



Aprender 2023 | Educación Primaria

➤ La incidencia de los factores familiares, escolares y personales en los resultados académicos de Lengua



**Ministerio de  
Capital Humano**  
República Argentina

**Secretaría  
de Educación**

# Autoridades

**PRESIDENTE**

Javier Gerardo Milei

**VICEPRESIDENTE**

Victoria Eugenia Villarruel

**JEFE DE GABINETE DE MINISTROS**

Guillermo Alberto Francos

**MINISTRA DE CAPITAL HUMANO**

Sandra Viviana Pettovello

**SECRETARIO DE EDUCACIÓN**

Carlos Horacio Torrendell

**SUBSECRETARIA DE INFORMACIÓN Y EVALUACIÓN EDUCATIVA**

María Ángela Cortelezzi



**Ministerio de  
Capital Humano**  
República Argentina

**Secretaría  
de Educación**



# Créditos

Este material fue producido por la Subsecretaría de Información y Evaluación Educativa

**DIRECTORA NACIONAL DE ANÁLISIS ESTRATÉGICO DE DATOS Y DIFUSIÓN DE LA INFORMACIÓN EDUCATIVA**

Florencia Sourrouille

**DIRECTORA NACIONAL DE EVALUACIÓN DE APRENDIZAJES Y ENSEÑANZAS**

Magdalena Benvenuto

**ELABORACIÓN DEL INFORME COORDINACIÓN GENERAL**

Florencia Sourrouille

**DISEÑO DE CONTENIDO, ANÁLISIS DE LA INFORMACIÓN, Y REDACCIÓN**

Augusto Hoszowski

Gabriela Itzcovich

**DISEÑO**

Natalia Acosta

**EDICIÓN**

Mercedes Mac Donnell



# Índice

<b>Introducción</b> .....	5
<b>1 / Metodología</b> .....	7
<b>2 / Resultados Aprender 2023. Análisis bivariado</b> .....	11
<b>3 / Determinantes del desempeño a nivel de los estudiantes</b> .....	15
<b>3.1 /</b> Modelo I: Características del alumno y sector de gestión .....	15
<b>3.2 /</b> Modelo II: Incorporación de variables del hogar y trayectorias .....	17
<b>3.3 /</b> Modelo III: Incorporación de variables de contexto escuela y contexto aula .....	21
<b>4 / Análisis multinivel</b> .....	27
<b>4.1 /</b> Modelos de análisis .....	27
<b>4.2 /</b> Dispersión entre escuelas y al interior de las escuelas .....	30
<b>Conclusiones</b> .....	34
<b>Bibliografía</b> .....	35
<b>Anexo metodológico</b> .....	36

## ► Introducción

Las encuestas de evaluación educativa estandarizadas<sup>1</sup> permiten analizar el desempeño de los estudiantes junto con las características de sus familias, el contexto educativo y los atributos propios de cada alumno. Por lo tanto, aportan evidencia acerca de los factores que pueden influir en los resultados y su importancia relativa.

En Argentina, inicialmente a través del Operativo Nacional de Evaluación (ONE), desde hace más de tres décadas se realizan evaluaciones estandarizadas de primaria y secundaria.

A partir de 2016 se lleva a adelante la Evaluación Aprender, un dispositivo de carácter nacional que releva información sobre el aprendizaje de los estudiantes en áreas prioritarias de la educación obligatoria. Esta evaluación consiste en un cuestionario cognitivo con 24 ítems (hay seis *modelos* de cuestionarios existentes, que totalizan 72 ítems cognitivos distintos) y un cuestionario de contexto donde se indagan características de los estudiantes y de sus hogares (educación formal de la madre y padre, bienes en el hogar, actividades fuera del ámbito escolar, entre otras). El operativo incluye además cuestionarios complementarios a docentes y directivos. La información obtenida permite realizar análisis estadísticos con diferentes niveles de agregación, contextualizar los logros de los alumnos, y aportar evidencia sobre cuáles pueden ser algunos de los determinantes de los desempeños educativos.

En 2023 se aplicó la Evaluación Nacional Aprender a los estudiantes de 6° grado de todo el país, con foco en las áreas de Lengua y Matemática<sup>2</sup>.

El informe de resultados<sup>3</sup> muestra que, para el área de Lengua, tres de cada diez estudiantes aún no alcanzan el nivel mínimo esperado. En el área de Matemática, el 50% de los estudiantes tampoco alcanza un nivel de desempeño aceptable. El informe presenta también un análisis descriptivo de posibles factores asociados al desempeño, considerando tanto variables intra como extraescolares.

Tomando en cuenta estos antecedentes, y partiendo del supuesto de que el aprendizaje de los niños es resultado de la interacción entre diversos elementos que conviven en los sistemas educativos, se plantea como objetivo principal analizar el efecto que tienen las características de los estudiantes y sus hogares, así como también del contexto escolar, en los resultados académicos de Lengua. Se trabaja específicamente con los resultados de Lengua dado que es un área transversal, en tanto condición necesaria para el aprendizaje de las demás materias, y porque la política de alfabetización es una prioridad en la agenda educativa.

Para dar cuenta del objetivo planteado, se elabora en primer lugar un modelo de regresión clásico que intenta establecer cuáles son los factores con mayor poder explicativo respecto de la variabilidad de resultados en Lengua. A diferencia de otros ámbitos de aplicación, en este caso no se pretende predecir el puntaje sino establecer un orden de magnitud de los coeficientes y, más precisamente, determinar el impacto de cada variable o *dimensión explicativa* en el puntaje. En segundo lugar, considerando la posibilidad de abordar los diferentes niveles de agregación de la información, se ensaya un modelo de análisis multinivel que permite cuantificar el “efecto colegio”, no captado por las variables explicativas, en los resultados alcanzados; asimismo, este análisis brinda la posibilidad de interpretar cuánto se va ganando en la explicación a partir de la incorporación de las nuevas variables.

<sup>1</sup> Se alude a las pruebas Aprender, PISA, ERCE, entre otras.

<sup>2</sup> La evaluación tuvo un carácter censal y participaron 19.272 escuelas (93,9%) y 614.817 estudiantes (82,2%) tanto del sector de gestión estatal como privada y de los ámbitos rural y urbano.

<sup>3</sup> Disponible en <https://www.argentina.gob.ar/educacion/evaluacion-informacion-educativa/aprender/aprender-2023>

Al respecto, es importante considerar las advertencias de algunos especialistas<sup>4</sup> acerca de las limitaciones de este tipo de modelizaciones; por ejemplo, la imposibilidad de incorporar algunas dimensiones o variables que podrían ser relevantes pero que no suelen incluirse ya sea porque son temáticas *sensibles* o porque resultan muy difíciles de medir (entre otras: estrategias didácticas empleadas por los docentes, clima propicio al aprendizaje en el aula, riqueza del lenguaje del hogar, nivel económico de la familia<sup>5</sup>, valorización del estudio por parte de la familia, alimentación del niño).

Más allá de esas consideraciones, un operativo de estas características es una **fuerza** de información invaluable, por lo que se espera que el presente informe, organizado en cuatro capítulos, constituya un aporte a la comprensión de los factores que inciden en los desempeños académicos de los estudiantes.

En el primer capítulo se explicitan las principales definiciones metodológicas consideradas. En el segundo capítulo se presenta un análisis descriptivo bivariado de los resultados de Aprender 2023 de 6° grado en el área de Lengua, considerando las principales variables utilizadas en la modelización. En el tercer capítulo se ajusta un modelo de regresión múltiple clásico. Finalmente, en el cuarto capítulo, se presenta el modelo multinivel, que compara los resultados obtenidos con el modelo clásico.

---

4 Una discusión sobre las lecturas de las evaluaciones estandarizadas, enfocado a PISA pero que puede aplicarse a otras evaluaciones estandarizadas, se halla en OCDE 2016, III. Qué hacer con PISA (Sandoval-Hernández 2016).

5 En este caso se consideran variables proxy del nivel económico del hogar.

## ► 1. Metodología

Se presentan a continuación las definiciones acerca del universo de estudio, el modelo de análisis, las variables seleccionadas y las fuentes de información.

El universo de estudio lo constituyen los alumnos de 6° grado de nivel primario participantes de la Evaluación Aprender, con puntaje en la prueba de Lengua<sup>6</sup>, con edad y sexo declarado<sup>7</sup>. Se trabaja con el 94% del total de casos.

**Tabla 1. Cobertura de casos de Aprender 2023**

	N	%
Con puntaje en Lengua	603.593	100
Con puntaje en Lengua y con edad y sexo declarado	568.883	94,2

**Fuente:** Evaluación Aprender 2023, DNEIEE-REFCEE | SSIEE | Secretaría de Educación | Ministerio de Capital Humano

Como en otras evaluaciones similares, Aprender procesa los resultados mediante la *Teoría de Respuesta al Ítem (TRI)*. La TRI asume que cada alumno tiene un nivel de habilidad y desempeño medido en una escala continua. En Aprender, la media de esta escala (para cada año y disciplina) es de 500 y su desvío estándar, 100; se trata de la escala fijada en el operativo Aprender 2016, que funciona como referencia desde entonces. Cabe destacar que se modelizaron los desempeños según un modelo logístico a dos parámetros. Este modelo supone que cada ítem (o pregunta del cuestionario cognitivo) tiene asociado un nivel de dificultad y un nivel de discriminación<sup>8</sup>.

El puntaje de cada alumno se obtiene a partir de este procesamiento. Al tratarse de una variable real continua, es factible la aplicación de modelos de regresión como estrategia de análisis. Se trabaja en primer lugar con un *modelo de regresión múltiple clásica*, a fin de entender cuáles son las variables que presentan mayor capacidad explicativa respecto de la variabilidad de resultados de los estudiantes en el área de Lengua.

En una segunda instancia, el informe aborda un modelo de *análisis multinivel*, considerando su pertinencia para el trabajo con datos de estructura jerárquica. Se parte del supuesto que, cuando los individuos se encuentran agrupados en unidades mayores, existe cierto grado de similitud entre ellos y, al mismo tiempo, cierto grado de diferencia respecto a quienes se encuentran en otras unidades; tal es el caso de los estudiantes y las escuelas. Este análisis permite dar cuenta de la variación de una variable dependiente en varios niveles de agrupamiento. En primer lugar, con un modelo vacío, a fin de discriminar la variabilidad entre y dentro de las escuelas; luego, con un modelo de análisis secuencial, introduciendo las variables del alumno (y el sector de gestión) y, posteriormente, las variables del contexto escolar.

<sup>6</sup> A los alumnos que respondieron menos de la mitad de los 24 ítems no se les asigna puntaje en esa disciplina.

<sup>7</sup> No todos los alumnos con puntaje tienen respuesta en las variables Sexo y Edad, debido a que completan esas variables y otras luego de realizar la prueba de las dos disciplinas; por lo tanto, la falta de respuesta se atribuye a diversos motivos (no querer completarlo, cansancio, problema en la lectura óptica, etc.).

<sup>8</sup> Para más precisiones sobre la *Teoría de Respuesta al Ítem* se puede consultar el trabajo de José Muñiz, «Introducción a la Teoría de Respuesta a los Ítems», Pirámide (1997); y la documentación de Aprender.

Respecto a las características de los alumnos, se trabaja con variables demográficas (sexo<sup>9</sup> y edad), de trayectorias educativas (asistencia a nivel inicial), y variables referidas a las características de los hogares como por ejemplo el clima educativo del hogar (nivel educativo de la madre, tenencia de libros), nivel económico del hogar (tenencia de bienes, hacinamiento, tenencia de un lugar para estudiar). Las variables referidas al nivel económico del hogar deben considerarse como variables *proxy*, no se pretende demostrar una relación causal entre la posesión de alguno de estos bienes y el desempeño en Lengua.

Respecto al contexto escolar, se diferencian dos grupos de variables. Las referidas al **contexto institucional** (de la escuela) como el sector de gestión, la edad del director, la antigüedad del director en el cargo en la escuela y el turno; y al **contexto del aula** (grupo de pares), que refiere a los datos promedio del alumno, agregados a nivel del aula, como por ejemplo la proporción de hogares con más de 50 libros en papel, la proporción de hogares que poseen auto, y la proporción de hogares con computadora. También se considera a nivel del aula la variable de clima escolar<sup>10</sup>.

