Universidad de Santander UDES
VIGILADA MINEDUCACIÓN | SNIES 2832
ACREDITADA ALTA CALIDAD Ministerio de Educación Nacional
ACREDITACIÓN INTERNACIONAL •MULTICAMPUS• EQUAA Education Quality Accreditation Agency
Planeación Institucional

Valor Agregado Académico

como Estrategia Institucional de Seguimiento y Mejora del Aprendizaje:

Saber TyT en la Universidad de Santander – Campus Bucaramanga

Documento de trabajo



Valor Agregado Académico como Estrategia Institucional de Seguimiento y Mejora del Aprendizaje: Saber TyT en la Universidad de Santander Campus Bucaramanga.

Elaboró
Miguel Oswaldo Pérez Pulido
Director de Planeación Institucional
direccionplaneacion@udes.edu.co
Universidad de Santander

Francisco Javier León
Coordinador de Analítica Académica
coordinador.analitica@udes.edu.co

Leonardo Andrés Pinto Guarguatí
Profesional de Investigación de Mercados
prof.investmercados@udes.edu.co

Angelica Maria Calderón Plata
Directora Administrativa y Financiera
dir.administrativayfinanciera@udes.edu.co

Revisó
Krystle Forero
Directora UDESTEC
direccionudestec@udes.edu.co

DOI: https://doi.org/10.5281/zenodo.23500324

Universidad de Santander (UDES)
https://www.udes.edu.co
Versión 1

Se desea obtener información estadística académica de la institución puede realizar la consulta a través del correo direccionplaneacion@udes.edu.co

Como citar este documento: Pérez Pulido, M. O., León, F. J., Pinto Guarguatí, L. A., & Calderón Plata, A. M. (2026). Valor agregado académico como estrategia institucional de seguimiento y mejora del aprendizaje: Saber TyT en la Universidad de Santander Campus Bucaramanga (Documento de Trabajo No. 21). Planeación Institucional, Universidad de Santander. https://doi.org/10.5281/zenodo.23500324

Bucaramanga, Santander, Colombia
19 de octubre de 2026

Valor Agregado Académico como Estrategia Institucional de Seguimiento y Mejora del Aprendizaje: Saber TyT en la Universidad de Santander – Campus Bucaramanga.

Documento de trabajo No 21
Planeación Institucional

Resumen

El valor agregado (VA) académico mide cuánto aporta un programa al desempeño de sus estudiantes en Saber TyT, una vez descontado el nivel con que ingresaron, medido en Saber 11. Es una medida relativa: compara a cada programa tecnológico de UDESTEC con lo esperado para estudiantes de igual punto de partida; cero equivale al programa promedio.

Se ajustaron modelos jerárquicos de intercepto aleatorio, con estudiantes anidados en programas, para cuatro competencias de Saber TyT. El VA de cada programa es su efecto aleatorio estimado, con intervalo de 95 %. Como las dos pruebas no están equiparadas, el VA se estima de forma condicional: es lo que el programa suma o resta a la predicción de Saber TyT a partir de Saber 11. La base cruza 1.207 registros de Saber TyT con el Saber 11 de cada estudiante. Los modelos usan los 411 estudiantes de 8 programas que presentaron Saber 11 entre 2014 y 2021.

La varianza entre programas es pequeña: va de 0,1 % en competencias ciudadanas a 6,7 % en inglés, y de los 37 VA estimados solo tres se distinguen de cero con 95 % de confianza. En razonamiento cuantitativo, Supervisión de Obras Civiles queda 7,9 puntos sobre lo esperado. En inglés, Diseño Gráfico Publicitario queda 4,9 puntos por encima y Supervisión de Obras Civiles 2,8 por debajo. En lectura crítica y competencias ciudadanas ningún programa se distingue de cero.

Palabras clave: valor agregado; modelos jerárquicos; examen Saber TyT; examen Saber 11; programas tecnológicos; calidad de la educación superior.

1. Introducción

Este documento estima cuánto aporta cada programa tecnológico de UDESTEC al desempeño de sus estudiantes en las competencias genéricas Saber TyT, una vez descontado el nivel con el que llegaron, medido en Saber 11. Ese aporte es el valor agregado.

Los puntajes brutos de Saber TyT mezclan dos cosas: lo que el programa enseñó y lo que el estudiante ya traía. Un programa que recibe aspirantes con mejor Saber 11 obtendrá mejores resultados aun sin enseñar más. El VA separa ambos efectos al comparar a cada programa con el resultado esperado para estudiantes de igual punto de partida.

