABER 11° y SABER PRO

SABER PRO	SABER 11°							
	2006-1	2006-2	2007-1	2007-2	2008-1	2008-2	2009-1	2009-2
2012-1	1,135	9,318	1,704	7,966	583	765	0	0
2012-2	1,356	10,973	2,204	14,872	1,894	3,169	0	0
2013-1	0	0	1,405	11,139	1,975	8,749	522	860
2013-2	0	0	2,301	17,862	3,583	24,063	3,374	7,433

2.1.1 Relación entre las pruebas

Las correlaciones a nivel de estudiantes entre los puntajes de las pruebas que se consideraron en el análisis se muestran en la tabla 2. Las pruebas SABER PRO, por lo general, se correlacionan más estrechamente entre sí (sus correlaciones oscilan entre 0,5 y 0,66) que las de SABER 11° (cuyas correlaciones oscilan entre 0,4 y 0,52). Esto se debe en parte a que las pruebas SABER 11°, para el periodo de análisis, tiene niveles de confiabilidad menores que llevan a correlaciones más atenuadas.

Tabla 2. Correlaciones entre puntajes

	SABER 11°				SABER PRO		
	Lenguaje	Matemáticas	C. sociales	Química	LC	RC	CC
Lenguaje	139,198						
Matemáticas	0,44	139,116					
Ciencias Sociales	0,49	0,43	139,190				
Química	0,41	0,51	0,47	139,201			
Lectura crítica (LC)	0,49	0,40	0,52	0,42	139,202		
Razonamiento cuantitativo (RC)	0,42	0,58	0.46	0,51	0,51	139,205	
Competencias ciudadanas (CC)	0,47	0,40	0,53	0.41	0,66	0,50	139,199

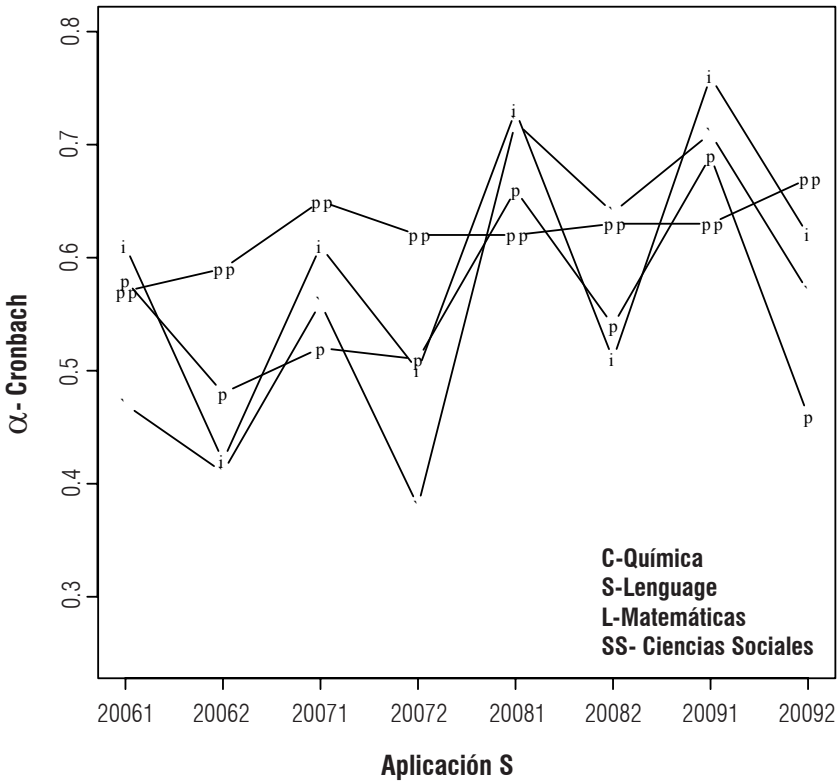
Las correlaciones más importantes para este estudio son entre las pruebas de SABER PRO y las de SABER 11°. Estas son relativamente bajas para los puntajes en las pruebas a nivel de estudiante, y oscilan entre 0,40 y 0,58. Si se incluyera solo el puntaje de una sola prueba previa para cada resultado de SABER PRO, probablemente se podría encontrar correlaciones más altas a nivel intraclases que los observados en estudios similares en otros contextos, debido a que el puntaje de la prueba previa explica poco la variabilidad a nivel de estudiante.

Como se mostrará en la sección 4, se incluirán los puntajes previos de las cuatro pruebas de SABER 11° para cada resultado de SABER PRO, por lo que niveles de confiabilidad relativamente bajos de las pruebas individuales de SABER 11° no deberían ser un problema.

2.1.2 Propiedades psicométricas de las medidas de SABER 11°

En la figura 1 se ilustra la magnitud de los errores de medición en las diferentes pruebas de SABER 11° que se tienen en cuenta para el estudio. Las confiabilidades oscilan generalmente entre 0,4 y 0,7 y hay una leve tendencia hacia el incremento de la confiabilidad con el tiempo. Las confiabilidades promedio son bastante similares en todas las pruebas: 0,56 en lenguaje y química, 0,6 en matemáticas y 0,62 en ciencias sociales.

Figura 1. Coeficiente de α- Cronbach para los puntajes previos de SABER 11°.



2.1.3 Índice de nivel socioeconómico (INSE)

En los datos utilizados para este análisis, el *INSE* tiene una media de 53 y una desviación estándar (*DE*) de 10. El *INSE* está débilmente correlacionado con los puntajes de las pruebas a nivel del estudiante, para los resultados de las tres aplicaciones de SABER PRO, las correlaciones son cercanas a 0,25. Cuando se toma el promedio de toda la institución por grupo de referencia (*IGR*), el *INSE* tiene una media de 52 y una desviación estándar de 6,7. Las correlaciones entre el promedio del *INSE* a nivel *IGR* y los resultados de las pruebas a nivel del estudiante (con correlaciones de 0,29), fueron ligeramente mayores que las correlaciones entre el *INSE* y el puntaje de las pruebas a nivel de estudiante.

2.1.4 Selección de los estudiantes

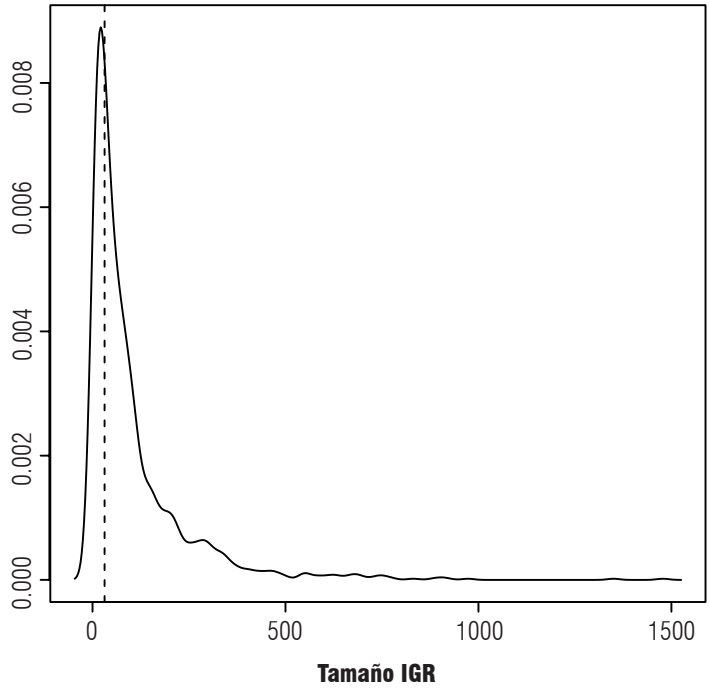
Las decisiones de los estudiantes sobre a cuál institución educativa asistir y cuál programa educativo cursar inciden sustancialmente en los efectos de la clasificación y en las diferencias en el número de estudiantes disponibles para ciertos análisis. La tabla 3 muestra el número

de estudiantes por grupo de referencia. En Ingeniería hay más de 30.000 estudiantes. Sin embargo, en la mayoría de los otros grupos de referencia la cantidad de estudiantes no se acerca a esta cifra. La figura 2 muestra la densidad estimada según el número de estudiantes por *IGR*. Se puede observar que la mayoría de las *IGR* tienen menos de 100 estudiantes (el número promedio de estudiantes por *IGR* es 55). De las 1347 *IGR*, 433 tienen menos de 30 estudiantes. En este informe se describen los resultados con base en la inclusión de todos los estudiantes (por ejemplo, estudiantes de *IGR* de menos de 35 estudiantes también se incluyen).

