

APRENDIZAJE  
COMPOSICIÓN  
ECONÓMICA  
REGRESIÓN  
MULTINIVEL  
HETEROGÉNEO  
SIMULACIÓN  
INVARIANZA  
CLASIFICACIÓN  
COMPONENTES  
PRINCIPALES

Desafíos en la medición y el análisis  
del estatus socioeconómico  
de los estudiantes peruanos

APRENDIZAJE  
COMPOSICIÓN  
ECONÓMICA  
REGRESIÓN  
MULTINIVEL  
HETEROGÉNEO  
SIMULACIÓN  
INVARIANZA  
CLASIFICACIÓN  
COMPONENTES  
PRINCIPALES

## Desafíos en la medición y el análisis del estatus socioeconómico de los estudiantes peruanos



PERÚ

Ministerio  
de Educación

**Daniel Alfaro Paredes**

Ministro de Educación

**José Carlos Chávez Cuentas**

Viceministro de Gestión Institucional

**Susana Helfer Llerena**

Viceministro de Gestión Pedagógica

**Carlos Castro Serón**

Secretario de Planificación Estratégica

**Humberto Pérez León Ibáñez**

Jefe de la Oficina de Medición de la Calidad de los Aprendizajes

**Desafíos en la medición y el análisis del estatus socioeconómico de los estudiantes peruanos**

**Equipo técnico**

Andrés Christiansen (coordinador)

Wilmer Hernández

Manuel Marcos

Giovanna Moreano

Juan Carlos Saravia

**Asesor del estudio**

Fernando Andrade Adaniya

Esta publicación es el producto final del esfuerzo institucional de la UMC a través de sus diferentes equipos de especialistas.

**Primera edición, mayo 2018**

**Tiraje:** 1500 ejemplares

**Impresión:** Corporación Creagrama E.I.R.L.

Av. Argentina 975, Lima - Lima

Mayo 2018

**ISBN: 978-9972-246-82-1**

Hecho el Depósito Legal en la Biblioteca Nacional del Perú N° 2018-07231

©Ministerio de Educación, 2018

Calle Del Comercio 193, San Borja

Lima, Perú

Teléfono: 615-5800

[www.minedu.gob.pe](http://www.minedu.gob.pe)

Citar esta publicación de la siguiente manera:

Ministerio de Educación. (2018). *Desafíos en la medición y el análisis del estatus socioeconómico de los estudiantes peruanos*. Lima: Oficina de Medición de la Calidad de los Aprendizajes.

En el presente documento, se utilizan de manera inclusiva términos como “el docente”, “el estudiante” y sus respectivos plurales (así como otras palabras equivalentes en el contexto educativo) para referirse a hombres y mujeres. Esta opción se basa en una convención idiomática y tiene por objetivo evitar las formas para aludir a ambos géneros en el idioma castellano (“o/a”, “los/las” y otras similares), debido a que implican una saturación gráfica que puede dificultar la comprensión lectora.

---



# Índice

<b>Lista de siglas y abreviaturas</b> . . . . .	<b>7</b>
<b>Presentación</b> . . . . .	<b>9</b>
<b>Capítulo 1: Introducción</b> . . . . .	<b>11</b>
1.1 Definición del estatus socioeconómico y su medición en la ECE . . . . .	12
1.2 Relación entre el estatus socioeconómico y el rendimiento académico . . . . .	14
1.3 Desafíos en el análisis del estatus socioeconómico . . . . .	15
<b>Capítulo 2: Elaboración del índice socioeconómico en la ECE</b> . . . . .	<b>17</b>
2.1 Codificación de las preguntas utilizadas . . . . .	19
2.1.1 Educación de los padres . . . . .	19
2.1.2 Material de construcción de la vivienda . . . . .	20
2.1.3 Servicios básicos en el hogar . . . . .	21
2.1.4 Activos y otros servicios en el hogar . . . . .	22
2.2 Imputación de valores perdidos . . . . .	23
2.3 Cálculo de indicadores del ISE . . . . .	24
2.3.1 Matrices de correlaciones . . . . .	24
2.3.2 Análisis de componentes principales . . . . .	26
2.4 Construcción del índice socioeconómico . . . . .	28
2.5 Análisis descriptivos del índice socioeconómico . . . . .	29
<b>Capítulo 3: Estabilidad del índice socioeconómico entre grados y años</b> . . . . .	<b>33</b>
3.1 Introducción . . . . .	34
3.2 Metodología . . . . .	34
3.3 Resultados . . . . .	36
3.3.1 Estabilidad del ISEP entre grados en una misma escuela . . . . .	36
3.3.2 Estabilidad del ISEP en el tiempo en una misma escuela . . . . .	39
3.4 Conclusiones . . . . .	41
<b>Capítulo 4: La heterogeneidad social en la construcción del índice socioeconómico</b> . . . . .	<b>43</b>
4.1 Introducción . . . . .	44
4.1.1 Desigualdad social y clasificación socioeconómica . . . . .	44
4.1.2 Construcción del ISE considerando heterogeneidad social . . . . .	45
4.2 Metodología . . . . .	47
4.3 Resultados . . . . .	48
4.3.1 Igualdad de las matrices de correlación por subpoblación . . . . .	48
4.3.2 Análisis de componentes principales . . . . .	48
4.3.3 Análisis de componentes principales comunes . . . . .	49
4.4 Conclusiones . . . . .	52

<b>Capítulo 5: Agregación del índice socioeconómico</b>	<b>55</b>
5.1 Introducción	56
5.1.1 Variabilidad del índice socioeconómico al interior de las escuelas	57
5.1.2 Centrado de variables predictivas	58
5.2 Metodología	59
5.2.1 Aleatorización de datos reales	60
5.2.2 Estudio de simulación de efectos contextuales	61
5.3 Resultados	61
5.3.1 Efectos contextuales en aleatorización de datos reales	61
5.3.2 Simulación de efectos contextuales	63
5.4 Conclusiones	64
<b>Capítulo 6: Variación del efecto del índice socioeconómico entre escuelas</b>	<b>65</b>
6.1 Introducción	66
6.2 Metodología	67
6.2.1 Población y muestra	67
6.2.2 Medición	67
6.2.3 Estrategia analítica	69
6.3 Resultados	70
6.4 Conclusiones	72
<b>Capítulo 7: Comentarios finales</b>	<b>73</b>
<b>Referencias</b>	<b>77</b>
<b>Anexos</b>	<b>83</b>
A1 Preguntas usadas en la elaboración del índice socioeconómico	84
A2 Densidad del ISE en 2.º grado de secundaria	86
B1 Densidad y correlación del ISEP entre años y entre grados	87
C1 Análisis de componentes principales comunes por área y región natural	95
D1 Modelos multinivel con datos aleatorizados	99
D2 Estudio de simulación - código en R	102
E1 Indicadores utilizados en el modelo de regresión multinivel	103

## Lista de siglas, abreviaturas y símbolos

AIC	Criterio de información de Akaike
ACP	Análisis de componentes principales
ACPC	Análisis de componentes principales comunes
CCI	Coefficiente de correlación intraclase
BIC	Criterio de información bayesiano
CFI	Índice de ajuste comparativo
DE	Desviación estándar
ECE	Evaluación Censal de Estudiantes
ESE	Estatus socioeconómico
ICSC	Índice de calidad del servicio y el contexto de la escuela
ISE	Índice socioeconómico del estudiante
ISEP	Índice socioeconómico promedio de la escuela
GM	Centrado a la gran media
$\mathcal{L}$	Función de máxima verosimilitud
matviv	Índice de material de la vivienda
maxedu	Máximo nivel educativo en el hogar en años de estudio
MC	Centrado a la media del clúster
otrsv	Índice de otros servicios en el hogar
RMSEA	Error cuadrático medio de aproximación
serbas	Índice de servicios básicos en el hogar
TLI	Índice de Tucker-Lewis
UMC	Oficina de Medición de Calidad de los Aprendizajes





## Presentación

El presente informe, elaborado por la Oficina de la Medición de la Calidad de los Aprendizajes, examina la relación entre el estatus socioeconómico y el logro académico dentro del contexto de la educación primaria y secundaria en el Perú. Este estudio goza, en mi modesta opinión, de un carácter único en el contexto peruano y latinoamericano, ya que son muy pocas las investigaciones que reflexionan detalladamente sobre la manera en que el estatus socioeconómico puede ser medido y exploran diferenciadamente su relación con el rendimiento académico. Desde esta perspectiva, contribuye al diálogo ya existente sobre el efecto del estatus socioeconómico en el rendimiento académico.

Subyacentes al contenido de este informe existen cuatro ideas clave. La primera es la distinción entre la conceptualización del estatus socioeconómico y su medición. Es así que en el capítulo 1, se presenta la definición del estatus socioeconómico utilizada, resaltando su carácter multidimensional. De otro lado, el capítulo 2 detalla cómo se elabora la medida del estatus socioeconómico de los estudiantes utilizada, a través de la construcción de un índice socioeconómico. El cambio de nombre refleja la diferencia entre la definición teórica del estatus socioeconómico y la forma en que es medido. De esta manera, se usa “estatus socioeconómico” para denominar a la parte conceptual, mientras que “índice socioeconómico” es utilizado para referirse a aspectos más concretos de la medición.

La segunda idea importante es la distinción entre el estatus socioeconómico del estudiante y la composición socioeconómica de la escuela. Cabe notar que el estatus socioeconómico es una característica individual, pues generalmente los estudiantes lo adquieren de los padres o de la familia con la que viven. Sin embargo, es muy común en ciencias sociales generar agregados de características individuales y acreditarlas como características representativas de un grupo, institución u organización. El agregado de estas características individuales a nivel grupal debe adquirir otro significado conceptual. Es así que las escuelas no tienen estatus socioeconómico, inteligencia o rendimiento académico; son los estudiantes quienes poseen estas características. Por ello, el capítulo 3 examina la estabilidad de la composición socioeconómica de una escuela entre grados y en el tiempo. Además, en el capítulo 5 se ve con mayor profundidad aspectos técnicos referidos a la relación de la composición socioeconómica de la escuela con el rendimiento académico.

En tercer lugar, se problematiza sobre la complejidad de la heterogeneidad social y económica en las distintas regiones del país. El Perú es un país que goza de una gran diversidad natural y cultural que es importante reflejar en este tipo de estudios. Por ejemplo, las diferencias culturales entre regiones o estratos poblacionales podrían afectar la conceptualización y, sobre todo, la estimación del estatus

socioeconómico. Es por ello que el capítulo 4 examina en detalle la construcción del índice socioeconómico, incorporando diferencias entre centros urbanos y rurales, así como diferencias entre regiones naturales (costa, sierra y selva).

Finalmente, se destaca la necesidad de considerar la complejidad de las escuelas cuando se quiere entender cómo es que algunas características o rasgos individuales se relacionan con logros académicos. Generalmente, se asume que el efecto del estatus socioeconómico en el rendimiento académico tiende a ser similar o idéntico para todos los estudiantes, al margen de otras características individuales o escolares. En otras palabras, las estimaciones del efecto del estatus socioeconómico sobre el rendimiento escolar no consideran otro tipo de relaciones. Por ello, el capítulo 6 explora en detalle cómo es que esta relación puede variar dependiendo de algunas características escolares.

En resumen, el lector encontrará en el primer capítulo la definición conceptual del estatus socioeconómico, además de una discusión de su relación teórica con el rendimiento académico, así como de los desafíos encontrados al analizar dicha relación. Los siguientes tres capítulos (del 2 al 4) examinan detalladamente aspectos técnicos en la construcción del índice socioeconómico. Esto incluye la metodología usada en su estimación, la estabilidad temporal de estos índices, la estabilidad dentro de escuelas y la inclusión de la heterogeneidad de algunos conglomerados poblacionales en la construcción del índice socioeconómico. El capítulo 5 examina con mayor profundidad aspectos técnicos relacionados con la estimación del índice de composición socioeconómica de la escuela y su relación con el rendimiento académico. Finalmente, el capítulo 6 presenta y discute resultados de la relación diferenciada del estatus socioeconómico en función a características escolares.

Por último, me gustaría hacer un reconocimiento al trabajo permanente que el equipo de análisis de la Oficina de Medición de la Calidad de los Aprendizajes viene desarrollando desde hace algunos años. La cantidad y calidad de las publicaciones que viene produciendo está continuamente aumentando, y estoy seguro que la constante presencia de sus publicaciones representa una contribución importante a nuestro sistema educativo.

Fernando H. Andrade Adaniya

Ann Arbor, Michigan. Marzo de 2018

# Introducción

## Capítulo 1

El estatus socioeconómico (ESE) ha sido identificado como uno de los factores más importantes en explicar la variación en el desempeño académico de los estudiantes (León & Collahua, 2016; Ministerio de Educación, 2017c; Sirin, 2005; White, 1982). Sus efectos sobre el rendimiento de los estudiantes hacen necesaria una exploración profunda, con el objetivo de identificar en qué medida los resultados académicos son producto de los procesos educativos o del contexto en que se encuentran.

Por ello, en los últimos años, la Oficina de Medición de la Calidad de los Aprendizajes (UMC) del Ministerio de Educación ha considerado necesario recoger información socioeconómica de los estudiantes en los grados en los que se aplica la Evaluación Censal de Estudiantes (ECE). En función de ello, se decidió elaborar un indicador socioeconómico que permitiera caracterizar a los estudiantes y a sus escuelas, contextualizar los logros de aprendizaje y visibilizar problemas del sistema educativo peruano, para lograr que los estudiantes desarrollen aprendizajes de calidad y con equidad. La elaboración y uso de este indicador ha generado el planteamiento de diversas preguntas que se intentan responder en este informe.

### **1.1. Definición del estatus socioeconómico y su medición en la ECE**

Si bien el estatus socioeconómico ha sido estudiado extensivamente por las ciencias sociales, muchos autores concuerdan en la complejidad para llegar a una definición única, sobre todo por su carácter multidimensional y su estrecha relación con el contexto de la realidad que pretende evaluar o medir (Caro & Cortés, 2012; Dickinson & Adelson, 2014; León & Collahua, 2016; National Center for Education Statistics, 2012; Raudenbush & Willms, 1995; Sirin, 2005; van Ewijk & Slegers, 2010; White, 1982). Sin embargo, existe cierto consenso en señalar que el estatus socioeconómico permite el ordenamiento relativo de individuos, a partir de sus características económicas y sociales dentro de una estructura social jerárquica (Caro & Cortés, 2012; León & Collahua, 2016; Mueller & Parcel, 1981).

Para aproximarse al ESE muchos estudios recogen indicadores sobre tres aspectos: el ingreso familiar, el nivel educativo de los padres y el prestigio ocupacional de los mismos, esquema que responde a la definición tripartita propuesta por O. D. Duncan, Featherman y Duncan (1972). Bajo este esquema, el ingreso familiar representa los recursos a los cuales potencialmente se puede acceder, mientras que el nivel educativo y la ocupación indican el prestigio que puede tener una familia, además de sus niveles de ingreso (Sirin, 2005).