Las variables originales del cuestionario de contexto se recategorizaron para evitar categorías con muy pocos casos, como puede verse en la siguiente **Tabla**.

---

9 En el cuestionario Aprender 2023 la pregunta sobre sexo (AP03) incluye la opción X en las respuestas. Debido a que el porcentaje es menor al 1%, eran pocos casos para tratarlos estadísticamente y por eso no se consideraron en el ejercicio

10 Clima escolar: clima escolar Bajo se asocia a mal clima escolar, y clima escolar Alto a buen clima escolar. Para la construcción de esta variable se utilizaron los resultados del módulo Clima escolar aplicado en dos modelos de los cuestionarios de contexto 2023.



Tabla 2. Variables seleccionadas en los modelos de análisis

NIVEL DE ANÁLISIS	DIMENSIÓN	VARIABLES/INDICADORES	CATEGORÍAS	
Características de los alumnos	Variables demográficas	Sexo	Varón Mujer	
		Edad	11 años o menos 12 años o más	
	Trayectorias educativas	Asistencia a nivel inicial	Asistió desde sala de tres Asistió desde sala de cuatro Asistió desde sala de cinco/No asistió	
	Clima educativo del hogar	Nivel educativo de la madre	Primario (Sin instrucción, primario incompleto, primario completo) Secundario (Secundario incompleto- secundario completo) Universitario (Terciario/universitario incompleto, terciario/universitario completo)	
		Libros de papel en el hogar	0 a 5 6 a 50 51 o más	
	Nivel económico del hogar	Tenencia de bienes	Auto Computadora (Computadora de escritorio, <i>laptop</i> , <i>netbook</i> , etc.)	
		Hacinamiento (personas por habitación)	0 a 1 1,1 a 2,3 Más de 2,3	
		Lugar tranquilo para estudiar	Sí No	
	Variables a nivel aula/escuela	Grupo de pares (Aula)	Clima escolar (ClimaE)	Bajo Medio Alto
			Media universitaria	Proporción de madres con educación terciaria/ universitaria en el colegio
Media más de 50 libros			Proporción de hogares con más de 50 libros en papel	
Media auto			Proporción de hogares que poseen auto	
Media computadora			Proporción de hogares con computadora	
Contexto institucional (Escuela)		Sector de gestión	Estatat Privada	
		Edad del Director	45 46 a 55 56 o más	
		Antigüedad en el cargo en la institución	No tiene antigüedad en ese cargo 5 años o menos Entre 6 y 10 años Entre 11 y 20 años Más de 20 años	
		Turno	Simple Doble	

Como fuente de información se utilizaron las bases de microdatos públicas de Aprender 2023 y la información pública del Relevamiento Anual (RA), publicadas en el sitio web de la Secretaría de Educación de la Nación<sup>11</sup>. De esta forma, podrán reproducirse los resultados del trabajo. Para las variables de contexto se consideraron las versiones imputadas por la Subsecretaría de Evaluación para corregir posibles sesgos producto de la no respuesta. Para las variables a nivel escuela se trabajó con la base de datos de Directivos<sup>12</sup>.

---

11 Un aporte de Aprender al sistema educativo argentino es la publicación de las bases de *microdatos*, para que los especialistas puedan analizar los resultados de la encuesta en todos los operativos. Para una descripción detallada de **Aprender**, puede consultarse el **Documento Técnico/2 (2016)** y el sitio correspondiente de la Secretaría de Educación.

12 Para mayor detalle, ver: Anexo Metodológico.

## ► 2. Resultados Aprender 2023: Análisis bivariado

En este apartado se presenta un análisis descriptivo de la relación entre las variables consideradas en los modelos de regresión y multinivel, y el puntaje promedio obtenido por los estudiantes en el área de Lengua.

Como ya se mencionó, Aprender procesa los resultados mediante la Teoría de Respuesta al Ítem, y de ese modo se construye el puntaje de cada alumno. A partir de esta variable se pueden calcular los promedios de los resultados obtenidos; debe aclararse que dicha escala es arbitraria (es decir, son posibles infinitas escalas). En Aprender 2023 se escalaron los valores  $\theta$  mediante una transformación lineal (multiplicando por un factor y sumando una constante) para obtener una media 500 y un desvío estándar 100, tomando Aprender 2016 como punto de referencia.

Respecto a la relación entre las características de los alumnos y el puntaje promedio de Lengua, se observan algunas diferencias, esperables, por sexo y edad. Como han mostrado otros estudios (Dabenigno V y otros: 2014) se verifica un mejor desempeño de las mujeres comparativamente con los varones en el área de Lengua. También se observa un mayor promedio entre aquellos estudiantes que tienen la edad teórica correspondiente al grado, comparativamente con quienes presentan mayor edad, lo que alude a la relación entre el rezago escolar y el desempeño académico.

Sin embargo, en términos de brechas, se observa una distancia mayor en las variables que refieren a atributos del hogar de los estudiantes. Uno de estos atributos es el clima educativo del hogar, representado en este caso por el nivel educativo de la madre y la cantidad de libros en el hogar. Los estudiantes cuyas madres han accedido al nivel terciario o universitario presentan casi 60 puntos de diferencia con aquellos cuyas madres cuentan solo con nivel primario. También se observa una diferencia importante respecto a la tenencia de libros de papel en el hogar, con mayores puntajes en Lengua en los alumnos que disponen de mayor cantidad de libros (la diferencia llega a los 48 puntos porcentuales). Asimismo, se observan diferencias importantes en variables que aluden al nivel económico del hogar, como la tenencia de computadora o la condición de hacinamiento, y a las trayectorias de los estudiantes, específicamente con relación a la edad de ingreso al nivel inicial (en el siguiente apartado se podrá observar con mayor detalle la incidencia de esta variable en los diferentes niveles socioeconómicos).

Es decir, las mayores brechas de desempeño se observan al considerar algunas de las variables que dan cuenta del nivel socioeconómico de los hogares (**Tabla 3**), corroborando uno de los aspectos sobre lo que más se ha tematizado en el análisis sobre factores asociados al desempeño académico.

Tabla 3. Puntaje promedio en Lengua y diferencia, según características de los alumnos y sus hogares

CARACTERÍSTICAS DEL ALUMNO Y SU HOGAR	PROMEDIO	DIFERENCIA (MÁX-MIN)
<b>SEXO</b>		
Mujer	502,0	9,3
Varón	492,7	
<b>EDAD</b>		
11 o menos	500,5	10,2
12 o más	490,3	
<b>NIVEL EDUCATIVO DE LA MADRE</b>		
Primario (incompleto/completo)	463,2	59,6
Secundario (incompleto/completo)	489,7	
Terciario/Universitario (incompleto/completo)	522,8	
<b>TENENCIA DE AUTOMÓVIL EN EL HOGAR</b>		
No	479,2	28,1
Sí	507,3	
<b>TENENCIA DE COMPUTADORA EN EL HOGAR</b>		
No	475,6	35,3
Sí	510,9	
<b>ESPACIO TRANQUILO PARA ESTUDIAR</b>		
No	483,1	16,9
Sí	500	
<b>PERSONAS POR HABITACIÓN</b>		
0 a 1	515,4	42,8
1,1 a 2,3	496,9	
Más de 2,3	472,6	
<b>CANTIDAD DE LIBROS EN PAPEL EN EL HOGAR</b>		
0 a 5	469,2	48,5
6 a 50	511,3	
51 o más	517,7	
<b>ASISTENCIA AL JARDÍN</b>		
Sala de tres	504,9	49,1
Sala de cuatro	496,3	
Sala de cinco	470,5	
No asistió	455,8	
<b>TOTAL</b>	<b>497,4<sup>13</sup></b>	

Fuente: Evaluación Aprender 2023, DNEIEE-REFCEE | SSIIEE | Secretaría de Educación | Ministerio de Capital Humano

Respecto a las variables relacionadas con la oferta educativa, la mayor diferencia en los puntajes se observa en el sector de gestión. Como se ha evidenciado en otros trabajos (de nivel nacional e internacional), se verifican mejores resultados entre estudiantes que asisten a establecimientos educativos del sector de gestión

<sup>13</sup> Se observa una leve diferencia respecto del promedio presentado en el Informe Aprender 2023, que se explica por la variación en el universo de estudio.

privada, con una diferencia de 59,5 puntos respecto a quienes asisten al sector de gestión estatal. Esto no significa que la diferencia sea consecuencia pura del sector; este aspecto se analizará con mayor profundidad en el siguiente apartado. Asimismo, el clima escolar positivo en el aula se relaciona con el puntaje promedio de Lengua, ya que los estudiantes insertos en contextos escolares más favorables presentan un mejor desempeño.

Otro aspecto interesante, menos abordado en trabajos anteriores, es la antigüedad del director en el cargo en el establecimiento escolar. Se selecciona dicha variable considerando la importancia del rol en la definición del proyecto institucional<sup>14</sup>. En la siguiente **Tabla** se observa un mayor puntaje en Lengua entre quienes asisten a escuelas cuyos directivos presentan una mayor antigüedad en el cargo en la institución, con una diferencia de 34 puntos respecto a quienes no la tienen.

**Tabla 4. Puntaje promedio en Lengua y diferencia, según características de la escuela/aula**

CARACTERÍSTICAS DE LA ESCUELA /AULA	PROMEDIO	DIFERENCIA (MÁX-MIN)
<b>SECTOR DE GESTIÓN</b>		
Estatal	481,7	59,5
Privada	541,2	
<b>CLIMA ESCOLAR</b>		
Bajo	475,7	31,3
Medio	501,6	
Alto	507	
<b>EDAD DEL DIRECTOR</b>		
Menos de 45	497,6	1,7
46 a 55	496,6	
56 o más	499,3	
<b>ANTIGÜEDAD EN EL CARGO DEL DIRECTOR</b>		
No tiene antigüedad en ese cargo	487	34
5 años o menos	493,3	
Entre 6 y 10 años	501,7	
Entre 11 y 20 años	509,2	
Más de 20 años	521	
<b>TURNO</b>		
Simple	498,1	1,8
Doble	496,3	
<b>TOTAL</b>	<b>497,4</b>	

Fuente: Evaluación Aprender 2023, DNEIEE-REFCEE | SSIEE | Secretaría de Educación | Ministerio de Capital Humano

Es decir, tanto los factores relativos a los estudiantes y sus hogares como la oferta educativa, parecieran estar relacionados con el desempeño de los estudiantes. El sector de gestión, la antigüedad del director en el cargo en la escuela, el capital cultural y económico del hogar, evidencian ser las variables más relevantes en este sentido.

<sup>14</sup> Puede profundizarse este aspecto en: Aprender Informe de Resultados 2023. Ministerio de Capital Humano. Secretaría de Educación. [https://www.argentina.gob.ar/sites/default/files/aprender\\_2023\\_final.pdf](https://www.argentina.gob.ar/sites/default/files/aprender_2023_final.pdf)

Sin embargo, es necesario señalar que las diferencias encontradas remiten a un análisis descriptivo del comportamiento de las variables. En los capítulos siguientes, al ajustar los modelos de regresión, podrá verse que las diferencias encontradas pueden estar influenciadas por otras variables que intervienen en la relación. Por lo tanto, es de suma importancia considerarlas simultáneamente en el análisis.

## ► 3. Determinantes del desempeño a nivel de los estudiantes

Para analizar y cuantificar la influencia de diferentes factores en los puntajes alcanzados en Lengua, se trabaja en primera instancia con una regresión lineal (aprovechando que el puntaje TRI es una variable continua y presenta una distribución aproximadamente normal).

Respecto del objetivo, no se busca en este caso generar un modelo que prediga los puntajes, sino que refleje el mecanismo que produce que algunos alumnos obtengan mejores resultados que otros. Se parte del supuesto de que existen algunas dimensiones que determinan los desempeños académicos.

