



INFORMES

¿Es posible aproximar el desempeño de las instituciones de educación superior?

Luisa Fernanda Bernat, Luis Fernando Gamboa

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 de 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, transformar¹, promocionar o realizar acción alguna de la cual se lucre directa o indirectamente con este material.

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 al ICFES como fuente de autor. Lo anterior siempre que los pasajes no sean tantos ni 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 usarlas 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.

¹ 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.

¿Es posible aproximar el desempeño de las instituciones de educación superior?

Luisa Fernanda Bernat
Luis Fernando Gamboa

El presente escrito contribuye a la discusión sobre los alcances de las medidas de valor agregado como instrumento para la medición de la efectividad de las instituciones de educación superior (IES), aclarando cuáles son sus alcances y sus limitaciones para el caso colombiano. En ese sentido, contribuye a dar respuesta de interrogantes como: i) ¿cómo se puede evaluar la efectividad de una institución de educación superior?; ii) ¿cuáles son los principales retos que enfrenta una medida de efectividad de este tipo?; y iii) ¿qué tan apropiadas son las medidas basadas en los rankings de desempeño de las instituciones, tan frecuentes hoy en día?

¿Para qué los indicadores? Colombia es un país de cerca de 40 millones de personas y una matrícula escolar aproximada de 11 millones de estudiantes en enseñanza básica y media. La cobertura en educación superior ha pasado de 24,4% en 2002 a 46% en 2014 con amplias diferencias a nivel regional (MEN, 2009). La matrícula en educación secundaria y media –siendo estos los futuros demandantes de la educación superior– viene creciendo también, pasando de 3,5 a 4,3 millones de estudiantes entre 2002 y 2012 según el Dane. Desafortunadamente no todos terminan el ciclo educativo, ya que la tasa de deserción por cohorte en educación superior en Colombia era de 49% en 2005 y sólo se redujo a 45% en 2012 según MEN (2009), evidenciando resultados desalentadores sobre cualquier medida de calidad en educación superior. Estos resultados sirven como alertas tempranas del sector educativo.

En los últimos años han surgido nuevos indicadores que pretenden proveer información sobre el sector educativo como el Modelo de Indicadores del Desempeño de la Educación (MIDE) del Ministerio de Educación, ideado para generar información sobre el desempeño de 74 universidades y 109 instituciones universitarias que utiliza distintas variables para la obtención de un indicador sintético. Los resultados generan alegrías en los primeros y críticas en los últimos por las variables incluidas y las formas de ponderación empleadas. A nivel internacional, varias instituciones como QS University Rankings o el Grupo Sapiens Research proveen clasificaciones de desempeño de las instituciones de educación superior utilizando diferentes criterios asociados a investigación científica, publicaciones, recursos y empleabilidad, y, por lo tanto, arrojan diferentes ordenamientos. En dichos rankings, los resultados para las instituciones colombianas no son alentadores: las posiciones están lejos de los primeros lugares tanto a nivel global como a nivel de América. No es de extrañar que un cambio de variables puede modificar las posiciones de las universidades y generar algún tipo de confusión. El ranking de Shanghái, por ejemplo, además de valorar publicaciones y su impacto, incluye el conteo de Premios Nobel.

La toma de decisiones de padres de familia, estudiantes, gobierno y gestores escolares se hace con base en indicadores o clasificaciones públicas de este tipo de iniciativas y de ahí la importancia que tienen. Pero, dadas las diferencias en los objetivos y las metodologías utilizadas, es común que arrojen resultados distintos y no del todo comparables, razón por la cual se requiere cautela en su interpretación. El principal factor que afecta la comparabilidad es la existencia de competencias complejas generadas por la educación superior que no siempre son observables o comparables. Las aproximaciones más cercanas son los puntajes en pruebas estandarizadas de competencias básicas o los ingresos laborales de sus egresados, pero aun así son medidas imperfectas de las competencias laborales de los egresados.