La construcción de un índice socioeconómico (ISE) que pueda aproximarse al estatus socioeconómico de las familias depende de la disponibilidad de información

y del contexto social que se evalúa. Por eso, muchas instituciones o investigaciones elaboran índices con características a las que tienen acceso o que consideran pertinentes para la realidad que intentan representar (León & Collahua, 2016). De esta forma, también se incluyen recursos o posesiones en el hogar, recursos de lectura en el hogar, ingresos familiares adicionales, características de la comunidad y recursos escolares (León & Collahua, 2016; National Center for Education Statistics, 2012; Sirin, 2005). Como señalan diversos autores (Gottfried & Ream, 2014; Sirin, 2005; White, 1982), la intensidad de la relación que se pueda encontrar entre el rendimiento académico y el ESE dependerá de la forma en que este último es medido.

En el caso peruano resulta conveniente ampliar el espectro de información recogida para tener una mejor aproximación al potencial socioeconómico de las familias. Dada la estructura económica peruana y sus características geográficas y sociales, es pertinente incluir variables que ayuden a diferenciar de manera más precisa estatus socioeconómicos bajos, como el acceso a servicios básicos y tipos de materiales de construcción de las viviendas.

La UMC ha desarrollado diversos indicadores socioeconómicos que acompañen sus evaluaciones nacionales e internacionales. Fue el caso de las ECE 2015 y 2016 donde se recogió información socioeconómica<sup>1</sup> mediante cuestionarios a los padres de familia (en primaria) y a los estudiantes (en secundaria). Los cuestionarios incluyeron preguntas sobre el máximo nivel educativo alcanzado por los padres o tutores, servicios básicos y no básicos en el hogar, material de construcción de paredes, pisos y techos, y activos o posesiones en el hogar del estudiante<sup>2</sup>. A pesar de su importancia, se excluyeron la jerarquización ocupacional de los padres y el ingreso familiar, debido a que el nivel de informalidad del mercado laboral peruano afecta la validez de las escalas ocupacionales internacionales, y la información autorreportada sobre el ingreso puede no ser fidedigna.

Finalmente, cabe mencionar que las variables incluidas en este índice guardan coherencia con otras mediciones peruanas del estatus socioeconómico desarrolladas para ámbitos educativos y no educativos (Asociación Peruana de Empresas de Investigación de Mercados, 2016, 2017; Caro, 2002; Centro de Investigación y Desarrollo del Instituto Nacional de Estadística e Informática, 2007). Además, este indicador ha sido objeto de una constante revisión y ajuste, lo que ha permitido mejorar su medición y conceptualización a través de diferentes evaluaciones (Ministerio de Educación, 2016a, 2017b, 2017c, 2017e).

<sup>1</sup>Esta información se recogió de manera muestral para 2.º grado de primaria en 2015 y 2016, y de manera censal para 4.º grado de primaria en 2016 y para 2.º grado de secundaria en 2015 y 2016.

<sup>2</sup>En revisiones anteriores del ISE se incluyó el hacinamiento en el hogar. Sin embargo, esta variable no contribuyó de forma significativa a la varianza explicada por el primer componente en un análisis de componentes principales.

## 1.2. Relación entre el estatus socioeconómico y el rendimiento académico

Como se mencionó anteriormente, en abundante literatura nacional (Caro, 2002; León & Collahua, 2016; Ministerio de Educación, 2016a, 2017b, 2017c) e internacional (Sirin, 2005; White, 1982) se hace referencia a la estrecha relación que existe entre el estatus socioeconómico y el rendimiento académico. Dicha relación se manifiesta tanto en el ámbito familiar como en el ámbito escolar del estudiante. En el primer caso, un alto estatus socioeconómico permitiría a los padres invertir en el desarrollo de habilidades de sus hijos, adquirir materiales de aprendizaje y tener más tiempo para dedicarse a su crianza (Gottfried & Ream, 2014). Por el contrario, estudiantes pertenecientes a familias con pocos recursos económicos tendrían menor acceso a capital económico, social y cultural y, por lo tanto, mayores restricciones para desarrollar sus aprendizajes (Coleman, 1988; Coleman et ál., 1966).

Respecto al ámbito escolar, el estatus socioeconómico de los estudiantes estaría asociado al tipo de escuelas y aulas al que los estudiantes asisten; por ejemplo, tener acceso a escuelas con mejores materiales y profesores más experimentados suele estar reservado para estudiantes pertenecientes a familias con mayores recursos socioeconómicos (National Center for Education Statistics, 2012). Más aún, escuelas con estudiantes de alto estatus socioeconómico tienden a ofrecer mejores condiciones de enseñanza y proveer ambientes cognitivamente más retadores, lo cual resulta favorable para el aprendizaje de los estudiantes (Caro, 2002; Gottfried & Ream, 2014; Sirin, 2005). Otros aspectos dentro del aula que pueden ser influenciados por el ESE de los estudiantes son el estilo de enseñanza de los docentes, el clima disciplinario, y las relaciones entre padres y docentes tal como ha sido advertido por diferentes estudios (Lareau, 1987; van Ewijk & Slegers, 2010).

Estas diferencias entre las condiciones donde se desarrollan los aprendizajes evidencian problemas de segregación escolar en un sistema educativo; es decir, la concentración de estudiantes de características económicas y sociales similares en determinado tipo de escuelas (Bellei, 2013; Murillo, 2016; Vazquez, 2012). En el caso peruano, este fenómeno ha afectado principalmente a estudiantes de bajo estatus socioeconómico, en su mayoría pertenecientes a escuelas estatales y rurales (Ministerio de Educación, 2016a, 2017b). Este hecho representa un doble riesgo para el logro de los aprendizajes de estas poblaciones, puesto que a las desventajas propias que implica tener bajo estatus socioeconómico se suman las desventajas de asistir a escuelas que en su mayoría carecen de condiciones adecuadas para su desarrollo académico y personal. Por ello, resulta pertinente analizar en qué medida y en qué contextos persiste esta segregación y cómo puede influir en el rendimiento de los estudiantes.

### 1.3. Desafíos en el análisis del estatus socioeconómico

La relación entre el ESE y el rendimiento, ampliamente tratada en la literatura, se puede ver evidenciada en diversas investigaciones publicadas por la UMC (Ministerio de Educación, 2016a, 2016b, 2017b, 2017c). Sin embargo, esta relación puede depender de la construcción del indicador socioeconómico (National Center for Education Statistics, 2012; Sirin, 2005), es decir, ser producto de las variables incluidas en el indicador o de la técnica utilizada para calcularlo. La relación también puede depender de la forma en que es especificada o de cómo se propone analizarla en un modelo de regresión. Sumado a esto, la constante revisión y ajuste en la medición y conceptualización del ESE por la UMC ha generado reflexiones y preguntas relacionadas al índice socioeconómico.

En primer lugar, en cuanto al índice socioeconómico construido, es posible preguntarse si varía en un período corto de tiempo y si se generan problemas para un correcto cálculo del análisis de componentes principales, considerando la heterogeneidad social presente en el país. Por otro lado, de la relación encontrada con el rendimiento queda analizar si esta puede estimarse correctamente con respecto a las características socioeconómicas de la escuela, y si el efecto encontrado es constante en los diferentes tipos de escuela.

Este informe intenta responder algunas de estas preguntas mediante cuatro estudios. El capítulo 3 compara el indicador socioeconómico escolar de diferentes años y grados, para determinar si existe un cambio en el tiempo o entre grados de una misma escuela. Para lo primero, se compara el indicador de 2.º grado de primaria y de 2.º grado de secundaria de 2015 y 2016. Para lo segundo, se comparan los indicadores de 2.º y 4.º grado de primaria de 2016. El análisis encuentra muy poca variabilidad entre años y entre grados, lo que mostraría que el indicador no varía de forma significativa en el período de un año, y que el indicador de un solo grado podría representar adecuadamente el indicador promedio de la escuela.

En el capítulo 4, se exploran las relaciones existentes entre las variables que componen el indicador socioeconómico en diferentes estratos: área y región natural. Se encuentra que algunas de las relaciones entre las variables son distintas en cada estrato. A partir de ello, se propone la utilización del análisis de componentes principales comunes, que busca una solución única para diagonalizar las distintas matrices de correlaciones de cada estrato, en vez de hacerlo con la matriz de correlaciones de toda la población.

En el capítulo 5 se pretende determinar si la agregación realizada para obtener un indicador socioeconómico promedio de la escuela, puede producir una estimación sesgada del coeficiente de regresión sobre el rendimiento. En este análisis se realiza una aleatorización de estudiantes en diferentes escuelas, obteniendo nuevos



indicadores promedio y aumentando la varianza del indicador individual dentro de la escuela. Se encontró que el efecto del indicador socioeconómico promedio desaparecería en las escuelas aleatorizadas, lo que se complementa con un estudio de simulación para datos con y sin efecto contextual. Los resultados de ambas técnicas sugieren una estimación insesgada del índice socioeconómico promedio de la escuela.

Finalmente, en el capítulo 6 se realiza un análisis de regresión multinivel sobre el rendimiento de Lectura en la ECE 2016, considerando el indicador socioeconómico y otras covariables. Se encontró que el efecto del indicador varía de forma importante entre escuelas, lo que sugiere que la relación entre el estatus socioeconómico y el rendimiento es distinta entre las escuelas.

# Elaboración del índice socioeconómico en la ECE

Capítulo 2

Los cuatro estudios presentados en este informe utilizan como aproximación al estatus socioeconómico un índice socioeconómico (ISE) calculado por la Oficina de Medición de la Calidad de los Aprendizajes. En este capítulo se detalla el cálculo de este índice usando la información obtenida mediante cuestionarios de los padres de familia o tutores y de los estudiantes a través de la Evaluación Censal de Estudiantes (ECE) de primaria y secundaria, respectivamente.

Se recogió información para calcular cinco indicadores que formen el ISE: años de estudio de los padres, materiales de construcción de la vivienda, servicios básicos en el hogar, activos en el hogar y otros servicios en el hogar. Estos indicadores, a excepción del primero, fueron calculados a través de un análisis de componentes principales (ACP) que permitió recoger la mayor parte de la varianza compartida entre los ítems que los componen. Luego, se realizó un ACP de segundo orden incorporando a los puntajes obtenidos en la primera etapa el máximo número de años de estudio de los padres. Con ello se pudo obtener un único indicador (ISE) que representara el estatus socioeconómico de la familia del estudiante.

A continuación, se describe cómo se transformaron las respuestas de los padres y de los estudiantes para obtener los indicadores que componen el ISE y qué método de imputación se siguió para tratar los valores perdidos, para especificar luego la construcción de cada indicador y los estadísticos correspondientes. Finalmente, se detalla el cálculo del ISE y sus principales medidas de tendencia central. Estos análisis fueron realizados separadamente con la información de 4.º grado de primaria y 2.º grado de secundaria, debido a que fueron recogidos de manera censal en 2016. Luego, para obtener el ISE de los otros grados evaluados estos se equipararon utilizando los estadísticos previamente calculados.

## 2.1. Codificación de las preguntas utilizadas

En esta sección se detalla la recodificación de las respuestas de padres y estudiantes a los cuestionarios respecto a los cinco indicadores utilizados en el cálculo del índice socioeconómico<sup>3</sup>.

### 2.1.1. Educación de los padres

Para recoger los años de estudio de los padres se preguntó por el máximo nivel educativo de cada uno, teniendo en cuenta 12 categorías en primaria y 10 en secundaria. Las respuestas de los padres de familia y de los estudiantes fueron recodificadas según los años de estudio promedio que implicarían cada una de estas categorías (tablas 2.1 y 2.2). Esta codificación se utilizó porque permitía precisar mejor las distancias entre las diferentes respuestas. Para el análisis final solo se consideró el máximo de años de estudio alcanzado por uno de los padres.

**Tabla 2.1. Códigos asignados a años de estudio en cuestionarios de primaria**

Respuesta	Código
Sin estudios	0
Primaria incompleta	3
Primaria completa	6
Secundaria incompleta	8,5
Secundaria completa	11
Educación ocupacional incompleta	7,5
Educación ocupacional completa	9
Superior no universitaria incompleta: pedagógica, técnica, artística o militar/policial	12,5
Superior no universitaria completa: pedagógica, técnica, artística o militar/policial	14
Superior universitaria incompleta o militar/policial (escuela de oficiales)	13,5
Superior universitaria completa o militar/policial (escuela de oficiales)	16
Posgrado (maestría y/o doctorado)	19

**Tabla 2.2. Códigos asignados a años de estudio en cuestionarios de secundaria**

Respuesta	Código
Sin estudios	0
Primaria incompleta	3
Primaria completa	6
Secundaria incompleta	8,5
Secundaria completa	11
Superior no universitaria incompleta	12,5
Superior no universitaria completa	14
Superior universitaria incompleta	13,5
Superior universitaria completa	16
Después de la universidad ha seguido estudiando (maestría o doctorado)	19

<sup>3</sup>Las versiones originales de las preguntas utilizadas se encuentran en el Anexo A1.

### 2.1.2. Material de construcción de la vivienda

En primaria y secundaria se preguntó de manera similar sobre el material de construcción de las paredes, los pisos y los techos. Las respuestas de los padres de familia y estudiantes fueron reasignadas a los códigos mostrados en las tablas 2.3, 2.4 y 2.5. El orden de los códigos asignados se determinó utilizando la media de ingresos en relación con cada material de la Encuesta Nacional de Hogares 2015 del Instituto Nacional de Estadística e Informática.

**Tabla 2.3. Códigos asignados a material predominante en las paredes**

Respuesta	Código
Ladrillo o bloque de cemento	5
Piedra, sillar con cal o cemento	2
Adobe o tapia	4
Quincha (caña con barro)	2
Piedra con barro	2
Madera o tablas	3
Esteras	1
Otro	1

**Tabla 2.4. Códigos asignados a material predominante en los pisos**

Respuesta	Código
Parquet o madera pulida	6
Pisos asfálticos, vinílicos o similares	5
Losetas, mayólicas, terrazos o similares	4
Madera (entablados)	3
Cemento	2
Tierra	1
Otro	1

**Tabla 2.5. Códigos asignados a material predominante en los techos**

Respuesta	Código
Concreto armado (cemento y ladrillo)	7
Madera	5
Tejas	6
Planchas de calamina, fibra de cemento (eternit) o similares	4
Caña o estera con barro	3
Esteras	2
Paja u hojas de palmera	4
Otro	1

### 2.1.3. Servicios básicos en el hogar

Sobre los servicios básicos en el hogar se recogió información sobre la procedencia del agua en la vivienda, la tenencia de desagüe y el tipo de alumbrado. Como se puede observar en las tablas 2.6, 2.7 y 2.8, los tres ítems fueron transformados en variables dicotómicas que informan sobre la conexión interna a la red pública de agua y desagüe, así como sobre el alumbrado eléctrico.

**Tabla 2.6. Códigos asignados a procedencia del agua**

Respuesta	Código
Del caño dentro de la casa	1
Del caño fuera de la casa	0
Del pilón de uso público	0
Del camión cisterna, aguatero u otro similar	0
De un pozo	0
Del río, acequia, riachuelos, manantial o similar	0
Otro	0

**Tabla 2.7. Códigos asignados a desagüe**

Respuesta	Código
El baño está dentro de la casa y podemos jalar la palanca o cadena.	1
El baño está fuera de la casa, es compartido con los vecinos y podemos jalar la palanca o cadena.	0
Tenemos un baño propio que no está conectado a un desagüe, pero tiene un tratamiento químico.	0
Tenemos un baño, compartido con los vecinos, que no está conectado a un desagüe, pero tiene un tratamiento químico.	0
Tenemos un baño propio que no está conectado a un desagüe, solo tiene un pozo.	0
Tenemos un baño, compartido con los vecinos, que no está conectado a desagüe, solo tiene un pozo.	0
No tenemos baño.	0

**Tabla 2.8. Códigos asignados a tipo de alumbrado**

Respuesta	Código
Electricidad	1
Generador y/o motor	0
Mechero o lámpara de kerosene	0
Lámpara de petróleo o gas	0
Vela	0
Otro	0

### 2.1.4. Activos y otros servicios en el hogar

En ambos cuestionarios (padres y estudiantes) se indagó también sobre diversos activos que podrían tenerse en el hogar, separándolos en dos categorías: activos en el hogar y otros servicios en el hogar<sup>4</sup>. Las respuestas de los padres de familia y estudiantes fueron reasignadas a los códigos mostrados en la tabla 2.9.