Debe aclararse que en el ejercicio no se trabaja con el universo de alumnos evaluados, sino con una muestra aleatoria *autoponderada*, es decir, con iguales factores de expansión y estratificada por sector de gestión, de 388 colegios con al menos cinco alumnos puntuados. Seleccionar una muestra *autoponderada* hace innecesario aplicar el factor de expansión en las estimaciones de los parámetros (aun cuando en los operativos censales de Aprender se tiene como objetivo corregir la no respuesta y tiene muy poca variabilidad)<sup>15</sup>.

**Tabla 5. Muestra de escuelas y alumnos**

SECTOR DE GESTIÓN	ESCUELAS	ALUMNOS
Estatad	293	9.744
Privada	95	4.350

Fuente: Evaluación Aprender 2023, DNEIEE-REFCEE | SSIEE | Secretaría de Educación | Ministerio de Capital Humano

Dado que la encuesta se realiza a todos los alumnos de los colegios seleccionados en el operativo censal, debe advertirse que esto produce un fuerte *efecto conglomeración* o *efecto contagio* entre alumnos, que debe ser tomado en cuenta al estimar la varianza de los estimadores de los parámetros del modelo.

Para analizar la significatividad estadística de los parámetros se usó el comando `svyglm` del paquete `survey` de `R`, que considera la estructura conglomerada de los datos y el tipo de muestra estratificada.

Se presentan a continuación los tres modelos de análisis.

### 3.1. Modelo I: Características del alumno y sector de gestión

En este primer modelo se considera como variable dependiente el puntaje en Lengua, y se seleccionan como variables explicativas el sexo y edad del alumno, y el sector de gestión al que asiste.

En el análisis de este primer modelo se puede ver que todos los parámetros son altamente significativos.

<sup>15</sup> Cuantiles del factor de expansión de la base de microdatos:

1%	5%	50%	95%	99%
1.00	1.00	1.14	1.73	2.57

Asimismo, el valor del  $R^2$  es relativamente bajo (0,08), es decir, el modelo en su conjunto no sería suficiente para explicar la varianza de los puntajes de Lengua. Esto es habitual en los modelos de regresión de datos educativos (Aprender, PISA, etc.).

En el **Gráfico 1** se visualiza la magnitud de los coeficientes para cada parámetro. Como las variables explicativas son todas categóricas, la importancia de cada categoría puede medirse por la magnitud en valor absoluto de su coeficiente<sup>16</sup>.

Las categorías “varón” y la edad de “12 años o más” presentan un efecto negativo sobre el puntaje de Lengua, aunque de menor magnitud que el efecto observado en la categoría “sector de gestión privada” donde se observa un coeficiente positivo y más alto (54,34), que indica un efecto positivo respecto del puntaje de Lengua.

#### ◆ Modelo I. Ajuste con svyglm (survey)<sup>17</sup>

<b>OBSERVACIONES</b>	15573
<b>VARIABLE DEPENDIENTE</b>	lpuntaje
<b>TIPO</b>	Survey-weighted linear regression

<b>R<sup>2</sup></b>	0,08
<b>R<sup>2</sup> AJUSTADO</b>	0,07

	ESTIMACIÓN	DESVÍO ESTÁNDAR	VALOR t	VALOR p
(Intercepto)	491,12	2,00	245,38	0,00
Sexo varón	-8,41	1,57	-5,35	0,00
Edad 12	-7,55	1,79	-4,22	0,00
Sector de gestión privada	54,34	4,12	13,19	0,00

Errores estándar: Robustos

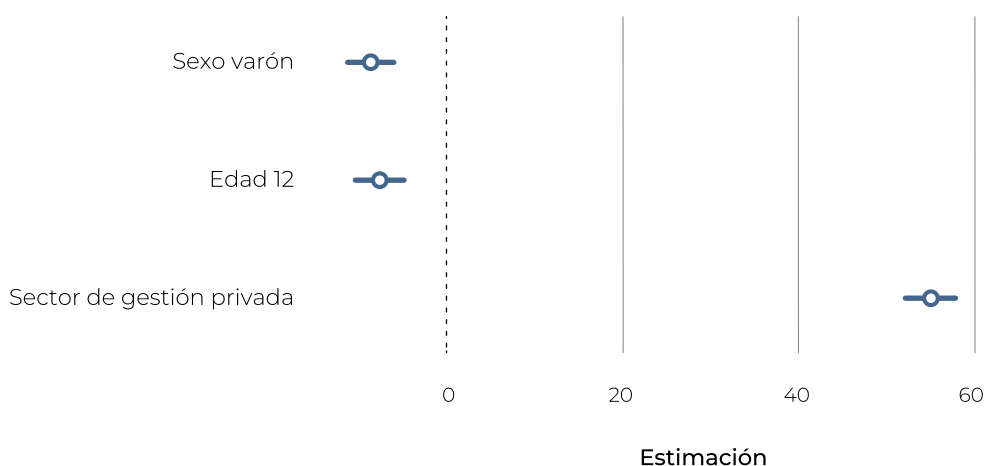
Fuente: Evaluación Aprender 2023, DNEIEE-REFCEE | SSIEE | Secretaría de Educación | Ministerio de Capital Humano

<sup>16</sup> En el caso de querer comparar variables continuas de diferente unidad de medida, una alternativa es estandarizarlas para luego comparar los coeficientes estimados.

<sup>17</sup> Para mayor detalle sobre el ajuste del modelo, ver Anexo metodológico.



Gráfico 1. Coeficientes Modelo I



Fuente: Evaluación Aprender 2023, DNEIEE-REFCEE | SSIEE | Secretaría de Educación | Ministerio de Capital Humano

### 3.2. Modelo II: Incorporación de variables del hogar y trayectorias

En este segundo modelo se incorporan variables relativas al clima educativo y el nivel económico del hogar. También se agrega la edad de ingreso del alumno al nivel inicial, que puede ser analizada no solo como una variable que da cuenta de la experiencia previa de escolarización (en términos de trayectorias escolares), sino también relativa a las posibilidades y decisiones diferenciales de las familias respecto a la edad de inicio de la escolarización de sus hijos<sup>18</sup>.

La incorporación de variables permite mostrar cómo el aparente efecto de una variable desaparece o disminuye fuertemente cuando se incluye otra. En este sentido, este tipo de análisis permite dar cuenta de los alcances y limitaciones de los análisis univariados.

Se observa nuevamente que la mayoría de las categorías incluidas siguen siendo altamente significativas. También puede verse que aumenta el  $R^2$  respecto del Modelo I (pasa de 0,08 a 0,14), aunque de todos modos sigue siendo bajo; por lo tanto, su capacidad explicativa también lo es.

En el **Gráfico 2** se observa que, al incluir las variables del hogar, se produce una disminución de los coeficientes estimados para las variables originales. Como era de esperarse, al incorporar variables relativas al nivel socioeconómico del hogar, el parámetro de **Sector de gestión privada** se reduce fuertemente: pasa de 54,34 (Modelo 1) a 34,62. Es decir, es posible pensar que lo que aparecía como un efecto del sector de gestión, en realidad estaba asociado al nivel socioeconómico de los hogares; esto puede observarse a partir de controlar las variables. De todos modos, el sector de gestión privada aún sigue presentando el coeficiente más alto, al igual que el nivel educativo de la madre (terciario/universitario). Asimismo, también las categorías que aluden a una escolarización temprana presentan los mayores coeficientes y de signo positivo, lo que indica su impacto sobre el puntaje de Lengua.

En contraposición, se puede observar que la variable “tenencia de automóvil” no es significativa: la presencia de otras variables *proxy* del nivel socioeconómico del hogar hacen que esta variable sea innecesaria.

18 En los datos de Aprender 2023 se analiza la desigualdad en la edad de ingreso al nivel inicial según el nivel socioeconómico del hogar. El 76,7% de los niños de hogares de NSE Alto comenzaron su escolarización a los tres años, mientras que entre los de hogares de NSE Bajo el porcentaje se reduce al 42,9%; además, casi un 20% comienza su escolarización a los cinco años. Aprender, Informe de resultados. 2023. Ministerio de Capital Humano. Secretaría de Educación. Pág. 59. Disponible en [https://www.argentina.gob.ar/sites/default/files/aprender\\_2023\\_final.pdf](https://www.argentina.gob.ar/sites/default/files/aprender_2023_final.pdf)

## ◆ Modelo II. Ajuste con svyglm (survey)

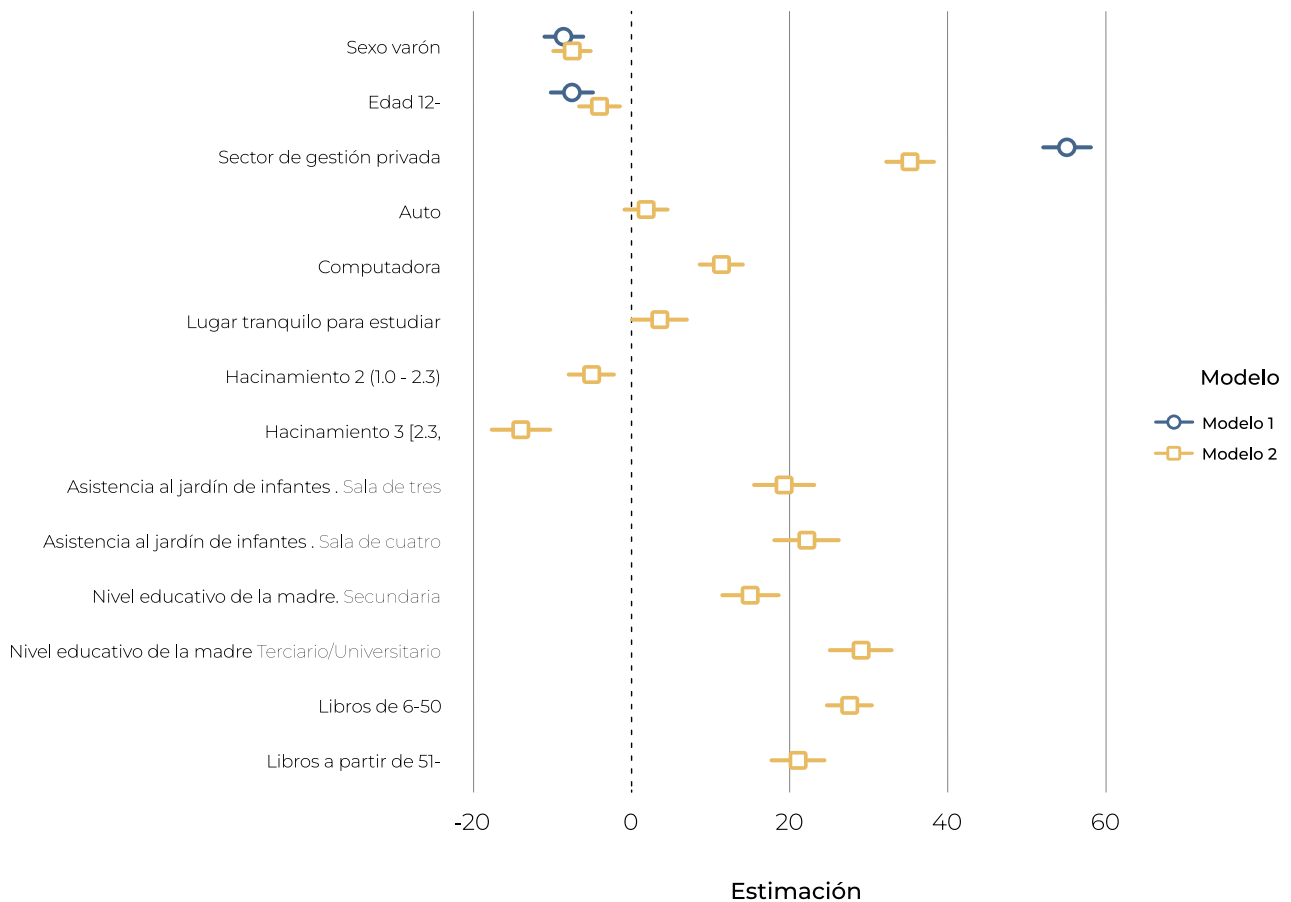
<b>OBSERVACIONES</b>	15573
<b>VARIABLE DEPENDIENTE</b>	lpuntaje
<b>TIPO</b>	Survey-weighted linear regression
<b>R<sup>2</sup></b>	0,14
<b>R<sup>2</sup> AJUSTADO</b>	0,11