La pregunta tiene uso institucional directo. Los resultados de aprendizaje de los programas se documentan en la autoevaluación con fines de acreditación, y el examen de Estado de calidad de la educación superior es una de las pocas fuentes comparables entre programas e instituciones (Ley 1324 de 2009; Decreto 3963 de 2009). El VA ofrece una lectura más justa que el puntaje bruto para la planeación académica.

Objetivo general. Estimar el VA de los programas tecnológicos de UDESTEC en cuatro competencias de Saber TyT mediante un modelo lineal jerárquico de dos niveles.

Objetivos específicos.

1. Cuantificar qué parte de la variabilidad de los puntajes está entre programas y cuánta dentro de ellos.
2. Estimar el efecto de los puntajes de Saber 11 sobre cada competencia de Saber TyT.
3. Obtener el VA de cada programa con su incertidumbre y señalar dónde es distinguible de cero.
4. Verificar la solidez de los resultados y documentar las limitaciones para su uso en gestión y acreditación.

El documento usa las bases de datos extraídas del sistema prisma ICFES, dataicfes y el código y salidas de R aportadas por Planeación Institucional.

2. Marco conceptual

El valor agregado es la contribución de una institución o programa al resultado de sus estudiantes que no se explica por sus características de ingreso. Es una medida relativa: compara al programa con lo esperado dado su alumnado, no con un estándar absoluto (Rubin et al., 2004).

En Colombia el punto de partida natural es Saber 11 y el de llegada Saber TyT. Las dos pruebas no están equiparadas verticalmente, así que no se pueden restar. Por eso el VA se estima condicionalmente: se predice el puntaje de Saber TyT a partir de Saber 11 y se interpreta como VA lo que el programa añade o resta a esa predicción.

Los estudiantes de un mismo programa comparten profesores, plan de estudios y entorno, de modo que sus resultados no son independientes. Los modelos jerárquicos (multinivel) incorporan esa dependencia y estiman el efecto de cada programa como un efecto aleatorio (Raudenbush y Bryk, 2002; Goldstein, 2011). Esto aporta tres ventajas frente a una regresión con variables indicadoras por programa:

- Separan la varianza entre programas de la varianza entre estudiantes del mismo programa.
- Contraen (encogen) las estimaciones de los programas con pocos estudiantes hacia la media, lo que evita premiar o castigar al azar a los programas pequeños (Tekwe et al., 2004).

- Reportan una medida de incertidumbre por programa.

El modelo de dos niveles con intercepto aleatorio se escribe así. Para el estudiante *i* del programa *j*:

$$Y_{ij} = \beta_{0j} + \sum_k \beta_k X_{kij} + e_{ij}, \qquad e_{ij} \sim N(0, \sigma^2)$$

$$\beta_{0j} = \gamma_{00} + u_{0j}, \qquad u_{0j} \sim N(0, \tau_{00})$$

Aquí *Y* es el puntaje genérico de Saber TyT, *X* son los puntajes genéricos de Saber 11, y el VA del programa *j* es el efecto aleatorio estimado (predictor empírico de Bayes):

$$\mathrm{VA}_j = \hat{u}_{0j}, \qquad \mathrm{CCI} = \frac{\tau_{00}}{\tau_{00} + \sigma^2}$$

El coeficiente de correlación intraclase (CCI) indica qué fracción de la varianza total está entre programas. Un CCI cercano a cero significa que, para esa competencia, conocer el programa casi no ayuda a predecir el puntaje.

3. Datos

La base de datos cruza 1.207 registros de Saber TyT de estudiantes de programas tecnológicos de la UDES (aplicaciones de 2016-1 a 2024-3, 13 programas) con su resultado de Saber 11 a partir. La información utilizada en el estudio proviene de las bases de datos del Examen de Estado de Calidad de la Educación Superior, Saber Pro, disponibles en el repositorio DataIcfes, sistema PRISMA del Instituto Colombiano para la Evaluación de la Educación (Icfes, 2025). Este repositorio proporciona acceso a microdatos anonimizados de las evaluaciones estandarizadas aplicadas en Colombia, lo que permite desarrollar análisis estadísticos sobre el desempeño académico de los estudiantes de educación superior. El cruce buscó a cada estudiante en 52 archivos de Saber 11 (2000-1 a 2025-2; 16,3 millones de filas) y vinculó 1.206 de los 1.207 registros. El año en que presentaron Saber 11 va de 2000 a 2021.

Variables. Los resultados de Saber TyT usan una escala de 0 a 200 (media cercana a 100); los de Saber 11, una escala de 0 a 100 (media cercana a 50).

