Índice de desempeño relativo pruebas Saber 11 – Saber Pro

Jhonniers Guerrero Erazo, Délany Ramírez del Río

Universidad Tecnológica de Pereira
Pereira, Colombia

Resumen

Desde la creación del Examen de Calidad de la Educación Superior, hoy llamado Saber PRO, liderado por el ICFES como institución, se ha fomentado la realización de análisis longitudinales que permitan identificar la trazabilidad de los avances en los resultados de los estudiantes, en particular contrastándolos contra los resultados de la Educación Básica y Media, medidas a partir de las pruebas SABER11; lo cual ha conllevado a establecer diferentes metodologías para medir este tipo de avances, donde incluso el mismo ICFES ha desarrollado estudios de valor agregado y aporte relativo, lo anterior en el entendido que la calidad de los resultados de salida (output) no se pueden adjudicar solamente a la Institución, en este se ven involucrados muchas variables adicionales, como por ejemplo, los mecanismos de selección, la demanda de la institución, su contexto geográfico, entre otros; de acuerdo con Muñoz Gómez (2016).

Dado lo anterior, se han realizado modelos de valor agregado los cuales "intentan evaluar la efectividad educativa haciendo énfasis en el progreso de los estudiantes en el tiempo" (ICFES, 2014). "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" (Martínez R., Gaviria J., Castro M., 2009). El modelo de valor agregado del ICFES toma variables académicas y socioeconómicas para determinar el modelo, incluyendo en ellas la relación entre las pruebas y un índice de nivel socioeconómico (ICFES, 2014). Los modelos de valor agregado levantan interés, dado que intentan separar los efectos del contexto sobre los resultados, permitiendo la identificación el desempeño de las instituciones frente a otras, permitiendo "aislar" los efectos de las escuelas sobre el rendimiento de los estudiantes (Martínez R., Gaviria J., Castro M., 2009).

De cara a la anterior, se desarrolló un modelamiento alternativo para definir una escala de desempeño relativo para los estudiantes que presentan las pruebas SABERPRO comparativo al momento de ingreso a la institución, medido a través de la prueba SABER11. Para esto se hace uso de la información consignada y publicada por el ICFES para investigadores, centrado en las competencias genéricas evaluadas en las pruebas SABERPRO, con sus análogos en las pruebas SABER11.

Esta medida busca brindar información insumo que permita revelar a la institución elementos que aportan a la calidad sobre los programas académicos impartidos, más allá de los resultados SABERPRO como única medida, dado que, no es comparable el desempeño de un estudiante que ingresó a una IES con un SABER11 en el quintil mayor de las calificaciones frente a uno que ingresó a la IES en el quintil inferior de calificaciones de SABER11. Sus resultados pueden ser utilizados en los procesos de autoevaluación de los programas académicos para mapear el estado del arte del programa y así levantar oportunidades de mejora que permitan trazar caminos al programa para mejorar la calidad de los procesos académicos.

Palabras clave: calidad; exámenes de estado; saber11; saberpro; desempeño; modelamiento

Abstract

Since the creation of the Higher Education Quality Examination, today called Saber PRO, led by ICFES as an institution, longitudinal analyzes have been encouraged to identify the traceability of progress in student results, in particular by contrasting them. against the results of Basic and Secondary Education, measured from the SABER11 tests; which has led to the establishment of different methodologies to measure this type of progress, where even the ICFES itself has developed value-added and relative contribution studies, the foregoing in the understanding that the quality of the output results cannot be judged only to the Institution, in this many additional variables are involved, such as, for example, the selection mechanisms, the demand of the institution, its geographical context, among others; according to Muñoz Gómez (2016).

In this sense, value-added models have been carried out which "try to evaluate educational effectiveness, emphasizing the progress of students over time" (ICFES, 2014). "In this sense, the added value of an institution is understood as the contribution to the net progress of students, towards established learning objectives, once the influence of other factors outside the institution that can contribute to that progress has been eliminated" (Martínez R., Gaviria J., Castro M., 2009). The ICFES value-added model takes academic and socioeconomic variables to determine the model, including the relationship between the tests and a socioeconomic level index (ICFES, 2014). The value-added models raise interest, since they try to separate the effects of the context on the results, allowing the identification of the performance of the institutions compared to others, allowing to "isolate" the effects of the schools on the performance of the students (Martínez R. ., Gaviria J., Castro M., 2009).