	ESTIMACIÓN	DESvíO ESTÁNDAR	VALOR t	VALOR p
(Intercepto)	438,73	3,88	113,14	0,00
Sexo varón	-7,41	1,52	-4,86	0,00
Edad 12	-4,10	1,69	-2,43	0,02
Sector de gestión privada	34,62	3,48	9,95	0,00
Auto	1,92	1,64	1,17	0,24
Computadora	11,08	1,64	6,74	0,00
Lugar tranquilo para estudiar	3,23	1,90	1,70	0,09
Hacinamiento 2 (1,0 – 2,3)	-4,86	1,65	-2,95	0,00
Hacinamiento 3 (Más de 2,3)	-14,10	2,09	-6,76	0,00
Asistencia al jardín de infantes. Sala de tres	19,35	2,41	8,03	0,00
Asistencia al jardín de infantes. Sala de cuatro	21,74	2,47	8,80	0,00
Nivel educativo de la madre. Secundaria	14,84	2,17	6,85	0,00
Nivel educativo de la madre. Terciario/Universitario	28,87	2,35	12,29	0,00
Libros de 6-50	27,49	1,61	17,08	0,00
Libros a partir de 51 y más	20,73	2,23	9,30	0,00

Errores estándar: Robustos

Fuente: Evaluación Aprender 2023, DNEIEE-REFCEE | SSIEE | Secretaría de Educación | Ministerio de Capital Humano

**Gráfico 2. Coeficientes modelo I y II**



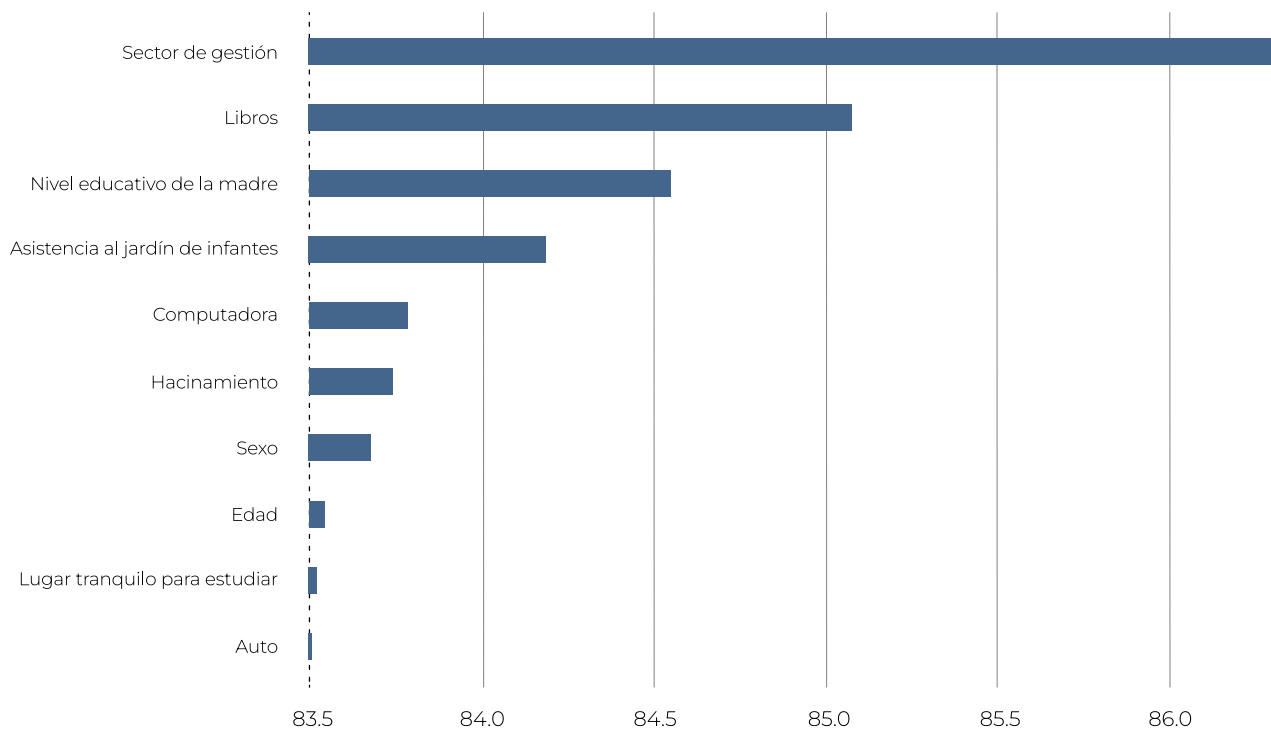
**Fuente:** Evaluación Aprender 2023, DNEIEE-REFCEE | SSIEE | Secretaría de Educación | Ministerio de Capital Humano

Ahora bien, debe advertirse que algunas de las variables incorporadas en el Modelo 2 no son dicotómicas (tienen más de dos categorías). Esto dificulta su comparación respecto de las variables dicotómicas; es decir, para evaluar la importancia de cada una de las variables no es posible considerar la magnitud de los coeficientes.

Una alternativa propuesta para sortear esta dificultad es evaluar la *importancia* relativa de cada variable (no de las categorías) sin suponer un modelo en particular (model-agnostic). La idea es que, para cierta variable, se permuten sus valores y se estime cómo se ve afectado el resultado de la predicción. El método permite evaluar la *importancia* para un conjunto de variables de diferentes escalas de medida y también para un agrupamiento de variables, que podrían pensarse como dimensiones. Para tal fin se utiliza el paquete DALEX<sup>19</sup> de R.

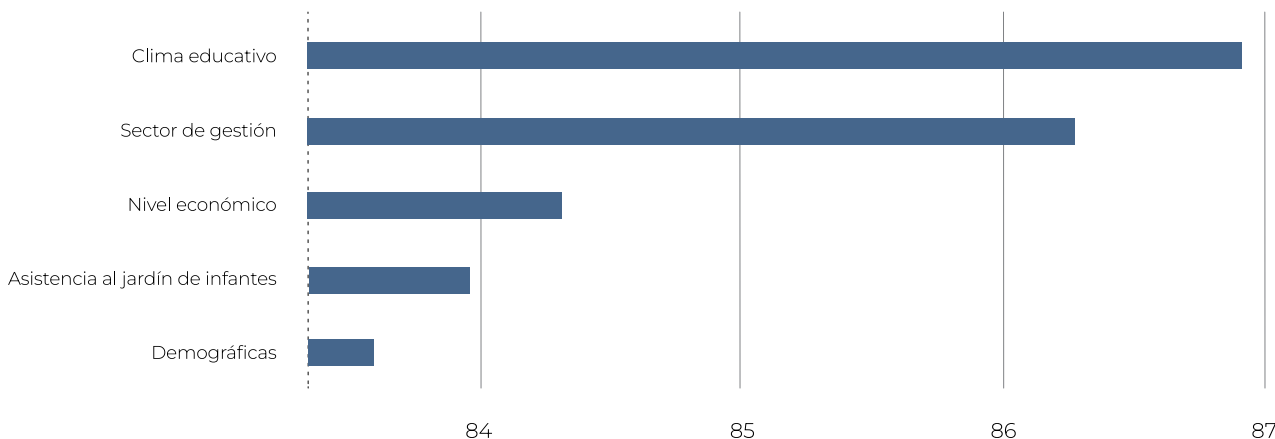
En los gráficos que se presentan a continuación se observa que la variable sector de gestión, tiene una mayor importancia relativa por sobre las demás variables (**Gráfico 3**).

<sup>19</sup> Para los detalles de este método, ver B. T. Biecek P 2020, Cp. 16.

**Gráfico 3: Importancia de variables. Modelo II**

**Fuente:** Evaluación Aprender 2023, DNEIEE-REFCEE | SSIEE | Secretaría de Educación | Ministerio de Capital Humano

Del mismo modo, si en lugar de comparar entre sí las variables se consideran las dimensiones (agrupamientos de variables), puede verse que el clima educativo del hogar (formado por las variables “tenencia de libros” y “nivel educativo de la madre”) presenta una importancia relativa mayor al sector de gestión y, a su vez, mayor a otras dimensiones como el nivel económico o las variables demográficas, reafirmando la observación anteriormente presentada.

**Gráfico 4: Importancia grupo de variables. Modelo II**

**Nota:** El clima educativo incluye las variables: nivel educativo de la madre y libros de papel en el hogar. El nivel económico incluye las variables: tenencia de bienes, hacinamiento, lugar tranquilo para estudiar. Demográficas incluye las variables: sexo y edad.

**Fuente:** Evaluación Aprender 2023, DNEIEE-REFCEE | SSIEE | Secretaría de Educación | Ministerio de Capital Humano

Este ejercicio muestra una posible ventaja de este método respecto a la construcción de índices sintéticos con variables de un mismo dominio.

### 3.3. Modelo III: Incorporación de variables de contexto escuela y contexto aula

Considerando que la capacidad explicativa de los modelos I y II es baja, se avanza en un tercer modelo de regresión que incorpora un conjunto de variables referidas a la escuela. En estas variables se pueden diferenciar las que remiten al **contexto institucional**, como la edad del director, la antigüedad del directivo en el cargo y el turno, de otro conjunto de variables que aluden al **contexto del aula (grupo de pares)**, entre las que se consideran el clima escolar, y algunos otros atributos del estudiante agregado a nivel del aula, como la media de hogares con computadora, la media de hogares con auto, la media de madres con estudios terciarios/ universitarios completos, y la media de hogares con más de 50 libros. Se incorporan estas variables a fin de considerar el impacto que puede tener la composición social de la escuela y, más específicamente, el perfil del grupo de pares con el cual interactúan los estudiantes, sobre los resultados de aprendizaje, teniendo en cuenta diferentes estudios que ilustran esta relación.

Asimismo, para analizar con un mayor detalle el efecto de la asistencia temprana al nivel inicial, se considera en el modelo el NSE de los estudiantes con relación a la edad de comienzo de la escolarización en el nivel inicial.

En principio, se observa que el  $R^2$  (0,16) aumenta levemente respecto del valor observado en el Modelo II; de todos modos, su capacidad explicativa sigue siendo baja.

En el **Gráfico 5** puede verse que, al incorporar las nuevas variables, disminuye de manera significativa la estimación correspondiente al sector de gestión privada que pasa de ser 34,62 a 12,57. Es decir, el parámetro de Sector de gestión privada continúa descendiendo al pasar a los sucesivos modelos.

De hecho, mientras que en los modelos anteriores se mantenía la mayor capacidad explicativa, en este modelo se observan parámetros más altos con respecto a la educación terciaria o universitaria de la madre.

Además, algunas de las categorías referidas a variables económicas (como la de tenencia de computadora) asumen un coeficiente más alto al ser consideradas como atributos agregados, a nivel del aula.

Por otra parte, se observa que la asistencia temprana al nivel inicial pareciera tener un impacto diferenciado en los resultados según el nivel socioeconómico del hogar de los estudiantes. Específicamente, la escolarización a los tres años de edad pareciera tener un impacto en los resultados en el nivel socioeconómico más alto, pero no necesariamente entre estudiantes de menor nivel socioeconómico; esto, en algún sentido, muestra la irreductibilidad de la posición en la estructura social.

Finalmente, las variables referidas al contexto institucional presentan una menor significación.