**Tabla 2.9. Códigos asignados a activos y otros servicios en el hogar**

Respuesta	Código
Radio	No = 0, Sí = 1
Licuada	No = 0, Sí = 1
Plancha eléctrica	No = 0, Sí = 1
Televisión	No = 0, Sí = 1
Celular	No = 0, Sí = 1
Reproductor de video	No = 0, Sí = 1
Computadora de escritorio	No = 0, Sí = 1
Laptop o computadora portátil	No = 0, Sí = 1
Tablet	No = 0, Sí = 1
Equipo de sonido	No = 0, Sí = 1
Consola de videojuegos (Xbox, Wii, etc.)	No = 0, Sí = 1
Horno microondas	No = 0, Sí = 1
Refrigeradora	No = 0, Sí = 1
Lavadora	No = 0, Sí = 1
Moto	No = 0, Sí = 1
Auto y/o camioneta	No = 0, Sí = 1
Teléfono fijo	No = 0, Sí = 1
Plan de datos o internet móvil	No = 0, Sí = 1
Conexión a internet	No = 0, Sí = 1
Servicio de televisión por cable	No = 0, Sí = 1

<sup>4</sup>Las variables Moto y Auto y/o camioneta fueron sumadas, quedando 0 si no poseía ninguno de estos vehículos, 1 si poseía uno de los dos y, 2 si poseía ambos.

## 2.2. Imputación de valores perdidos

Con el fin de no perder una importante cantidad de casos para generar el índice socioeconómico, se hizo una imputación simple para cada una de las variables. Solo fueron considerados para la imputación los casos que respondieron más de dos tercios de las variables utilizadas, es decir, 19 de las 28.

Como se puede observar en la tabla 2.10, el porcentaje de casos que cumplía el requisito fue significativamente mayor en secundaria que en primaria, lo que puede deberse a que el cuestionario de padres de familia (primaria) fue enviado al hogar y el estudiante estuvo encargado de devolverlo el segundo día de aplicación de la ECE. En cambio, el cuestionario del estudiante (secundaria) fue recogido en el aula junto con la prueba.

**Tabla 2.10. Número de casos imputados por grado**

	4.º grado de primaria	2.º grado de secundaria
Cuestionarios	539 669	504 076
≥ 19 respuestas	410 818	490 676
≥ 1 valor perdido	169 804	71 408
Porcentaje de casos considerados	76,1 %	97,3 %
Porcentaje de casos imputados sobre el total de cuestionarios	31,5 %	14,2 %

La imputación simple fue realizada de forma multivariada, utilizando únicamente las variables consideradas en el ISE mediante la librería *mice* (van Buuren & Groothuis-Oudshoorn, 2011) del lenguaje *R* (R Core Team, 2017) y considerando la semilla 343. Las variables con dos categorías fueron imputadas mediante una regresión logística, y aquellas con más de dos categorías mediante una regresión logística multinomial<sup>5</sup>. Los estadísticos de tendencia central y dispersión de los diferentes estratos analizados en el informe, no presentan diferencias estadísticamente significativas entre los datos imputados y no imputados, sugiriendo que la imputación no modificaría las distribuciones originales de los datos. Además, análisis previos confirmaron que la relación con el rendimiento no presenta diferencias estadísticamente significativas entre el ISE calculado con y sin imputación.

<sup>5</sup>Para mayor detalle del proceso utilizado por esta librería puede revisarse el manual de referencia (van Buuren & Groothuis-Oudshoorn, 2011).



### 2.3. Cálculo de indicadores del ISE

Como se detalló anteriormente, cuatro de las cinco dimensiones que forman el ISE recogen información de múltiples preguntas respondidas por los padres o los estudiantes. Por ello, fue necesario transformar estas preguntas en indicadores que las agruparan mediante análisis de componentes principales. En cada agrupación, el primer componente fue utilizado para la construcción final del ISE.

#### 2.3.1. Matrices de correlaciones

Para producir los indicadores de material de construcción de la vivienda, de servicios básicos en el hogar, así como de activos y otros servicios en el hogar, se obtuvieron las matrices de correlaciones policóricas de los ítems para cada indicador. En las tablas 2.11, 2.12, 2.13 y 2.14 se muestran estas correlaciones para 4.º grado de primaria y 2.º grado de secundaria<sup>6</sup>. Como se puede observar, todas las correlaciones al interior de cada dimensión son positivas y, a excepción de radio en 2.º grado de secundaria, medianas o altas.

**Tabla 2.11. Matrices de correlaciones - Material de la vivienda**

		2.º de secundaria		
		pared	piso	techo
4.º de primaria	pared			0,471
	piso	0,682		0,525
	techo	0,546	0,577	

**Tabla 2.12. Matrices de correlaciones - Servicios básicos en el hogar**

		2.º de secundaria		
		luz	agua	desagüe
4.º de primaria	luz		0,588	0,629
	agua	0,656		0,772
	desagüe	0,698	0,858	

**Tabla 2.13. Matrices de correlaciones - Otros servicios en el hogar**

		2.º de secundaria			
		1	2	3	4
4.º de primaria	1. teléfono fijo		0,548	0,735	0,653
	2. televisión por cable	0,658		0,593	0,570
	3. internet	0,853	0,718		0,848
	4. internet móvil	0,749	0,697	0,883	

<sup>6</sup>Las tablas resumen las correlaciones intra-grado. Por debajo de la diagonal se encuentran las correlaciones entre ítems de 4.º grado de primaria, por encima de la diagonal las correspondientes a 2.º grado de secundaria.

Tabla 2.14. Matrices de correlaciones - Activos en el hogar

	2.° de secundaria														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. radio	0,336	0,158	0,115	0,169	0,155	0,227	0,188	0,140	0,162	0,301	0,173	0,157	0,091	0,128	0,080
2. licuadora	0,322	0,818	0,754	0,754	0,524	0,615	0,627	0,570	0,537	0,538	0,528	0,642	0,731	0,542	0,330
3. plancha	0,320	0,788	0,760	0,699	0,529	0,574	0,624	0,580	0,546	0,490	0,529	0,646	0,717	0,587	0,310
4. televisor	0,350	0,567	0,544	0,621	0,535	0,816	0,592	0,526	0,536	0,626	0,520	0,604	0,689	0,503	0,362
5. celular	0,360	0,652	0,648	0,811	0,572	0,551	0,548	0,528	0,522	0,474	0,527	0,528	0,566	0,487	0,282
6. rep. de video	0,329	0,678	0,693	0,623	0,478	0,607	0,527	0,484	0,486	0,558	0,549	0,529	0,557	0,468	0,286
7. computadora	0,323	0,648	0,656	0,596	0,530	0,604	0,624	0,552	0,542	0,484	0,549	0,593	0,656	0,631	0,311
8. laptop	0,302	0,652	0,670	0,643	0,537	0,641	0,647	0,563	0,445	0,445	0,544	0,570	0,582	0,577	0,345
9. tablet	0,449	0,644	0,618	0,679	0,492	0,660	0,606	0,699	0,465	0,465	0,568	0,587	0,596	0,589	0,299
10. equipo de sonido	0,292	0,541	0,556	0,531	0,467	0,641	0,603	0,599	0,624	0,600	0,484	0,528	0,542	0,478	0,324
11. videojuegos	0,320	0,695	0,699	0,635	0,475	0,629	0,673	0,638	0,662	0,640	0,652	0,591	0,570	0,588	0,316
12. microondas	0,261	0,780	0,746	0,753	0,491	0,612	0,704	0,639	0,692	0,646	0,605	0,786	0,737	0,723	0,287
13. refrigeradora	0,290	0,686	0,708	0,622	0,471	0,622	0,720	0,696	0,708	0,622	0,664	0,812	0,819	0,734	0,306
14. lavadora	0,251	0,433	0,421	0,439	0,406	0,422	0,444	0,487	0,440	0,475	0,476	0,447	0,441	0,451	0,279
15. vehículo															

### 2.3.2. Análisis de componentes principales

A partir de estas matrices se realizó un análisis de componentes principales para cuatro indicadores: material de la vivienda (matviv), servicios básicos en el hogar (serbas), activos en el hogar (activos) y otros servicios en el hogar (otrsv). La tabla 2.15 muestra que en todos los casos el primer componente logra captar más del 50 % de varianza entre los ítems.

**Tabla 2.15. Varianza explicada por los ítems de cada indicador**

Indicador	4.º grado de primaria	2.º grado de secundaria
	Varianza explicada	Varianza explicada
matviv	73,51 %	69,20 %
serbas	82,62 %	77,65 %
activos	61,70 %	54,30 %
otrsv	82,14 %	74,66 %

En las tablas 2.16, 2.17, 2.18 y 2.19 se presentan los pesos<sup>7</sup>, las cargas en el componente, las medias y las desviaciones estándar de cada variable con respecto a cada indicador. Se puede observar que todos los ítems cargan de forma importante en el primer componente.

**Tabla 2.16. Pesos y cargas en el primer componente - Material de la vivienda**

Variable	4.º grado de primaria				2.º grado de secundaria			
	Peso	Carga	Media	DE	Peso	Carga	Media	DE
pared	0,395	0,870	4,300	0,981	0,404	0,840	4,311	1,047
piso	0,401	0,884	4,272	1,534	0,417	0,865	4,427	1,506
techo	0,371	0,817	3,681	1,594	0,380	0,789	4,065	1,673

**Tabla 2.17. Pesos y cargas en el primer componente - Servicios básicos en el hogar**

Variable	4.º grado de primaria				2.º grado de secundaria			
	Peso	Carga	Media	DE	Peso	Carga	Media	DE
luz	0,345	0,856	0,940	0,238	0,356	0,830	0,954	0,209
agua	0,374	0,927	0,687	0,464	0,385	0,898	0,683	0,465
desagüe	0,380	0,942	0,657	0,475	0,392	0,914	0,704	0,457

<sup>7</sup>Los pesos del componente son la ponderación de las cargas en el componente por la varianza explicada.

**Tabla 2.18. Pesos y cargas en el primer componente - Activos en el hogar**

Variable	4.º grado de primaria				2.º grado de secundaria			
	Peso	Carga	Media	DE	Peso	Carga	Media	DE
radio	0,047	0,438	0,795	0,404	0,029	0,238	0,873	0,333
licuadora	0,093	0,857	0,762	0,426	0,102	0,829	0,858	0,349
plancha	0,092	0,853	0,690	0,463	0,100	0,817	0,773	0,419
televisor	0,092	0,847	0,881	0,323	0,102	0,832	0,939	0,240
celular	0,073	0,675	0,872	0,334	0,088	0,716	0,846	0,361
rep. de video	0,088	0,813	0,610	0,488	0,093	0,761	0,791	0,406
computadora	0,088	0,812	0,373	0,484	0,096	0,782	0,454	0,498
laptop	0,087	0,808	0,307	0,461	0,091	0,741	0,426	0,495
tablet	0,089	0,828	0,357	0,479	0,091	0,740	0,409	0,492
equipo de sonido	0,086	0,798	0,511	0,500	0,086	0,703	0,696	0,460
videojuegos	0,083	0,764	0,178	0,382	0,091	0,741	0,287	0,452
microondas	0,092	0,850	0,350	0,477	0,100	0,816	0,436	0,496
refrigeradora	0,094	0,866	0,623	0,485	0,105	0,856	0,673	0,469
lavadora	0,093	0,856	0,429	0,495	0,095	0,777	0,476	0,499
vehículo	0,063	0,586	0,479	0,625	0,054	0,442	0,604	0,648

**Tabla 2.19. Pesos y cargas en el primer componente - Otros servicios en el hogar**

Variable	4.º grado de primaria				2.º grado de secundaria			
	Peso	Carga	Media	DE	Peso	Carga	Media	DE
teléfono fijo	0,274	0,902	0,304	0,460	0,285	0,852	0,362	0,481
televisión por cable	0,256	0,841	0,515	0,500	0,258	0,770	0,608	0,488
internet	0,291	0,957	0,383	0,486	0,311	0,929	0,510	0,500
internet móvil	0,281	0,922	0,336	0,472	0,301	0,898	0,497	0,500

Finalmente, para hallar el puntaje individual en cada indicador, los ítems fueron estandarizados y luego multiplicados por el peso respectivo. A continuación se muestra la fórmula utilizada, donde  $i$  representa al estudiante y  $j$  las variables del indicador:

$$\text{Indicador}_i = \sum_{j=1}^{\text{número de ítems}} \left( \frac{\text{ítem}_{ij} - \text{media}_j}{\text{DE}_j} \right) \times \text{peso}_j \quad (2.1)$$

## 2.4. Construcción del índice socioeconómico

Mediante los puntajes obtenidos con la fórmula (2.1) y el máximo nivel educativo alcanzado por los padres (maxedu), se construyó una matriz de correlaciones de Pearson (tabla 2.20) que permitiera el cálculo del ACP de segundo orden para obtener el índice socioeconómico. Como puede verse, existe una correlación mediana entre todos los indicadores que forman el ISE, cuyo primer componente, en ambos grados, explica el 62 % de la varianza entre indicadores, tanto en primaria como en secundaria. Además, la tabla 2.21 muestra cómo todos los indicadores presentan cargas adecuadas en el primer componente.

**Tabla 2.20. Matrices de correlaciones - Índice socioeconómico**

		2.º de secundaria				
		maxedu	matviv	serbas	activos	otrserv
4.º de primaria	maxedu		0,427	0,372	0,525	0,478
	matviv	0,465		0,509	0,599	0,550
	serbas	0,439	0,534		0,594	0,453
	activos	0,513	0,562	0,534		0,742
	otrserv	0,456	0,533	0,433	0,772	

**Tabla 2.21. Pesos y cargas en el primer componente - Índice socioeconómico**

Variable	4.º grado de primaria				2.º grado de secundaria			
	Peso	Carga	Media	DE	Peso	Carga	Media	DE
maxedu	0,230	0,716	10,665	4,018	0,223	0,695	11,520	4,295
matviv	0,252	0,785	0,000	0,950	0,251	0,783	0,000	0,945
serbas	0,237	0,737	0,000	0,861	0,236	0,737	0,000	0,855
activos	0,280	0,871	0,000	0,794	0,285	0,890	0,000	0,783
otrserv	0,265	0,824	0,000	0,892	0,266	0,829	0,000	0,890

Para estimar el ISE de cada individuo, cada indicador fue estandarizado y luego multiplicado por el peso respectivo. Esta fue la fórmula utilizada:

$$\text{Índice socioeconómico (ISE)}_i = \sum_{j=1}^5 \left( \frac{\text{ítem}_{ij} - \text{media}_j}{\text{DE}_j} \right) \times \text{peso}_j \quad (2.2)$$

Con el fin de obtener índices comparables entre grados y años de un mismo nivel escolar que permitieran realizar análisis de estabilidad del ISE, se utilizaron las medias, desviaciones estándar y pesos obtenidos en los ACP de primer orden y de segundo orden. En el caso de primaria, se equiparó la información de 2.º grado de primaria en 2015 y 2016 con lo hallado en 4.º grado de primaria. En secundaria, se equipararon las respuestas al cuestionario de 2.º grado de secundaria en 2015 con las respuestas mostradas de 2016. Los índices equiparados se utilizan en los análisis de los capítulos 3 y 5.