Tabla 3. Número de estudiantes por grupo de referencia.

Grupo de referencia	Número de estudiantes
INGENIERÍA	32582
ADMINISTRACIÓN Y AFINES	24663
EDUCACIÓN	12076
DERECHO	11990
CONTADURÍA Y AFINES	9150
SALUD	7526
COMUNICACIÓN, PERIODISMO Y PUBLICIDAD	6625
BELLAS ARTES Y DISEÑO	5371
PSICOLOGÍA	5196
CIENCIAS SOCIALES	4480
ENFERMERÍA	3835
MEDICINA	3108
ARQUITECTURA Y URBANISMO	2866
CIENCIAS NATURALES Y EXACTAS	2556
CIENCIAS AGROPECUARIAS	2392
HUMANIDADES	930
CIENCIAS MILITARES Y NAVALES	483

Figura 2. Densidad estimada de los tamaños de muestras para todas los IGR.



Otro aspecto para considerar es que la selección de estudiantes en las instituciones no es aleatoria. Con ese propósito, la tabla 4 compara las correlaciones intraclase *(CCI)* calculadas a partir de modelos de efectos aleatorios en donde la variable dependiente son los resultados de la prueba de SABER 11° que se especifican en cada columna y las otras variables se presentan por fila, representando el nivel de agrupación de cada modelo. La idea de esta tabla proviene de Hedges y Hedberg (2007), en la cual se calculan coeficientes similares para los colegios de primaria y de secundaria en EE.UU.

Tabla 4. Coeficiente de correlación intraclase (CCI), modelo vacío para diferentes niveles de agrupación.

Nivel de agrupación	Lenguaje	Matemáticas
Grupo de referencia (GR)	0,09	0,11
Programa	0,18	0,19
Institución (IES)	0,15	0,16
Institución - Grupo de referencia (IGR)	0,22	0,26
Institución - Programa	0,24	0,31

En EE.UU. existe aproximadamente un 22% de variabilidad en el logro académico entre colegios. En Colombia, aproximadamente el 10% de la variabilidad de los estudiantes se explica por la elección que hacen los estudiantes del grupo de referencia. Aproximadamente, el 16% de la variabilidad se explica por la elección de los estudiantes de la institución. Cuando se consideran ambos factores, la elección que hace el estudiante de su institución y el grupo de referencia, se aprecia un nivel de clasificación de habilidades más o menos comparable (o un poco más alto, dependiendo de la prueba) con respecto a lo que ocurre en los colegios de primaria y secundaria de EE.UU.

2.2 Métodos

En este estudio se definió que el tratamiento corresponde al estudiante i que ha cursado al menos el 75% de los créditos del programa académico (siendo así elegible para presentar las pruebas de SABER PRO) en el grupo k de referencia dentro de la universidad j. Es crucial el hecho de que los modelos se estiman dentro de los grupos de referencia. Esto ayuda a garantizar que el tratamiento sea de una sola forma (en un intento por acercarnos más al cumplimiento de la hipótesis del SUTVA) y que el enfoque se centre en la comparación que se considera más relevante para la toma de decisiones políticas y para los estudiantes: *de qué manera un grupo de referencia se compara entre universidades.*

Como consecuencia de este enfoque, los resultados no son comparables entre diferentes grupos de referencia. No se deben comparar las estimaciones de VA del grupo de referencia x y y dentro de la universidad z, porque los tratamientos no son comparables. Se reportarán los resultados de tres modelos, los cuales se han denominado en los reportes institucionales medidas de Aporte Relativo. La forma general del modelo de valor agregado se presenta a continuación, para el estudiante i en el colegio j:

$$y_{ij} \text{ Normal } [\beta' X_{ij} + \gamma_{j,} \sigma^2] \qquad (1)$$

Donde y_{ij} es el resultado de la medición para el individuo i y X_{ij} es un vector de covariables que puede o no depender de i (por ejemplo, puede ser una media a nivel del grupo). El signo γ_j son los efectos de interés. Se asume además que $\gamma_j \sim$ Normal [0; σ^2] y se usan técnicas de estimación de efectos aleatorios (específicamente REML vía Bates, Maechle y Bolker, 2013). Para el modelo 1, X_{ij} representa únicamente un vector de los puntajes obtenidos por el estudiante en las pruebas previas (matemáticas, lenguaje, química y ciencias sociales de SABER 11°).

El modelo es:

$$\text{M1}: y_{ij} \sim \text{N}[\beta_0 + \beta_1 \text{mat}_{ij} + \beta_2 \text{leng}_{ij} + \beta_3 \text{quím}_{ij} + \beta_4 \text{cien soci}_{ij} + \gamma_{j,} \sigma^2] \qquad (2)$$

Este sería un modelo tipo *A* en la tipología de Raudenbush y Willms (1995), porque no están especificados efectos del contexto a nivel *IGR* (covariables a nivel del promedio de los estudiantes del *IGR*). También se hicieron estimaciones de dos modelos tipo *B*. En el modelo 2, X_{ij} contiene el puntaje promedio del *INSE* a nivel de *IGR* además de los puntajes a nivel del estudiante en SABER 11°:

$$\text{M2}: y_{ij} \sim \text{N}[\beta_0 + \beta_1 \text{mat}_{ij} + \beta_2 \text{ leng}_{ij} + \beta_3 \text{quím}_{ij} + \beta_4 \text{cien soc}_{ij} + \beta_5 \text{INSE}_j + \gamma_j, \sigma^2] \quad (3)$$

En el modelo 3, X_{ij} no incluye el promedio del *INSE*, pero se agrega el promedio del puntaje de SABER 11° en el *IGR*, considerando la prueba de matemáticas como predictor para la prueba de razonamiento cuantitativo (*RC*) y lenguaje como predictor de lectura crítica (*LC*):

$$\text{M3}: y_{ij} \sim \text{N}[\beta_0 + \beta_1 \text{mat}_{ij} + \beta_2 \text{ leng}_{ij} + \beta_3 \text{quím}_{ij} + \beta_4 \text{cien soc}_{ij} + \beta_5 \text{SB11}_j + \gamma_j, \sigma^2] \quad (4)$$

Los modelos 2 y 3 no están anidados (por ejemplo, el promedio del *INSE* no se incluye en el modelo 3). Los ajustes en los modelos 2 y 3 buscan controlar los efectos de pares, es decir, los que se atribuyen a los compañeros de clases, en la producción de las estimaciones del *VA*. Es importante tener presente la advertencia de Raudenbush (2004) sobre los efectos de tipo *B*: la práctica de las instituciones no está definida u observada, así que es difícil estimar su efecto. Aún más importante, con base en Raudenbush y Willms (1995), las estimaciones tipo *B* estarán sesgadas para las instituciones con contextos favorables si hay una correlación positiva entre práctica y el contexto, aproximándose a este último a través del promedio del *INSE* o los puntajes de SABER 11°. Se considera que tal correlación positiva es un supuesto razonable para las *IES*.

Un aspecto importante es la proporción de la varianza total en los resultados del estudiante que se explica por la varianza de las *IGR* dentro de un modelo, conocido como el coeficiente de correlación intraclase (*CCI*):

$$\frac{\sigma_\gamma^2}{\sigma_\gamma^2 + \sigma^2} \quad (5)$$

Este coeficiente es proporción de la varianza total explicada por la varianza residual de la *IGR*. Usando esta medida, se pueden comparar los hallazgos con otros trabajos sobre los efectos de la institución (Hedges y Hedberg, 2007; Steedle, 2012), determinando esencialmente cuánta variabilidad hay en la efectividad de las IES en Colombia con relación a otros contextos, como la escolaridad en primaria o la educación superior en EE.UU.