¿Por qué una medida de valor agregado? La medición del desempeño de una IES de acuerdo al resultado de una prueba académica o un indicador de ingreso laboral es imprecisa e injusta por el desconocimiento de los factores que afectan al estudiante y su entorno. Si las condiciones de partida de los estudiantes difieren notoriamente entre instituciones, no es de extrañar que se mantengan a lo largo de los estudios superiores independiente de la gestión institucional en la IES. El ejercicio de medir la contribución o efecto de una IES sobre una cohorte en particular requiere tener control sobre todos aquellos factores que afectan el resultado de un estudiante cualquiera. Los estudios recientes sobre el efecto causal sugieren el uso de métodos experimentales o pseudo-experimentales para la estimación de efectos, en los que se compara el resultado de una población tratada con respecto a una población control. Este tipo de ejercicios no es fácilmente implementable en educación por temas éticos y altos costos económicos. Recientemente se han observado varios enfoques para abordar su medición. Algunos enfoques como el de Kane y Staiger (2008) sugieren modelos de emparejamiento (*Propensity score matching*) para comparar individuos de acuerdo a características similares. Otra alternativa es el uso de los modelos de valor agregado. Ambos enfoques tienen restricciones en sus alcances y en sus interpretaciones. En este trabajo nos centramos en el caso del valor agregado.

La medida de valor agregado cuantifica el logro obtenido por los estudiantes que no es explicado por sus condiciones académicas y socioeconómicas previas; es decir, es la diferencia entre lo que logra y lo que se espera del estudiante de acuerdo a sus características. En un escenario donde factores no observables están distribuidos de manera aleatoria entre las instituciones, esta es una medición altamente acorde con el resultado causado por el quehacer de la institución. La medición del valor agregado reconoce que la ganancia en competencias dentro de un mismo programa de estudio puede llegar a ser muy distinta entre un estudiante y otro. Sin embargo, la metodología del valor agregado se sustenta en reconocer que los estudiantes ingresan a las instituciones de educación superior con condiciones académicas y entornos socioeconómicos diferentes.

La modelación del valor agregado se ha implementado a través de dos enfoques: los modelos de regresión lineal de mínimos cuadrados ordinarios y los modelos jerárquicos (Clark et al., 2010; Kim y Lalancette, 2013). Los modelos de regresión lineal minimizan la diferencia entre los valores observados y los estimados (modelos mínimos cuadrados ordinarios, OLS), pero estos modelos OLS adolecen de una debilidad consistente en desconocer que las unidades sujeto de estudio, los estudiantes, están agrupados en aulas y las aulas a su vez en planteles educativos. De esta manera, los logros individuales no son totalmente independientes de lo que ocurre en el aula o el plantel al que pertenecen, y esto va en contra de los supuestos necesarios para la validez de los resultados porque los resultados pueden estar contaminados por las características propias de la cohorte o de las interacciones a su interior. La solución a este problema se ha intentado alcanzar con la introducción de controles de estudiante o plantel en las estimaciones (efectos fijos), pero esta alternativa es compleja en la práctica por el número de coeficientes que deberían ser estimados. El segundo enfoque utiliza los modelos multinivel o modelos jerárquicos (conocidos como HLM por su nombre en inglés). Estos modelos proveen una estimación de la varianza

que se puede descomponer entre la parte que es explicada por un nivel (estudiantes) y la otra parte que es explicada por un segundo nivel (institución), con lo cual se obtiene una contribución de la institución, pero se sacrifica la información sobre el aporte de otros factores. Aunque es posible utilizar más niveles en el análisis, es frecuente el uso de dos niveles en la literatura sobre efectividad de las instituciones educativas por restricciones informativas o disponibilidad de información.

Las tendencias recientes sobre los estudios de efecto causal también han motivado el uso de modelos experimentales y no experimentales para la medición del efecto de una institución. En particular, el Ministerio de Educación ha trabajado en una propuesta consistente en utilizar la metodología de Kane y Staiger (2008), basada en el emparejamiento de estudiantes para estimar el efecto de la institución de educación superior utilizando la clasificación de las universidades de acuerdo al MIDE y los resultados de las pruebas Saber 11 y Saber Pro. Esta metodología pretende responder la pregunta ¿cuál sería el logro de un estudiante determinado si estudia en una institución X en lugar de hacerlo en una Y? Este es un trabajo que aún está en construcción y por lo tanto los resultados no son públicos actualmente, pero plantean una visión alternativa de la efectividad de las IES. Como se puede observar, la tendencia hacia nuevas mediciones ha ido incrementando los resultados disponibles de diferentes estudios y ahí es donde es necesario profundizar en los alcances de cada uno de estos indicadores.