Faced with the above, an alternative modeling was developed to define a relative performance scale for students who take the comparative SABERPRO tests at the time of admission to the institution, measured through the SABER11 test. For this, use is made of the information recorded and

published by the ICFES for researchers, focused on the generic competencies evaluated in the SABERPRO tests, with their analogs in the SABER11 tests.

This measure seeks to provide input information that allows the institution to reveal elements that contribute to the quality of the academic programs taught, beyond the SABERPRO results as the only measure, since the performance of a student who entered an HEI is not comparable. with a SABER11 in the highest quintile of qualifications compared to one who entered the HEI in the lowest quintile of SABER11 qualifications. Its results can be used in the self-assessment processes of the academic programs to map the state of the art of the program and thus raise opportunities for improvement that allow the program to trace paths to improve the quality of the academic processes.

Keywords*: quality; state exams; know11; knowpro; performance; modeling*

1. Introducción

El objetivo del modelamiento es la definición de una métrica de desempeño relativo entre los resultados de los estudiantes en la Prueba SABERPRO, en contraste con los resultados de la prueba SABER11, que permita a una institución identificar frente a una métrica nacional diferencias significativas para su institución o un programa académico específico.

Esta medida busca brindar información insumo que permita revelar a la institución elementos de calidad sobre los programas académicos impartidos, más allá de los resultados SABERPRO como única medida, dado que, no es comparable el desempeño de un estudiante que ingresó a una IES con un SABER11 en el quintil mayor de las calificaciones frente a uno que ingresó a la IES en el quintil inferior de calificaciones de SABER11. Sus resultados pueden ser utilizados en los procesos de autoevaluación de los programas académicos para mapear el estado del arte del programa y así levantar oportunidades de mejora que permitan trazar caminos al programa para mejorar la calidad de los procesos académicos.

2. Revisión de la literatura y estado del arte

Desde la creación del Examen de Calidad de la Educación Superior (ECAES), hoy llamado Saber PRO, desde el ICFES como institución, se ha fomentado la realización de análisis longitudinales que permitan realizar trazabilidad de los avances en los resultados de los estudiantes, en particular contrastándolos contra los resultados de la Educación Básica y Media, medidas a partir de las pruebas ICFES (al día de hoy Saber 11), lo cual ha conllevado a diferentes metodologías para medir este avance, por ejemplo el mismo ICFES ha desarrollado estudios denominados de valor agregado y aporte relativo. De acuerdo con Muñoz Gómez (2016), la calidad de los resultados de salida (output) no se pueden adjudicar solamente a la Institución, en este se ven involucrados muchas variables adicionales, por ejemplo, los mecanismos de selección, la demanda de la institución, su contexto geográfico, entre otros.

En este sentido, se han realizado modelos de valor agregado los cuales "*intentan evaluar la efectividad educativa haciendo énfasis en el progreso de los estudiantes en el tiempo*" (ICFES, 2014). "*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*" (Martínez R., Gaviria J., Castro M., 2009). El modelo de valor agregado del ICFES toma variables académicas y socioeconómicas para determinar el modelo, incluyendo en ellas la relación entre las pruebas y un índice de nivel socioeconómico (ICFES, 2014). Los modelos de valor agregado levantan interés, dado que intentan separar los efectos del contexto sobre los resultados, permitiendo la identificación de buenas instituciones relativamente a otras, permitiendo "aislar" los efectos de las escuelas sobre el rendimiento de los estudiantes (Martínez R., Gaviria J., Castro M., 2009).

Kim & Lalancette (2013), realiza una agrupación de los modelos de valor agregado en cuatro categorías, así:

- Modelos de regresión lineal por mínimos cuadrados.
- Modelos con efectos fijos y aleatorios.
- Modelos de curva de crecimiento.
- Modelos de efectos aleatorios multivariados.