También se consideran las correlaciones entre las estimación de *VA* y los promedios de las *IGR* en SABER 11° y en SABER PRO. Estas correlaciones son útiles porque describen en qué medida las clasificaciones basadas en estimaciones del *VA* podrían diferir de clasificaciones basadas únicamente en el promedio de SABER PRO sin ajustar los puntajes (la cual es simplemente una medida del estatus, no del cambio).

3. Resultados

Esta sección recoge los principales resultados de los análisis de valor agregado descritos anteriormente, con los efectos estimados para todos los grupos de referencia y para las pruebas de razonamiento cuantitativo (RC) y lectura crítica (LC) de SABER PRO. En el anexo, se relacionan las tablas que contienen la información de las estimaciones de los modelos que se presentan en este capítulo.

3.1 Modelos estimados: efectos fijos y componentes de varianza

Las tablas A1 a A6 muestran los efectos fijos estimados para los grupos de referencia que tienen al menos cinco (5) instituciones por grupo de referencia (IGR). En primer lugar, las tablas A1 y A2 muestran las estimaciones para el modelo 1 (ecuación 2). Para la prueba de LC, los puntajes previos en lenguaje, química y ciencias sociales fueron predictores importantes comparados con el puntaje de matemáticas (coeficientes estimados promedio entre todos los grupos de referencia: 0,028, 0,020 y 0,030 comparados con 0,009, respectivamente). En RC disminuye la importancia de los puntajes previos en lenguaje y ciencias sociales (coeficiente estimados promedio de 0,013 y 0,018 comparados con 0,027 y 0,025 en matemáticas y química) lo cual tiene sentido dadas las correlaciones presentadas en la tabla 2.

En segundo lugar, las tablas A3 y A4 muestran las estimaciones para el modelo 2 (ecuación 3). El comportamiento de los coeficientes estimados para los puntajes de SABER 11° es similar al observado en el modelo 1. Adicionalmente, se observa que el promedio del $INSE$ es un predictor más fuerte para la prueba de LC que para RC (coeficiente estimado promedio de 0,010 comparado con 0,008).

Finalmente, las tablas A5 y A6 presentan los resultados del modelo 3 (ecuación 4). En estas se observa un cambio leve en la magnitud de los coeficientes para los puntajes individuales en SABER 11°. Tanto para LC como para RC, el puntaje promedio en SABER 11° fue un predictor más fuerte que el promedio del $INSE$. Este efecto fue mayor para LC que para RC (coeficiente estimado promedio de 0,44 en LC comparado con 0,30 en RC).

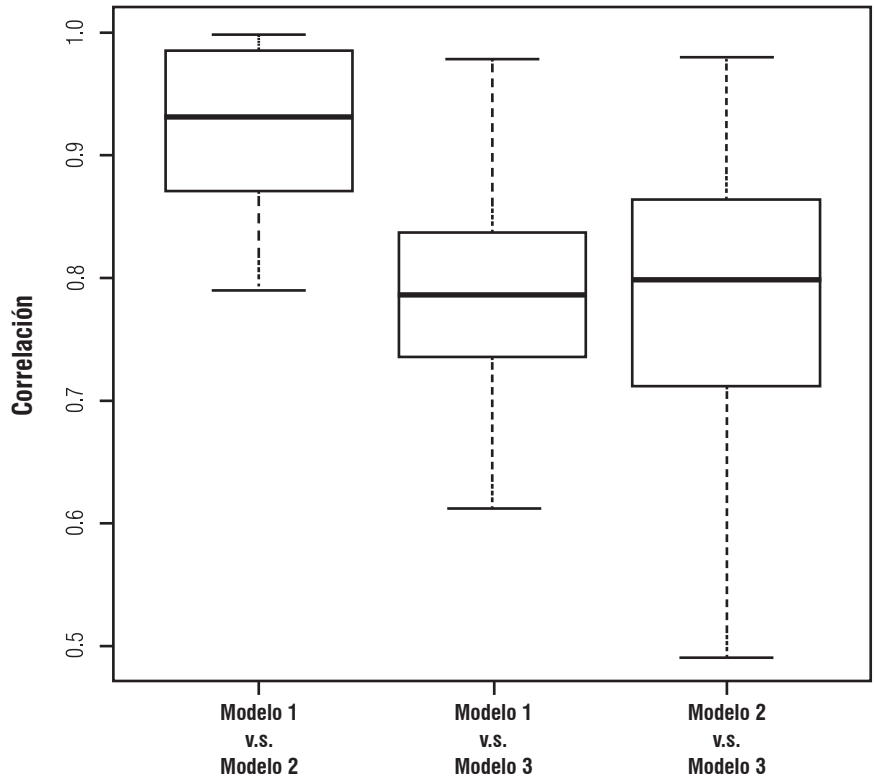
Las tablas A7 y A8 muestran las estimaciones para los componentes de la varianza y los *coeficientes de correlación intraclase* (CCI) para todos los modelos. En términos generales, los CCI del modelo 3 son más pequeños que los observados en modelos 1 y 2. Para LC, el promedio del CCI en el modelo 3 fue 0,026 comparado con 0,057 y 0,048 en los modelos 1

y 2, respectivamente (*véase,* tabla A7). En RC los CCI fueron más altos, aunque se observa el mismo patrón (0,069, 0,061 y 0,040, respectivamente). Esto se asemeja a los hallazgos de otros estudios. Por ejemplo, Briggs and Domingue (2011) encontraron más variabilidad en la efectividad a nivel del profesor en matemáticas que en lectura. En este contexto, este hallazgo podría tomarse como una evidencia de que las IGR influyen más en la capacidad del estudiante en RC que en LC. Sin embargo, esta afirmación podría requerir mayores evidencias, puesto que son posibles otras interpretaciones (por ejemplo, podría atribuirse a diferencias en los tipos de ítems que se emplean en los dos exámenes y en la capacidad de respuesta de acuerdo con la madurez del estudiante).

3.2 Relaciones entre modelos

La figura 3 muestra las correlaciones de Spearman entre las estimaciones de VA de los tres diferentes modelos. Los modelos 1 y 2, por lo general, están altamente correlacionados (de las 34 correlaciones, 33 están por encima de 0,8).Por otra parte, los resultados del modelo 3 son un poco diferentes a los de los modelos 1 y 2. Por ejemplo, la mayoría de las correlaciones entre los coeficientes de estos últimos con los del modelo 3 están por debajo de 0,8.

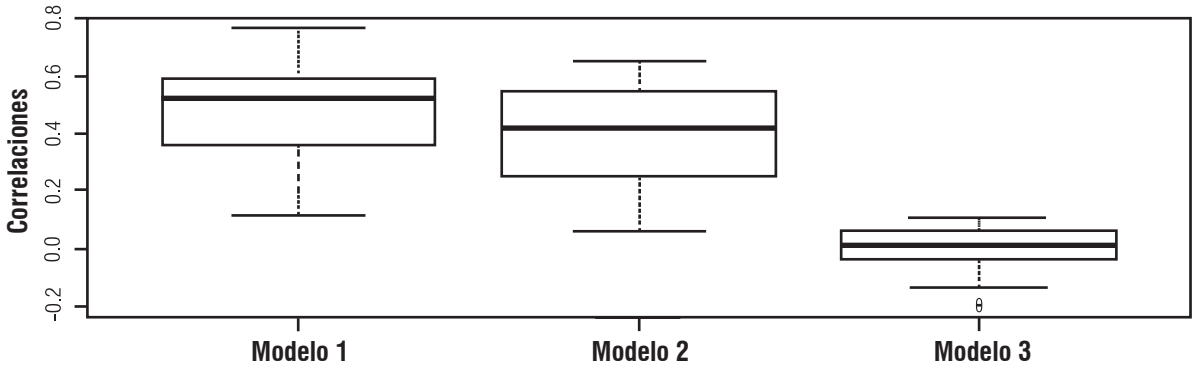
Figura 3. Correlaciones de Spearman entre modelos (tomando todos los GR).