Tabla 1. Descripción de variables predictivas y predictoras del modelo

Variable	Descripción	Papel en el modelo
PLCSPRO	Lectura crítica, Saber TyT	Respuesta
PRCSPRO	Razonamiento cuantitativo, Saber TyT	Respuesta
PINGSPRO	Inglés, Saber TyT	Respuesta
PCCSPRO	Competencias ciudadanas, Saber TyT	Respuesta
PLCS11	Lectura crítica, Saber 11	Control (nivel 1)
PMATS11	Matemáticas, Saber 11	Control (nivel 1)
PSCS11	Sociales y ciudadanas, Saber 11	Control (nivel 1)

Variable	Descripción	Papel en el modelo
PCNS11	Ciencias naturales, Saber 11	Control (nivel 1)
PINGS11	Inglés, Saber 11	Control (nivel 1)
Program	Programa Académico	Nivel 2 (agrupador)

Muestra analítica. Solo 411 estudiantes (34 %) presentaron las pruebas saber TyT entre 2014 y 2021; los otros 796 lo presentaron entre 2000 y 2014.

Tabla 2. Descripción de la muestra analizada en los modelos

Programa	Registros en la base	En los modelos	% en los modelos en cada programa académico
Tecnología en Diseño Gráfico Publicitario	304	125	41 %
Tecnología en Supervisión de Obras Civiles	404	120	30 %
Tecnología en Gestión Industrial	177	58	33 %
Tecnología en Producción Gráfica Publicitaria	36	32	89 %
Tecnología en Comunicación Gráfica	30	28	93 %
Tecnología en Gestión Administrativa	60	22	37 %
Tecnología en Gestión de Mercadotecnia	53	18	34 %
Tecnología en Desarrollo de Aplicaciones Web	11	8	73 %
Excluidos: Tecnología Industrial (83), Mercadotecnia y Publicidad (38), Gestión Financiera y de Sistemas (9), Sistemas (1), Producción (1)	132	0	0 %

Descriptivos de la muestra analítica (n = 411).

Tabla 3. Estadísticos descriptivos de las variables analizadas en el modelo

Variable	Media	DE	Mínimo	Máximo
Lectura crítica, TyT	101,67	22,63	40	159
Razonamiento cuantitativo, TyT	94,10	22,57	41	158
Inglés, TyT	107,74	24,71	0	200
Competencias ciudadanas, TyT	97,15	27,08	0	162
Lectura crítica, Saber 11	55,15	8,43	27	76

Variable	Media	DE	Mínimo	Máximo
Matemáticas, Saber 11	55,21	9,47	28	79
Sociales y ciudadanas, Saber 11	52,21	8,80	24	76
Ciencias naturales, Saber 11	54,24	8,53	31	81
Inglés, Saber 11	47,94	9,89	20,62	94

4. Metodología

Para cada competencia se ajustaron dos modelos con el paquete lme4 de R (Bates et al., 2015), por máxima verosimilitud (`REML = FALSE`), de modo que sus devianzas sean comparables.

- **Modelo nulo (M0):** solo intercepto y efecto aleatorio de programa. Da el CCI "bruto".
- **Modelo condicional (M1):** agrega los puntajes de Saber 11 como efectos fijos. Sus efectos aleatorios son el VA.

A continuación, se presentan las competencias que actúan como predictivas y predictoras:

Tabla 4. Descripción de las variables predictivas vs predictoras.

Competencia (respuesta)	Controles de Saber 11
Lectura crítica (PLCSPRO)	Lectura crítica, Sociales y ciudadanas, Ciencias naturales
Razonamiento cuantitativo (PRCSPRO)	Matemáticas, Ciencias naturales
Inglés (PINGSPRO)	Inglés, Sociales y ciudadanas
Competencias ciudadanas (PCCSPRO)	Lectura crítica, Sociales y ciudadanas, Ciencias naturales

Los predictores entran sin centrar, por lo que el intercepto representa un puntaje con todos los controles en cero y no tiene lectura sustantiva. No afecta al VA.

Cómo se leen los resultados.

1. **VA por programa:** efecto aleatorio estimado (BLUP). Se expresa en puntos de Saber TyT sobre o bajo lo esperado; cero equivale al programa promedio de UDESTEC.
2. **Incertidumbre:** error estándar condicional del efecto aleatorio. Un VA es distinguible de cero si su intervalo de 95 % (±1,96 errores estándar) no incluye el cero.
3. **¿Hay diferencias entre programas?** Prueba de razón de verosimilitud de τ_{00} frente al mismo modelo sin efecto aleatorio. Como la varianza se prueba en el borde del espacio

paramétrico, el valor p se calcula con la mezcla 50:50 de χ^2 con 0 y 1 grados de libertad (Self y Liang, 1987).

