

## Uso Software Lenguaje Imactiva y SIMCE 2022

### Resumen

Un análisis estadístico de los resultados de la prueba SIMCE Lenguaje rendida el año 2022 por los 4tos Básicos muestra los siguientes aspectos relevantes a nivel agregado.

- Los principales determinantes de los resultados promedios del SIMCE son: la dependencia administrativa, el nivel de ingresos y la cantidad de días en que el establecimiento estuvo abierto durante la pandemia.
- Un establecimiento público con Ingresos Bajos y que estuvo cerrado el mayor número de días obtuvo un puntaje promedio de 242.8. Si el establecimiento era particular subvencionado obtuvo 4.7 puntos adicionales y si era particular pagado, 11.6 puntos adicionales.
- Si era de Ingresos Medios obtuvo 13.1 puntos adicionales y si era de Ingresos Altos 38.4 puntos más. Se debe advertir que estas variables son agregadas y pueden indicar varios efectos que están resumidos por lo que se debe ser cauto al extraer conclusiones.
- Un establecimiento que estuvo abierto el máximo número de días durante la pandemia obtuvo 8.4 puntos adicionales.
- Respecto a los software de lenguaje de Imactiva, un 50% de uso por parte de los estudiantes de un establecimiento implica un aumento de 8 puntos en el promedio del SIMCE. Entre los establecimientos que usaron los software, el uso promedio alcanzó un 32,8% lo cual les habría permitido un aumento de 5,2 puntos en el SIMCE.

### Variables

El siguiente análisis está basado en los resultados para el SIMCE Lenguaje aplicado a los 4tos Básicos de todo el país por el Ministerio de Educación en noviembre de 2022. Los resultados promedio por establecimiento fueron publicados en Octubre de 2023 por la Agencia de Calidad de la Educación y están disponibles en <https://informacionestadistica.agenciaeducacion.cl/#/bases>.

Todas las variables usadas en el análisis provienen de la base de datos SIMCE anterior con excepción de las del uso de los software de Imactiva y los días de apertura del establecimiento. La siguiente es la definición de cada variable de la base de datos SIMCE:

Simce\_Lenguaje: promedio SIMCE Lenguaje 4tos Básicos por RBD

Particular\_Subvencionado: valor 1 si el establecimiento es particular subvencionado y 0 si no lo es.

Particular\_Pagado: valor 1 si el establecimiento es particular pagado y 0 si no lo es.

Ingreso\_Alto: valor 1 si establecimiento es catalogado de ingreso alto, 0 si no lo es.

Ingreso\_Medio: valor 1 si establecimiento es catalogado de ingreso medio alto, medio o medio bajo, 0 si no es catalogada en ninguna de ellas.

Las siguientes variables no pertenecen a la base de datos SIMCE:

Dias\_Abierto: suma de días en que el establecimiento estuvo abierto entre octubre de 2020 y diciembre de 2021 de acuerdo a las estadísticas del Ministerio de Educación en <https://datosabiertos.mineduc.cl/estado-apertura-de-establecimientos/>

Uso\_Bartolo: porcentaje promedio de uso de los software Imactiva Lenguaje (Aprendiendo a Leer con Bartolo, Leer Te Da Poder e Historias Fabulosas con Bartolo) para los estudiantes de los establecimientos que rindieron el SIMCE 2020. La medición toma en cuenta el uso de los software en 2020, 2021 y 2022, es decir, cuando los estudiantes estaban en 2do. Básico, en 3ero Básico y en 4to. Básico respectivamente.

La muestra consiste de 6555 establecimientos que rindieron la prueba en 2022. Se excluyeron 27 establecimientos cuya variable de Días Abierto presentaba más de un valor por RBD con lo que la muestra definitiva contiene 6527 observaciones.

A continuación se presenta un resumen de las variables.

Statistic	N	Mean	St. Dev.	Min	Max
Simce_Lenguaje	6,527	260.657	25.275	149	372
Particular_Subvencionado	6,527	0.419	0.493	0	1
Particular_Pagado	6,527	0.072	0.258	0	1
Ingreso_Medio	6,527	0.696	0.460	0	1
Ingreso_Alto	6,527	0.075	0.264	0	1
Dias_Abierto	6,527	43.673	18.717	0	125
Uso_Bartolo	6,527	0.006	0.058	0.000	0.981

Como se puede observar, el puntaje Simce tiene un mínimo de 149 y un máximo de 372 puntos y una media de 260,7 para todos los establecimientos. Un 50% de los establecimientos tiene un puntaje promedio menor que 262 y el 75% uno menor que 278.