## 2.5. Análisis descriptivos del índice socioeconómico

Como se muestra en las tablas 2.22 y 2.23, existen diferencias importantes entre el ISE nacional en función a las estratificaciones por área, gestión de la escuela y región geográfica<sup>8</sup>. En 4.º grado de primaria y 2.º grado de secundaria, los estudiantes que asisten a escuelas urbanas tienen en promedio 0,15 de una desviación estándar (DE) más que la media nacional, mientras que los estudiantes de escuelas rurales tienen 1,2 DE menos que la media nacional. Además, existe una diferencia cercana de 1 DE del ISE entre estudiantes que asisten a escuelas estatales y no estatales. Estas tendencias pueden verse también en las figuras 2.1 y 2.2 para 4.º grado de primaria<sup>9</sup> y son consistentes con datos nacionales provenientes de investigaciones educativas y no educativas (Asociación Peruana de Empresas de Investigación de Mercados, 2016, 2017; Centro de Investigación y Desarrollo del Instituto Nacional de Estadística e Informática, 2007; Ministerio de Educación, 2016a)<sup>10</sup>.

**Tabla 2.22. Estadísticos descriptivos del ISE - 4.º grado de primaria**

Estrato	Media	DE	Percentil 25	Percentil 50	Percentil 75
Nacional	-0,023	0,999	-0,809	-0,048	0,807
Urbano	0,151	0,926	-0,564	0,142	0,923
Rural	-1,198	0,610	-1,608	-1,250	-0,868
Estatal	-0,347	0,888	-1,013	-0,405	0,292
No estatal	0,852	0,720	0,393	1,003	1,424
Costa	0,281	0,931	-0,409	0,338	1,065
Sierra	-0,428	0,925	-1,119	-0,559	0,182
Selva	-0,616	0,893	-1,277	-0,720	-0,020

**Tabla 2.23. Estadísticos descriptivos del ISE - 2.º grado de secundaria**

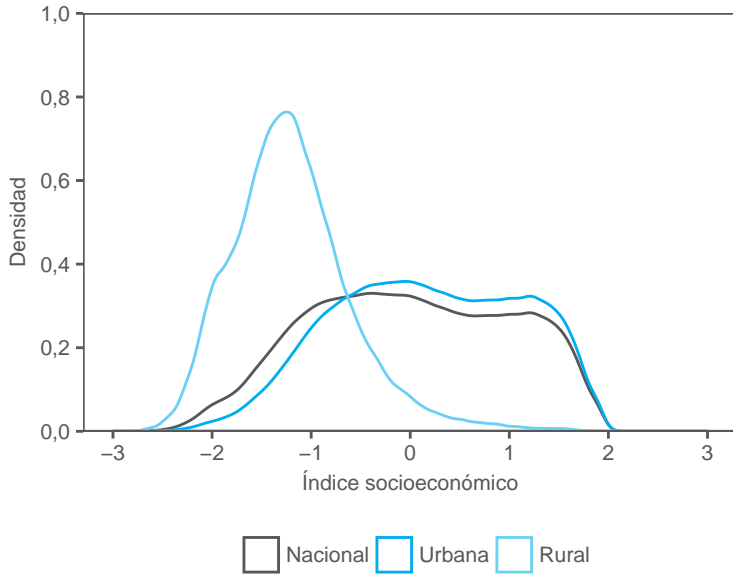
Estrato	Media	DE	Percentil 25	Percentil 50	Percentil 75
Nacional	0,000	1,000	-0,768	0,100	0,850
Urbano	0,153	0,920	-0,523	0,266	0,919
Rural	-1,258	0,699	-1,704	-1,298	-0,863
Estatal	-0,279	0,938	-0,987	-0,263	0,472
No estatal	0,860	0,624	0,578	1,006	1,302
Costa	0,377	0,854	-0,164	0,547	1,053
Sierra	-0,469	0,956	-1,185	-0,566	0,233
Selva	-0,612	0,958	-1,307	-0,637	0,085

<sup>8</sup>Se utiliza la clasificación propuesta por Ministerio de Educación (2016b).

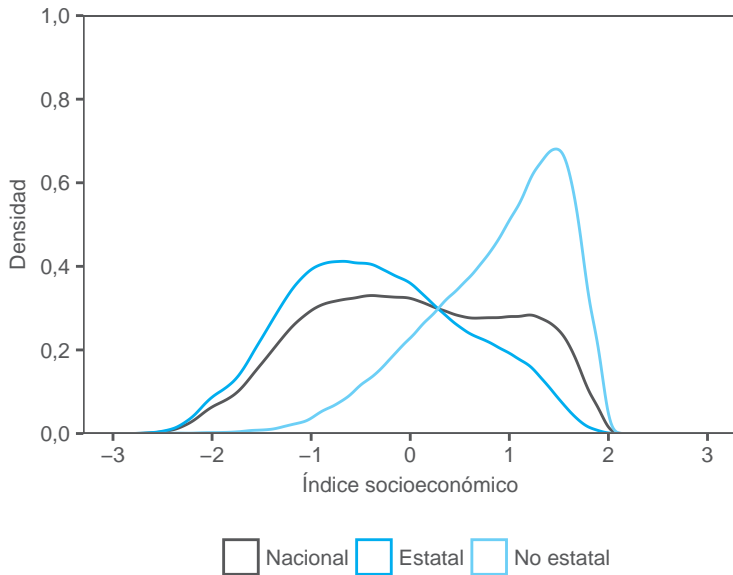
<sup>9</sup>Para los gráficos de densidad de 2.º grado de secundaria ver el Anexo A2.

<sup>10</sup>La clasificación por regiones geográficas se detalla en el capítulo 4.

**Figura 2.1. Densidad del ISE de 4.º grado de primaria según área de la escuela**



**Figura 2.2. Densidad del ISE de 4.º grado de primaria según gestión de la escuela**



Según lo encontrado, el ISE parece estimar de forma aproximada la composición socioeconómica de los estratos que caracterizan a la población peruana. Esto, sumado al alto porcentaje de varianza explicada en los ACP, sugieren la solidez de la construcción de este índice. Finalmente, como se verá en los capítulos 5 y 6, el índice socioeconómico calculado con la información de la ECE se relaciona de forma significativa con los resultados de rendimiento académico, tal como lo muestran muchos otros estudios (León & Collahua, 2016; Sirin, 2005; White, 1982).





# Estabilidad del índice socioeconómico entre grados y años

Capítulo 3

### 3.1. Introducción

La Evaluación Censal de Estudiantes (ECE) evalúa sólo algunos grados de la educación básica regular, ya que su aplicación se vuelve compleja en términos logísticos a medida que aumenta la cantidad de grados y áreas evaluadas. Por lo tanto, sería relevante examinar si el indicador socioeconómico calculado en base a los estudiantes de un grado específico es representativo para todo el nivel educativo<sup>11</sup> de ese grado. Además, debido a su posible uso en estudios longitudinales, es importante verificar si este indicador es estable en el tiempo. No se esperan mayores diferencias en el corto plazo, pero a largo plazo la composición socioeconómica de la escuela podría variar. Por consiguiente, también sería pertinente monitorear estas diferencias con la finalidad de obtener patrones longitudinales.

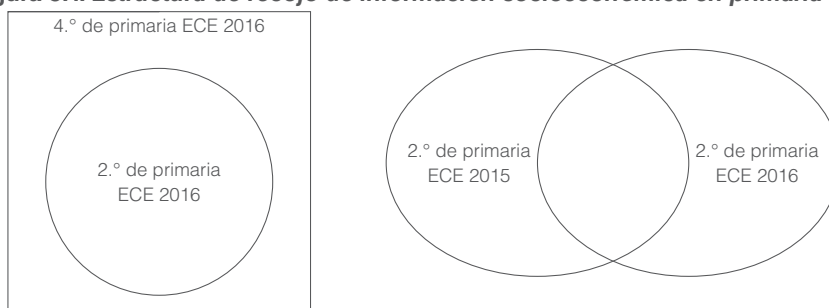
Asimismo, conocer la variación o la estabilidad de la composición socioeconómica de las escuelas permitiría entender mejor el fenómeno de la segregación escolar, teniendo mayores indicios para analizar si las escuelas estarían atrayendo estudiantes con similares características socioeconómicas a través de los años. Dicho de otro modo, se haría evidente si esta segregación escolar persiste en el tiempo.

Es por ello que en el presente estudio se plantean dos objetivos principales: 1) analizar la estabilidad del índice socioeconómico entre grados de una misma escuela; 2) analizar la estabilidad del índice socioeconómico en el tiempo dentro una misma escuela.

### 3.2. Metodología

Para el análisis de la estabilidad del indicador socioeconómico entre grados en una misma escuela, se utilizó un grupo de comparación formado por las escuelas de la ECE 2016 para 2.º y 4.º grado de primaria. Para examinar la estabilidad del ISE en el tiempo se utilizaron dos grupos de comparación: 1) escuelas que fueron evaluadas en las ECE 2015 y 2016 para 2.º grado de primaria; 2) escuelas que fueron evaluadas en las ECE 2015 y 2016 para 2.º grado de secundaria. Dado que la información socioeconómica en 2.º grado de primaria se recogió de manera muestral, solo se consideraron las escuelas evaluadas que coinciden en las comparaciones mencionadas (figura 3.1).

<sup>11</sup> La Educación Básica Regular en el contexto peruano se compone de tres niveles educativos: inicial, primaria y secundaria.

**Figura 3.1. Estructura de recojo de información socioeconómica en primaria**

La estabilidad del indicador socioeconómico, entre grados y en el tiempo dentro de las escuelas, se analizó mediante: 1) la distribución de la varianza del indicador entre escuelas y estudiantes; 2) análisis visuales. Para este propósito se utiliza como principal insumo el índice socioeconómico promedio de la escuela (ISEP), que resulta de un promedio simple de los valores del índice socioeconómico individual (ISE) de los estudiantes pertenecientes a un grado o escuela.

En primer lugar, usando el ISE como variable de criterio, se calculó el coeficiente de correlación intraclase (CCI) estimando modelos multinivel nulos para cada uno de los grupos. En estos modelos se consideró al estudiante y su escuela como los niveles de análisis. De esta manera, se pudo obtener la distribución de la varianza del índice socioeconómico por nivel, para cada grado o año de comparación y para la población en conjunto. La distribución de la varianza del ISE entre escuelas y estudiantes refleja el comportamiento similar del índice socioeconómico entre grados o en el tiempo en una misma escuela, tanto para el ámbito nacional como para los estratos empleados en este estudio (área rural y urbana; gestión estatal y no estatal; y región costa, sierra y selva).

Para mostrar la estabilidad del ISEP entre grados y en el tiempo para el ámbito nacional y cada uno de los estratos, se realizaron análisis visuales mediante dos tipos de figuras: distribución de densidad y dispersión del ISEP a nivel de escuela. Esto permite la comparación del comportamiento general del ISEP y de su correspondencia, calculados en una escuela en dos grados o años.

### 3.3. Resultados

#### 3.3.1. Estabilidad del ISEP entre grados en una misma escuela

La tabla 3.1 muestra la distribución de la varianza del ISE entre grados, donde se puede observar que a nivel nacional el CCI<sup>12</sup> de 2.º y 4.º grados de primaria en la ECE 2016 son semejantes. Además, el CCI muestra que la varianza del ISE corresponde entre 31 % y 65 % a la escuela, es decir, a pesar de que esta es una variable individual existe mucha similitud entre los estudiantes de una misma escuela. El CCI para ambos grados en conjunto es similar al calculado por separado. La tendencia se mantiene al analizar este coeficiente por estratos, aunque se presenten algunas pequeñas diferencias. Estos resultados son un primer indicio de que el ISEP, para ambos grados, podría tener un comportamiento similar en el ámbito nacional y para cada uno de los estratos considerados.

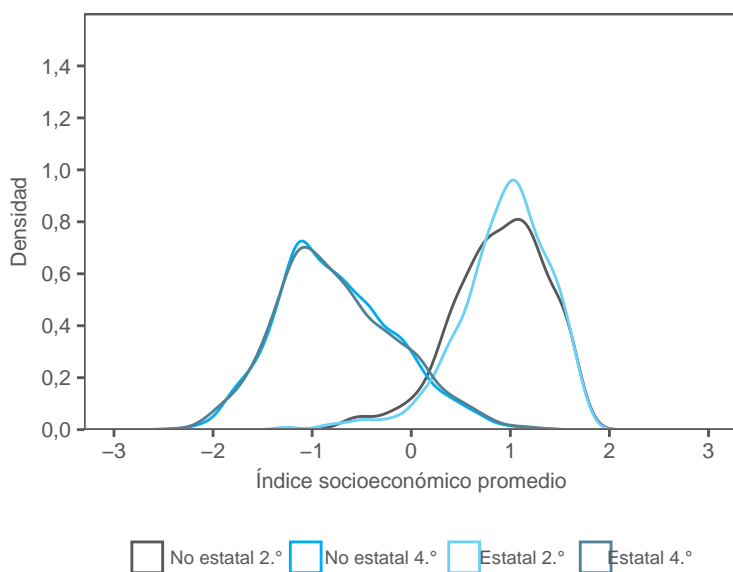
**Tabla 3.1. Distribución de la varianza del ISE de 2.º y 4.º grado de primaria - ECE 2016**

Estrato	Grupo	Estudiante	Escuela
<b>Nacional</b>	2.º grado	0,394	0,606
	4.º grado	0,394	0,606
	2.º y 4.º grado	0,394	0,606
<b>Urbana</b>	2.º grado	0,451	0,549
	4.º grado	0,451	0,549
	2.º y 4.º grado	0,451	0,549
<b>Rural</b>	2.º grado	0,686	0,314
	4.º grado	0,688	0,312
	2.º y 4.º grado	0,694	0,306
<b>Estatad</b>	2.º grado	0,541	0,459
	4.º grado	0,540	0,460
	2.º y 4.º grado	0,542	0,458
<b>No Estadad</b>	2.º grado	0,532	0,468
	4.º grado	0,549	0,451
	2.º y 4.º grado	0,536	0,464
<b>Costa</b>	2.º grado	0,351	0,649
	4.º grado	0,350	0,650
	2.º y 4.º grado	0,349	0,651
<b>Sierra</b>	2.º grado	0,510	0,490
	4.º grado	0,509	0,491
	2.º y 4.º grado	0,512	0,488
<b>Selva</b>	2.º grado	0,477	0,523
	4.º grado	0,489	0,511
	2.º y 4.º grado	0,486	0,514

<sup>12</sup>Este estadístico describe la proporción de varianza que corresponde al segundo nivel de análisis, en este caso la escuela.

Para analizar la estabilidad del índice socioeconómico promedio de las escuelas, se debe observar la distribución de densidad del ISEP. Estas distribuciones se encuentran superpuestas, tanto para el ámbito nacional como para cada uno de los estratos que fueron empleados en el estudio (figura 3.2). Esta superposición reafirma que el ISEP de 2.º y 4.º grados de primaria se distribuirían de manera similar.