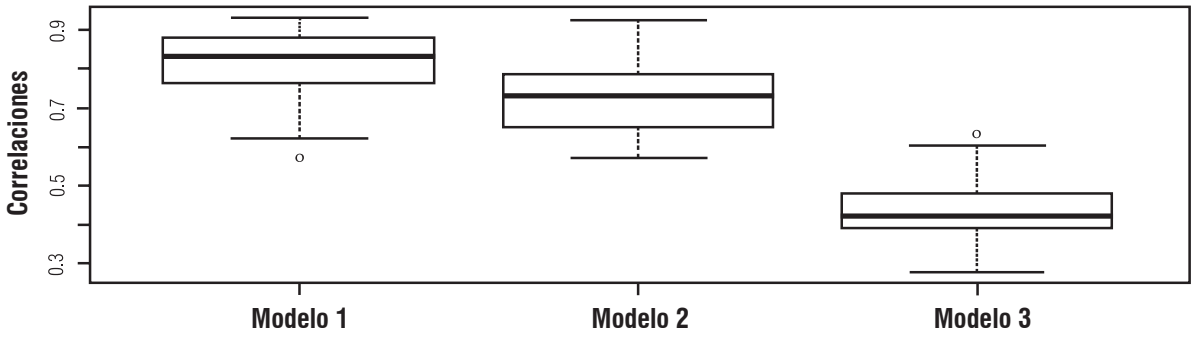
La figura 4 muestra las correlaciones de Spearman entre las estimaciones de *VA* para cada modelo y las tres covariables: el promedio de SABER 11° (prueba usada como el promedio del puntaje previo para cada resultado en el modelo 3), el promedio de SABER PRO y el promedio del *INSE*. En el promedio SABER 11° se observa una disminución, desde modelo 1 hasta el 3, en la correlación de las estimaciones del *VA* y el promedio de la habilidad del estudiante a la entrada. Las estimaciones de modelo 3 no están correlacionadas con los puntajes de SABER 11°, pero esto se debe a la inclusión del promedio de SABER 11° como un predictor en el modelo.

Figura 4. Correlaciones entre modelos y covariables (tomando de todos los GR).

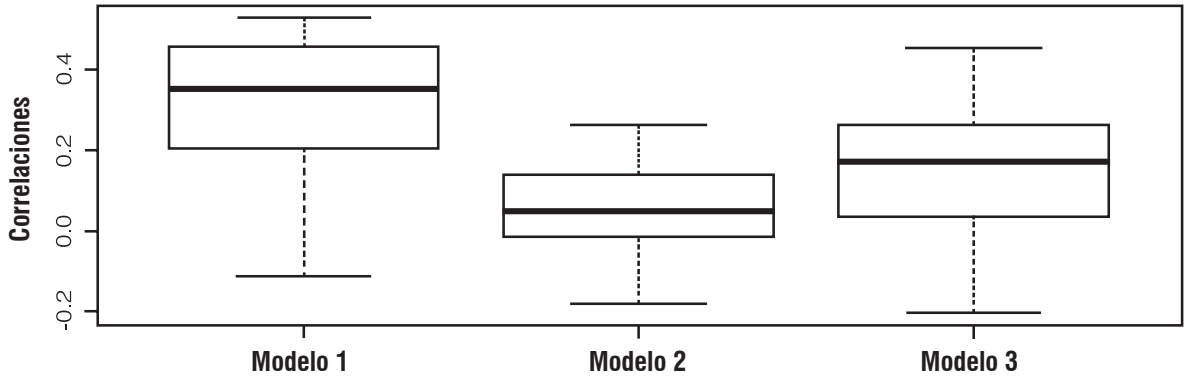
a. Correlación con promedio SABER 11°



b. Correlación con promedio SABER PRO



c. Correlación con promedio del INSE



Las correlaciones con SABER PRO son importantes debido a que, para efectos de rendición de cuentas, los rankings del puntaje promedio en esta prueba son una alternativa natural a las estimaciones de VA. Las estimaciones del modelo 1, comúnmente se encuentran altamente correlacionadas con el promedio de SABER PRO (correlación promedio de 0,8). Las estimaciones del modelo 2 presentan una correlación ligeramente más baja con SABER PRO (correlación promedio alrededor de 0,7), mientras que las estimaciones del modelo 3 tienen una correlación promedio inferior a 0,5. Las estimaciones de VA son relativamente independientes del promedio del $INSE$, especialmente para el modelo 2 (los cuales son, por diseño, cercanas a cero).

4. Análisis de sensibilidad

En esta sección se plantean preguntas específicas sobre el análisis de sensibilidad de las estimaciones según ciertos supuestos. Para entender este análisis, se expondrá un ejemplo usando los resultados de la prueba de razonamiento crítico (RC) y el grupo de referencia "CONTADURÍA Y AFINES", ya que este tiene un tamaño de muestra moderado (pero no extremo).

4.1 Efectos heterogéneos del tratamiento

La primera pregunta es: ¿qué tan sensibles son las estimaciones el supuesto de que el efecto del tratamiento es homogéneo? Es posible atenuar este supuesto incluyendo un coeficiente aleatorio en la prueba de matemáticas a nivel del estudiante (obsérvese que es el mayor coeficiente estimado en la tabla A2). Entonces, el modelo 1 se transforma en:

$$y_{ij} \sim \mathrm{N}[\beta_0+(\beta_1 + \delta j)\mathrm{mat}_{ij} + \beta_2\mathrm{leng}_{ij} + \beta_3\,\mathrm{quím}_{ij} + \beta_4\,\mathrm{cien\ soc}_{ij} + \gamma_j, \sigma^2] \qquad (6)$$

Haciendo énfasis en la inclusión del término δ_j[8], esto permite que el efecto de la prueba de matemáticas en SABER 11° sobre RC en SABER PRO varíe por colegio. Si se piensa en un colegio cuyos estudiantes tienen altas habilidades al ingresar a la educación superior, en ese colegio se espera que j sea positiva, es decir, se esperaría que estudiantes con los más altos puntajes de SABER 11° se desempeñen aún mejor en SABER PRO que los estudiantes de otras instituciones.