Los cuales varían en complejidad y dificultades de implementación, se puede afirmar que a mayor complejidad menor interpretabilidad de los resultados, complementariamente, uno de los grandes retos de los modelos de valor agregado es su interpretabilidad, *"la presentación de la información y resultados de valor agregado de las escuelas debe tener en cuenta la cuestión de cómo exponer de la mejor manera posible asuntos estadísticos complicados a un público lego en estadística"* (OECD, 2011), lo anterior para un público objetivo no necesariamente versado en el lenguaje estadístico, dado que *"los resultados de valor agregado son numéricos y se presentan de diversas formas, como puntuaciones continuas o una puntuación específica superior o inferior a un resultado de desempeño esperado de la escuela"* (OECD, 2011); sin embargo, conocerlos es de utilidad dado que *"las posiciones relativas de varias escuelas pueden servir como punto de partida para los análisis del desarrollo de la escuela"* (OECD, 2011).

El modelo de desempeño relativo aquí propuesto se encuentra catalogado en aquellos basados en la regresión lineal, con esfuerzos metodológicos posteriores para facilitar la interpretabilidad de resultados, estableciendo un modelo complementario al abanico de estudios longitudinales. Éste considera la relación entre las pruebas, sin embargo, se realiza un proceso de agrupamiento de acuerdo con diferentes niveles de entrada y salida con el uso de criterios estadísticos, que promuevan la comparabilidad entre grupos de desempeño comparabilidad entre grupos de desempeño relativo a sus grupos de origen. De este modo, los resultados que se obtengan permiten realizar comparaciones entre grupos de desempeño, más allá del resultado de salida aislado, como mala praxis de socialización. Lo que facilita la generación acciones de mejora entre programas, instituciones o incluso el sistema educativo y el uso de la información de manera inteligente. Adicionalmente, esta propuesta busca el desarrollo de un aplicativo que permita la consulta y toma de decisiones de los interesados, este componente es importante, para poder realizar la apropiación social del conocimiento, de acuerdo con la afirmación sobre el que "*el tablero ejecutivo es muy necesario*

para la junta directiva de la universidad como una de las herramientas en el proceso de toma de decisiones para ganar el mercado cada vez más competitivo" (Meyliana, Widjaja & Santoso , 2014).

3. Metodología y datos

La metodología de abordaje de los datos se realiza en dos momentos de análisis:

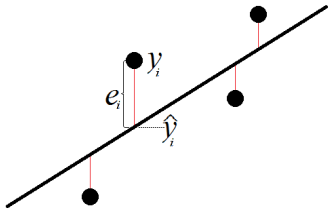
3.1. El análisis de residuales como medida de desempeño

Para proponer una metodología que busque medir el desempeño relativo, debe entenderse como la medición de los resultados relativos a la prueba inicial (Saber 11), en el entendido que el aprendizaje y el mejoramiento de las competencias no tiene un comportamiento lineal en el tiempo.

Considerando esto, sería injusto realizar una simple resta entre el resultado final del estudiante en la prueba de educación en contraste con la prueba de la educación media, incluso cuando la escala de ambas pruebas se encuentre estandarizada. Por este motivo, en esta aproximación se propone un enfoque que sopese los resultados de todos los estudiantes con trazabilidad de ambas pruebas a nivel nacional. Con este objetivo se propone la realización de un análisis que tome como base los residuales de la regresión lineal de ambas pruebas.

Al aplicar un modelo de regresión lineal, tomando como variable independiente los resultados Saber 11, y como variable dependiente los resultados Saber PRO, se obtiene la línea recta que se encuentra más cerca de todos los puntos observados (es decir, todos los estudiantes) relativa a su desempeño en ambas pruebas.