**Figura 3.2. Densidad del ISEP de 2.º y 4.º grado de primaria según gestión de la escuela - ECE 2016**

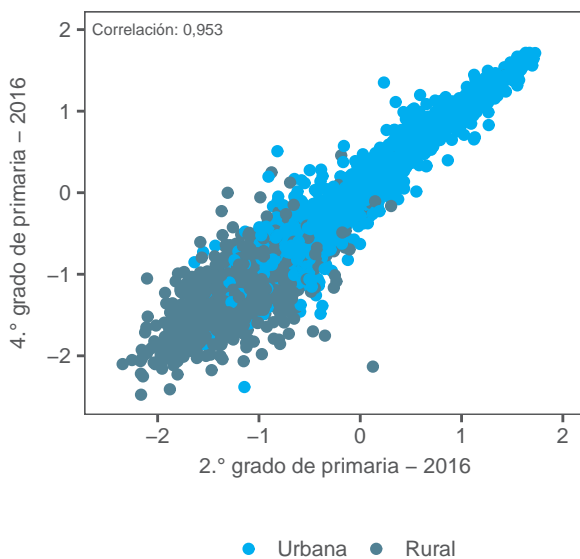


En el estrato estatal se observa que la distribución de densidad del ISEP para 2.º y 4.º grados de primaria a nivel de escuela presenta una ligera asimetría positiva, lo que también se observa en la distribución de densidad del ISEP para el estrato urbano. Esto puede ser explicado porque las escuelas en el Perú se diferencian por tamaño y cantidad de población atendida.

Se puede afirmar que las distribuciones entre grados, nacionales y dentro de cada estrato son similares, pero esto no indica necesariamente la existencia de una alta relación del ISEP para ambos grados de una misma escuela. Para poder realizar esta afirmación se necesitan otras herramientas; en ese sentido, los gráficos de dispersión son útiles para determinar la existencia de una alta relación del ISEP entre ambos grados de una misma escuela.

Como se puede observar en la figura 3.3, existe una alta correlación entre los ISEP de los grados de una misma escuela, tanto a nivel nacional como en los estratos urbano y rural<sup>13</sup>. La concentración de los puntos en los gráficos de dispersión muestra comportamientos diferenciados que van de acuerdo a las características de su estrato.

**Figura 3.3. Dispersión del ISEP de 2.º y 4.º grado de primaria según área de la escuela - ECE 2016**



<sup>13</sup>Esto se repite en los otros estratos analizados y puede verse en el Anexo B1.

### 3.3.2. Estabilidad del ISEP en el tiempo en una misma escuela

En el caso de la comparación del ISEP del mismo grado en los años 2015 y 2016, los resultados son similares a los encontrados entre los grados de una misma escuela<sup>14</sup>. En primer lugar, el CCI guarda correspondencia entre años a nivel nacional y por cada uno de los estratos analizados (tabla 3.2).

**Tabla 3.2. Distribución de la varianza del ISE de 2.º grado de primaria 2015 y 2016**

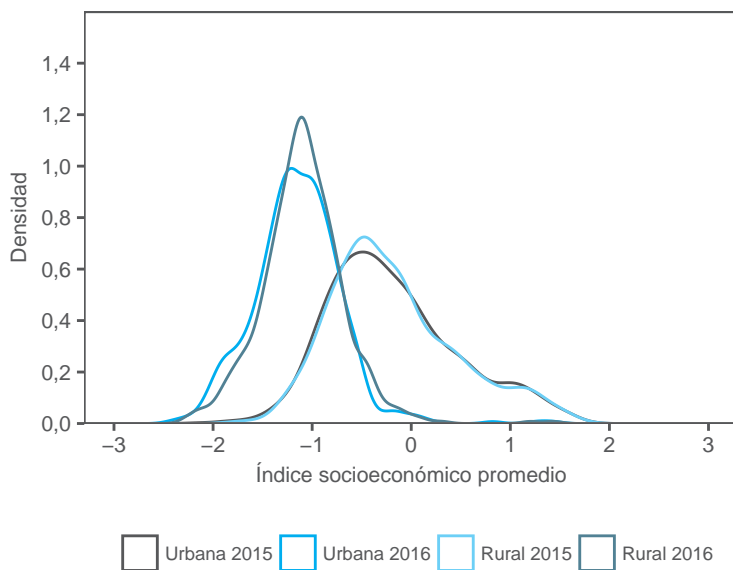
Estrato	Grupo	Estudiante	Escuela
<b>Nacional</b>	ECE 2015	0,446	0,554
	ECE 2016	0,438	0,562
	ECE 2015 - 2016	0,441	0,559
<b>Urbana</b>	ECE 2015	0,514	0,486
	ECE 2016	0,498	0,502
	ECE 2015 - 2016	0,506	0,494
<b>Rural</b>	ECE 2015	0,633	0,367
	ECE 2016	0,663	0,337
	ECE 2015 - 2016	0,666	0,334
<b>Estatal</b>	ECE 2015	0,577	0,423
	ECE 2016	0,568	0,432
	ECE 2015 - 2016	0,570	0,430
<b>No Estatal</b>	ECE 2015	0,568	0,432
	ECE 2016	0,551	0,449
	ECE 2015 - 2016	0,559	0,441
<b>Costa</b>	ECE 2015	0,458	0,542
	ECE 2016	0,444	0,556
	ECE 2015 - 2016	0,448	0,552
<b>Sierra</b>	ECE 2015	0,523	0,477
	ECE 2016	0,518	0,482
	ECE 2015 - 2016	0,523	0,477
<b>Selva</b>	ECE 2015	0,508	0,492
	ECE 2016	0,489	0,511
	ECE 2015 - 2016	0,497	0,503

<sup>14</sup>En esta sección se detallan los resultados para 2.º grado de primaria. En el Anexo B1 se exponen los resultados de 2.º grado de secundaria, los cuales mantienen la misma tendencia.



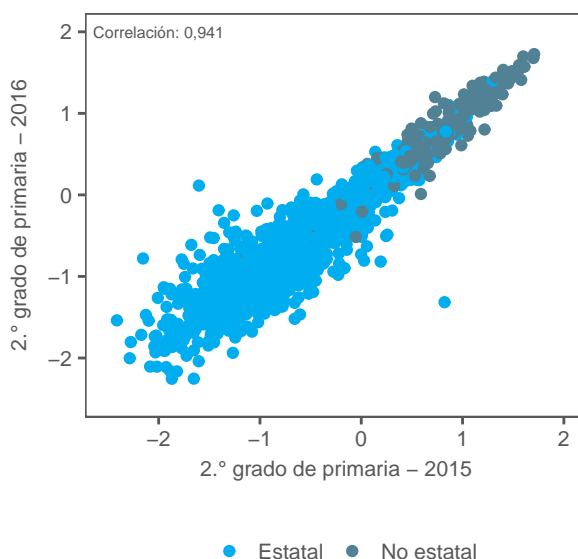
Las distribuciones de densidad del ISEP de 2.º grado de primaria para las ECE 2015 y 2016 también se encuentran superpuestas para el ámbito nacional y para cada uno de los estratos. En la figura 3.4 se muestra su distribución para las escuelas urbanas y rurales.

**Figura 3.4. Densidad del ISEP de 2.º grado de primaria según área de la escuela - ECE 2015 y 2016**



Finalmente, la dispersión del ISEP en la figura 3.5 muestra la alta correspondencia entre años para las escuelas estatales y no estatales. Al igual que en la sección anterior, el comportamiento del ISEP por estrato se corresponde con lo descrito en el capítulo 2. Estos análisis permiten concluir que no existirían diferencias importantes entre los ISEP de un mismo grado en dos años consecutivos.

**Figura 3.5. Dispersión del ISEP de 2.º grado de primaria según gestión de la escuela - ECE 2015 y 2016**



### 3.4. Conclusiones

Los resultados encontrados a lo largo de este capítulo muestran que la estimación del ISEP para un grado podría resultar representativa de todo el nivel educativo (primaria o secundaria). Además, el ISEP recogido en un determinado año puede ser representativo del año siguiente. La estabilidad del ISEP no debe ser entendida como la uniformidad del estatus socioeconómico dentro de una escuela, sino como la conservación, entre grados y años, de una composición socioeconómica producto de la segregación escolar que podría perjudicar a los estudiantes de bajos recursos para alcanzar mejores resultados educativos.

La estabilidad del ISEP entre grados y en el tiempo probadas en este estudio no está exenta de limitaciones. Si bien los resultados sugieren que existe estabilidad del ISEP entre grados de primaria, no se puede asegurar que esta se mantenga para las escuelas con secundaria. Además, contar con una medición del ISEP para dos años consecutivos no permite determinar en cuánto tiempo el ISEP podría sufrir un cambio relevante. A su vez, cambios bruscos en la composición de la población también pueden generar afectar la estabilidad del ISEP. Por ello, es necesario continuar generando mediciones de las características socioeconómicas de los estudiantes y las escuelas.



# La heterogeneidad social en la construcción del índice socioeconómico

Capítulo 4

### 4.1. Introducción

El índice socioeconómico (ISE) de la Oficina de Medición de la Calidad de los Aprendizajes (UMC) es calculado de forma estandarizada para todos los estudiantes evaluados en la ECE. Sin embargo, esta aproximación puede presentar problemas en un contexto de desigualdad y heterogeneidad social, lo cual hace necesario considerar la forma de incluir estas características contextuales dentro de su estimación.

En este estudio se presenta una forma alternativa de cálculo del ISE que podría recoger esta heterogeneidad. Para ello se analiza cómo la fragmentación y heterogeneidad social afectaría la construcción de indicadores socioeconómicos. Se explora luego la posibilidad de incluir parte de la heterogeneidad social en el cálculo del ISE mediante el método de análisis de componentes principales comunes (ACPC) propuesto por Flury (1984) y extendido por Trendafilov (2003).

#### 4.1.1. Desigualdad social y clasificación socioeconómica

El fenómeno de la desigualdad social implica una distribución de bienes y servicios inequitativa a lo largo del territorio nacional (Banerjee & Duflo, 2012), lo que propicia que la sociedad se aglomere y estratifique en grupos poblacionales con distintos niveles de acceso a oportunidades de desarrollo (Escobal & Torero, 2000; Martínez, 2010; Rivas, 2015). Las diferencias entre estos grupos poblacionales se debe principalmente a aspectos externos a las características de las familias, tales como la segregación urbana y la geografía, que ejercerían una influencia importante en la calidad de vida de sus habitantes (Piketty, 2014; Rivas, 2015; Therborn, 2015).

De acuerdo al Instituto Nacional de Estadística e Informática (2016), existe un comportamiento diferenciado en el acceso a bienes y servicios según el ámbito geográfico en el que se encuentran las familias. Por ejemplo, en cuanto a la presencia de servicios públicos, el área urbana tiene acceso en un 92,8 %, 88,8 % y 99,4 % a las redes públicas de agua, desagüe y electricidad, respectivamente. Por otro lado, el ámbito rural presenta niveles de acceso de 60,3 %, 18,9 % y 77,7 % a dichos servicios. Además, las regiones de la costa concentran en mayor medida a las familias con acceso a redes de servicios públicos, mientras que en la sierra y la selva dicho acceso se reduce considerablemente. Estas diferencias entre ámbitos geográficos se deberían a factores de accesibilidad del territorio.

Adicionalmente, estas diferencias entre regiones naturales podrían explicarse, en parte, por un componente cultural que generaría una valoración diferenciada de bienes. Es decir, aspectos idiosincráticos y de costumbres propias del contexto se encontrarían mediando las elecciones que realizan las familias al momento de adquirir bienes y servicios para satisfacer sus necesidades.

En consecuencia, es importante considerar las diferencias contextuales de la población analizada para construir un índice socioeconómico que refleje adecuadamente a la sociedad. No hacerlo de esta manera podría generar una sobreestimación o subestimación del índice socioeconómico asignado a algunos grupos poblacionales (Yang & Gustafsson, 2004).

#### **4.1.2. Construcción del ISE considerando heterogeneidad social**

Los índices socioeconómicos propuestos por la UMC han sido calculados a través de análisis de componentes principales (ACP) (Ministerio de Educación, 2016a, 2017b, 2017c). Este método consiste en la diagonalización de una matriz de correlaciones o de covarianzas, de tal manera que se obtenga un conjunto mínimo de vectores que maximicen la cantidad de varianza explicada por cada uno de ellos (Alkan, Atakam & Alkan, 2015; Véliz, 2016). Esto se logra mediante la descomposición de la matriz de relaciones en una matriz de autovectores que represente la dirección de los componentes, así como una matriz diagonal de autovalores que contenga la escala del componente.

Dada la alta heterogeneidad social del país, resulta conveniente analizar si las relaciones encontradas entre los componentes del ISE son similares en diversas subpoblaciones. El cálculo del ACP sin considerar estas diferencias puede ocasionar: (a) sesgos en la estimaciones de los puntajes del ISE, (b) sesgos en las relaciones del ISE con otros factores como el rendimiento académico, y (c) problemas en la interpretación de los efectos del ISE en dichos factores.

De este modo, incluir las condiciones de heterogeneidad social en el cálculo del ISE podría significar un avance en la forma de aproximarse a un contexto socioeconómico complejo utilizando herramientas de estimación. Una manera de incorporar estas diferencias es analizar las distintas relaciones que puedan tener las variables en cada de las subpoblaciones. Esto se puede lograr a través del método de ACPC, que incorpora tantas matrices de correlaciones o de covarianzas como grupos se pretenda analizar. El ACPC permite la diagonalización simultánea de estas matrices, obteniendo una matriz única de autovectores y una de autovalores para cada uno de los grupos (Juneja, 2012).

La formalización de este procedimiento considera  $k$  subpoblaciones, siendo  $k = 1, 2, \dots, n$ , y teniendo  $j$  variables, donde  $j = 1, 2, \dots, p$ , con matrices de correlaciones semidefinidas y positivas:

$$A_1 = \begin{bmatrix} A_{11} & \dots & A_{1p} \\ \vdots & \ddots & \vdots \\ A_{p1} & \dots & A_{pp} \end{bmatrix} \neq A_2 = \begin{bmatrix} A_{11} & \dots & A_{1p} \\ \vdots & \ddots & \vdots \\ A_{p1} & \dots & A_{pp} \end{bmatrix} \neq \dots \neq A_n = \begin{bmatrix} A_{11} & \dots & A_{1p} \\ \vdots & \ddots & \vdots \\ A_{p1} & \dots & A_{pp} \end{bmatrix} \quad (4.3)$$

A través de un proceso de diagonalización se obtiene una única matriz de autovectores  $U$  de tamaño  $p \times p$ .

$$U = \begin{bmatrix} U_{11} & \dots & U_{1p} \\ U_{21} & \dots & U_{2p} \\ \vdots & \ddots & \vdots \\ U_{p1} & \dots & U_{pp} \end{bmatrix} \quad (4.4)$$

Y tantos vectores de autovalores  $\Lambda$ , de tamaño  $p \times 1$ , como subpoblaciones existan.

$$\Lambda_1 = \begin{bmatrix} \lambda_{11} \\ \lambda_{12} \\ \vdots \\ \lambda_{1p} \end{bmatrix} \neq \Lambda_2 = \begin{bmatrix} \lambda_{21} \\ \lambda_{22} \\ \vdots \\ \lambda_{2p} \end{bmatrix} \neq \dots \neq \Lambda_n = \begin{bmatrix} \lambda_{n1} \\ \lambda_{n2} \\ \vdots \\ \lambda_{np} \end{bmatrix} \quad (4.5)$$

Los datos hallados se utilizan para el cálculo de las cargas al componente ( $C$ ) de las variables consideradas.

$$C_k = U \times \sqrt{\Lambda_k} \quad (4.6)$$

En el caso del ISE, se considera solo el primer componente. Por lo tanto, los pesos de las variables se calculan de la siguiente manera:

$$P_k = \frac{U \times \sqrt{\Lambda_k}}{\lambda_{k1}} \quad (4.7)$$

Con esta información se procede al cálculo de los componentes mediante la siguiente fórmula:

$$Y_j = \sum_{j=1}^p \left( \frac{X_j - \bar{X}_j}{\sigma_{X_j}} \right) \times P_k \quad (4.8)$$

## 4.2. Metodología

Para poder incorporar la heterogeneidad social en el cálculo del ISE, se consideraron cinco subpoblaciones agrupadas por criterios de ámbito geográfico (urbano, rural) y por regiones naturales (costa, sierra y selva)<sup>15</sup>. La información analizada proviene de los cuestionarios de 4.º grado de primaria y 2.º grado de secundaria en la ECE 2016. En la sección de resultados se muestran los análisis para 4.º grado de primaria, mientras que los de 2.º grado de secundaria se encuentran en el Anexo C1.