4. **Varianza explicada:** R^2 marginal y condicional (Nakagawa y Schielzeth, 2013) y reducción proporcional de σ^2 y τ_{00} entre M0 y M1.

Se reajustaron los cuatro modelos condicionales de forma independiente (verosimilitud perfilada con un intercepto aleatorio) sobre la misma base. Coeficientes, varianzas, devianzas y VA coinciden con las salidas de R a dos decimales. Las pruebas de verosimilitud, los intervalos de 95 % y la sensibilidad son cálculos propios de esta versión.

5. Resultados

De los 32 VA estimados (8 programas × 4 competencias), tres se distinguen de cero con 95 % de confianza: uno en razonamiento cuantitativo y dos en inglés. En lectura crítica y competencias ciudadanas no hay ninguno.

5.1 Efectos de Saber 11 sobre Saber TyT

Todos los controles de Saber 11 son significativos al 5 % en las cuatro competencias. Ciencias naturales es el predictor más constante (0,49 a 0,73 puntos de TyT por punto de Saber 11), y Sociales y ciudadanas pesa más en competencias ciudadanas (0,92), como cabría esperar por afinidad de contenido.

Tabla 5. Estimación de modelos de valor agregado-Competencias Genéricas

Predictor	Lectura crítica	Razonamiento cuantitativo	Inglés	Competencias ciudadanas
Intercepto	18,33 (7,23)*	10,12 (6,37)	24,47 (8,12)**	−1,01 (8,50)
Lectura crítica, Saber 11	0,56 (0,15)***	—	0,45 (0,16)**	0,43 (0,18)*
Matemáticas, Saber 11	—	0,78 (0,13)***	—	—
Inglés, Saber 11	—	—	1,17 (0,07)***	—
Sociales y ciudadanas, Saber 11	0,32 (0,16)*	—	0,42 (0,16)**	0,92 (0,19)***
Ciencias naturales, Saber 11	0,66 (0,15)***	0,73 (0,14)***	0,67 (0,16)***	0,49 (0,18)**
Observaciones	411	411	410	411

Estimación (error estándar). * p < 0,05; ** p < 0,01; *** p < 0,001. Fuente: salidas compiladas de R.

5.2 Varianza entre y dentro de programas

La varianza entre programas es pequeña y solo es estadísticamente distinguible de cero en tres competencias. En inglés alcanza 6,7% de la varianza total; en competencias ciudadanas es prácticamente nula.

Tabla 6. Descomposición de la varianza de los puntajes de Saber TyT entre estudiantes y programas, modelos nulo y con controles de Saber 11.

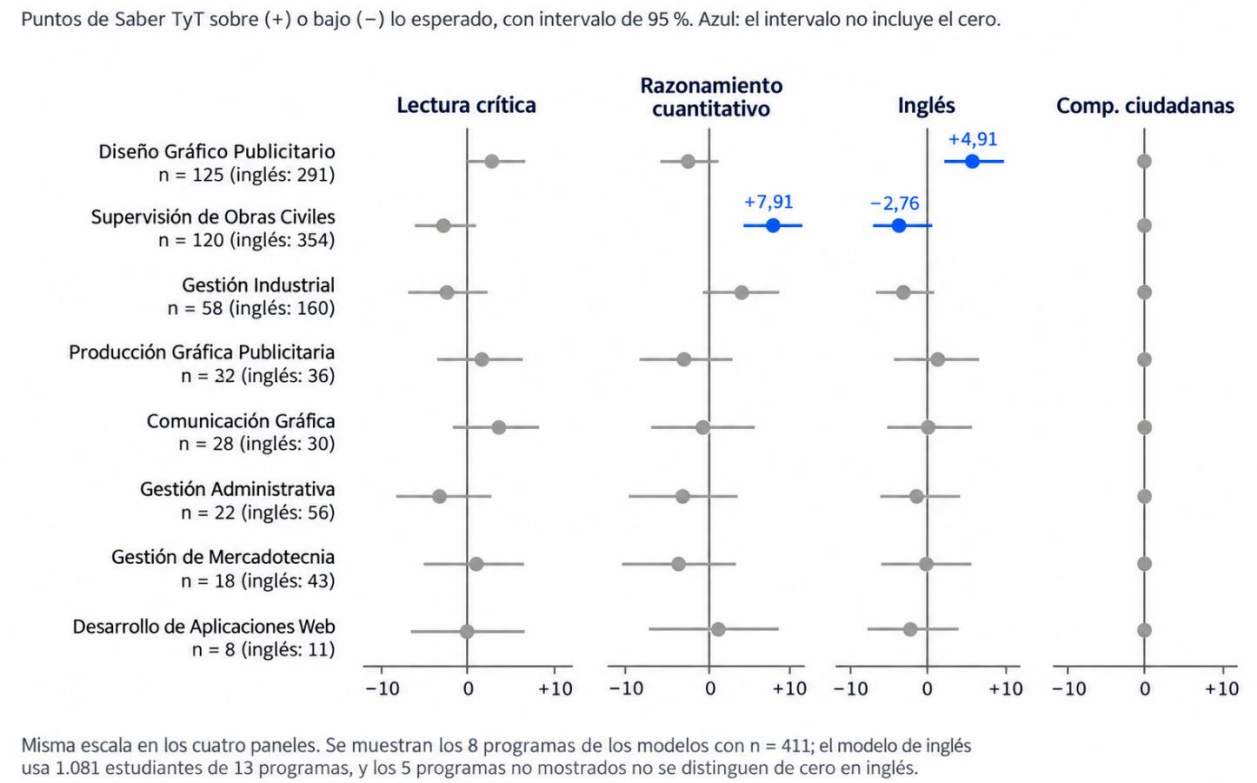
Competencia	σ^2 nulo	τ_{00} nulo	CCI nulo	σ^2 con controles	τ_{00} con controles	R^2 marginal / condicional	Prueba de τ_{00} (χ^2; p)
Lectura crítica	495,51	19,48	3,8 %	369,77	10,07	0,254 / 0,274	4,00; 0,023
Razonamiento cuantitativo	475,55	22,95	4,6 %	323,79	20,77	0,307 / 0,349	21,35; < 0,001
Inglés	561,04	40,16	6,7 %	403,46	8,41	0,296 / 0,310	21,83; < 0,001
Competencias ciudadanas	731,17	0,37	0,1 %	538,49	0,14	0,264 / 0,264	0,00; 0,50