Del total de la muestra un 41,9% corresponde a colegios Particular Subvencionados (2733), un 7.2% a Particular Pagados (468) y un 51% a establecimientos públicos: Corporaciones Municipales (610), DAEM (2295) y Servicios Locales (SLEP) (421).

<b>Dependencia</b>		
<i>Particular Subvencionado</i>	2733	41.87%
<i>Particular Pagado</i>	468	7.17%
<i>DAEM</i>	2295	35.16%
<i>Corporación Municipal</i>	610	9.35%
<i>SLEP</i>	421	6.45%
<b>Total</b>	<b>6527</b>	<b>100.00%</b>

Por nivel de ingreso, los establecimientos son catalogados en la base de datos del SIMCE en Ingresos Altos, Medios y Bajos. La distribución de los establecimientos por nivel de ingreso se muestra en la siguiente tabla. El 70% es de Ingresos Medios y el 23% de Ingresos Bajos.

<b>Nivel de Ingreso</b>		
<i>Ingreso Alto</i>	492	7.54%
<i>Ingreso Medio</i>	4541	69.57%
<i>Ingreso Bajo</i>	1494	22.89%
<b>Total</b>	<b>6527</b>	<b>100.00%</b>

Respecto al uso de los software de lenguaje de Imactiva (Uso\_Bartolo), 129 colegios hicieron uso de los software. Entre los establecimientos que usaron los softwares, el uso fluctuó entre un 0,4% y un 98,1% de uso promedio por colegio con una media de 31,28%.

La siguiente tabla muestra una comparación entre la muestra completa y aquellos colegios que usaron Bartolo. Se desprende que las muestras son similares con un mayor porcentaje de colegios públicos y de Ingresos Medios entre los establecimientos que usaron Bartolo.

<b>Muestra Uso Software vs Total</b>		
	Total Muestra	Uso Bartolo
Particular Subvencionado	41.87%	36.43%
Particular Pagado	7.17%	0.00%
Público	50.96%	63.57%
Ingreso Alto	7.54%	0.00%
Ingreso Medio	69.57%	82.17%
Ingreso Bajo	22.89%	17.83%
Simce Max	372	337
Simce Min	149	193
Simce Promedio	260.66	263.89

Esto es importante puesto que los resultados obtenidos más adelante no se deben a la elección de una muestra muy distinta al total de la población.

### **Análisis Estadístico**

La siguiente Tabla muestra el resultado de un modelo lineal entre las variables. La variable dependiente es el puntaje promedio por RBD de la prueba Simce Lenguaje rendida por los 4tos básicos el año 2022. Naturalmente, el grueso de la varianza está explicado por las características individuales de los estudiantes y sus familias y por las características de sus establecimientos (profesores, directores, infraestructura, etc.). A nivel agregado se ha observado en varias investigaciones<sup>1</sup> que la dependencia del establecimiento es un determinante importante de los resultados académicos. De la misma manera tanto en Chile como a nivel de otros países el nivel de ingreso es un predictor clave de los puntajes en pruebas estandarizadas de rendimiento.

Como hemos usado variables para establecimientos Particulares Subvencionados y Particulares Pagados, el resto de los colegios se agrupa en establecimientos públicos, lo que incluye a los DAEM, a las Corporaciones Municipales y a los SLEP. Debemos notar que los resultados que se presentan a continuación no cambian si es que diferenciamos a los establecimientos dentro de esta agrupación, es decir, los resultados no son significativamente distintos para los DAEM, Corporaciones o SLEP.

Respecto a los ingresos hemos agrupado los Ingresos Medios Altos, Ingresos Medios Medios e Ingresos Medios Bajos en una sola variable de Ingresos Medios. Por lo tanto, usamos las categorías de Ingreso

<sup>1</sup>Ver por ejemplo:

Gallego, F.A., 2002. Competencia y resultados educativos: teoría y evidencia para Chile. Cuadernos de Economía, 39(118), pp.309-352.

Alejandra Mizala & Pilar Romaguera, 2000. "Determinación de factores explicativos de los resultados escolares en educación media en Chile," Documentos de Trabajo 85, Centro de Economía Aplicada, Universidad de Chile.