En primer lugar, se verificó que las matrices de correlaciones entre los componentes del ISE fueran diferentes para cada subpoblación mediante el test de igualdad de matrices de correlación propuesto por Jennrich (1970). Esto evidencia la presencia de heterogeneidad entre los estratos analizados.

Luego, se realizó el cálculo del índice socioeconómico a través de un ACP para cada subpoblación de manera independiente. Después de obtenido el índice socioeconómico para cada grupo poblacional, se analizó la carga en el componente<sup>16</sup> para cada uno de los indicadores considerados. Asimismo, se analizó la forma en que los valores de las cargas en el componente variaban entre estratos.

Se calculó el índice socioeconómico integrando las diferencias por subpoblaciones utilizando un ACPC mediante la librería *CPCA* (Ziyatdinov, Kanaan-Izquierdo, Trendafilov & Perera-Lluna, 2014) del lenguaje *R* (R Core Team, 2017). Una vez calculadas las cargas en el componente por grupo poblacional, se compararon con las estimadas mediante el ACP. Finalmente, se contrastaron las distribuciones del índice socioeconómico calculado mediante el ACP y el ACPC, a nivel nacional y por cada estrato o subpoblación.

<sup>15</sup>No se considera el estrato de gestión de la escuela (estatal y no estatal), debido a que la asistencia del estudiante a una u otra escuela, en la medida de lo posible, suele estar bajo la decisión de la familia y no se muestra como una característica contextual independiente a esta.

<sup>16</sup>La carga es el efecto directo que tiene cada indicador sobre el componente (denominado índice) sin considerar la varianza explicada por el componente.



### 4.3. Resultados

#### 4.3.1. Igualdad de las matrices de correlación por subpoblación

Las matrices de correlaciones de las subpoblaciones analizadas son distintas entre sí y distintas a la matriz de correlaciones nacional (tabla 4.1), resultado que sugeriría que las relaciones entre los componentes del ISE son diferentes entre subpoblaciones. Por lo tanto, se justifica la utilización del ACPC para el cálculo del ISE, ya que puede recoger estas relaciones diferenciadas.

**Tabla 4.1. Prueba de igualdad de matrices de correlaciones de 4.º grado de primaria - ECE 2016**

Matrices analizadas	$\chi^2$	Significancia
Nacional vs. Urbana	4108,573	<0,001
Nacional vs. Rural	18 515,110	<0,001
Nacional vs. Costa	2431,882	<0,001
Nacional vs. Sierra	3180,602	<0,001
Nacional vs. Selva	2340,000	<0,001
Urbana vs. Rural	15 136,610	<0,001
Costa vs. Sierra	4614,556	<0,001
Costa vs. Selva	2952,641	<0,001
Sierra vs. Selva	1231,893	<0,001

#### 4.3.2. Análisis de componentes principales

En primer lugar, es importante examinar la construcción del ISE usando el ACP dentro de cada uno de los estratos, para luego compararlos con las estimaciones del índice construido usando ACPC. En los ACP independientes por subpoblación se encuentran diferencias en las cargas en el componente entre los ámbitos geográficos urbano y rural para el grado analizado, en mayor medida para los indicadores de nivel educativo en el hogar y materiales de la vivienda (tabla 4.2). Esto hace suponer que las limitaciones del acceso a la educación, en sus distintos niveles y según el ámbito geográfico, podría estar distorsionando el cálculo del ISE. De igual manera, la presencia de ciertos materiales de construcción que pueden no estar presentes de manera uniforme a lo largo del territorio nacional, podría estar influenciando los resultados de la estimación.

**Tabla 4.2. Estadísticos del ACP del índice socioeconómico de 4.º grado de primaria - ECE 2016**

Cargas en el componente	Área		Región natural		
	Urbana	Rural	Costa	Sierra	Selva
Máximo nivel educativo	0,686	0,474	0,708	0,694	0,671
Material de la vivienda	0,768	0,560	0,775	0,763	0,700
Servicios básicos	0,683	0,618	0,699	0,717	0,736
Activos en el hogar	0,854	0,835	0,856	0,863	0,862
Otros servicios	0,821	0,727	0,827	0,781	0,796
Varianza explicada	0,59	0,43	0,60	0,59	0,57

### 4.3.3. Análisis de componentes principales comunes

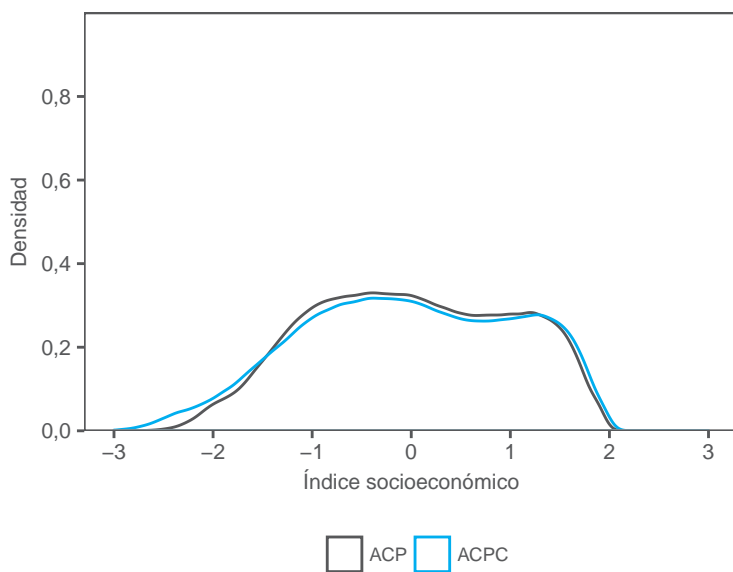
De acuerdo a los resultados expuestos, se puede concluir que la heterogeneidad presente en la sociedad peruana puede ejercer una influencia importante al momento de realizar una clasificación socioeconómica. En la tabla 4.3 se presentan las cargas en el componente y la varianza explicada de dos modelos que estiman el ISE a través del método de ACPC de manera independiente, por área y región natural.

**Tabla 4.3. Estadísticos del ACPC del índice socioeconómico de 4.º grado de primaria - ECE 2016**

Cargas en el componente	Área		Región natural		
	Urbana	Rural	Costa	Sierra	Selva
Máximo nivel educativo	0,628	0,538	0,700	0,691	0,682
Calidad de la vivienda	0,718	0,615	0,756	0,747	0,737
Servicios básicos	0,701	0,600	0,726	0,718	0,708
Otros servicios	0,831	0,711	0,811	0,802	0,791
Activos en el hogar	0,910	0,779	0,871	0,860	0,849
Varianza explicada	0,58	0,43	0,60	0,59	0,57

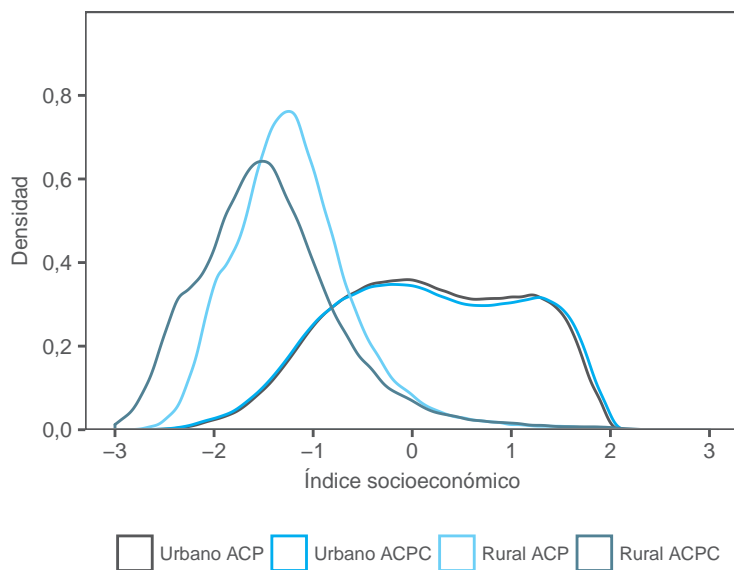
Al comparar la distribución del ISE calculado por ACP y ACPC que consideran el ámbito geográfico (figura 4.1), se observan pequeñas diferencias en valores medios y bajos del ISE. Sin embargo, ambos ISE parecen mantener la misma distribución, habiendo entre ellos una correlación de 0,998. Esto sugiere que no habría una diferencia notable en términos de valores del ISE asignados a los estudiantes y sus familias por cada método.

**Figura 4.1. Densidad del ISE calculado mediante ACP y ACPC considerando diferencias por área de la escuela - 4.º grado de primaria**



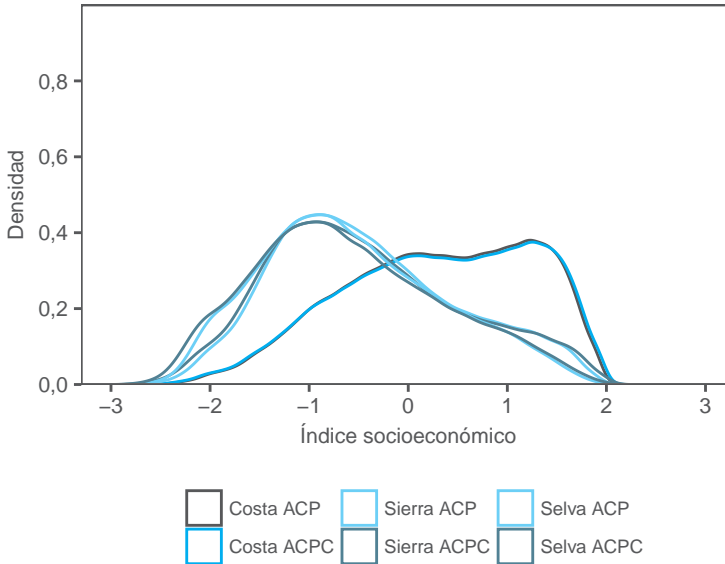
Al analizar por separado las distribuciones de los índices por ámbito geográfico, en la figura 4.2, no se encuentran diferencias para el ámbito urbano. Sin embargo, en el ámbito rural se observa una marcada diferencia entre los resultados por método de estimación. Esto se refleja en un traslado hacia la izquierda de la curva de densidad de la estimación ACPC, aunque conservando la forma de la distribución.

**Figura 4.2. Densidad según área del ISE calculado mediante ACP y ACPC considerando diferencias por área de la escuela - 4.º grado de primaria**



Para el caso de la estimación con distintas matrices de correlación por regiones naturales, no se encuentran diferencias en los resultados de ambos tipos de estimación. Asimismo, al comparar las distribuciones del ISE calculadas con ACPC y ACP, tampoco se encuentran diferencias importantes al interior de dichos estratos (figura 4.3).

**Figura 4.3. Densidad según región del ISE calculado mediante ACP y ACPC considerando diferencias por región natural - 4.º grado de primaria**



#### 4.4. Conclusiones

Si bien se encuentran pequeñas diferencias en los resultados obtenidos al estimar el índice socioeconómico a través de los métodos de ACP y ACPC, se puede afirmar que existe una varianza parcial de medida considerando ámbito geográfico, y una invarianza de medida por regiones naturales. La varianza de medida se podría estar generando por efecto de la dispersión geográfica, así como por la poca presencia de vías de comunicación que limita el acceso a bienes y servicios. De otro lado, la invarianza de medida por región natural podría estar siendo ocasionada por una posible homogeneidad respecto al acceso bienes y servicios al interior de las regiones naturales.

A pesar de no presentar diferencias marcadas en los resultados entre ambos tipos de estimación, el método de ACPC considera e incorpora la heterogeneidad contextual por tipos de subpoblación dentro del análisis. Esta aproximación metodológica podría, desde el punto de vista teórico, ser la más adecuada al

momento de construir un índice socioeconómico en un contexto tan desigual como el peruano.

En este sentido, se debe continuar investigando acerca de cómo la heterogeneidad social puede influir en la construcción de índices socioeconómicos. Deben considerarse también ítems e indicadores que puedan dar mayores luces acerca de las diferencias socioeconómicas, adicionalmente a los que ya se utilizan. Más allá de la tenencia o acceso a bienes y servicios, la calidad de estos debe ser tomada en cuenta dentro de los análisis. Asimismo, la identificación de subpoblaciones con características más acotadas podría contribuir a una mejor especificación del ISE.

De modo complementario, se recomienda analizar las diferencias contextuales a través de la construcción del ISE por ACPC, considerando estratos de ámbito rural y regiones naturales de manera combinada (costa urbana, costa rural; sierra urbana, sierra rural; selva urbana, selva rural). Estos resultados podrían mostrar configuraciones diferenciadas del ISE al considerar contextos más específicos, sobre todo en las áreas rurales. Además de ello, se podrían utilizar otras metodologías que incorporen de mejor manera a la heterogeneidad al momento de calcular un índice como el socioeconómico.



# Agregación del índice socioeconómico

Capítulo 5



### 5.1. Introducción

En diversos estudios realizados por la UMC (Ministerio de Educación, 2016a, 2017b, 2017c) se pueden encontrar modelos de regresión explicativos del rendimiento académico donde el indicador socioeconómico tiene un efecto importante en la variable criterio. En ellos también se observa que cuando el índice socioeconómico se promedia por escuela y es incluido como un factor escolar en estos modelos, el coeficiente asociado a este promedio suele presentar un mayor poder explicativo, al igual que en otros países de la región.

Sin embargo, una preocupación constante en la literatura ha sido saber si realmente existe un efecto de la composición socioeconómica escolar sobre el desempeño de los estudiantes (Balarin, 2016; Bellei, 2013; Castellano, Rabe-Hesketh & Skrondal, 2014; Ministerio de Educación, 2017c; Murillo, 2016; Nash, 2003; Raudenbush & Willms, 1995). Como detallan Harker y Tymms (2004), los procedimientos estadísticos utilizados pueden indicar la existencia de un efecto contextual, generando *efectos fantasma* que sobrestimen el verdadero efecto de la covariable en el rendimiento. Esto puede suceder por una inadecuada especificación del modelo (Harker & Tymms, 2004), o también por no observar o modelar los errores (Pokropek, 2015; Televantou et ál., 2015).