Los controles de Saber 11 reducen la varianza entre estudiantes (σ^2) entre 25 % (lectura crítica) y 32 % (razonamiento cuantitativo), y en inglés 28 %. Sobre la varianza entre programas el efecto es dispar: la reduce 79 % en inglés, 48 % en lectura crítica y solo 9,5 % en razonamiento cuantitativo. En inglés, casi toda la diferencia entre programas del modelo nulo corresponde al inglés con que ingresan los estudiantes. En razonamiento cuantitativo, en cambio, las diferencias entre programas no se explican por quién ingresa según Saber 11. En competencias ciudadanas la reducción de τ_{00} (62 %) no es interpretable, porque parte de un valor casi cero. La prueba de τ_{00} de lectura crítica (p = 0,023) es el resultado más frágil de la tabla.

5.3 Valor agregado por programa

La figura 1 muestra el VA de cada programa con su intervalo de 95 %, en un mismo orden en las cuatro competencias (de mayor a menor n). Un intervalo que no toca el cero indica un VA distinguible de cero. Las figuras originales de la salida de R se reproducen y comentan en el anexo C.

Figura 1. Estimación de valor agregado de cada programa académico por cada competencia genérica.



Solo 3 de 32 valores agregados se distinguen de cero: 1 en razonamiento cuantitativo y 2 en inglés:

- Razonamiento cuantitativo: Obras Civiles (+7,91) es el resultado más fuerte, cerca de 0,35 desviaciones estándar de Saber TyT. Gestión Industrial (+3,70) queda cerca, pero su intervalo toca el cero.
- Inglés: Diseño Gráfico (+4,91) y Obras Civiles (−2,76) están en lados opuestos. Son magnitudes menores, de 0,2 y 0,1 desviaciones estándar, pero con intervalos más cortos porque el modelo tiene 291 y 354 estudiantes en esos programas.
- Lectura crítica y competencias ciudadanas: ningún programa se distingue. Diseño Gráfico (+2,82) queda al borde en lectura crítica.

6. Discusión

Los programas de UDESTEC se separan entre sí en razonamiento cuantitativo e inglés, y no en lectura crítica ni en competencias ciudadanas. En ninguna de las cuatro competencias el programa explica más de 7 % de la varianza; la mayor parte de la variación está entre estudiantes del mismo programa.

Competencias transversales. En competencias ciudadanas la varianza entre programas es prácticamente nula (τ_{00} = 0,14) y en lectura crítica es pequeña (CCI de 3,8 % en el modelo nulo y 2,7 % con controles); en ninguna se distingue un VA de cero. Con estos datos el VA no permite diferenciar programas en estas dos competencias y no hay base para focalizar acciones de mejora en un programa específico.