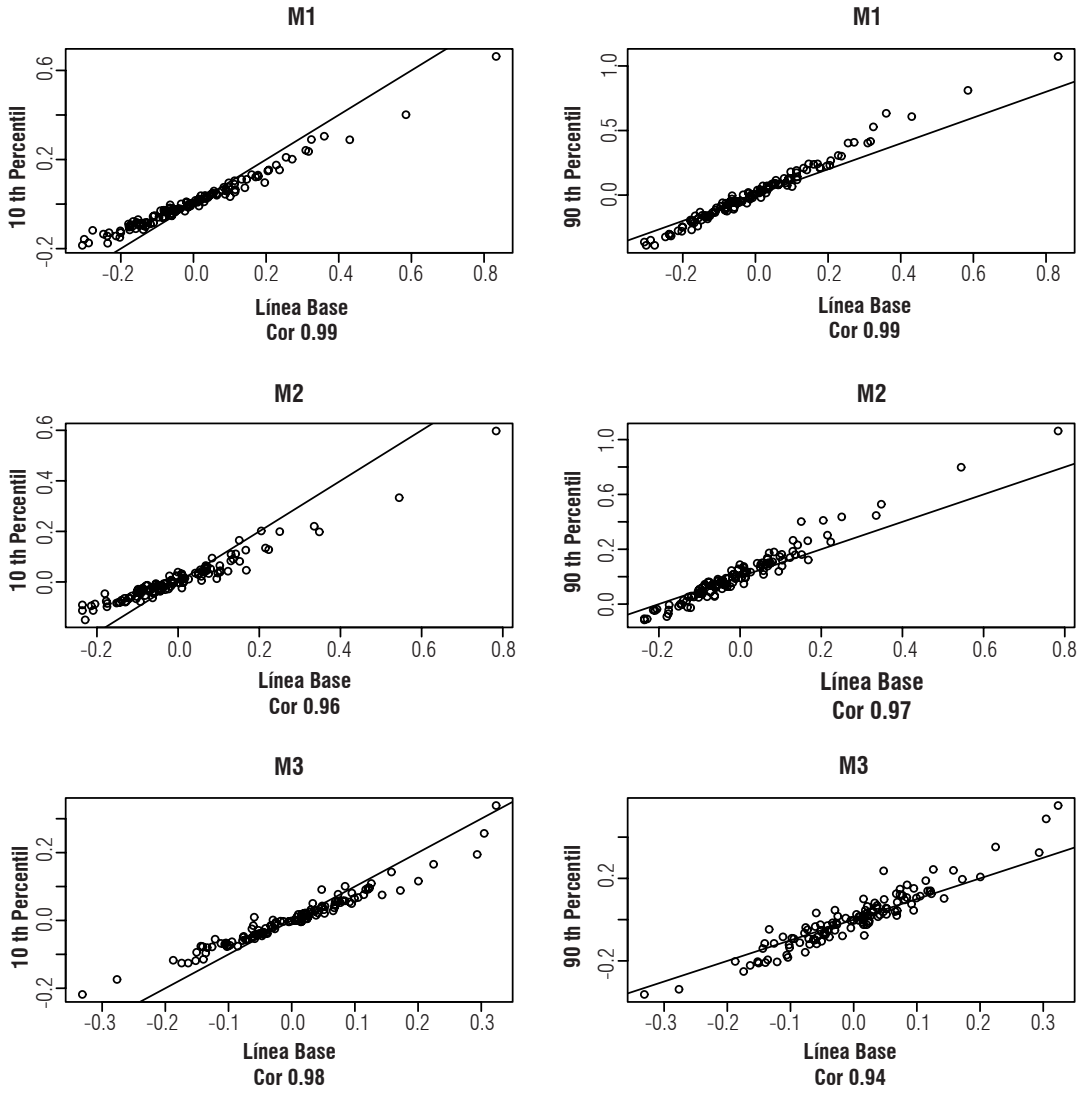
Modelos similares a la ecuación 6 se estimaron en los tres modelos. En primer lugar, se observa que la varianza de los coeficientes aleatorios (el término j) fue pequeña. Para el modelo 1, σ^2 es de 0,000042. En segundo lugar, los coeficientes aleatorios presentaron una fuerte correlación negativa con el intercepto aleatorio ($\sigma\gamma\delta$ mayor a 0,85 para los tres modelos).

Las estimaciones individuales relativas a la línea de base, mostradas en la sección 3 se presentan en la figura 5. El eje x siempre presenta las estimaciones de la sección 3, mientras que el eje y tiene dos diferentes tipos de estimaciones.

8 Asumiendo, además, que $\begin{bmatrix} \gamma j \\ \delta j \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\gamma^2, \sigma_\gamma, \delta \\ \sigma_\gamma, \delta, \sigma_\delta^2 \end{bmatrix}\right)$

A la izquierda se muestran los calculados para un estudiante en el percentil 10 de los puntajes de matemáticas en SABER 11° que asistió a esa IGR[9]; a la derecha, las estimaciones para un estudiante en el percentil 90. Las correlaciones de Spearman se presentan debajo de cada gráfico. En general, se observan ordenamientos similares según cada enfoque. Del mismo modo, al considerar el ordenamiento para los percentiles 10 y 90, se observan que están altamente correlacionados (superiores a 0,94 para todos los modelos).

Figura 5. Líneas base estimadas comparadas con los efectos obtenidos cuando se permite que varíe como una función de las habilidades previas en matemáticas.



9 Esto se calcula aplicando: $\gamma_j + \delta_j Q_j^{0.1}$ donde Q_j es el percentil 10 de los puntajes estudiantiles en matemáticas en SABER 11 para la institución j.

4.2 Términos de orden superior

En los modelos de *VA* a veces se incluyen términos de órdenes superiores (por ejemplo, términos cuadráticos para los puntajes de pruebas previas). Proceder o no de esta forma es una decisión difícil, como puede verse en Lockwood y McCaffrey (*s.f.*). En síntesis, si la verdadera relación es lineal, la inclusión de términos de un orden superior no sustituye las correcciones de errores de medición y probablemente, resulte una solución menos deseable a la de incluir múltiples puntajes previos.

Si la verdadera relación no es lineal, entonces el problema es más complicado. Dada la falta de alineamiento entre los puntajes en este contexto, el hecho de asumir relaciones no lineales y la necesidad de múltiples puntajes previos podría ser un supuesto exigente. Sin embargo, para considerar el impacto potencial de incluir términos de un orden superior, los modelos se estimaron de tal manera que incluyeran términos lineales y términos cuadráticos para cada uno de las pruebas de SABER 11°. Las correlaciones entre las estimaciones de *VA* y los de los modelos de la línea de base fueron todas mayores o iguales a 0,99 y, de esta manera, los modelos fueron bastante sólidos a la inclusión de los términos cuadráticos adicionales.

4.3 Diferentes efectos contextuales

Este análisis se enfocó en incluir efectos contextuales para la habilidad previa. En este punto, interesa estudiar la consistencia de los resultados cuando se incluyen, a nivel de cada *IGR,* el promedio de los cuatro puntajes de SABER 11° para cada estudiante o si se agregan estos promedios para todos los estudiantes.

Las motivaciones subyacentes que se producen de la observación que resulta pueden ser bastante sensibles a la especificación del término del nivel superior. Dado que la especificación inicial fue de alguna forma *adhoc*, hay motivos para preguntarse si especificaciones alternativas podrían producir resultados dramáticamente diferentes. Estas modificaciones en la especificación no llevan a grandes cambios. Por ejemplo, respecto a la línea de base, el modelo 3 está correlacionado con el modelo 1 y 2 en 0,75 y 0,61.

La especificación alternativa que introduce el promedio de la prueba tiene una correlación de 0,65, 0,56 y 0,92 con los modelos 1, 2 y 3, respectivamente. Cuando se incluye el promedio para el puntaje de cada prueba a nivel de institución, las estimaciones del *VA* tienen una correlación de 0,66, 0,57 y 0,90. Las dos alternativas tienen una correlación de 0,99. La mayor desviación podría ser la correlación 0,56 entre el modelo 2 y la alternativa que incluye el promedio de todos los puntajes previos. Este es un descenso de menos de 0,1 respecto a la correlación entre el modelo 2 y 3, el cual es relativamente pequeño.

5. Consideraciones finales

En muchos sentidos, los resultados que se muestran en este documento son similares a lo que se han mostrado en trabajos previos. Un punto importante es que las clasificaciones generalmente son sensibles a la inclusión del promedio de SABER 11°, pero, en general, los resultados son bastante sólidos para las especificaciones alternativas exploradas en la sección 4. Esto es alentador, puesto que la sensibilidad a especificaciones alternativas podría complicar la elección de un modelo. Un hallazgo que quizá merece un estudio posterior es que las pruebas de razonamiento cuantitativo indican más variabilidad en la efectividad de la *IGR* que las pruebas de lectura crítica.

Hay diversos escenarios posibles consistentes con este hallazgo. Uno de estos sería que las *IGR* son más efectivas diferencialmente en la enseñanza de razonamiento cuantitativo que en lectura crítica. Pero se necesitaría más investigación para sustentar esta hipótesis.