¿Cuáles son sus restricciones? Si bien la medida de valor agregado permite solventar algunas debilidades de otras mediciones, aún no es perfecta. Para incrementar su precisión requiere varios elementos: muestras de tamaño considerable, tener mecanismos para reducir el efecto de factores no observables (el esfuerzo, la motivación, la existencia de contextos familiares complejos, la autoselección de los estudiantes frente a un determinado programa¹) y tener en cuenta efectos de corto y largo plazo. Es decir, la ausencia de una asignación aleatoria de los estudiantes a las instituciones lleva a que la medición del “efecto del programa” esté contaminada por las variables que inducen la asignación y, por lo tanto, debe ser corregida para no sobreestimar el verdadero valor generado. Desde el punto de vista operativo, los niveles de deserción universitaria (completa e interna) impiden tener un cruce completo de información entre la población entrante y la población saliente de un programa².

El Icfes ha desarrollado una medida de valor agregado reconociendo que es más equitativo evaluar el desempeño de una IES o de un programa académico por medio de la comparación de los resultados de la población egresada de sus programas y las condiciones académicas con que ingresan. Para controlar la influencia de los factores previos se emplean variables de condiciones de entrada, tales como los puntajes en Saber 11, las condiciones socioeconómicas representadas por el nivel educativo de los padres o las condiciones del grupo de referencia aproximadas por el puntaje promedio de la cohorte en Saber 11. Cualquiera de estas variables permite controlar el grado de homogeneidad de entrada en estas dimensiones.

El modelo básico valora como parte del aporte relativo de la institución el efecto que tiene estudiar con un grupo de alumnos con unas características que favorecen el aprendizaje. Por su parte, los modelos que incluyen contexto socioeconómico y contexto académico penalizan a las instituciones que trabajan con grupos de estudiantes con alto nivel socioeconómico y a las instituciones que trabajan con grupos de estudiantes de alto nivel académico. Esta situación se da porque al contar con estudiantes de mayor

1 A finales de los años 90, medicina era la carrera más demandada y por lo tanto la selección llevaba a que los estudiantes que entraban tuvieran el promedio más alto en la prueba de Estado.

2 Según el MEN (2009), la deserción total es la diferencia entre estudiantes a la salida y a la entrada del programa. La deserción interna corresponde a la migración interna de estudiantes dentro de los programas de una IES.

capital socioeconómico, debe realizar mayores esfuerzos para aportar valor al estudiante. Desde el punto de vista operativo, la estimación de modelos con este conjunto de controles no es intuitiva y en muchos casos implica cierto grado de subjetividad en la selección de variables o formas funcionales. De hecho, estas elecciones pueden afectar la claridad, la interpretación de los resultados y su magnitud (Goldhaber et al., 2013).

Operativamente, el ejercicio empieza con la información de las pruebas de lectura crítica y razonamiento cuantitativo de Saber Pro de 2012, ya que es el primer año comparable entre Saber 11 y Saber Pro. La medición se lleva a cabo sobre 17 grupos de referencia de las instituciones de educación superior universitaria, entendiéndolos como un subconjunto de programas de estudio con características similares de sus egresados y campos de acción laboral sin que ello sea necesariamente sinónimo de facultad³. Teniendo en cuenta el número de grupos de referencia y de instituciones de educación superior, se obtienen resultados para aproximadamente 1.300 instituciones-grupo de referencia (IGR)⁴. La disponibilidad de información y el tamaño de algunas carreras universitarias que reciben y gradúan cohortes pequeñas de estudiantes ha inducido a realizar la evaluación agrupando a los estudiantes de dos años consecutivos (2012-2013 y 2013-2014), para los cuales no debe haber cambiado las condiciones institucionales del grupo de referencia.

Algunos resultados interesantes emergen de estas estimaciones iniciales. La primera medida de valor agregado se calcula sobre los puntajes obtenidos en razonamiento crítico o sobre lectura crítica de la Prueba Saber Pro de final de carrera para el periodo 2012-2013. Los resultados muestran que aquellas áreas donde hay mayor demanda de cupos, como el caso de medicina, al haber una mayor competencia de entrada, hay un proceso de selección más fuerte (promedio en Saber 11 más alto) y por lo tanto se tiende a generar valor agregado con mayor dificultad con respecto a otras áreas del conocimiento. En el extremo opuesto están los grupos de referencia *ciencias sociales y humanidades*, en los cuales se obtiene un menor puntaje. Para tener una idea de la heterogeneidad de los grupos de referencia, se encuentra que *administración de empresas* y afines tiene 165 IGR mientras que en *humanidades* solamente hay 21. En términos de estudiantes, el tamaño promedio de egresados del IGR *administración de empresas* es el doble que el de psicología o el de arquitectura. Estas diferencias son importantes para cuantificar los desempeños de cada institución-grupo de referencia por la variabilidad que puede haber al interior de cada programa.