Razonamiento cuantitativo. Tecnología en Supervisión de Obras Civiles aporta cerca de 8 puntos sobre lo esperado (0,35 desviaciones estándar), y los controles de Saber 11 explican poco de esa diferencia (9,5 % de τ_{00}). Una hipótesis razonable, por contrastar con el plan de estudios y los docentes del programa, es que su plan de estudios tiene mayor carga cuantitativa que el de los programas gráficos y de mercadotecnia, cuyos VA son negativos aunque no distinguibles de cero.

Inglés. Diseño Gráfico Publicitario aporta cerca de 5 puntos sobre lo esperado y Tecnología en Supervisión de Obras Civiles cerca de 3 por debajo; Gestión Industrial (−2,6) tiene el mismo signo, pero su intervalo incluye el cero. La magnitud (0,1 a 0,2 desviaciones estándar de Saber TyT) es menor que la de razonamiento cuantitativo. Con los controles (inglés y Sociales y ciudadanas de Saber 11) la varianza entre programas baja de 40,2 a 8,4 puntos² (79 %), de modo que buena parte de la diferencia entre programas viene del inglés con que ingresan los estudiantes. Que Diseño Gráfico Publicitario quede por encima es una hipótesis para validar con su coordinación, por ejemplo, con la exposición al inglés en el plan de estudios.

Perfiles contrapuestos. La Tecnología en Obras Civiles combina el mayor VA en razonamiento cuantitativo (+7,9) con el menor en inglés (−2,8). La Tecnología en Diseño Gráfico Publicitario es el contraste: +4,9 en inglés y un VA negativo, sin distinguirse de cero, en razonamiento cuantitativo. Puede deberse a una especialización curricular, pero los datos no lo prueban, y el modelo de inglés usa otra muestra (291 estudiantes de Diseño frente a 125 en los demás), de modo que los dos VA de un mismo programa no son del todo comparables. Un mismo programa puede ser “alto” o “bajo” según la competencia: no corresponde resumirlo en un solo indicador.

Uso en gestión y mejora continua. Estos resultados sirven como evidencia complementaria de resultados de aprendizaje, no como ranking. Se recomienda reportar siempre el VA con su intervalo y el n del programa, no calificar a ningún programa como “mejor” o “peor” a partir de

un VA no distinguible de cero, y no usar el VA como indicador de desempeño docente: mide el programa en conjunto y de forma relativa a UDESTEC, no frente a un estándar externo.

7. Conclusiones

El valor agregado de los programas tecnológicos de UDESTEC es medible y distinto de cero solo en razonamiento cuantitativo e inglés; en lectura crítica y competencias ciudadanas los programas no se diferencian de forma sólida.

1. La varianza entre programas es 3,8 % (lectura crítica), 4,6 % (razonamiento cuantitativo) y 6,7 % (inglés) de la varianza total; la mayor parte de la variación está entre estudiantes del mismo programa. En competencias ciudadanas es prácticamente nula (0,1 %).
2. Saber 11 predice Saber TyT en las cuatro competencias. En inglés pesa sobre todo el inglés de Saber 11 (1,17 puntos por punto); en las otras tres, Ciencias naturales es el predictor más constante (0,49 a 0,73).
3. Solo tres de 37 VA se distinguen de cero (95 %): Obras Civiles en razonamiento cuantitativo (+7,9) y, en inglés, Diseño Gráfico (+4,9) y Obras Civiles (−2,8). En lectura crítica y competencias ciudadanas ningún programa se distingue.
4. En inglés, los controles de Saber 11 reducen la varianza entre programas de 6,7 % a 2,0 % de la total (τ_{00} baja 79 %), de modo que buena parte de la diferencia entre programas corresponde al inglés con que ingresan los estudiantes.
5. El VA es evidencia complementaria de resultados de aprendizaje. Se sugiere reportarlo siempre con su intervalo y el n del programa, y no usarlo como ranking ni como indicador de desempeño docente.

8. Referencias

Bates, D., Mächler, M., Bolker, B., y Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Colombia, Congreso de la República. (2009). *Ley 1324 de 2009, por la cual se fijan parámetros y criterios para organizar el sistema de evaluación de resultados de la calidad de la educación, se dictan normas para el fomento de una cultura de la evaluación, en procura de facilitar la inspección y vigilancia del Estado y se transforma el Icfes*.

Colombia, Presidencia de la República. (2009). *Decreto 3963 de 2009, por el cual se reglamenta el Examen de Estado de Calidad de la Educación Superior*.

Goldstein, H. (2011). *Multilevel statistical models* (4.ª ed.). Wiley.