### ◆ Modelo III. Ajuste con svyglm (survey)

<b>OBSERVACIONES</b>	15309
<b>VARIABLE DEPENDIENTE</b>	lpuntaje
<b>TIPO</b>	Survey-weighted linear regression
<b>R<sup>2</sup></b>	0,16
<b>R<sup>2</sup> AJUSTADO</b>	0,07

	ESTIMACIÓN	DESVÍO ESTÁNDAR	VALOR t	VALOR p
(Intercepto)	424,01	8,40	50,46	0,00
Sexo varón	-7,48	1,51	-4,96	0,00
Edad 12-	-3,52	1,69	-2,09	0,04
Sector de gestión privada	12,57	3,93	3,20	0,00
Auto	-3,56	1,65	-2,16	0,03
Computadora	7,62	1,58	4,82	0,00
Lugar tranquilo para estudiar	2,88	1,85	1,56	0,12
Hacinamiento 2.(1.0 - 2.3)	-3,12	1,67	-1,86	0,06
Hacinamiento 3.[2.3,	-10,77	2,08	-5,17	0,00
Nivel educativo de la madre. Secundaria	12,09	2,09	5,78	0,00
Nivel educativo de la madre. Terciario/Universitario	21,35	2,34	9,12	0,00
Libros de 6-50	25,47	1,58	16,15	0,00
Libros a partir de 51-	15,49	2,18	7,10	0,00
Clima Escolar 2.Medio	11,35	2,81	4,03	0,00
Clima Escolar 3.Alto	16,21	3,84	4,22	0,00
Media universitaria 0,15-0,48	3,42	3,06	1,12	0,27
Media universitaria 0,48-1,00	14,19	5,82	2,44	0,02
Media computadora	29,24	12,10	2,42	0,02

	ESTIMACIÓN	DESVÍO ESTÁNDAR	VALOR t	VALOR p
Media auto	25,98	8,87	2,93	0,00
Media más de 50 libros	17,99	10,87	1,66	0,10
Edad del Director 46 - 55	-4,70	3,59	-1,31	0,19
Edad del Director 56 -	-3,84	4,54	-0,84	0,40
Antigüedad en el cargo Entre 11 y 20 años	2,13	4,47	0,48	0,63
Antigüedad en el cargo Entre 6 y 10 años	1,35	3,13	0,43	0,67
Antigüedad en el cargo Más de 20 años	6,30	6,06	1,04	0,30
Antigüedad en el cargo No tiene antigüedad	-1,44	3,07	-0,47	0,64
Turno director Turno doble	-2,19	2,54	-0,86	0,39
Asistencia al jardín de infantes desde sala de tres: Nivel socioeconómico Bajo	-9,96	6,35	-1,57	0,12
Asistencia al jardín de infantes desde sala de cuatro: Nivel socioeconómico Bajo	-0,23	6,53	-0,04	0,97
Asistencia al jardín de infantes desde sala de cinco/ No asistió: Nivel socioeconómico Bajo	-10,74	7,60	-1,41	0,16
Asistencia al jardín de infantes desde sala de tres: Nivel socioeconómico Medio	-1,94	4,74	-0,41	0,68
Asistencia al jardín de infantes desde sala de cuatro: Nivel socioeconómico Medio	4,37	4,91	0,89	0,37
Asistencia al jardín de infantes desde sala de cinco/ No asistió: Nivel socioeconómico Medio	-16,25	5,08	-3,20	0,00
Asistencia al jardín de infantes desde sala de tres: Nivel socioeconómico Alto	10,31	5,64	1,83	0,07
Asistencia al jardín de infantes desde sala de cinco/ No asistió: Nivel socioeconómico Alto	-29,10	8,44	-3,45	0,00

Errores estándar: Robustos

Fuente: Evaluación Aprender 2023, DNEIIE-REFCEE | SSIEE | Secretaría de Educación | Ministerio de Capital Humano

La interacción entre el nivel socioeconómico y la asistencia al jardín puede visualizarse en la siguiente tabla, donde se presentan los parámetros relativos a la categoría “No asistió / Asistencia desde sala de cinco”:

El modelo sugiere un aumento de la penalización por no asistir o empezar en sala de cinco, conforme aumenta el nivel socioeconómico.

	ESTIMACIÓN	DESVÍO ESTÁNDAR	VALOR t	VALOR p
Asistencia al jardín de infantes desde sala de cinco/ No asistió: Nivel socioeconómico Bajo	-10,74	7,60	-1,41	0,16
Asistencia al jardín de infantes desde sala de cinco/ No asistió: Nivel socioeconómico Medio	-16,25	5,08	-3,20	0,00
Asistencia al jardín de infantes desde sala de cinco/ No asistió: Nivel socioeconómico Alto	-29,10	8,44	-3,45	0,00

Gráfico 5. Coeficientes modelo I, II y III

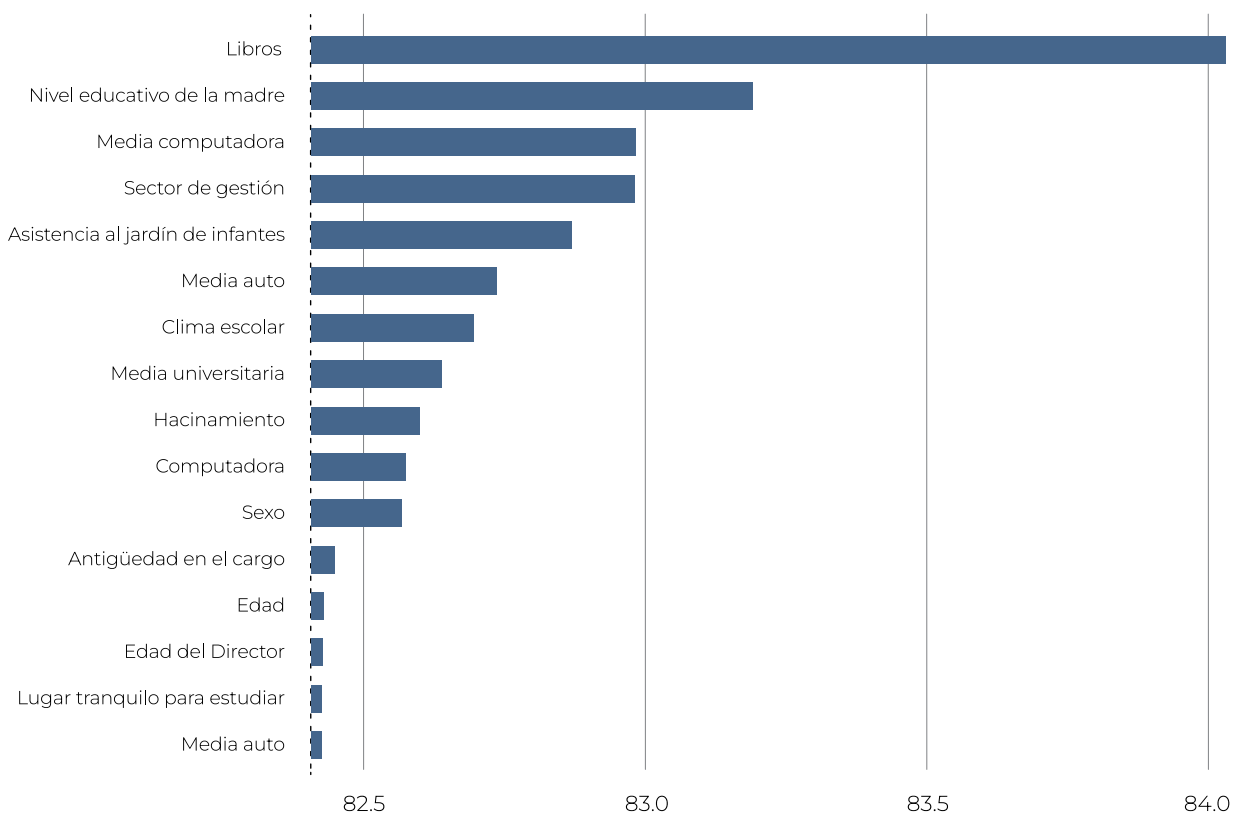


Fuente: Evaluación Aprender 2023, DNEIEE-REFCEE | SSIEE | Secretaría de Educación | Ministerio de Capital Humano

Si al igual que en el modelo II, se analiza la importancia relativa de las variables, puede verse que “Libros en el hogar” pasa a ser la principal variable explicativa.



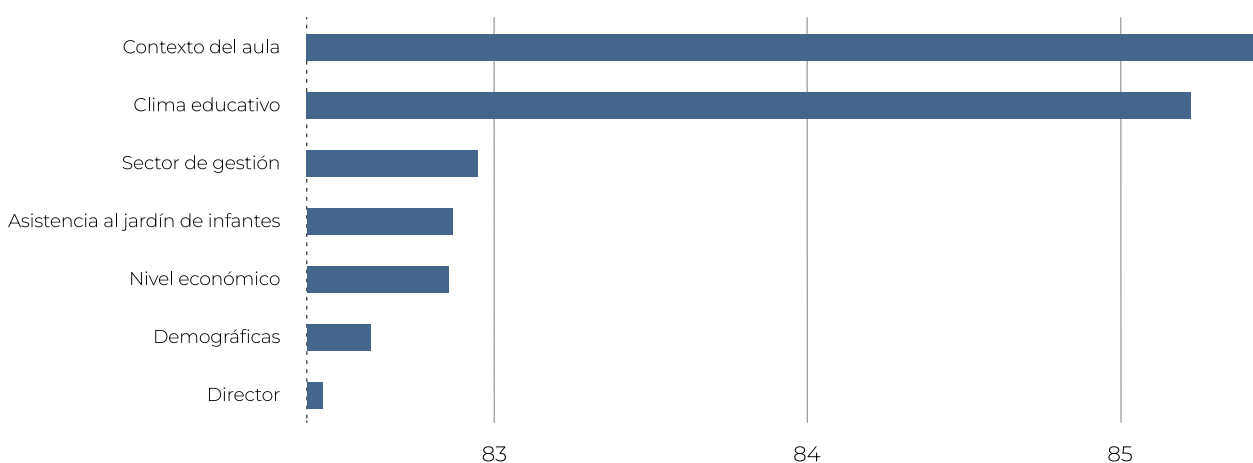
**Gráfico 6. Importancia de las variables. Modelo III**



Fuente: Evaluación Aprender 2023, DNEIEE-REFCEE | SSIEE | Secretaría de Educación | Ministerio de Capital Humano

Si se agrupan las variables en dimensiones, se observa en este caso que adquieren una mayor importancia las variables referidas al contexto del aula.

**Gráfico 7. Importancia grupo de variables A. Modelo III**

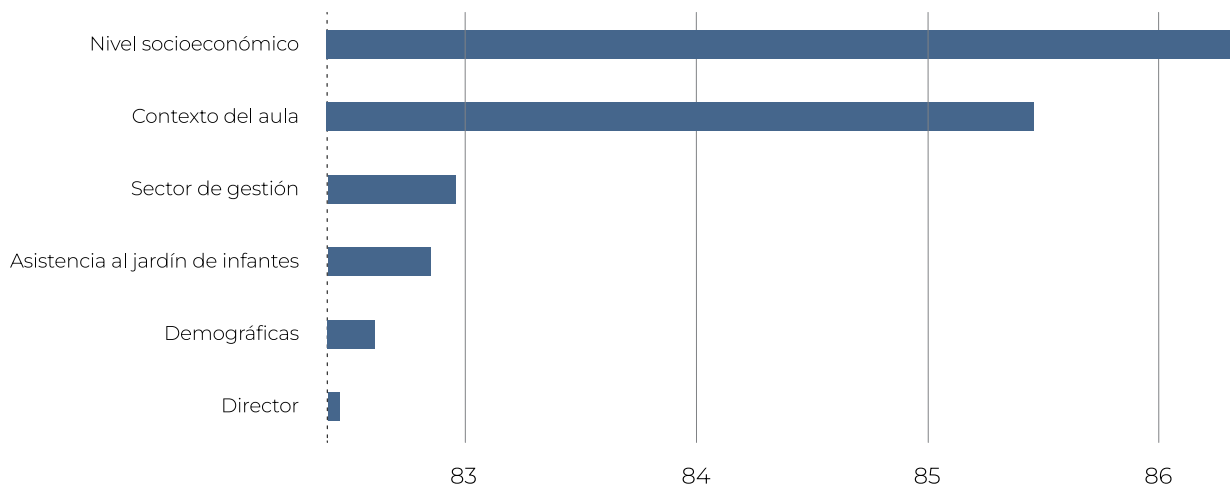


**Nota:** "Contexto del aula" incluye las variables: clima escolar, media de madres con educación terciaria/ universitaria, media de hogares con más de 50 libros en papel, media de hogares con auto, media de hogares con computadora. "Clima educativo" incluye las variables: nivel educativo de la madre y libros de papel en el hogar. El nivel económico incluye las variables tenencia de bienes, hacinamiento, lugar tranquilo para estudiar. "Demográficas" incluye las variables: sexo y edad.