PISA 2018 Results: <https://www.oecd.org/pisa/publications/pisa-2018-results.htm>.

Alto e Ingreso Medio. Lo anterior implica que el resultado por defecto corresponde a los establecimientos públicos y al Ingreso Bajo.

La variable `Días_Abierto` persigue aislar el efecto de las pérdidas de horas de clase durante el Covid, efecto documentado en muchas investigaciones al respecto<sup>2</sup>.

Finalmente incluimos en el análisis la variable `Uso Bartolo` que resume el uso de los softwares de Lenguaje de Imactiva y que fue descrita anteriormente.

---

<sup>2</sup> Jack, Rebecca, Clare Halloran, James Okun, and Emily Oster. 2023. "Pandemic Schooling Mode and Student Test Scores: Evidence from US School Districts." *American Economic Review: Insights*, 5 (2): 173-90.

**TABLA N°1**

*Dependent variable:*

	Simce_Lenguaje		
	OLS	coefficient	
	(1)	(2)	(3)
Constant	242.833*** (0.839)	242.833*** (0.899)	242.833*** (0.920)
Particular_Subvencionado	4.674*** (0.600)	4.674*** (0.615)	4.674*** (0.117)
Particular_Pagado	11.558*** (2.956)	11.558*** (2.155)	11.558*** (3.193)
Ingreso_Medio	13.131*** (0.679)	13.131*** (0.801)	13.131*** (0.004)
Ingreso_Alto	38.409*** (2.915)	38.409*** (2.130)	38.409*** (3.814)
Uso_Bartolo	15.883*** (4.700)	15.883*** (3.414)	15.883*** (0.149)
Dias_Abierto	0.067*** (0.015)	0.067*** (0.015)	0.067*** (0.023)
Observations	6,527		
R <sup>2</sup>	0.231		
Adjusted R <sup>2</sup>	0.231		
Residual Std. Error	22.171 (df = 6520)		
F Statistic	326.958*** (df = 6; 6520)		

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

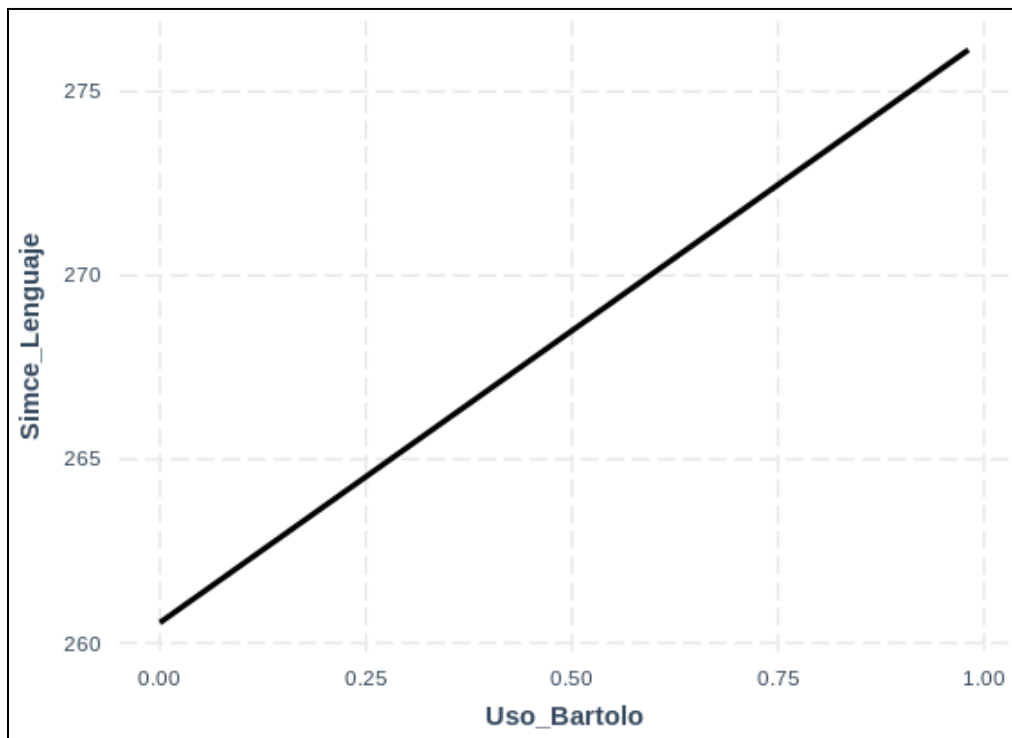
En la columna (1) se muestran los resultados de Mínimos Cuadrados Ordinarios. En paréntesis la desviación estándar. Todos los parámetros estimados tienen una alta significancia estadística, la probabilidad de que cualquiera de los efectos sea nulo es menos de un 1%.