Instituto Colombiano para la Evaluación de la Educación. (2025, 26 de junio). Data Icfes [Repositorio de datos]. https://www.icfes.gov.co/investigaciones/data-icfes/

Nakagawa, S., y Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Raudenbush, S. W., y Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2.ª ed.). Sage.

Rubin, D. B., Stuart, E. A., y Zanutto, E. L. (2004). A potential outcomes view of value-added assessment in education. *Journal of Educational and Behavioral Statistics, 29*(1), 103–116. https://doi.org/10.3102/10769986029001103

Self, S. G., y Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association, 82*(398), 605–610. https://doi.org/10.1080/01621459.1987.10478472

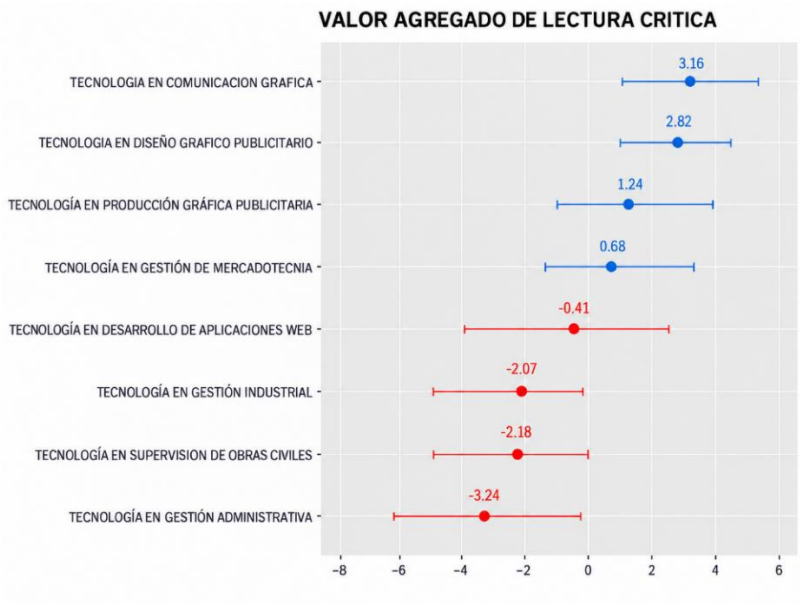
Tekwe, C. D., Carter, R. L., Ma, C.-X., Algina, J., Lucas, M. E., Roth, J., Ariet, M., Fisher, T., y Resnick, M. B. (2004). An empirical comparison of statistical models for value-added assessment of school performance. *Journal of Educational and Behavioral Statistics, 29*(1), 11–36. https://doi.org/10.3102/10769986029001011

Anexos

Anexo A. Figuras originales de la salida de R

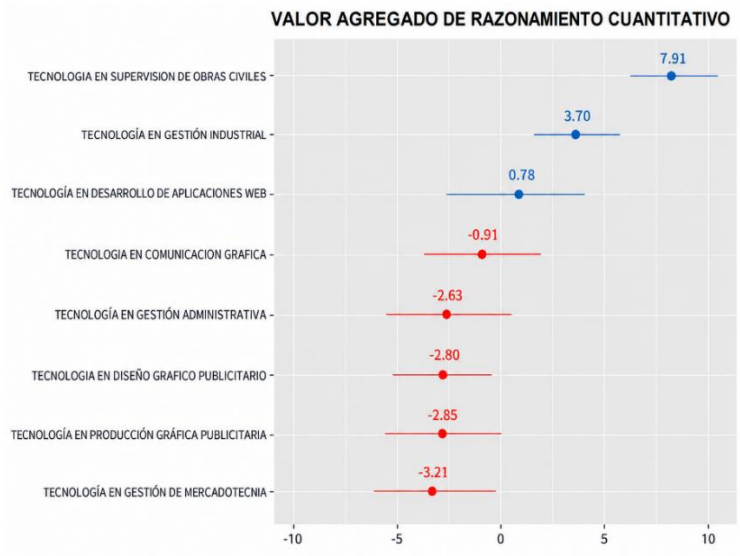
Las cuatro figuras provienen de `sjPlot::plot_model(type = "re")` y ordenan los programas de mayor a menor VA (azul: positivo; rojo: negativo). Reproducen los mismos VA de este documento, pero sus barras equivalen a ±1 error estándar y no a 95 % (sección 7, alerta 1). Sirven para ilustrar el orden de los programas; no para inferir diferencias. Se recomienda mantenerlas en anexo y usar en el cuerpo la figura de la sección 5.3.