Fuente: Evaluación Aprender 2023, DNEIEE-REFCEE | SSIEE | Secretaría de Educación | Ministerio de Capital Humano

Sin embargo, si se agrupan en una sola dimensión las variables referidas al clima educativo y el nivel económico del hogar, esta dimensión es la que pasa a tener la mayor importancia relativa (este resultado también está presente en otras evaluaciones de diferentes países).

**Gráfico 8. Importancia de grupo de variables B. Modelo III**



**Fuente:** Evaluación Aprender 2023, DNEIEE-REFCEE | SSIEE | Secretaría de Educación | Ministerio de Capital Humano

En síntesis, si bien en el análisis descriptivo se observaba una relación entre el sector de gestión y el puntaje promedio de Lengua, en el modelo de regresión que pone a jugar diferentes variables simultáneamente, pareciera suavizarse la importancia del sector de gestión privada en este puntaje, y adquieren una mayor envergadura las variables referidas al capital cultural y económico del hogar del estudiante. Asimismo, en el Modelo III se observa que las variables del contexto del aula (o grupo de pares) tienen una mayor significación que los atributos referidos al contexto institucional de la escuela.

Sin embargo, esta idea debe ser matizada. En todos los modelos analizados, la parte de la varianza explicada es baja. Es decir, hay cuestiones que los modelos no llegan a explicar con relación a la variación de los puntajes. A modo de hipótesis se puede plantear que existen aspectos relativos a la oferta que no están siendo considerados y podrían ser relevantes en la variabilidad de resultados. En este sentido, el modelo de regresión permitió ver el grado de significación de algunos atributos estructurales medidos a nivel del aula, que podrían dar una pista de la necesidad de profundizar en este nivel de análisis.

## ► 4. Análisis multinivel

### 4.1. Modelos de análisis

Se presenta a continuación el análisis multinivel, apropiado para el tratamiento de datos con estructura jerárquica.

Se parte del supuesto de que los individuos agrupados en unidades mayores (en este caso, alumnos en escuelas) tienen un grado de similitud, es decir, tienden a parecerse entre sí. Bajo este supuesto, las unidades mayores condicionan el comportamiento de los individuos.

En el caso planteado, a diferencia del análisis de regresión, es posible considerar el efecto del colegio en los puntajes de los alumnos.

A continuación, se presenta un modelo de análisis secuencial de introducción de variables: el Modelo 0 (Modelo Vacío), sin variables predictoras; el Modelo I, con las variables referidas a las características de los alumnos y sus hogares; y el Modelo II, en el cual se introducen las variables relativas al contexto educativo (del aula y la escuela). Se intenta establecer cuánto se gana en el poder explicativo al incorporar las variables de contexto educativo.

El modelo 0, como se anticipó, es un modelo no condicional, donde no se toman en cuenta las variables predictoras (se considera el error aleatorio asociado a cada colegio y el error aleatorio asociado al alumno).

La Ecuación del modelo es

$$TEL_{ij} = \alpha_0 + u_i + \epsilon_{ij}$$

Donde:

$TEL_{ij}$  es el puntaje del alumno  $j$  del colegio  $i$ .

$\alpha_0$  es la media general de puntajes.

$u_i$  es un error aleatorio asociado al colegio, se supone con una distribución normal  $N(0, \sigma_u)$ . Igual desvío para todos los colegios (efecto del colegio  $i$  en el puntaje).

$\epsilon_{ij}$  es un error aleatorio asociado al alumno  $j$  del colegio  $i$ , se supone con una distribución normal  $N(0, \sigma_e)$ . Igual desvío para todos los alumnos.

Es decir, se supone:

- ◇ Los errores  $u$  tienen todos media cero e igual varianza.
- ◇ Los errores  $e$  tienen todos media cero e igual varianza.
- ◇ Los errores  $u$  y  $e$  son independientes. Esto no ocurriría si, por ejemplo, los colegios con mejor plantel docente o mejor organización seleccionan por examen de ingreso a los alumnos más talentosos o preparados.

La varianza total en el puntaje es  $DS(e)^2 + DS(u)^2$

Un valor importante para el análisis es el *coeficiente de correlación intraclase  $\rho$* , que representa la correlación de la variable dependiente entre dos unidades distintas *de un mismo conglomerado* (colegio, en este caso). Se interpreta como la parte de la varianza total que corresponde a un *efecto colegio* no captado por las variables explicativas (como se mencionó, las variables pueden incluir variables a nivel colegio).

En el Modelo 0 (modelo multinivel sin variables explicativas) se realiza un ajuste con **lme4** sobre la muestra de alumnos. Se obtiene un valor del ICC de 0,191. En la Tabla 6 puede verse que, a medida que se agregan variables explicativas, este valor disminuye. Es decir, a medida que se agregan nuevas variables, se reduce la proporción de varianza no explicada a nivel colegio a un nivel relativamente bajo. Se trata de un resultado común en las encuestas educativas.

**Tabla 6. Índice correlación intraclase según modelo**

MODELO	ICC
Modelo Vacío	0,191
Modelo I: Variables del alumno y del hogar	0,088
Modelo II: Variables del alumno, del hogar y del contexto educativo	0,073

Fuente: Evaluación Aprender 2023, DNEIEE-REFCEE | SSIEE | Secretaría de Educación | Ministerio de Capital Humano

En el Modelo I se incorporan las variables referidas al alumno y su hogar, y al sector de gestión. Al igual que lo observado en el modelo de regresión, nuevamente el sector de gestión presenta el mayor poder explicativo, así como también, algunos atributos del hogar de los estudiantes como la tenencia de libros o un mayor nivel de instrucción de la madre. Además, se observa un coeficiente negativo alto en la no asistencia al nivel inicial, o asistencia a la edad de cinco años, específicamente en el nivel socioeconómico más alto, que en este universo representa el impacto de la escolarización tardía en el desempeño de Lengua, (de modo similar a lo encontrado en el análisis de regresión).

En el Modelo II, al incorporar las variables del contexto escolar, se observa que disminuye aún más la capacidad explicativa del sector de gestión privada. Asimismo, la escolarización temprana en el nivel inicial mantiene su magnitud. Entre las variables de contexto educativo presentan mayores coeficientes las que refieren al contexto del aula (y en particular, el promedio de alumnos con computadora en el hogar).

Tabla 7. Coeficientes Modelos multinivel 0, I y II

VARIABLE	MODELO 0	MODELO I	MODELO II
(Intercepto)	496	463,958	436,42
Sexo varón		-7,72	-7,55
Edad 12-		-2,53	-2,43
Sector de gestión privada		39,67	18,25
Nivel educativo de la madre. Secundaria		12,03	11,59
Nivel educativo de la madre. Terciario/Universitario		21,67	20,63
Libros de 6-50		25,58	25,26
Libros a partir de 51-		16,69	16,07
Auto		-2,13	-2,39
Computadora		8,64	7,71
Hacinamiento 2.(1,0 – 2,3)		-3,59	-3,29
Hacinamiento 3.[2,3,		-10,89	-10,38
Lugar tranquilo		2,57	2,38
Asistencia al jardín de infantes desde sala de tres: Nivel socioeconómico Bajo		-8,93	-8,64
Asistencia al jardín de infantes desde sala de cuatro: Nivel socioeconómico Bajo		0,46	1,31
Asistencia al jardín de infantes desde sala de cinco/ No asistió: Nivel socioeconómico Bajo		-11,73	-10,48
Asistencia al jardín de infantes desde sala de tres: Nivel socioeconómico Medio		-2,32	-2,34
Asistencia al jardín de infantes desde sala de cuatro: Nivel socioeconómico Medio		4,83	5,18
Asistencia al jardín de infantes desde sala de cinco/ No asistió: Nivel socioeconómico Medio		-16,15	-15,55
Asistencia al jardín de infantes desde sala de tres: Nivel socioeconómico Alto		10,62	9,84
Asistencia al jardín de infantes desde sala de cinco/ No asistió: Nivel socioeconómico Alto		-29,03	-28,44
Antigüedad en el cargo. No tiene antigüedad			-4,41
Antigüedad en el cargo. Entre 6 y 10 años			3,40
Antigüedad en el cargo. Entre 11 y 20 años			1,22
Antigüedad en el cargo. Más de 20 años			2,50
Edad del Director 46 - 55			-4,76
Edad del Director 56 -			-2,42
Clima Escolar 2. Medio			9,57
Clima Escolar 3. Alto			15,49
Media computadora			33,53
Media más de 50 libros			17,94
Media universitaria 0,15-0,48			2,42
Media universitaria 0,48-1,00			15,79

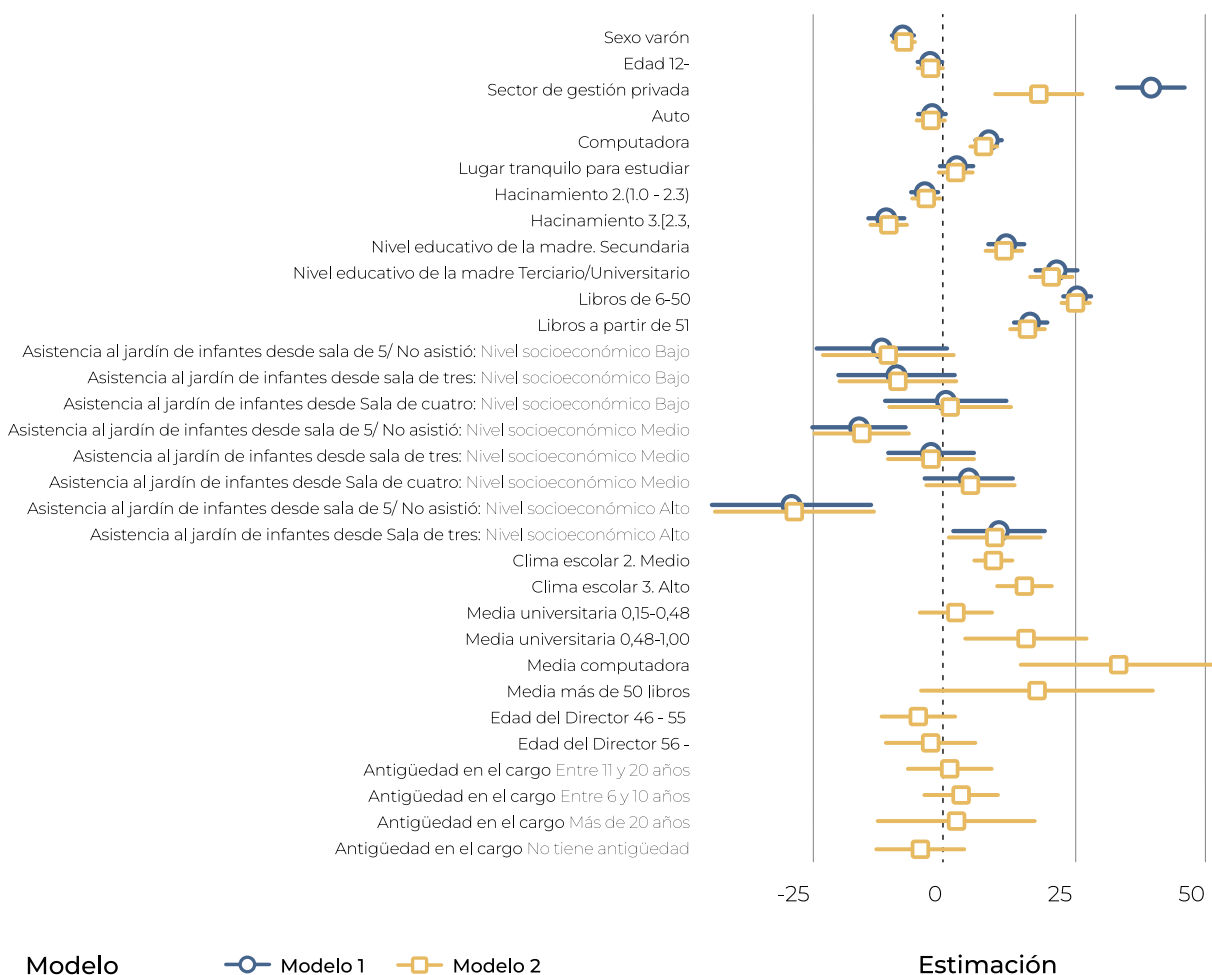
Nota: Ver significación en el Modelo III Ajuste con svyglm