La medida de aporte relativo. Teniendo en cuenta los resultados simples de valor agregado, para tener una medida más aproximada de valor agregado entre instituciones con estudiantes similares, el Icfes plantea la construcción de una medida de aporte relativo. Esta medida incluye comparaciones *uno a uno* entre instituciones-grupos de referencia que reciben estudiantes con *condiciones de entrada similares* (medidas a través del índice Saber 11). Las mediciones se llevan a cabo al interior de la vecindad de la IGR. Los integrantes de la vecindad de una IGR son las 15 IGR que 5 años atrás recibieron estudiantes con promedios Saber 11, similares a aquellos que recibió la institución analizada. Con este replanteamiento se obtiene una medida más precisa de la efectividad de la institución con respecto a otras y le permite identificar experiencias exitosas y fallidas. Las medidas de aporte relativo muestran el desarrollo de competencias durante la educación superior para grupos similares y dado que las IGR

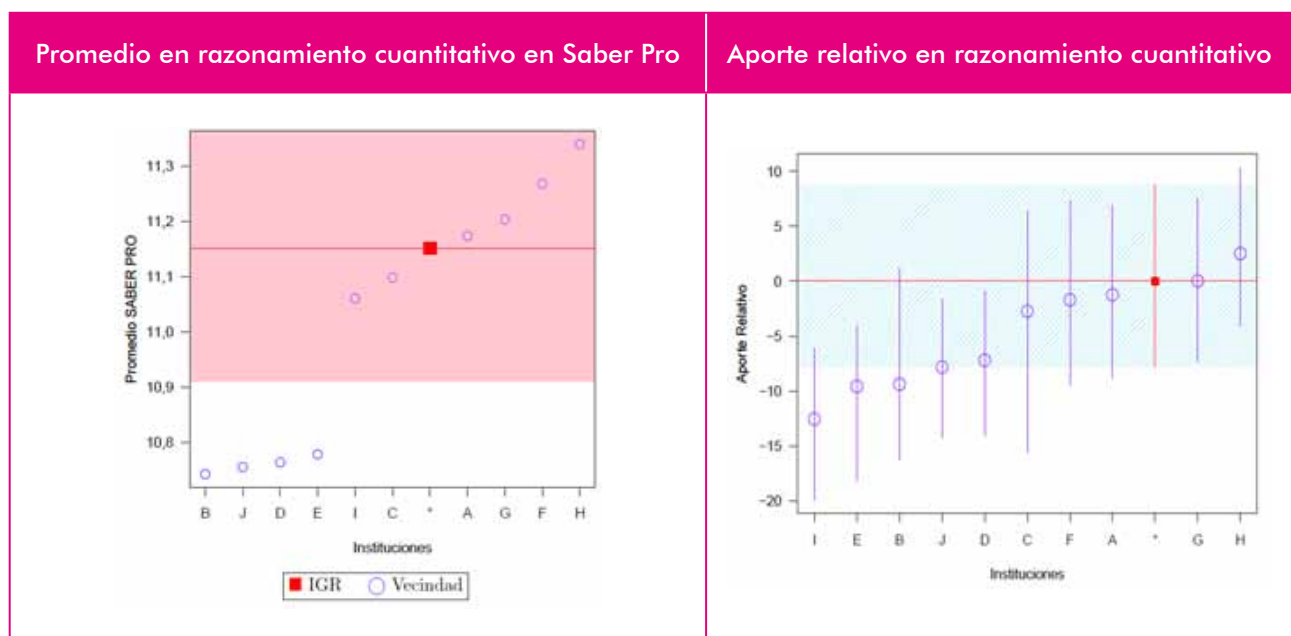
3 Los grupos de referencias son: administración y afines; arquitectura y urbanismo; bellas artes y diseño; ciencias agropecuarias; ciencias naturales y exactas; ciencias sociales; comunicación periodismo y publicidad; contaduría y afines; derecho; economía; educación; enfermería; humanidades; ingeniería; y medicina, psicología y salud.