C1. Lectura crítica



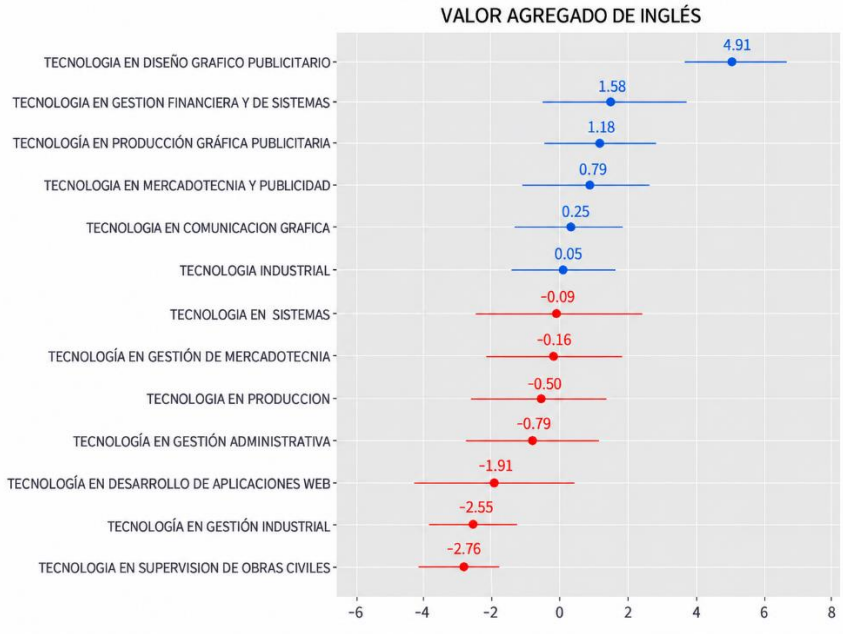
Los VA van de −3,24 (Gestión Administrativa) a +3,16 (Comunicación Gráfica). La distancia entre el programa más alto y el más bajo es de 6,4 puntos, 0,28 desviaciones estándar de Saber TyT. Con 95 % todos los intervalos incluyen el cero.

C2. Razonamiento cuantitativo



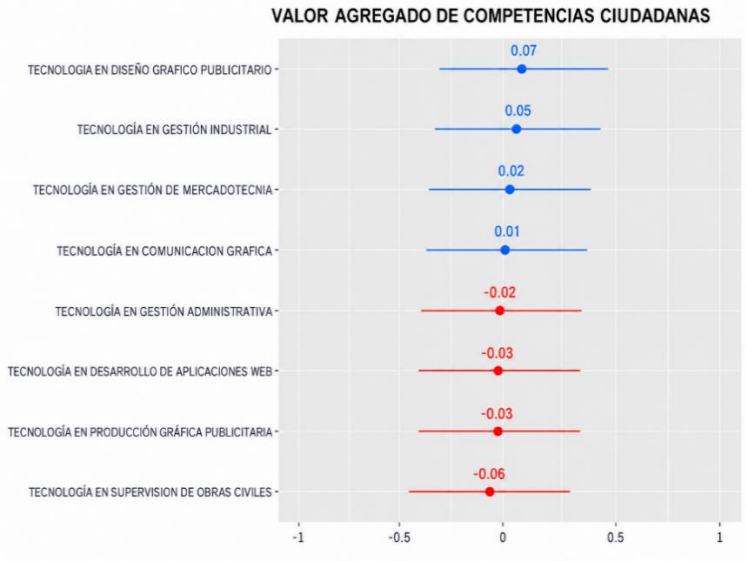
Supervisión de Obras Civiles (+7,91) se separa del resto; Gestión Industrial (+3,70) y Aplicaciones Web (+0,78) también son positivos, y los cinco programas restantes, negativos. La amplitud es de 11,1 puntos (0,49 desviaciones estándar). Solo Obras Civiles es distinguible de cero con 95 %.

C3. Inglés



Diseño Gráfico Publicitario (+4,91) encabeza el orden y Supervisión de Obras Civiles (−2,76) lo cierra. La amplitud es de 7,7 puntos (0,31 desviaciones estándar). Con 95 % solo esos dos programas se distinguen de cero. Sistemas y Producción, con un estudiante cada uno, quedan cerca de cero por contracción

C4. Competencias ciudadanas



Los VA están entre −0,06 y +0,07 puntos, con un eje de −1 a 1. El color separa a los programas por el signo aunque las diferencias son nulas (τ_{oo} = 0,14; CCI ≈ 0). Esta figura puede inducir a error si se publica sola; conviene citarla junto con la nota de que no hay diferencias entre programas.