Otra conclusión importante se muestra en la figura 2. La línea vertical en esa figura está en el umbral de 30 estudiantes considerados en el trabajo previo. La mayoría de las *IGR* no cumplen ese umbral. Aunque ciertamente hay un motivo de preocupación en cuanto al mayor "error" en la estimación de *VA* para *IGR* más pequeñas, también es importante pensar cuántas *IGR* no obtendrán estimaciones de *VA* con base en el tamaño de su muestra. En lugar de imponer un umbral de 30 estudiantes en forma relativamente aleatoria, valdría la pena establecer otros criterios que permitan la estimación de estos modelos para un número mayor de *IGR*.

Referencias

- **Ballou, D.** (2009). Test scaling and value-added measurement. Education Finance and Policy,4(4), 351-383. Retrieved from *http://www.mitpressjournals.org/doi/abs/10.1162/edfp.2009.4.4.351*

- **Ballou, D., Sanders, W., & Wright, P.** (2004). Controlling for student background in valueadded assessment of teachers. *Journal of Educational and Behavioral Statistics*, 29(1), 37–65.

- **Bates, D., Maechler, M., & Bolker, B.** (2013). lme4: Linear mixed-effects models using s4 classes [Computer software manual]. Retrieved from http://CRAN.R-project.org/package=lme4 (R package version 0.999999-2)

- **Briggs, D., & Domingue, B.** (2011). Due diligence and the evaluation of teachers: A review of the value-added analysis underlying the effectiveness rankings of Los Angeles unified school district teachers by the" Los Angeles Times". National Education Policy Center.

- **Castellano, K., Rabe-Hesketh, S., & Skrondal, A.** (s.f.).Composition, context, and endogeneity in school and teacher comparisons.

- **Chetty, R., Friedman, J. N., & Rockoff, J. E.** (2011).The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood (*Working Paper* No. 17699). Cambridge, MA: National Bureau of Economic Research.

- **Cook, J., &Stefanski, L.** (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, 89, 1314–1328.

- **Ehlert, M., Koedel, C., Parsons, E., & Podgursky, M.** (2012, August). Selecting growth measures for school and teacher evaluations (*Working Paper* No. 80). Washington, D.C.: CALDER.

- **Ewell, P. T.** (2012). A world of assessment: Oecd'sahelo initiative. Change: *The Magazine of Higher Learning*, 44(5), 35–42.

- **Fisher V. University of Texas at Austin** (Vol. 631) (No. No. 09-50822). (2011). Court of Appeals, 5th Circuit.

- **Goldhaber, D., & Chaplin, D.** (2012, January).Assessing the “Rothstein falsification test”: Does it really show teacher value-added models are biased (*Working Paper* No. 2012-1.3). Seattle, WA: Center for Education Data and Research

- **Hausman, J. A., & Taylor, W. E.** (1981).Panel data and unobservable individual effects. Econometrica: *Journal of the Econometric Society*, 1377–1398.

- **Hedges, L. V., & Hedberg, E. C.** (2007).Intraclass correlation values for planning group randomized trials in education. Educational Evaluation and Policy Analysis, 29(1),

- **Ho, D. E., & Rubin, D. B.** (2011). Credible causal inference for empirical legal studies. Annual Review of Law and Social Science, 7, 17–40.

- **Holland, P. W.** (1986, December). Statistics and causal inference.*Journal of the American Statistical Association,*81, 945-960.

- **Imai, K., King, G., & Stuart, E. A.** (2008).Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society*: series A (statistics in society), 171(2), 481–502.

- **Kane, T. J., & Staiger, D. O.** (2008, December).Estimating teacher impacts on student achievement: An experimental evaluation (Working Paper No. 14607). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w14607

- **Koedel, C., & Betts, J. R.** (2011). Does student sorting invalidate value-added models of teacher effectiveness? an extended analysis of the rothstein critique. *Education*,6(1), 18–42

- **Leckie, G., & Goldstein, H.** (2009). The limitations of using school league tables to inform school choice. *Journal of the Royal Statistical Society*, A, 172, 835–851.

- **Lockwood, J., & McCaffrey, D.** (s.f.). Should nonlinear functions of test scores be used as covariates in a regression model? In R.

- **Lissitz** (Ed.), Value-added modeling and growth modeling with particular application to teacher and school effectiveness. Charlotte, NC: Information Age Publishing.

- **Lockwood, J., & McCaffrey, D.** (2014). Correcting for test score measurement error in ancova models for estimating treatment effects. *Journal of Educational and Behavioral Statistics*,39(1), 22-52. Retrieved from http://jeb.sagepub.com/content/39/1/22.abstract

- **Manzi, J., San Martin, E., & Van Bellegem, S.** (2014). School system evaluation by value added analysis under endogeneity. Psychometrika, 79(1), 130-153. Retrieved from *http://dx.doi.org/10.1007/s11336-013-9338-0*

- **Newton, X. A., Darling-Hammond, L., Haertel, E., & Thomas, E.** (2010). Value-added modeling of teacher effectiveness: An exploration of stability across models and contexts. *Education Policy Analysis Archives*, 18(23).

- **Raudenbush, S.** (2004). What are value-added models estimating and what does this imply for statistical practice? *Journal of Educational and Behavioral Statistics*, 29, 121–130.

- **Raudenbush, S., &Willms, J.** (1995).The estimation of school effects. *Journal of Educational and Behavioral Statistics*, 20, 307–335.

- **Reardon, S. F., & Raudenbush, S. W.** (2009).Assumptions of value-added models for estimating school effects. *Education*, 4(4), 492–519.

- **Rothstein, J.** (2010). Teacher quality in educational production: Tracking, decay, and student achievement*. *Quarterly Journal of Economics*, 125(1), 175-214. Retrieved from *http://www.mitpressjournals.org/doi/abs/10.1162/qjec.2010.125.1.175*

- **Rubin, D. B., Stuart, E. A., & Zanutto, E. L.** (2004). A potential outcomes view of value added assessment in education. *Journal of Educational and Behavioral Statistics*,29, 103–116.

- **Sacerdote, B.** (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far? *Handbook of the Economics of Education*, 3, 249–

- **Shavelson, R.** (2010). Measuring college learning responsibly: Accountability in a new era. Stanford University Press.

- **Spence, M.** (1973).Job market signaling.The Quarterly Journal of Economics,87(3), 355–374.

- **Steedle, J. T.** (2012). Selecting value-added models for postsecondary institutional assessment.Assessment & Evaluation in Higher Education, 37(6), 637–652.

- **Wainer, H.** (2004).Introduction to a special issue of the journal of educational and behavioral statistics on value-added assessment.Journal of Educational and Behavioral Statistics, 29(1), 1–3.

- **Winston, G., & Zimmerman, D.** (2004). Peer effects in Higher Education. In College Choices: The economics of where to go, when to go, and how to pay for it(pp. 395–424). University of Chicago Press.

Anexo

Tabla A1. Coeficientes estimados para lectura crítica, Modelo 1

Grupo de referencia	Intercepto	Matemáticas	Lenguaje	Química	C.Sociales
ADMINISTRACIÓN Y AFINES	5,5484	0,0096	0,0284	0,0227	0,0313
ARQUITECTURA Y URBANISMO	5,3258	0,0098	0,0312	0,0236	0,0324
BELLAS ARTES Y DISEÑO	5,6844	0,0086	0,0300	0,0218	0,0309
CIENCIAS AGROPECUARIAS	5,7337	0,0081	0,0276	0,0217	0,0303
CIENCIAS NATURALES Y EXACTAS	6,3056	0,0056	0,0273	0,0191	0,0273
CIENCIAS SOCIALES	6,4126	0,0090	0,0217	0,0186	0,0318
COMUNICACIÓN, PERIODISMO Y PUBLICIDAD	5,9508	0,0116	0,0239	0,0206	0,0335
CONTADURÍA Y AFINES	5,3398	0,0110	0,0322	0,0211	0,0308
DERECHO	6,0539	0,0065	0,0304	0,0181	0,0320
ECONOMÍA	6,2163	0,0110	0,0275	0,0182	0,0270
EDUCACIÓN	5,2511	0,0080	0,0341	0,0213	0,0352
ENFERMERÍA	5,9232	0,0083	0,0286	0,0205	0,0282
HUMANIDADES	6,5200	0,0093	0,0266	0,0147	0,0303
INGENIERÍA	5,9083	0,0089	0,0269	0,0185	0,0302
MEDICINA	6,7782	0,0091	0,0253	0,0165	0,0223
PSICOLOGÍA	5,9994	0,0074	0,0302	0,0196	0,0316
SALUD	5,8493	0,0071	0,0286	0,0203	0,0300