Por ello, este estudio busca analizar si la agregación del índice socioeconómico individual (ISE) puede producir una sobrestimación del coeficiente de regresión del índice socioeconómico promedio de la escuela (ISEP). Para comprobarlo, se realiza un análisis de simulación que aleatoriza estudiantes en diferentes escuelas, obteniendo nuevos indicadores promedio y aumentando la varianza del indicador individual dentro de la escuela. La técnica utilizada permite observar si el coeficiente de regresión del ISEP es un subproducto de la agregación del ISE. Si la utilización del promedio del ISE por escuela produjera una estimación de efectos contextuales en datos que estructuralmente no los tuvieran, dada la alta variabilidad socioeconómica, probaría que estos son productos de la agregación.

A continuación, se expone brevemente sobre la poca variabilidad del ISE al interior de las escuelas, primer indicio de la existencia de un ISEP representativo de la realidad escolar; y se detallan las transformaciones que se suelen seguir para asignar significado a los coeficientes estimados.

### 5.1.1. Variabilidad del índice socioeconómico al interior de las escuelas

El índice socioeconómico calculado con los datos de las ECE 2015 y 2016 muestra poca variabilidad entre estudiantes de una misma escuela. Una forma de analizar esta variabilidad es a través del cálculo de los coeficientes de correlación intraclase mostrados en el capítulo 3. Dichos análisis evidencian que una parte importante de la variabilidad del ISE se debe a diferencias entre las escuelas, a pesar de ser una característica individual.

Otra forma de analizar la variabilidad del ISE dentro de las escuelas es compararla con la variabilidad de este indicador a nivel nacional. La desviación estándar (DE) nacional del ISE es cercana a 1 en todos los grados y años; sin embargo, las DE dentro de las escuelas suelen ser menores. En la tabla 5.1 se exponen los resultados de calcular la DE del ISE dentro de cada escuela y los percentiles de este estadístico por escuela. Como se observa, la DE de la mitad de las escuelas varía entre 0,260 y 0,303, menor a la DE nacional. Además, en todos los casos el percentil 95 de escuelas tiene una variabilidad de dos tercios o menos de la DE nacional.

**Tabla 5.1. Desviaciones estándar del ISE al interior de la escuela según percentiles**

Grado y año	DE nacional	DE de las escuelas				
		P25	P50	P75	P90	P95
2.º grado de primaria 2015	1,018	0,143	0,303	0,453	0,577	0,675
2.º grado de primaria 2016	0,915	0,129	0,260	0,403	0,513	0,596
4.º grado de primaria 2016	0,999	0,140	0,271	0,419	0,553	0,651
2.º grado de secundaria 2015	0,988	0,180	0,289	0,417	0,539	0,631
2.º grado de secundaria 2016	1,000	0,176	0,285	0,414	0,542	0,638

Este indicio reflejaría que los estudiantes de una misma escuela tienen valores de ISE con mucha menor variabilidad que la existente a nivel nacional y que, por lo tanto, el ISEP podría representar características socioeconómicas comunes de los estudiantes dentro de una misma escuela. Por ello, en la mayoría de investigaciones realizadas por la UMC (Ministerio de Educación, 2016a, 2017b, 2017c) se incluye el ISEP, o una medida similar, como covariable predictiva del rendimiento individual.

### 5.1.2. Centrado de variables predictivas

En el contexto de modelos explicativos del rendimiento académico, muchos de los predictores utilizados son variables con escalas intervalares donde el 0 es asignado de manera arbitraria y, por lo tanto, no indica la ausencia de lo que se mide (Véliz, 2016). Es por ello que muchas veces se hace necesario transformar las variables para que los parámetros hallados sean fácilmente interpretables (Castellano et ál., 2014; Enders & Tofighi, 2007; Raudenbush & Willms, 1995).

Una de las transformaciones más utilizadas y que permite asignar 0 a la media de las variables es el centrado. En el caso del análisis de regresión multinivel, se trabaja usualmente con dos tipos de centrado: a la gran media y a la media del clúster (Enders & Tofighi, 2007). El primero consiste en restar al valor de la variable utilizada la media aritmética de todos los individuos; de esta manera, el coeficiente de regresión representará el efecto que tiene el aumento en una unidad de la variable sobre la media del conjunto de datos. El segundo tipo de centrado, a la media del clúster, consiste en restar a la variable individual la media aritmética del grupo. Esto permite calcular el efecto de la covariable centrada sobre las diferencias entre los individuos y la composición del grupo al pertenece (Enders & Tofighi, 2007; Hofmann & Gavin, 1998; Paccagnella, 2006).

En la presente investigación, el centrado es de vital importancia para la interpretación de los resultados, pues lo que se pretende determinar es si el ISEP tiene un efecto propio sobre el rendimiento de los estudiantes. Es decir, si es que, independientemente del efecto del ISE de los estudiantes, las características socioeconómicas promedio de la escuela explican el resultado individual (Diez Roux, 2017). Esto se captura técnicamente mediante la inclusión de ambos predictores (ISE e ISEP) en el modelo de regresión y eligiendo algún tipo de centrado para aislar o aglomerar ambos efectos en uno o dos coeficientes.

En el caso de variables derivadas<sup>17</sup>, la inclusión a nivel de estudiante capturará en todos los casos los efectos de estas sobre la variable criterio. Sin embargo, cuando la covariable individual es centrada a la media del clúster, el coeficiente asociado a la variable agrupada de segundo nivel recoge el efecto total de ambas variables (Castellano et ál., 2014; Enders & Tofighi, 2007). Por lo tanto, agrupa tanto el efecto individual como el efecto contextual (Castellano et ál., 2014; C. Duncan, Jones & Moon, 1998; Feaster, Brincks, Robbins & Szapocznik, 2011). Esto se demuestra a continuación, considerando un modelo multinivel con efectos contextuales de una covariable endógena:

$$y_{ij} = \alpha + \beta x_{ij} + \gamma \bar{x}_{.j} + u_j + \varepsilon_{ij}. \quad (5.9)$$

<sup>17</sup>Variable grupal matemáticamente construida por características individuales (Diez Roux, 2017).

El modelo en (5.9) puede reparametrizarse considerando el centrado a la media del clúster. En este caso se utilizan los parámetros  $\delta$  para el efecto de la diferencia entre la variable y su media, y  $\eta$  para el efecto de la media grupal.

$$y_{ij} = \alpha + \delta (x_{ij} - \bar{x}_{.j}) + \eta \bar{x}_{.j} + u_j + \varepsilon_{ij}. \quad (5.10)$$

La ecuación (5.10) se puede reordenar para que tenga la misma estructura que (5.9):

$$y_{ij} = \alpha + \delta x_{ij} + (\eta - \delta) \bar{x}_{.j} + u_j + \varepsilon_{ij}. \quad (5.11)$$

Si se comparan ambas ecuaciones se puede ver que  $\delta = \beta$ , y que el efecto de la media de la variable es  $(\eta - \beta)$ :

$$y_{ij} = \alpha + \beta x_{ij} + (\eta - \beta) \bar{x}_{.j} + u_j + \varepsilon_{ij}. \quad (5.12)$$

Finalmente, es fácil obtener que el efecto  $\eta$ , calculado en la ecuación (5.10), es la suma de los efectos de la primera ecuación, de tal manera que:  $\eta = \beta + \gamma$ . Por ello, cuando se incluye el centrado a la media del clúster, el coeficiente de regresión de la variable agrupada recoge el efecto total. Por lo tanto, si el efecto individual tiene la misma dirección que el efecto grupal, el coeficiente  $\eta$  siempre será mayor que  $\delta$  y será necesario restar  $\beta$  para obtener el efecto contextual.

En resumen, cuando se utilice solamente el centrado a la gran media, los coeficientes estimados recogerán el aporte de la variable individual o derivada sobre el rendimiento. A su vez, permitirán conocer cuál es el efecto de estas variables conforme se aumente en una unidad sobre el promedio de la población. Por otro lado, la inclusión del centrado a la media del clúster agrupará ambos efectos en el coeficiente asociado a la variable derivada. Las distintas combinaciones de centrado han sido utilizadas en esta investigación para ilustrar sus efectos y observar su comportamiento en los datos aleatorizados.

## 5.2. Metodología

Este estudio utilizó la información socioeconómica y de rendimiento recogida en las ECE 2015 y 2016. Las variables incluidas en el análisis fueron el índice socioeconómico del estudiante, el índice socioeconómico promedio de la escuela y los resultados de la ECE para Lectura, Matemática e Historia, Geografía y Economía (HGE), según el grado y el año. El estudio consta de dos partes; en la primera se aleatorizan datos reales y seguidamente se realiza un análisis de simulación de efectos contextuales.

### 5.2.1. Aleatorización de datos reales

Para la primera parte se extrajeron 1000 muestras aleatorias de escuelas que reunieran aproximadamente el 10% del total de estudiantes evaluados. Por este motivo, la cantidad de escuelas seleccionadas para cada análisis puede variar, dependiendo de la cantidad de estudiantes que fueron evaluados en cada una de ellas.

Dado que el objetivo era determinar la posible sobrestimación del ISEP, cada muestra fue separada en dos conjuntos de datos. En el primero, se conservaron los datos reales de los estudiantes: rendimiento, ISE e ISEP; en el segundo, se aleatorizó a los estudiantes en cualquier escuela de la muestra. Así, en vez de que se conservara la información de la escuela a la que pertenecían, se les asignó una escuela aleatoria de la muestra. Para estos datos se calculó un nuevo ISEP con la media aritmética de los estudiantes asignados a cada escuela. Esto permite comparar la estimación de efectos contextuales con datos de baja y de alta variabilidad socioeconómica.

Con cada conjunto de datos se estimaron once modelos de regresión multinivel utilizando como variable criterio el rendimiento. Como covariables se incluyeron el ISE y el ISEP con distintas combinaciones de centrado (ver tabla 5.2). Para el caso del ISE, se probó el centrado a la gran media<sup>18</sup> (GM) y el centrado a la media del clúster<sup>19</sup> (MC). En cuanto al ISEP, dado que no es una variable individual, solo se centró a la gran media.

**Tabla 5.2. Especificación de modelos estimados**

Modelo	Fórmula
1: ISE	$y_{ij} = \beta_{0j} + \beta_1(ISE_{ij}) + u_j + e_{ij}$
2: ISE(GM)	$y_{ij} = \beta_{0j} + \beta_1(ISE_{ij} - \overline{ISE}) + u_j + e_{ij}$
3: ISE(MC)	$y_{ij} = \beta_{0j} + \beta_1(ISE_{ij} - \overline{ISEP}_j) + u_j + e_{ij}$
5: ISEP	$y_{ij} = \beta_{0j} + \beta_2(ISEP_j) + u_j + e_{ij}$
4: ISEP(GM)	$y_{ij} = \beta_{0j} + \beta_2(ISEP_j - \overline{ISEP}) + u_j + e_{ij}$
6: ISE + ISEP	$y_{ij} = \beta_{0j} + \beta_1(ISE_{ij}) + \beta_2(ISEP) + u_j + e_{ij}$
7: ISE + ISEP(GM)	$y_{ij} = \beta_{0j} + \beta_1(ISE_{ij}) + \beta_2(ISEP_j - \overline{ISEP}) + u_j + e_{ij}$
8: ISE(GM) + ISEP	$y_{ij} = \beta_{0j} + \beta_1(ISE_{ij} - \overline{ISE}) + \beta_2(ISEP_j) + u_j + e_{ij}$
9: ISE(GM) + ISEP(GM)	$y_{ij} = \beta_{0j} + \beta_1(ISE_{ij} - \overline{ISE}) + \beta_2(ISEP_j - \overline{ISEP}) + u_j + e_{ij}$
10: ISE(MC) + ISEP	$y_{ij} = \beta_{0j} + \beta_1(ISE_{ij} - \overline{ISEP}_j) + \beta_2(ISEP_j) + u_j + e_{ij}$
11: ISE(MC) + ISEP(GM)	$y_{ij} = \beta_{0j} + \beta_1(ISE_{ij} - \overline{ISEP}_j) + \beta_2(ISEP_j - \overline{ISEP}) + u_j + e_{ij}$

$y_{ij}$  = Rendimiento en Lectura, Matemática o HGE

Para cada grado y área se obtuvieron 22 modelos de regresión multinivel, considerando los datos reales y los aleatorizados. Los resultados agregados de las 1000 muestras obtenidas se detallan en la sección 5.3.1.

<sup>18</sup>Para este tipo de centrado se le resta a la variable el promedio de toda la población.

<sup>19</sup>En este tipo de centrado se le resta a la variable el promedio del grupo.

## 5.2.2. Estudio de simulación de efectos contextuales

Para complementar la aleatorización de datos reales, se realizó un estudio de simulación que pudiera examinar la correcta recuperación de parámetros mediante un análisis de regresión multinivel. Con este objetivo se construyó una variable criterio que considerara en su estructura efectos individuales ( $y_1$ ), así como otra que solo considerara efectos individuales y contextuales ( $y_2$ ). Se esperaba que en el caso de  $y_2$  se recuperasen correctamente los parámetros asignados. Por otro lado, para  $y_1$  se esperaba encontrar un efecto nulo asociado a la variable agregada, lo que confirmaría la estimación insesgada del parámetro de nivel 2 en datos sin efectos contextuales.

Este análisis consideró una estructura de datos anidados con diferentes tamaños y número de clústeres. Para cada caso, dentro de cada conglomerado, se generó la variable  $x$  con distribución normal cuya media tuviera  $N(0; 1)$  y con desviación estándar con distribución  $N(\frac{1}{2}; \frac{3}{20})$  entre los conglomerados<sup>20</sup>. Con  $x$  se construyeron las dos variables criterio, con y sin efectos contextuales. La variable  $y_1$  se construyó de la siguiente forma:

$$y_{1ij} = \beta_{0j} + \beta_1 x_{ij} + u_j + e_{ij} \quad (5.13)$$

Y para  $y_2$ :

$$y_{2ij} = \beta_{0j} + \beta_1 x_{ij} + \beta_2 \bar{x}_{.j} + u_j + e_{ij} \quad (5.14)$$

En ambos casos se consideraron efectos aleatorios para el intercepto con  $\beta_{0j} \sim U(1; 2)$ , además de efectos fijos para las variables  $x$  y  $\bar{x}$ ,  $\beta_1 = \beta_2 = 1$ . Finalmente, se incluyeron errores con  $u_j \sim N(0; 1)$  y  $e_{ij} \sim N(0; 1)$ .

A continuación se realizaron modelos de regresión multinivel para explicar las variables  $y_1$  y  $y_2$ , utilizando en ambos casos  $x$  y  $\bar{x}$  como variables predictoras. Como se señaló anteriormente, se esperaba encontrar un coeficiente de regresión nulo asociado a la variable agregada  $\bar{x}$  para  $y_1$ . Este proceso se realizó con 1000 réplicas, cuyos resultado se detallan en la sección 5.3.2<sup>21</sup>.