Fuente: Evaluación Aprender 2023, DNEIEE-REFCEE | SSIEE | Secretaría de Educación | Ministerio de Capital Humano

**Gráfico 9. Coeficientes Modelo I y II**

Modelo I: Variables explicativas sin contexto

Modelo II: Variables explicativas que incluyen variables de contexto



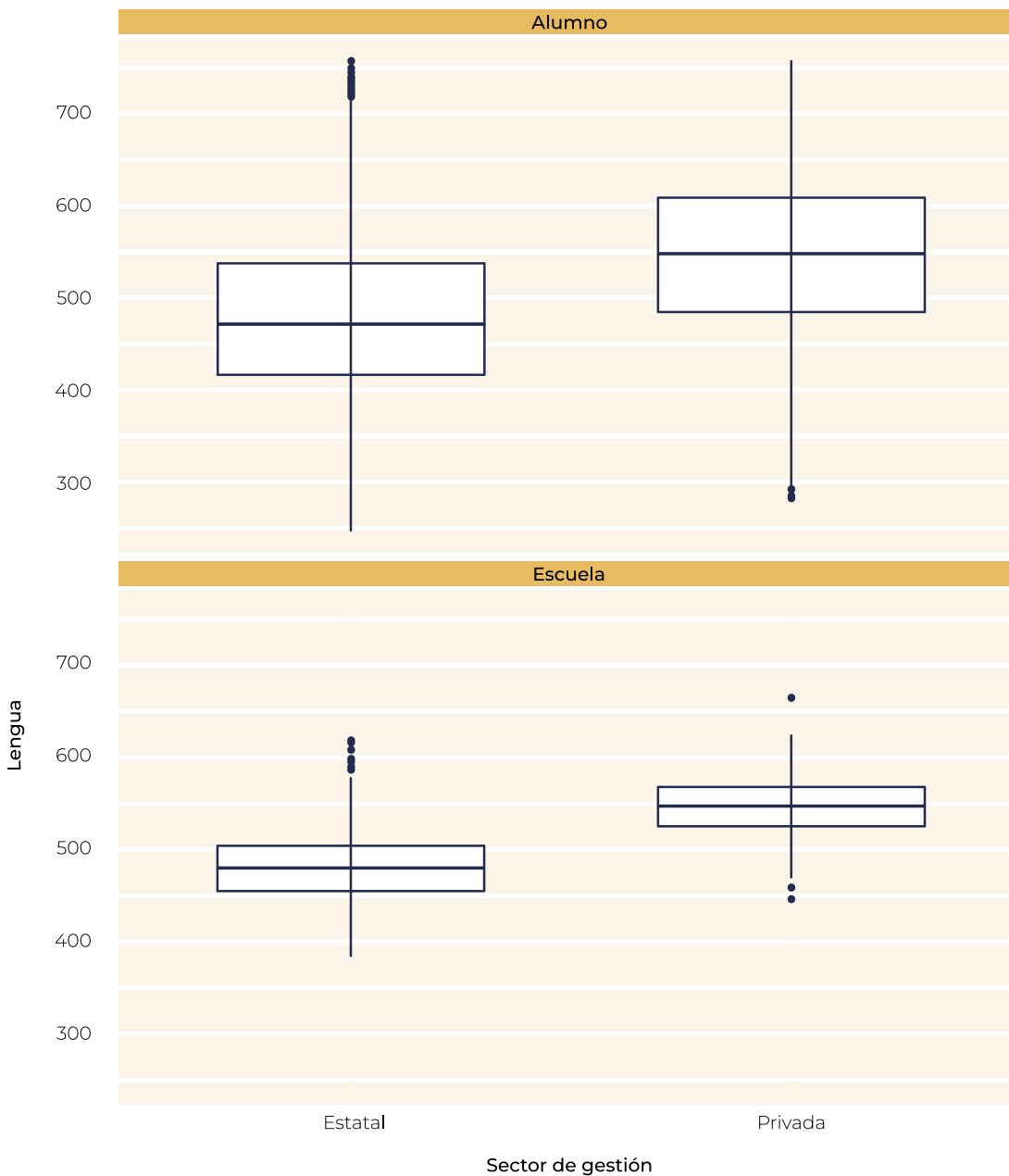
Fuente: Evaluación Aprender 2023, DNEIEE-REFCEE | SSIEE | Secretaría de Educación | Ministerio de Capital Humano

## 4.2. Dispersión entre escuelas y al interior de las escuelas

Si bien se ha evidenciado de algún modo el efecto del contexto escolar sobre los puntajes de los estudiantes, el interrogante que se plantea es cuál es el nivel de dispersión al interior de las escuelas.

En los gráficos *Box Plot* que se presentan a continuación, aunque se evidencia una mayor dispersión a nivel alumno, también se verifica una alta variabilidad a nivel colegio, sobre todo al considerar la posición de los valores extremos. Se observa en mayor medida en el sector de gestión estatal, dando cuenta de situaciones más heterogéneas.

Gráfico 10. Puntaje en Lengua a nivel alumno y escuela



Fuente: Evaluación Aprender 2023, DNEIIE-REFCEE | SSIEE | Secretaría de Educación | Ministerio de Capital Humano

A modo de ejemplo, y en relación con la variabilidad entre y al interior de las escuelas, es posible considerar el puntaje en Lengua en una muestra aleatoria de escuelas.

Con tal fin, se seleccionó una muestra sistemática del total de escuelas puntuadas en Aprender 2023, ordenando ese total según puntaje medio en Lengua, y seleccionando el punto medio en cada tramo. Así, puede suponerse que cada escuela seleccionada representa una franja de 6.25% de escuelas.

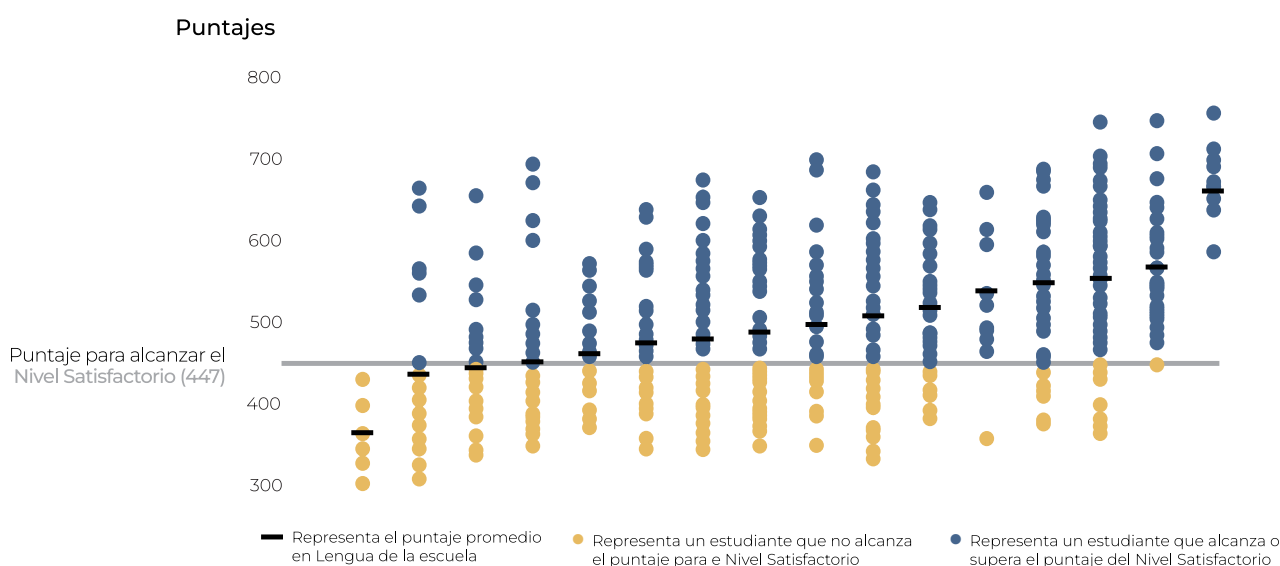
Nuevamente se observan dos situaciones: una alta variabilidad **entre escuelas**, que se visualiza claramente en el promedio por escuela (que va de aproximadamente 350 puntos a casi 700 puntos) y, al mismo tiempo, una variabilidad **al interior de cada escuela**.

Se observan dos casos extremos: en la primera escuela todos los alumnos tienen puntaje inferior al primer punto de corte, es decir: todos los alumnos están en el nivel de desempeño *Por debajo del nivel básico*. En contraposición, en la escuela que representa a la franja con el 6.25% con mayor puntaje, todos los alumnos están en el nivel *Avanzado* (se trabajó para esto con escuelas con al menos 10 alumnos puntuados para tener cierta estabilidad en el promedio).

Probablemente estos casos extremos tengan relación con situaciones de desigualdad en la estructura social.

Respecto a las escuelas intermedias, ubicadas entre ambos casos extremos, hay una alta dispersión al interior de las escuelas, con diferencias importantes entre los valores mínimos y máximos registrados.

**Gráfico 11. Dispersión de puntajes de Lengua dentro de colegio y entre medias de colegios**  
Escuelas con al menos 10 alumnos puntuados. Lengua. Aprender 2023.



**Nota:** Se incluyeron únicamente escuelas con al menos 10 alumnos evaluados. Se calculó el promedio en Lengua y se excluyó el 5% superior e inferior para evitar casos extremos. Las escuelas restantes se ordenaron según su puntaje promedio en Lengua y se seleccionaron 16 escuelas equidistantes mediante muestreo sistemático.

**Fuente:** Evaluación Aprender 2023 - Primaria | Secretaría de Educación | Ministerio de Capital Humano

Otra forma de analizar el grado de homogeneidad o dispersión con relación a los puntajes de Lengua es a través de las medidas de homogeneidad de los *conglomerados* (colegios). En este caso se calcula una de ellas, el  $\rho$  de Lohr.

En la **Tabla 8** se visualiza un valor de 0,198. Considerando que estos índices van de 0 a 1 (de menor a mayor homogeneidad) el valor presentado estaría indicando una baja homogeneidad en los colegios, con relación a los puntajes de Lengua (lpuntaje) y Matemática (mpuntaje). En contraposición, al calcular el índice respecto de la variable de nivel socioeconómico<sup>20</sup> (NSE) se verifica un sensible aumento del indicador, mostrando un mayor grado de homogeneidad, el índice asume un valor de 0,348.

<sup>20</sup> Refiere solo a colegios con al menos cinco alumnos evaluados. Se utilizó el paquete fishmethods para el cálculo de la homogeneidad.



**Tabla 8. Homogeneidad de las escuelas según  $\rho$  de Lohr en el puntaje de Lengua y Matemática y en el puntaje de NSE**

	PUNTAJE DE LENGUA	PUNTAJE DE MATEMÁTICA	PUNTAJE NSE
Lohr rho	0,198	0,207	0,348

Fuente: Evaluación Aprender 2023, DNEIEE-REFCEE | SSIEE | Secretaría de Educación | Ministerio de Capital Humano

Estos datos muestran una casi idéntica homogeneidad en los puntajes de Lengua y Matemática, ambos sensiblemente menores a la homogeneidad en el nivel socioeducativo de los estudiantes. Puede interpretarse como que estudiantes con similar nivel socioeconómico (NSE) e insertos en el mismo contexto educativo, obtienen puntajes disímiles.

## ► Conclusiones

En el presente trabajo se abordó el análisis de los factores asociados al rendimiento académico de los estudiantes de 6° grado del nivel primario en el área de Lengua.