Es preciso recordar que la recta de regresión es una recta de promedios, es decir, el valor estimado en la prueba Saber PRO (*y*) corresponde al valor promedio estimado para la prueba Saber PRO de acuerdo con un resultado Saber 11 de entrada, es decir, que dicha estimación arroja un valor medio esperado de Saber PRO para un Saber 11 particular. Por tanto, al calcular los residuos de cada estudiante (ver formula) frente a esta recta, se podría interpretar como el mejor desempeño (en el caso de ser positivo) o menor desempeño (en el caso de ser negativo) frente al promedio del desempeño de todos los datos, en especial frente a sus pares cercanos (en términos de la prueba de entrada Saber 11).



$$e_i = y_i - \hat{y}_i$$

Figura 1 Residuales de la Regresión Lineal Simple

Al realizar un promedio de los residuales sobre todos los datos, el valor esperado es 0, esto dado que la recta de regresión (estimada por mínimos cuadrados), está lo más cerca de todos los puntos, por tanto, la suma de los residuales positivos tiene la misma magnitud de los residuales negativos. Sin embargo, si realizamos agrupaciones, sean estas por grupos de referencia, programas académicos o instituciones, el promedio no será 0, y este valor nos ayudará a determinar si el avance relativo se encuentra por encima del nivel nacional o por debajo, de igual manera nos permitirá realizar comparaciones de acuerdo con las variables de agrupación que se requiera analizar.

3.2. Cálculo del Desempeño Relativo

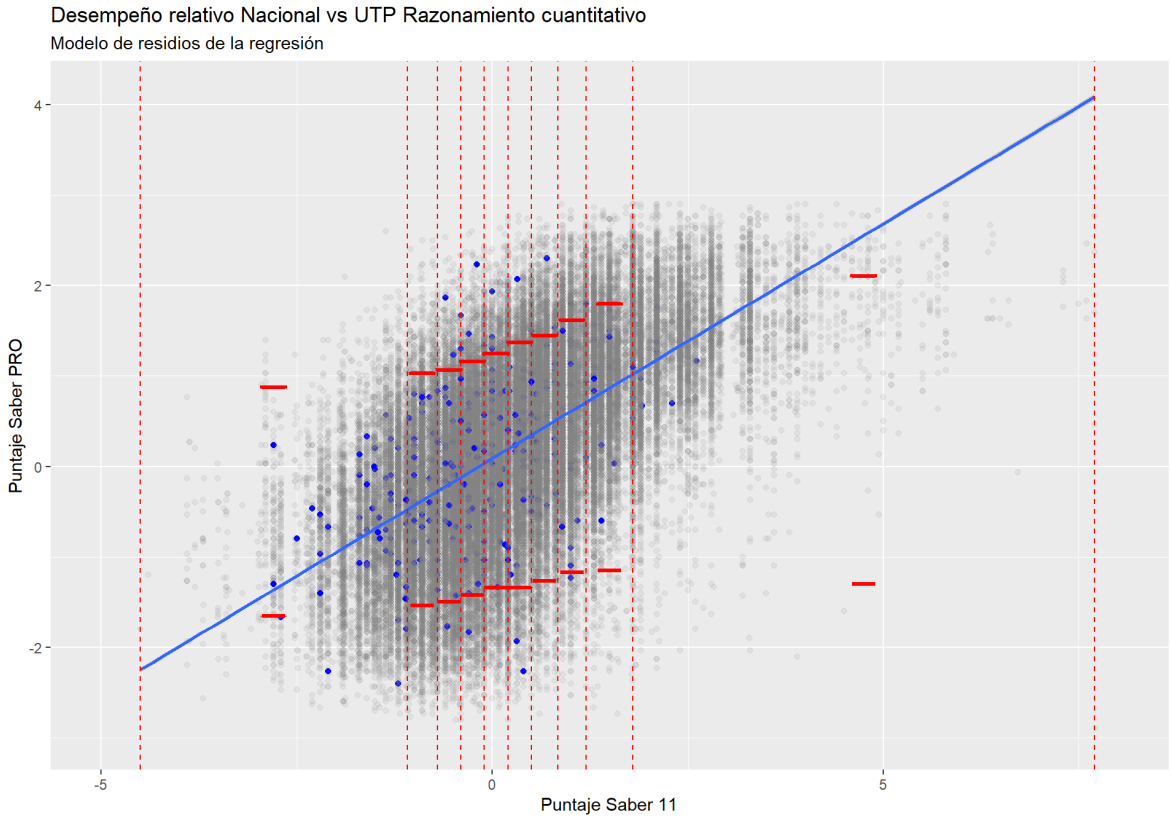
Si bien el valor del residual nos da una idea acerca del desempeño de un estudiante o un grupo frente al valor promedio dado por la recta de regresión, se requiere tener un referente de comparación para definir el desempeño frente a un nivel deseado observado. Sin embargo, resultaría injusto tomar como referente el estudiante con el mayor residual positivo observado, teniendo en cuenta, que a mayor puntaje de entrada (Saber 11) la recta de regresión sería más estricta frente al resultado medio esperado en Saber Pro, además los estudiantes con menor Saber 11 tienen un rango mucho más amplio de posibilidad de mejorar que un estudiante con un resultado Saber 11 superior. Con el fin de considerar que los diferentes niveles de entrada Saber 11 pueden poseer un residual relativo diferente, se realiza un ejercicio de agrupamiento según su Saber 11, utilizando el método de cálculo de deciles, que dividen la distribución de observaciones de Saber 11 en 10 grupos diferentes. Una vez separados los resultados de cada estudiante en los deciles, se procede a determinar el punto de referencia para la comparación. Una aproximación sería tomar el valor de desempeño más alto para cada grupo decil, sin embargo, este valor correspondería a un solo estudiante, adicionalmente, podría ser un dato atípico dentro de la distribución, en este sentido se propone establecer un punto de referencia que dependa de la distribución de los datos realizando los siguientes pasos:

1. Para cada grupo decil, se divide en 6 grupos de acuerdo a la concentración de los datos, percentil 33%, percentil 66% y percentil 100% para los datos positivos, y de la misma manera para los datos negativos. Esto nos generaría las siguientes clasificaciones:

Etiqueta	Descripción
-P3	Percentil 33% de los desempeños relativos negativos.
-P2	Percentil 66% de los desempeños relativos negativos.
-P1	Percentil 100% de los desempeños relativos negativos.
P1	Percentil 33% de los desempeños relativos positivos.
P2	Percentil 66% de los desempeños relativos positivos.
P3	Percentil 100% de los desempeños relativos positivos.

2. Se calcula la media de resultados de desempeño para el grupo del percentil más alto en el caso de los desempeños positivos, este sería el punto de referencia para los desempeños positivos.
3. Se calcula la media de resultados de desempeño para el grupo del percentil más bajo en el caso de los desempeños negativos, este sería el punto de referencia para los desempeños negativos.

En el siguiente gráfico se ilustra el resultado de este procedimiento:



De este modo, para cada estudiante o agrupación podemos identificar lo siguiente:

Desempeño	**Valor de su residual frente a la recta de medias**
Grupo	Décil de clasificación SABER 11
Cuántil	Grupo de desempeño SABER PRO.

Una vez identificados los valores referentes por cada grupo se procede a calcular el desempeño relativo para cada individuo o grupo con la siguiente ecuación:

$$dr_i = \begin{cases} \dfrac{e_i}{o_j - \hat{y}_i}, & e_i < o_j \\ 1, & e_i \geq o_j \end{cases}$$

donde, dr_i, es el desempeño relativo para la i-ésima observación, e_i, es el desempeño de la i-ésima observación, o_j es el óptimo para el j-ésimo decil y $\hat{y}_i$ es el valor esperado en la recta de medias para la i-ésima observación. El valor de 1 se asigna siempre cuando la observación sea uno de los atípicos identificados. La fórmula considera un desempeño positivo, pero existe su contraparte para cuando un desempeño es negativo así:

$$dr_i = \begin{cases} \dfrac{e_i}{o_j - \hat{y}_i}, & e_i \geq o_j \\ -1, & e_i < o_j \end{cases}$$