Tabla A2. Coeficientes estimados para R, Modelo 1

Grupo de referencia	Intercepto	Matemáticas	Lenguaje	Química	C.Sociales
ADMINISTRACIÓN Y AFINES	5,6842	0,0286	0,0155	0,0270	0,0192
ARQUITECTURA Y URBANISMO	5,7460	0,0269	0,0144	0,0273	0,0192
BELLAS ARTES Y DISEÑO	5,8412	0,0303	0,0079	0,0278	0,0167
CIENCIAS AGROPECUARIAS	5,9998	0,0284	0,0151	0,0240	0,0185
CIENCIAS NATURALES Y EXACTAS	6,0286	0,0338	0,0118	0,0259	0,0165
CIENCIAS SOCIALES	5,8356	0,0232	0,0123	0,0253	0,0207
COMUNICACIÓN, PERIODISMO Y PUBLICIDAD	5,8839	0,0241	0,0137	0,0219	0,0200
CONTADURÍA Y AFINES	5,9126	0,0265	0,0190	0,0249	0,0162
DERECHO	6,0095	0,0239	0,0122	0,0261	0,0174
ECONOMÍA	5,8763	0,0343	0,0143	0,0228	0,0215
EDUCACIÓN	6,1869	0,0222	0,0118	0,0255	0,0135
ENFERMERÍA	6,6576	0,0183	0,0123	0,0241	0,0120
HUMANIDADES	5,4283	0,0287	0,0132	0,0289	0,0183
INGENIERÍA	5,8140	0,0326	0,0145	0,0270	0,0190
MEDICINA	5,9642	0,0358	0,0083	0,0270	0,0167
PSICOLOGÍA	6,0816	0,0192	0,0159	0,0242	0,0167
SALUD	6,4888	0,0203	0,0110	0,0226	0,0155

Tabla A3. Coeficientes estimados para lectura crítica, Modelo 2

Grupo de referencia	Intercepto	Matemáticas	Lenguaje	Química	C.Sociales	Promedio INSE
ADMINISTRACIÓN Y AFINES	5,0633	0,0095	0,0283	0,0227	0,0312	0,0097
ARQUITECTURA Y URBANISMO	4,8709	0,0095	0,0309	0,0234	0,0322	0,0090
BELLAS ARTES Y DISEÑO	5,8355	0,0086	0,0300	0,0218	0,0309	-0,0028
CIENCIAS AGROPECUARIAS	5,1031	0,0079	0,0273	0,0220	0,0301	0,0134
CIENCIAS NATURALES Y EXACTAS	5,6297	0,0052	0,0270	0,0191	0,0271	0,0142
CIENCIAS SOCIALES	5,7434	0,0088	0,0215	0,0185	0,0317	0,0137
COMUNICACIÓN, PERIODISMO Y PUBLICIDAD	5,8979	0,0116	0,0239	0,0206	0,0334	0,0010
CONTADURÍA Y AFINES	4,4904	0,0109	0,0322	0,0212	0,0307	0,0176
DERECHO	5,6410	0,0065	0,0303	0,0181	0,0320	0,0078
ECONOMÍA	5,8302	0,0107	0,0271	0,0180	0,0269	0,0081
EDUCACIÓN	4,0739	0,0081	0,0340	0,0211	0,0353	0,0257
ENFERMERÍA	5,7114	0,0083	0,0286	0,0205	0,0282	0,0043
HUMANIDADES	5,9705	0,0091	0,0262	0,0145	0,0298	0,0120
INGENIERÍA	5,3874	0,0088	0,0268	0,0185	0,0302	0,0102
MEDICINA	6,1224	0,0092	0,0251	0,0165	0,0223	0,0113
PSICOLOGÍA	5,3228	0,0073	0,0301	0,0193	0,0315	0,0132
SALUD	5,3626	0,0072	0,0286	0,0203	0,0299	0,0093

Tabla A4. Coeficientes estimados para razonamiento cuantitativo, Modelo 2

Grupo de referencia	Intercepto	Matemáticas	Lenguaje	Química	C.Sociales	Promedio INSE
ADMINISTRACIÓN Y AFINES	5,0002	0,0285	0,0154	0,0270	0,0192	0,0134
ARQUITECTURA Y URBANISMO	5,8817	0,0269	0,0145	0,0273	0,0192	-0,0025
BELLAS ARTES Y DISEÑO	5,8695	0,0303	0,0079	0,0278	0,0167	-0,0005
CIENCIAS AGROPECUARIAS	5,3775	0,0282	0,0148	0,0243	0,0184	0,0133
CIENCIAS NATURALES Y EXACTAS	5.1309	0.0336	0.0114	0.0260	0.0164	0.0182
CIENCIAS SOCIALES	5,3202	0,0230	0,0118	0,0250	0,0203	0,0114
COMUNICACIÓN, PERIODISMO Y PUBLICIDAD	5,8006	0,0241	0,0137	0,0219	0,0200	0,0015
CONTADURÍA Y AFINES	4,8960	0,0265	0,0189	0,0251	0,0161	0,0210
DERECHO	5,4493	0,0238	0,0120	0,0260	0,0173	0,0107
ECONOMÍA	5,0321	0,0339	0,0138	0,0227	0,0215	0,0164
EDUCACIÓN	5,9864	0,0222	0,0118	0,0255	0,0135	0,0045
ENFERMERÍA	6,7692	0,0183	0,0123	0,0241	0,0120	-0,0022
HUMANIDADES	5,0685	0,0285	0,0128	0,0286	0,0178	0,0082
INGENIERÍA	5,0449	0,0326	0,0143	0,0270	0,0190	0,0150
MEDICINA	5,9194	0,0357	0,0082	0,0270	0,0167	0,0009
PSICOLOGÍA	5,8250	0,0192	0,0159	0,0242	0,0167	0,0050
SALUD	6,1303	0,0203	0,0110	0,0226	0,0155	0,0069

Tabla A5. Coeficientes estimados para lectura crítica, Modelo 3

Grupo de referencia	Intercepto	Matemáticas	Lenguaje	Química	C.Sociales	Promedio lenguaje SABER 11
ADMINISTRACIÓN Y AFINES	3,7170	0,0092	0,0279	0,0225	0,0311	0,0378
ARQUITECTURA Y URBANISMO	4,2018	0,0087	0,0296	0,0231	0,0319	0,0253
BELLAS ARTES Y DISEÑO	4,9197	0,0083	0,0294	0,0217	0,0307	0,0156
CIENCIAS AGROPECUARIAS	3,5414	0,0075	0,0264	0,0213	0,0298	0,0460
CIENCIAS NATURALES Y EXACTAS	4,1261	0,0042	0,0254	0,0182	0,0267	0,0459
CIENCIAS SOCIALES	3,6982	0,0082	0,0201	0,0182	0,0308	0,0567
COMUNICACIÓN, PERIODISMO Y PUBLICIDAD	3,4451	0,0111	0,0230	0,0206	0,0332	0,0501
CONTADURÍA Y AFINES	2,5531	0,0105	0,0313	0,0211	0,0305	0,0578
DERECHO	3,9749	0,0060	0,0296	0,0177	0,0317	0,0420
ECONOMÍA	5,0461	0,0102	0,0265	0,0176	0,0268	0,0247
EDUCACIÓN	2,2062	0,0076	0,0334	0,0211	0,0350	0,0642
ENFERMERÍA	3,1967	0,0076	0,0275	0,0199	0,0277	0,0569
HUMANIDADES	4,9045	0,0086	0,0250	0,0149	0,0296	0,0327
INGENIERÍA	4,0431	0,0085	0,0265	0,0182	0,0300	0,0376
MEDICINA	4,7415	0,0079	0,0239	0,0144	0,0214	0,0411
PSICOLOGÍA	3,8663	0,0066	0,0287	0,0191	0,0314	0,0447
SALUD	2,6916	0,0067	0,0278	0,0200	0,0298	0,0645