Un establecimiento público con Ingreso Bajo y que estuvo cerrado el mayor número de días obtuvo un puntaje promedio de 242.8. Si el establecimiento es Particular Subvencionado obtiene 4.7 puntos adicionales y si es particular pagado, 11.6. Si es de Ingresos Medios obtiene 13.1 puntos adicionales y si es de Ingresos Altos, 38.4. Se debe advertir que estas variables son agregadas y pueden indicar varios efectos que están resumidos por lo que se debe ser cauto al extraer conclusiones.

Un establecimiento que estuvo abierto el máximo número de días durante la pandemia obtuvo 8.4 puntos adicionales.

El coeficiente del % promedio de uso de los softwares implica que un 100% de uso por parte de los estudiantes de un establecimiento implica un aumento de 15.9 puntos en el promedio del Simce. Se debe notar que 100% de uso significa que el estudiante realiza todos los desafíos del software lo cual no debiera implicar más de 30 minutos semanales durante el año académico. Entre los establecimientos que usaron los software, el uso promedio alcanzó un 32,8% lo cual les habría permitido un aumento de 5,2 puntos en el SIMCE.

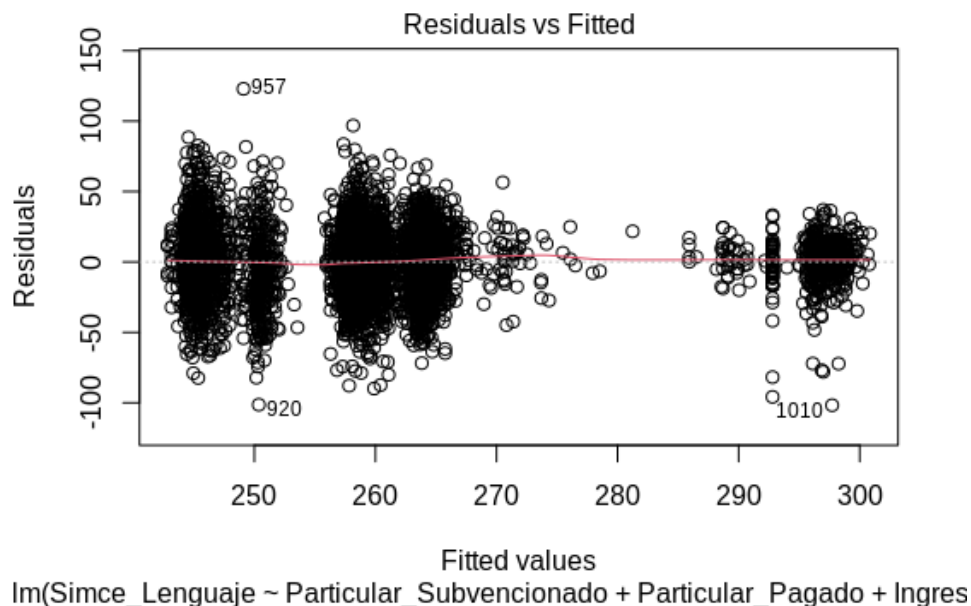
El siguiente Gráfico muestra la relación estimada entre el Uso\_Bartolo y el resultado del Simce Lenguaje para el valor promedio de las variables.



Por ejemplo, un 25% de uso promedio habría permitido al establecimiento promedio pasar de 260 a 264 puntos en el Simce, un 50% de uso le habría permitido llegar a 268.

### Errores Estándar Robustos

Los resultados anteriores sugieren un problema de heteroscedasticidad en los errores, lo cual se observa en el siguiente Gráfico que muestra los errores ordenados por los valores predichos por la regresión. Los errores tienden a disminuir con el nivel del puntaje, siendo manifiestamente menores para los mayores valores del Simce que corresponden a los establecimientos particulares pagados. Esto puede distorsionar la significancia estadística de los coeficientes haciendo que parezcan más relevantes de lo que en realidad son.