4. Aportes, limitaciones y posibles sesgos del estudio

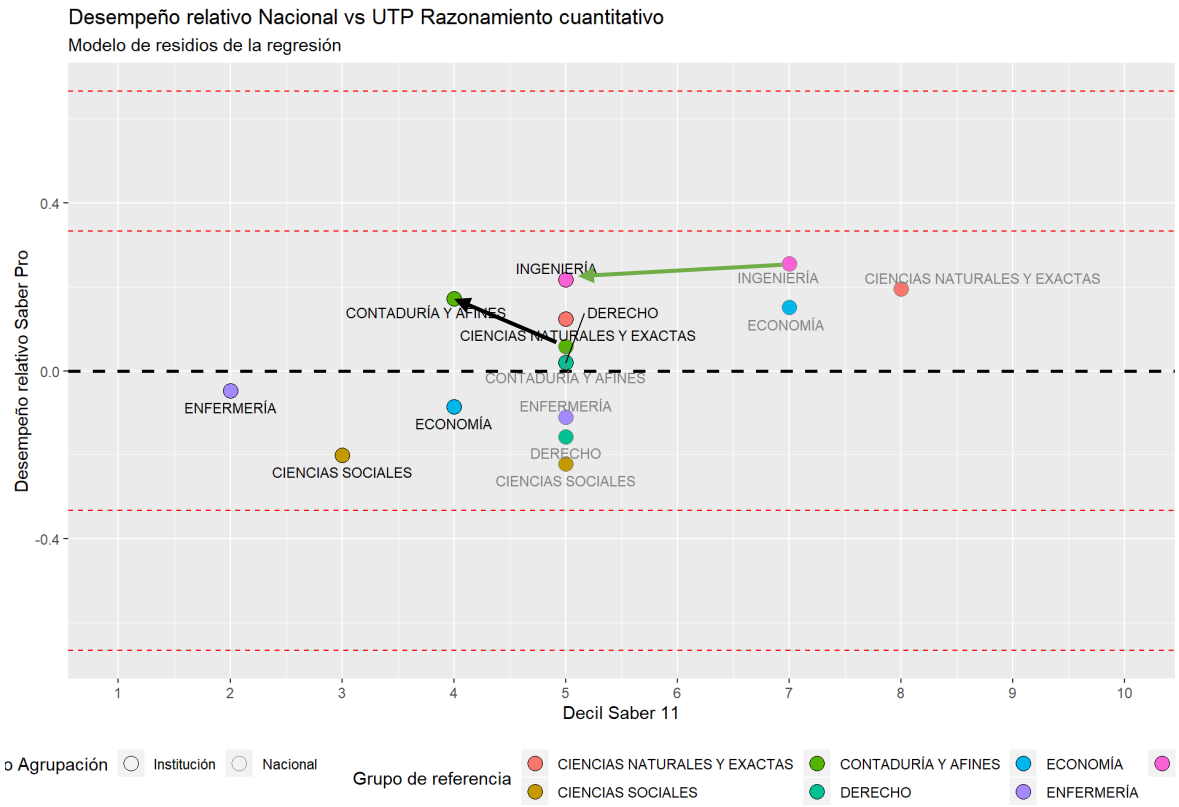
Entre los aportes más relevantes de la metodología planteada se encuentra una manera alternativa de calcular el desempeño de una Institución o Programa académico de los resultados de sus estudiantes frente al nivel de entrada del mismo a la institución, que obedece a condiciones geográficas, socioeconómicas de los estudiantes, como también a los mismos procesos de selección de los mismos.

Se estima que se tengan las siguientes limitaciones:

- Con respecto a las bases de datos de SABER PRO se presentará una disminución de la muestra disponible, a razón del emparejamiento de los estudiantes con la prueba SABER 11, eso a razón del cambio de documento de los estudiantes que genere esta situación, adicionalmente, la falta de comparabilidad entre las pruebas intrínsicamente y extrínsecamente, la primera a razón de que ha cambiado la metodología de una prueba al pasar de los años, y la segunda a razón de supuestos de homologación de las competencias entre las pruebas SABER 11 y SABER PRO.
- El modelo tendrá problemas de sensibilidad a los datos atípicos presentes en las bases de datos insumos, para lo cual se estima se debe ahondar en las rutinas de exploración de datos que permita detectar los mismos y tomar acciones de ajuste o limpieza que permita atenuarlos.