Tabla A6. Coeficientes estimados para razonamiento cuantitativo, Modelo 3

Grupo de referencia	Intercepto	Matemáticas	Lenguaje	Química	C.Sociales	Promedio matemáticas SABER 11º
ADMINISTRACIÓN Y AFINES	4,1314	0,0280	0,0152	0,0268	0,0191	0,0333
ARQUITECTURA Y URBANISMO	5,0619	0,0261	0,0138	0,0270	0,0188	0,0152
BELLAS ARTES Y DISEÑO	5,2023	0,0298	0,0073	0,0277	0,0165	0,0137
CIENCIAS AGROPECUARIAS	4,3611	0,0276	0,0145	0,0238	0,0183	0,0356
CIENCIAS NATURALES Y EXACTAS	4,5190	0,0326	0,0111	0,0254	0,0160	0,0307
CIENCIAS SOCIALES	4,6257	0,0225	0,0110	0,0248	0,0199	0,0289
COMUNICACIÓN, PERIODISMO Y PUBLICIDAD	5,4610	0,0240	0,0136	0,0219	0,0199	0,0092
CONTADURÍA Y AFINES	3,7142	0,0257	0,0187	0,0247	0,0160	0,0472
DERECHO	5,0124	0,0234	0,0116	0,0258	0,0171	0,0220
ECONOMÍA	4,6392	0,0331	0,0134	0,0221	0,0213	0,0259
EDUCACIÓN	4,4221	0,0218	0,0114	0,0253	0,0133	0,0401
ENFERMERÍA	4,7361	0,0176	0,0116	0,0233	0,0118	0,0433
HUMANIDADES	4,8836	0,0281	0,0126	0,0287	0,0178	0,0124
INGENIERÍA	4,1727	0,0322	0,0141	0,0266	0,0189	0,0327
MEDICINA	4,8154	0,0345	0,0073	0,0256	0,0160	0,0239
PSICOLOGÍA	4,8240	0,0186	0,0152	0,0239	0,0166	0,0284
SALUD	4,1490	0,0199	0,0106	0,0224	0,0154	0,0501

Tabla A7. Componentes de varianza de lectura crítica, ordenados por CCI, Modelo 1

		M1			M2			M3	
Grupos de referencia	σ^2_γ	σ^2	$\frac{\sigma^2_\gamma}{\sigma^2+\sigma^2_\gamma}$	σ^2_γ	σ^2	$\frac{\sigma^2_\gamma}{\sigma^2+\sigma^2}$	σ^2_γ	σ^2	$\frac{\sigma^2_\gamma}{\sigma^2+\sigma^2_\gamma}$
CIENCIAS SOCIALES	0,073	0,604	0,108	0,055	0,605	0,083	0,023	0,603	0,036
ENFERMERÍA	0,059	0,519	0,102	0,060	0,519	0,104	0,025	0,520	0,047
HUMANIDADES	0,055	0,594	0,085	0,046	0,595	0,071	0,025	0,598	0,040
EDUCACIÓN	0,051	0,593	0,079	0,030	0,592	0,049	0,019	0,593	0,031
SALUD	0,046	0,534	0,079	0,042	0,534	0,073	0,013	0,534	0,023
DERECHO	0,040	0,594	0,063	0,034	0,594	0,055	0,019	0,594	0,032
CIENCIAS AGROPECUARIAS	0,034	0,561	0,058	0,025	0,561	0,043	0,014	0,563	0,023
PSICOLOGÍA	0,033	0,562	0,056	0,029	0,562	0,049	0,017	0,562	0,029
CONTADURÍA Y AFINES	0,031	0,545	0,054	0,022	0,545	0,039	0,014	0,545	0,025
COMUNICACIÓN, PERIODISMO Y PUBLICIDAD	0,029	0,598	0,046	0,029	0,598	0,047	0,017	0,597	0,027
CIENCIAS NATURALES Y EXACTAS	0,026	0,588	0,043	0,016	0,588	0,027	0,007	0,587	0,012
INGENIERÍA	0,023	0,576	0,039	0,018	0,576	0,031	0,011	0,576	0,019
ADMINISTRACIÓN Y AFINES	0,024	0,598	0,038	0,019	0,598	0,031	0,015	0,598	0,025
MEDICINA	0,022	0,560	0,038	0,016	0,561	0,027	0,004	0,561	0,008
ECONOMÍA	0.,22	0,568	0,038	0,020	0,568	0,034	0,018	0,568	0,031
BELLAS ARTES Y DISEÑO	0,019	0,640	0,028	0,019	0,640	0,029	0,018	0,640	0,027
ARQUITECTURA Y URBANISMO	0,010	0,572	0,018	0,009	0,572	0,015	0,006	0,571	0,011

Tabla A8. Componentes de varianza de razonamiento cuantitativo, ordenados por CCI, Modelo 1

		M1			M2			M3	
Grupos de referencia	σ_γ^2	σ^2	$\frac{\sigma_\gamma^2}{\sigma^2+\sigma_\gamma^2}$	σ_γ^2	σ^2	$\frac{\sigma_\gamma^2}{\sigma^2+\sigma_\gamma^2}$	σ_γ^2	σ^2	$\frac{\sigma_\gamma^2}{\sigma^2+\sigma_\gamma^2}$
SALUD	0,064	0,461	0,121	0,061	0,461	0,117	0,028	0,461	0,056
CIENCIAS NATURALES Y EXACTAS	0,092	0,699	0,117	0,072	0,700	0,093	0,051	0,699	0,068
ENFERMERÍA	0,050	0,442	0,102	0,051	0,442	0,104	0,027	0,442	0,058
CIENCIAS AGROPECUARIAS	0,054	0,553	0,089	0,045	0,553	0,075	0,036	0,553	0,062
CONTADURÍA Y AFINES	0,041	0,504	0,076	0,030	0,504	0,056	0,018	0,504	0,035
MEDICINA	0,056	0,688	0,075	0,057	0,688	0,076	0,026	0,689	0,036
ECONOMÍA	0,055	0,696	0,073	0,045	0,695	0,061	0,033	0,696	0,045
PSICOLOGÍA	0,033	0,455	0,067	0,032	0,455	0,065	0,022	0,455	0,046
INGENIERÍA	0,047	0,683	0,064	0,034	0,683	0,048	0,019	0,683	0,027
AMINISTRACIÓN Y AFINES	0,037	0,565	0,062	0,027	0,565	0,046	0,018	0,565	0,030
HUMANIDADES	0,035	0,580	0,057	0,032	0,581	0,052	0,032	0,581	0,052
CIENCIAS SOCIALES	0,028	0,516	0,052	0,021	0,516	0,039	0,011	0,517	0,020
DERECHO	0,029	0,540	0,051	0,025	0,540	0,044	0,020	0,540	0,035
EDUCACIÓN	0,024	0,478	0,047	0,023	0,478	0,047	0,010	0,478	0.021
ARQUITECTURA Y URBANISMO	0,025	0,535	0,045	0,026	0,535	0,047	0,017	0,535	0,032
COMUNICACIÓN, PERIODISMO Y PUBLICIDAD	0,019	0,518	0,036	0,020	0,518	0,036	0,019	0,518	0,035
BELLAS ARTES Y DISEÑO	0,019	0,562	0,032	0,019	0,562	0,033	0,015	0,562	0,027

Calle 17 No. 3-40 • Teléfono:(57-1)338 7338 • Fax:(57-1)283 6778 • Bogotá - Colombia
www.icfes.gov.co