Considerando que los desempeños de los estudiantes se encuentran condicionados por la interacción de diversos elementos, intra y extraescolares, se planteó un abordaje a partir de diferentes estrategias de análisis.

En el análisis descriptivo pudo observarse la relación del sector de gestión con los resultados de aprendizaje de Lengua, que muestran (al igual que en otros estudios) puntajes más altos entre los estudiantes del sector de gestión privada.

Sin embargo, los modelos secuenciales de regresión y de análisis multinivel han demostrado que el peso del sector de gestión como variable explicativa pierde fuerza cuando se incorporan variables representativas del clima educativo del hogar (como el nivel de instrucción de la madre) y el nivel económico.

También se ha podido observar que este tipo de variables adquieren mayor capacidad explicativa al ser consideradas a nivel agregado como dato promedio del aula, lo que evidencia la importancia del grupo de pares en los resultados de aprendizaje de los estudiantes. Por otra parte, el análisis multinivel ha mostrado que el *efecto colegio* no explicado, medido por el *coeficiente de correlación intraclass* ( $\rho$ ), a medida que se agregan variables explicativas disminuye hasta llegar a menos del 10%. Este resultado también se presenta en encuestas similares de otros países (Carabaña 2016), (Snijders Tom, 2004, pág.52).

Finalmente, a través de las medidas de homogeneidad de los conglomerados se evidenció un bajo nivel de homogeneidad en los resultados de Lengua, al mismo tiempo que se observó una mayor similitud en la composición socioeconómica de los alumnos.

Es decir, estudiantes que tienen un nivel económico y social similar, y están insertos en el mismo contexto educativo, de todos modos, obtienen diferentes puntajes en las evaluaciones.

Esto reafirma de algún modo la perspectiva planteada en el Informe Aprender 2023, que muestra la existencia de algunas variables al interior de cada nivel socioeconómico que marcan diferencias en los resultados alcanzados, como el clima escolar, la asistencia al grado a la edad teórica esperada y la escolarización temprana al nivel inicial (en particular entre los estudiantes de mayor nivel socioeconómico).

Debe advertirse que el R2 de los modelos a nivel alumno es bajo (menor a 0.2) aún incorporando todas las variables explicativas, lo cual muestra que hay características que no han podido ser medidas.

Para entender con mayor profundidad qué otros factores explican la diferencia de resultados académicos entre estudiantes de un origen social similar y que comparten un mismo contexto, será necesario avanzar en otro tipo de abordajes que analicen aspectos más sutiles, como por ejemplo diferencias de actitud, intereses, preferencias, gustos, vínculos, capacidad de atención, etc. En síntesis, aspectos relacionados con la subjetividad de los estudiantes, que desafía permanentemente las prácticas docentes y estrategias educativas.

## ► Bibliografía

Agresti, Alan; Kateri, Maria. *Foundations of Statistics for Data Scientists*. CRC Press, 2022.

Biecek P.; Baniecki, H.; Maksymiuk, S. *Package DALEX*, 2023. <https://cran.r-project.org/web/packages/DALEX/DALEX.pdf>

Biecek, P.; Burzykowski, T. *Explanatory Model Analysis*, 2020. <https://ema.drwhy.ai/>.

Bressoux, Pascal. *Modélisation Statistique Appliquée Aux Sciences Sociales*. De Boeck, 2014. Carabaña, Julio. *“El Informe Coleman, 50 años después”*. Revista de Sociología de la Educación, Vol. 9 Núm. 1, España, 2016.

Cervini, Rubén. *“Desigualdades en el logro académico y reproducción cultural en Argentina. Un modelo de tres niveles”*. Revista mexicana de investigación educativa, Vol. 7 Núm. 16. México, 2002.

———. *“Diferencias de resultados cognitivos y no-cognitivos entre estudiantes de escuelas públicas y privadas en la educación secundaria de Argentina: Un análisis multinivel”*. Education Policy Analysis Archives, Vol. 11, Núm. 6, 2003.

Dabenigno, Valeria; Larripa, Silvina; Austral, Rosario; Iñigo Luisa. *“Análisis multinivel de los resultados de las pruebas PISA 2012 de Matemática en la Ciudad Autónoma de Buenos Aires”*. Dirección General de Evaluación de la Calidad Educativa (DGECE), Ministerio de Educación de la ciudad de Buenos Aires, 2014.

Evaluación e Información Educativa, Subsecretaría de. *“Pruebas Aprender”*, 2024. <https://www.argentina.gob.ar/educacion/evaluacion-informacion-educativa/aprender>

Fernández, José Muñiz. *“Introducción a la Teoría de Respuesta a los Ítems”*. Pirámide. Madrid, 1997.

Llach, Juan; Cornejo, Magdalena. *“Factores condicionantes de los aprendizajes. Primaria y Secundaria”*. Serie de informes de investigación/3. Ministerio de Educación de la Nación, 2018.

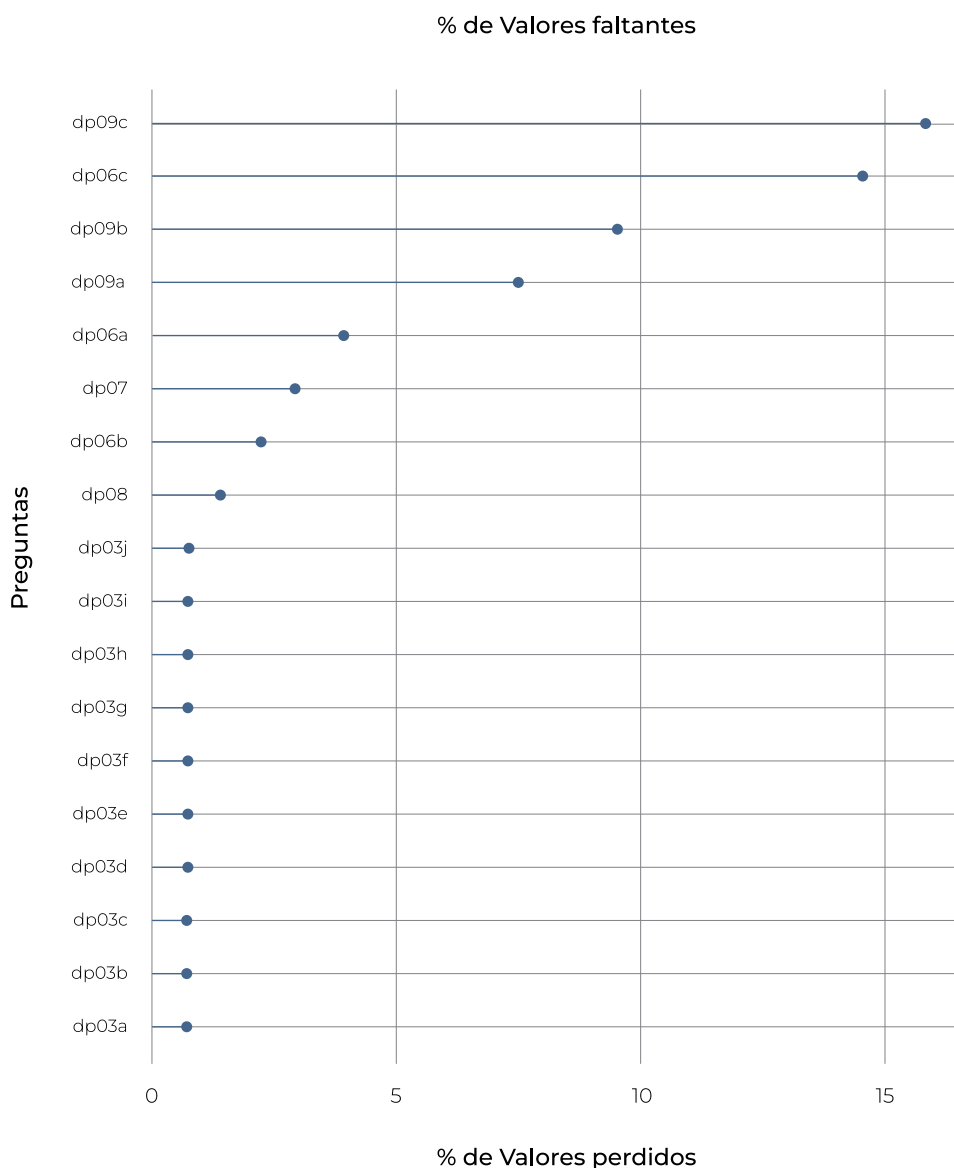
OCDE. *“El Programa PISA de la OCDE. Qué es y para qué sirve”*, 2016. <https://www.oecd.org/pisa/39730818.pdf>

Serie de Documentos Técnicos/2. *“Aprender 2016: Notas Metodológicas”*, 2016. [https://www.argentina.gob.ar/sites/default/files/notas\\_metodologicas\\_0.pdf](https://www.argentina.gob.ar/sites/default/files/notas_metodologicas_0.pdf)

Snijders, Tom; Bosker, Roel. *“Multilevel Analysis”*. Sage Publishers, 2004.

## ► Anexo Metodológico

Aprender administra un cuestionario a los directivos de cada colegio encuestado. Para el presente trabajo se exploraron un conjunto de variables que pueden estar asociadas a la gestión de la institución, observándose un bajo nivel de no respuesta, tal como puede verse en el **Gráfico**. Se imputaron mediante *hotdeck* las no respuestas para no tener que eliminar esos colegios del análisis.



Fuente: Evaluación Aprender 2023. Base Directivos. DNEIEE-REFCEE | SSIEE | Secretaría de Educación | Ministerio de Capital Humano

## Distribución de frecuencia de las variables seleccionadas

Alumnos según sector de gestión (%)

Estatal	73
Privada	27
Total	100

Fuente: Evaluación Aprender 2023, DNEIEE-REFCEE | SSIEE | Secretaría de Educación | Ministerio de Capital Humano

Características del alumno y hogar

VARIABLES	CATEGORÍAS	SECTOR DE GESTIÓN	
		Estatal	Privada
Sexo	Mujer	49,8	51,1
	Varón	50,2	48,9
Edad	11 o menos	67,2	74,3
	12 o más	32,8	25,7
Nivel educativo de la madre	Primario	21,6	5,5
	Secundaria	49,8	33,1
	Terciario/ universitario	28,6	61,4
Asistencia al jardín	Sala de tres	51,4	73,2
	Sala de cuatro	31,6	20,7
	No asistió/ Sala de cinco	16,9	6,1
Cantidad de libros en el hogar	0 a 5	43,4	18,4
	6 a 50	36,6	43,6
	51 o más	20	38
Habitaciones	1 a 2	39,1	25,1
	3	29,2	30,6
	4 o más	31,7	44,3
Hacinamiento (personas por habitación)	0 a 1	24,8	45,8
	1 a 2,3	49,5	45,9
	Más de 2,3	25,7	8,3
Bienes en el hogar	Lugar tranquilo para estudiar	82,8	89,3
	Auto	57,2	84,9
	Baño	95,6	99,2
	Lavarropas	94,1	98,2
	Cable	91,9	97,2
	Agua	89,2	92,4
	Computadora	54	83,3
	Consola	44,5	63,7
Tablet	37,4	50,5	

Fuente: Evaluación Aprender 2023, DNEIEE-REFCEE | SSIEE | Secretaría de Educación | Ministerio de Capital Humano

## Características de la escuela/aula

VARIABLES	CATEGORÍAS	SECTOR DE GESTIÓN	
		Estatal	Privada
Clima escolar	Bajo	23,2	8,8
	Medio	63	75,3
	Alto	13,9	15,8
Edad del Director	Hasta 45	21,1	22,7
	46 a 55	58,6	55,6
	56 o más	20,2	21,7
Antigüedad en el cargo del director	No tiene antigüedad	14,1	5,3
	5 años o menos	51,8	38,6
	Entre 6 y 10 años	18,8	27
	Entre 11 y 20 años	11,5	22,6
	Más de 20 años	3,8	6,5

Fuente: Evaluación Aprender 2023, DNEIEE-REFCEE | SSIEE | Secretaría de Educación | Ministerio de Capital Humano