5. Interpretación del análisis

Con el desarrollo de este modelo, se construye un gráfico tipo dispersión en el cual encontramos en el eje horizontal el grupo decil Saber 11 en el que su ubico el promedio de la agrupación (en este caso grupo de referencia), y en el eje vertical el desempeño relativo calculado del promedio de desempeño en Saber Pro de la agrupación frente al punto de referencia calculado para el decil de Saber 11 correspondiente. Con texto en color gris se posicionan los resultados a nivel nacional para el grupo de referencia, y en color negro aquellos correspondientes a la institución. En este ejemplo particular, la institución tuvo un desempeño relativo mayor en el grupo de referencia de CONTUDARÍA Y AFINES frente al nivel nacional, incluso con estudiantes que en promedio tuvieron un desempeño de entrada Saber 11 menor, un segundo ejemplo sería el caso de INGENIERÍA, donde la institución tuvo un desempeño relativo similar al grupo de referencia, a pesar que los estudiantes se recibieron con un desempeño menor en la prueba Saber 11.



6. Conclusiones y trabajos futuros

Este modelamiento, permite a una institución conocer el desempeño de los resultados de sus estudiantes en diferentes niveles de agregación (Institucional, grupo de referencia o programa académico) ajustado al contexto de los estudiantes de acuerdo a su desempeño de entrada al proceso de formación. Se considera una herramienta útil para inferir los aportes del programa académico en torno a las variables de calidad de la instrucción impartida. Sin embargo, este modelamiento inicial podría robustecerse en el futuro incluyendo variables adicionales tanto académicas como sociodemográficas que permitan trascender de la descripción a la estimación de los resultados o la identificación de variables críticas de impacto en los resultados.

Por otra parte, también se considera importante poder migrar este modelamiento a una plataforma de consulta web, que permita la visualización de los resultados para cada institución bajo diferentes niveles de agregación.

7. Referencias

- Medición de los efectos de la educación superior en Colombia sobre el aprendizaje estudiantil – Informe Técnico. ICFES. 2014. Pag.5.
- OECD (2011), "El diseño de modelos de valor agregado", in La Medición del Aprendizaje de los Alumnos: Mejores Prácticas para Evaluar el Valor Agregado de las Escuelas, OECD Publishing, Paris, https://doi.org/10.1787/9789264090170-3-es.

- Martínez Arias, M., Gaviria Soto, J. y Castro Morera, M. (2009). Concepto y evolución de los modelos de valor añadido en educación. Revista de educación, ISSN 0034-8082, N° 348, 2009 (Ejemplar dedicado a: El valor añadido en educación), pags. 15-46
- Muñoz Gómez, I. (2016). Modelo de Valor Agregado: una implementación para el caso de la educación superior en Colombia. Universidad Nacional de Colombia, Facultad de Ciencias Económicas, Maestría en Ciencias Económicas.
- Kim, H., & Lalancette, D. (2013). Literature review on the value-added measurement in higher education. OECD. Retrieved from http://www.oecd.org.proxyiub.uits.iu.edu/edu/skills-beyondschool/Litterature%20Review%20VAM. pdf Lau, RR (1989). Individual and contextual influences on group identification. Social Psychology Quarterly, 52, 220-231.
- Meyliana, H. A. E. Widjaja and S. W. Santoso, "University dashboard: An implementation of executive dashboard to university," 2014 2nd International Conference on Information and Communication Technology (ICoICT), Bandung, 2014, pp. 282-287, doi: 10.1109/ICoICT.2014.6914080.

Sobre los autores

- **Jhonniers Guerrero Erazo**: Ingeniero Sanitario de la Universidad del Valle, Magister en Ingeniería Sanitaria y Ambiental de la misma universidad y Doctor en Ciencias de la Ingeniería en la universidad de Universidad Técnica de Berlín. Vicerrector Académico. jhguerre@utp.edu.co
- **Délany Ramírez del Río**: Ingeniero Industrial de la Universidad Tecnológica de Pereira, Magister en Investigación Operativa y Estadística de la Universidad Tecnológica de Pereira, Estudiante Doctorado en Ingeniería. Profesional Planeación y Profesor catedrático. delram@utp.edu.co

---

Los puntos de vista expresados en este artículo no reflejan necesariamente la opinión de la Asociación Colombiana de Facultades de Ingeniería.

Copyright © 2022 Asociación Colombiana de Facultades de Ingeniería (ACOFI)