Las columnas (2) y (3) de la Tabla N°1 intentan corregir este problema. En la columna (2) se presentan los errores estándar robustos calculados con el método de White<sup>3</sup> y en la (3) se agrupan por dependencia. Se puede observar que estas correcciones hacen más precisa la estimación del Uso Bartolo, el error estándar baja de 4.7 a 0.15. Por lo anterior descartamos que la significancia de los coeficientes se deba a heteroscedasticidad.

<sup>3</sup> White, H. (1980) A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity. *Econometrica*, 48, 817-838.



## Causalidad

Otro problema que podría tener la estimación anterior es lo que se denomina "causalidad inversa". Se podría pensar que los establecimientos que usan Bartolo pueden tener una característica que está correlacionada con los resultados del Simce y que también hace que usen Bartolo. Por ejemplo, sus profesores pueden estar más motivados lo cual se refleja en los resultados de sus estudiantes y en que usen más elementos pedagógicos como Bartolo. El Uso de Bartolo sería en realidad un indicador de la motivación de los profesores y no representaría un efecto real.

Si los establecimientos que usan Bartolo y los que no lo usan fueran muy distintos, el riesgo de una causalidad inversa es mayor. Como observamos, este no es el caso en esta muestra. Una manera estadísticamente correcta de solucionar este problema es el uso de variables instrumentales. Para ello se requiere de una variable que esté correlacionada con el Uso Bartolo, pero no con el error o la variable dependiente (Simce). En general encontrar una variable con esas características es complejo. Afortunadamente en este caso existe una variable que tiene las propiedades estadísticas requeridas. El precio de los softwares de Imactiva es fijo por establecimiento y no por número de usuarios. Así, mientras más estudiantes tenga el establecimiento, más barato resulta el software por alumno. Es razonable suponer que a mayor número de usuarios, por lo tanto, más probable es que, si todo lo demás permanece constante, el establecimiento contrate los software. Al mismo tiempo, uno no debería esperar que el resultado del Simce se vea afectado por el número de usuarios de los software.

<b>Correlación entre las Variables</b>			
	Simce_Lenguaje	Uso_Bartolo	N_Usuarios
Simce_Lenguaje	1	0.030	0.032
Uso_Bartolo	0.030	1	0.717
N_Usuarios	0.032	0.717	1

La Tabla N°2 muestra la correlación entre las variables donde N\_Usuarios es el número de usuarios de los software. Se puede observar que en principio se cumplen los supuestos para una variable instrumental. Intuitivamente lo que se busca es que el coeficiente de Uso Bartolo solo refleje las variaciones en el instrumento y no toda la variación que puede contener el efecto de otras variables que influyan en dicha variable y en el Simce.

La Tabla N°3 presenta los resultados de una estimación con variables instrumentales. El coeficiente del Uso Bartolo es mayor que en los resultados anteriores, pero es menos preciso. Esto puede deberse a errores en la medición de las variables.

Los test de diagnóstico sugieren que N\_Usuarios es un buen instrumento, pero que el problema de endogeneidad o causalidad inversa no es importante.

TABLA N°3

<i>Dependent variable:</i>			
	Simce_Lenguaje		
	<i>instrumental</i>	<i>coefficient</i>	
	<i>variable</i>	<i>test</i>	
	(1)	(2)	(3)
Constant	242.822*** (0.839)	242.822*** (0.913)	242.822*** (1.230)
Uso_Bartolo	20.106*** (6.559)	20.106*** (5.785)	20.106*** (0.111)
Particular_Subvencionado	4.691*** (0.600)	4.691*** (0.620)	4.691*** (0.168)
Particular_Pagado	11.585*** (2.956)	11.585*** (2.106)	11.585*** (3.539)
Ingreso_Medio	13.107*** (0.680)	13.107*** (0.803)	13.107*** (0.003)
Ingreso_Alto	38.405*** (2.915)	38.405*** (2.055)	38.405*** (4.336)
Dias_Abierto	0.066*** (0.015)	0.066*** (0.015)	0.066** (0.031)
Observations	6,527		
R <sup>2</sup>	0.231		
Adjusted R <sup>2</sup>	0.230		
Residual Std. Error	22.172 (df = 6520)		

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Las columnas (2) y (3) corrigen por los mismos potenciales problemas de error estándar. La columna (3) modela los errores en función a la dependencia y presenta una estimación del Uso\_Bartolo extraordinariamente precisa.

Dado los antecedentes creemos que la estimación con un efecto de los software de 15.9 puntos del Simce es la más correcta estadísticamente.