4 Un IGR es un programa académico de una IES particular o la unión de un par de programas similares como periodismo y comunicación social.

pueden ser muy heterogéneas al interior de una misma institución universitaria, el aporte relativo permite identificar qué grupos de referencia y qué instituciones son más exitosos en el fortalecimiento de una o varias competencias.

Las medidas de aporte relativo constituyen un avance importante para comprender la efectividad de un IGR, pero también enfrentan ciertas dificultades. En primer lugar, la medida provee una medición relativa frente a otros IGR que son considerados sus vecinos naturales de acuerdo a las condiciones de entrada, pero estos vecinos pueden cambiar por factores exógenos entre un año y otro, como por ejemplo cambios en los requisitos de entrada de un programa y oscilaciones importantes en el tamaño de la cohorte que afectan el desempeño promedio del programa (Condie et al., 2014). En segundo lugar, el resultado puede ser sensible al uso de un determinado control de condiciones de entrada en lugar de otro.

En el siguiente gráfico se tienen los resultados promedio en razonamiento cuantitativo (izquierda) para un conjunto de instituciones pertenecientes a la vecindad de una IGR (marcada con asterisco) en un año cualquiera⁵. En términos de los promedios (panel izquierdo) se encuentra que la institución de referencia tiene un desempeño promedio mayor al de los de B, J, D y E. En el lado derecho se tiene el aporte relativo de la IGR frente a sus vecinos y se observa que la institución grupo de referencia no es significativamente mejor que ninguna (la zona sombreada representa el intervalo dentro del cual se observa que no existen diferencias significativas) y además pareciera que el ordenamiento cambiara.



Una posible explicación de este resultado podría provenir del hecho que la variable utilizada para la configuración de la vecindad afecta de manera diferencial a las instituciones y ello cambia el ordenamiento; es decir, las características de los estudiantes que ingresan permiten mejores desempeños en algunas de

5 Los resultados corresponden a un IGR particular pero no se agrega el nombre de sus integrantes para concentrar la atención en el fenómeno antes que en las particularidades de las instituciones.

estas instituciones frente a otras. Esto muestra que el resultado puede ser sensible al tipo de variable de condiciones de entrada porque la manera como afectan las condiciones socioeconómicas de los estudiantes puede ser diferente a la manera como lo hacen las condiciones académicas de los estudiantes que comparten las aulas de clase. El nivel socioeconómico explica una proporción del logro de los hijos, pero no existe una correlación perfecta entre estas dos variables.

Ahora bien, si se utilizara un modelo más completo, que controlara a la vez ambos tipos de variables, probablemente se tendrían ordenamientos más estables (menores fluctuaciones entre un periodo y otro), pero el costo asumido sería una menor claridad sobre los resultados obtenidos y sus causas en estos indicadores para los interesados. Es decir, dado que la provisión de educación es un fenómeno complejo con múltiples interacciones, entre mayores controles se incluyan en estos modelos será más difícil entender los fenómenos que subyacen entre estos controles y por tanto se dificulta la comprensión de los resultados por parte del público en general. Esto se da porque no es posible separar del todo el efecto de cada control, el cambio de un año a otro y el cambio en la composición de la vecindad que puede provenir entre estas instituciones.

Referencias

- Clark, P.; Crawford, C.; Steele, F. & Vignoles, A. (2010). "The choice between fixed and random effects models: Some considerations for educational research". University of Bristol Centre for Market and Public Organisation. Working paper 10/240, pp. 1-33.
- Condie S.; Lefgren, L. y Sims, D. (2014). Teacher heterogeneity, value added and education policy. *Economics of Education Review*, 40, 76–92.
- Goldhaber, D.; Walch, J. y Gabele, B. (2013). Does the Model Matter? Exploring the Relationship between Different Student Achievement-Based Teacher Assessments. *Statistics and Public Policy* 1 (1), pp. 28-39.
- Kane, T y Staiger, D. (2008). Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation, National Bureau of Economic Research, NBER 14607, Dic., p. 51.
- Kim, H. y Lalancette, D (2013). Literature Review on the Value-Added Measurement in Higher Education. OCDE Publishing.
- Ministerio de Educación Nacional (2009). Deserción estudiantil en la educación superior colombiana. Metodología de seguimiento, diagnóstico y elementos para su prevención. Disponible en http://www.mineducacion.gov.co/sistemasdeinformacion/1735/articles-254702_libro_desercion.pdf.