## 5.3. Resultados

### 5.3.1. Efectos contextuales en aleatorización de datos reales

En la tabla 5.3 se muestra el promedio de los coeficientes de regresión de las 1000 muestras analizadas para Lectura y Matemática en 2.º grado de secundaria en la ECE 2016<sup>22</sup>. Se detallan las estimaciones para los datos reales (que considera la verdadera

<sup>20</sup>Se utilizaron estas distribuciones y parámetros por su semejanza con los datos reales del ISE.

<sup>21</sup>Para el código del lenguaje R utilizado ver el Anexo D2.

<sup>22</sup>Los resultados de las demás áreas y grados se muestran en el Anexo D1.

escuela del estudiante) y los datos simulados (que aleatoriza a los estudiantes en distintas escuelas).

**Tabla 5.3. Modelos con datos aleatorizados de 2.º grado de secundaria - ECE 2016**

Modelo	Datos	Lectura			Matemática		
		Intercepto	ISE	ISEP	Intercepto	ISE	ISEP
Modelo 1:	Aleatorizados	570,61	31,13		561,87	26,99	
ISE	Reales	561,99	14,51		553,70	12,81	
Modelo 2:	Aleatorizados	570,85	31,13		562,08	26,99	
ISE(GM)	Reales	562,10	14,51		553,79	12,81	
Modelo 3:	Aleatorizados	570,85	31,14		562,08	26,99	
ISE(MC)	Reales	557,63	11,99		549,96	10,61	
Modelo 4:	Aleatorizados	570,61		31,12	561,87		26,91
ISEP	Reales	569,39		41,79	559,75		35,29
Modelo 5:	Aleatorizados	570,85		31,12	562,08		26,91
ISEP(GM)	Reales	569,71		41,79	560,02		35,29
Modelo 6:	Aleatorizados	570,61	31,14	-0,02	561,87	26,99	-0,08
ISE+ISEP	Reales	569,38	11,99	29,79	559,74	10,61	24,67
Modelo 7:	Aleatorizados	570,61	31,14	-0,02	561,87	26,99	-0,08
ISE+ISEP(GM)	Reales	569,62	11,99	29,79	559,94	10,61	24,67
Modelo 8:	Aleatorizados	570,85	31,14	-0,02	562,08	26,99	-0,08
ISE(GM)+ISEP	Reales	569,47	11,99	29,79	559,82	10,61	24,67
Modelo 9:	Aleatorizados	570,85	31,14	-0,02	562,08	26,99	-0,08
ISE(GM)+ISEP(GM)	Reales	569,70	11,99	29,79	560,01	10,61	24,67
Modelo 10:	Aleatorizados	570,61	31,14	31,12	561,87	26,99	26,91
ISE(MC)+ISEP	Reales	569,38	11,99	41,77	559,74	10,61	35,28
Modelo 11:	Aleatorizados	570,85	31,14	31,12	562,08	26,99	26,91
ISE(MC)+ISEP(GM)	Reales	569,70	11,99	41,77	560,01	10,61	35,28

En cuanto al coeficiente de regresión del ISE, se puede observar que es menor en los modelos con datos reales que en aquellos con datos aleatorizados. En el primer caso, varía entre 11,99 y 14,51 en Lectura y 10,61 y 12,81 en Matemática. Esto representa alrededor de 18,5% y 14,7%, respectivamente de las desviaciones estándar de Lectura y Matemática<sup>23</sup>. Con respecto a los modelos de datos aleatorizados, este valor se incrementa hasta 44,3% (31,14 puntos en la prueba) para Lectura y 33,9% (26,99 puntos en la prueba) para Matemática.

Por otro lado, en el caso del índice socioeconómico promedio, se observa una disminución importante del coeficiente de regresión en los datos aleatorizados. Por ejemplo, para el caso de Lectura, el coeficiente del ISEP con datos reales fluctúa entre 29,79 y 41,79 (alrededor de 50,9% de una DE), reduciéndose a 31,12 para los modelos que solo consideran el ISEP o que centran el ISE a la media del clúster y volviéndose nulo en los modelos que utilizan otro tipo de centrado. Estos resultados son consistentes con lo encontrado en los demás grados y áreas (ver Anexo D1). Asimismo, en todos los casos se observa un aumento importante del coeficiente de regresión del ISE y la reducción o anulación del coeficiente del ISEP.

<sup>23</sup>La desviación estándar para Lectura y Matemática en la ECE 2016 de 2.º grado de secundaria fue 70,22 y 79,71, respectivamente (Ministerio de Educación, 2017d).

### 5.3.2. Simulación de efectos contextuales

Como se puede apreciar en las tablas 5.4 y 5.5, la regresión recupera de forma insesgada los parámetros  $\beta_0$  y  $\beta_1$ . Esto sucede en ambos escenarios; en el caso de  $y_1$  (sin efectos contextuales) y en el caso de  $y_2$  (con efectos contextuales), para todas las combinaciones de tamaños de clúster y número de clústeres  $\hat{\beta}_0 \approx 1,5$  (el valor esperado de la distribución uniforme asignada) y  $\hat{\beta}_1 \approx 1$  (el valor con el que se construyeron las variables criterio).

**Tabla 5.4. Estimación de  $\beta_0$  - medias de coeficientes de regresión**

Total de clústeres	Sin efectos contextuales					Con efectos contextuales				
	Tamaño del clúster									
	5	10	20	40	80	5	10	20	40	80
20	1,519	1,505	1,496	1,498	1,503	1,519	1,505	1,496	1,498	1,503
40	1,502	1,507	1,502	1,503	1,491	1,502	1,507	1,502	1,503	1,491
100	1,495	1,503	1,496	1,504	1,499	1,495	1,503	1,496	1,504	1,499
1000	1,500	1,500	1,501	1,501	1,500	1,500	1,500	1,501	1,501	1,500

**Tabla 5.5. Estimación de  $\beta_1$  - medias de coeficientes de regresión**

Total de clústeres	Sin efectos contextuales					Con efectos contextuales				
	Tamaño del clúster									
	5	10	20	40	80	5	10	20	40	80
20	1,003	1,001	1,000	1,002	1,001	1,003	1,001	1,000	1,002	1,001
40	0,993	0,998	1,003	1,002	0,998	0,993	0,998	1,003	1,002	0,998
100	0,998	0,996	1,000	1,000	1,000	0,998	0,996	1,000	1,000	1,000
1000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

Por otro lado, el coeficiente de  $\beta_2$  también se estima insesgadamente en ambos escenarios y según las diferentes combinaciones de tamaños de clúster y de población. Como se muestra en la tabla 5.6, para la variable construida con efectos contextuales, la regresión recupera correctamente el parámetro. Además, para  $y_1$ ,  $\hat{\beta}_2 \approx 0$ , es decir, en el modelo sin efectos contextuales, no se aprecia una sobrestimación del verdadero coeficiente de regresión.

**Tabla 5.6. Estimación de  $\beta_2$  - medias de coeficientes de regresión**

Total de clústeres	Sin efectos contextuales					Con efectos contextuales				
	Tamaño del clúster									
	5	10	20	40	80	5	10	20	40	80
20	-0,007	-0,007	-0,004	0,001	-0,010	0,993	0,993	0,996	1,001	0,990
40	0,002	-0,001	-0,002	0,003	0,009	1,002	0,999	0,998	1,003	1,009
100	0,000	-0,005	0,002	0,003	-0,003	1,000	0,995	1,002	1,003	0,997
1000	0,001	0,000	-0,002	0,001	-0,001	1,001	1,000	0,998	1,001	0,999



#### 5.4. Conclusiones

Como se puede apreciar en la sección previa, el coeficiente de regresión relacionado al ISEP desaparece en la aleatorización de datos reales. Solo es diferente a 0 en los modelos que incluyen el centrado a la media del clúster para el ISE (modelos 10 y 11). Sin embargo, el coeficiente estimado del ISEP en estos modelos es la suma de los coeficientes individuales y grupales sin MC (ver sección 5.1.2). Se pueden calcular los efectos contextuales en estos casos, resultando ser idénticos a los encontrados en los modelos 8 y 9, para Lectura ( $31,12 - 31,14 = -0,02$ ) y para Matemática ( $26,91 - 26,99 = -0,08$ ). Por lo tanto, en todos los casos de datos aleatorizados, el coeficiente de regresión asociado al ISEP es nulo, lo que sería un primer indicio de que el coeficiente encontrado con los datos reales no es una sobrestimación producto de la ponderación de la variable individual.

Sin embargo, debido al proceso de aleatorización, la variabilidad del ISEP entre escuelas puede haberse reducido, ocasionando que sean muy parecidos entre sí. Esto reduciría la varianza entre los grupos y podría ocasionar un coeficiente de regresión nulo en el segundo nivel. Para resolver este problema, el estudio de simulación conserva la estructura original de los datos, baja variabilidad al interior de los clústeres y alta variabilidad entre ellos. Los resultados apuntan en la misma dirección que la aleatorización: solo se encuentra un coeficiente de nivel 2 cuando los datos poseen una estructura con efectos contextuales.

Estos dos análisis aportan evidencias de que el procedimiento utilizado en estudios anteriores estaría recogiendo insesgadamente el efecto del ISEP (Ministerio de Educación, 2016a, 2017b, 2017c). Sin embargo, tal como muestran las estimaciones de los modelos 10 y 11, es importante tomar en cuenta el centrado de las variables para establecer correctas interpretaciones de los resultados. De otro lado, sería importante considerar la aleatorización de datos reales en diferentes cuantiles de rendimiento, con el fin de observar si se mantienen los resultados encontrados. Además, para aislar de forma efectiva el efecto del ISE y del ISEP, los modelos deben especificar otras variables que puedan explicar el rendimiento de los estudiantes. Estos modelos son explorados en el siguiente capítulo.

# Variación del efecto del índice socioeconómico entre escuelas

Capítulo 6

### 6.1. Introducción

La investigación en el área de educación y las ciencias sociales ha documentado la importancia del efecto que tiene el estatus socioeconómico (ESE) sobre el rendimiento académico (León & Collahua, 2016; Sirin, 2005). Por ejemplo, la evidencia empírica ha mostrado que los estudiantes que tienen padres con un bajo ingreso económico, menores niveles educativos o están en situación de desempleo, tienen mayor probabilidad de obtener un menor desarrollo cognitivo en comparación con estudiantes que tienen padres con situaciones más acomodadas (Hertzman, 1994; Hertzman & Weins, 1996). Asimismo, es menos probable que estos estudiantes se comprometan activamente en actividades escolares y extracurriculares (Datcher, 1982; Finn & Rock, 1997; Johnson, Crosnoe & Elder, 2001; Voelkl, 1995). Esto podría ocasionar que los estudiantes no logren construir los aprendizajes adecuados para su edad, aumentando la probabilidad de deserción escolar (Cairns, Cairns & Neckerman, 1989; Crane, 1991; Ensminger & Slusarcick, 1992; Janosz, LeBlanc, Boulerice & Tremblay, 1997; Rumberger, 1995). Consecuentemente, también podría afectar el acceso de los estudiantes a la educación superior y limitaría sus oportunidades laborales (Raudenbush & Kasim, 1998).

Sin embargo, estas condiciones adversas no necesariamente tienden a afectar el rendimiento escolar de la misma manera en todos los contextos. Existen estudiantes resilientes, que a pesar de un contexto desfavorable y de tener un bajo estatus socioeconómico, logran desarrollar aprendizajes adecuados para su edad. Del mismo modo, el efecto del ESE también puede no ser constante en todas las escuelas. Por ejemplo, otros factores como la infraestructura escolar, la gestión de la escuela o las prácticas pedagógicas, además de relacionarse con el rendimiento académico (Ministerio de Educación, 2017a, 2017b), podrían atenuar o acentuar el efecto del ESE.

Tomando en cuenta todo lo mencionado anteriormente, en un contexto como el peruano donde se manifiestan diferentes realidades, resulta importante explorar si es que el efecto del ESE en el rendimiento escolar varía dependiendo de algunas características de las escuelas. Dado esto, el presente estudio plantea los siguientes objetivos: 1) analizar cuánto varía el efecto del estatus socioeconómico en el

rendimiento académico por escuela y dentro de ella<sup>24</sup>; 2) observar cuánto varía el efecto del estatus socioeconómico en el rendimiento académico en función a las características del estudiante o escuela; y 3) observar cuánto varía el efecto de la composición socioeconómica de la escuela en el rendimiento, en función a las características del estudiante o de la escuela.

## 6.2. Metodología

### 6.2.1. Población y muestra

La población objetivo de este estudio está conformada por estudiantes de 2.º grado de secundaria y que rindieron la prueba de Lectura de la ECE 2016<sup>25</sup>. Dieron esta evaluación 500 098 estudiantes distribuidos en 12 875 escuelas. El 75,4 % de estos estudiantes asistían a escuelas estatales y el 89,1 % a escuelas de zona urbana. Además, el 10,7 % tiene como lengua materna una lengua originaria. El modelo final fue calculado con una muestra de 167 749 estudiantes de 3025 escuelas<sup>26</sup>.

### 6.2.2. Medición

En este estudio se consideran diferentes variables relacionadas a los estudiantes y a las escuelas a las que asisten. Las variables explicativas utilizadas en cada uno de estos niveles y su descripción se muestran en las tablas 6.1 y 6.2. Se utiliza como variable criterio el rendimiento de los estudiantes de 2.º grado de secundaria en Lectura en la ECE 2016.

<sup>24</sup> Este modelo también es llamado como modelo de efectos aleatorios donde las pendientes de la relación entre el ISE y el rendimiento varían según la escuela (Snijders & Bosker, 1999).

<sup>25</sup> Se estima encontrar resultados similares con los resultados de Matemática.

<sup>26</sup> Esta muestra se adecua al diseño muestral de la ECE 2016 (Ministerio de Educación, 2017d). A pesar de la reducción de casos, el diseño muestral no se ha visto afectado de manera significativa.

**Tabla 6.1. Características del estudiante**

Variable	Descripción
Rendimiento en Lectura	Puntaje obtenido en la prueba de Lectura de la ECE 2016 de 2.º grado de secundaria (Ministerio de Educación, 2017d).
Sexo	-1: hombre +1: mujer
Lengua materna	-1: lengua originaria +1: castellano
Repetición	-1: no ha repetido de grado +1: sí ha repetido de grado
Asistencia a educación inicial	-1: no ha asistido a educación inicial +1: sí ha asistido a educación inicial
Índice socioeconómico (ISE)	Puntaje que representa el estatus socioeconómico de la familia.
No violencia en la escuela	Puntaje factorial que mide la ausencia de conductas agresivas o violentas contra los estudiantes en la escuela (ej. agresión física, insultos, amenazas, etc.) (ver Anexo E1).
Exposición a diferentes tipos de pregunta sobre comprensión de textos	Puntaje factorial que mide la frecuencia con la que los estudiantes resuelven diferentes tipos de preguntas en las clases de Comunicación (ej. para encontrar información en el texto, deductivas, de opinión, etc.) (ver Anexo E1).
Expectativas de máximo nivel educativo	No terminará la secundaria Terminará la secundaria Terminará una carrera técnica Terminará una carrera universitaria Terminará una maestría o doctorado

**Tabla 6.2. Características de la escuela**

Variable	Descripción
Gestión	-1: escuela estatal +1: escuela no estatal
Tamaño de la escuela	Escuela grande (más de 574 evaluados) Escuela mediana (entre 150 y 574 evaluados) Escuela pequeña (menos de 150 evaluados)
Índice socioeconómico promedio de la escuela (ISEP)	Promedio del ISE de los estudiantes de la escuela
Índice de calidad del servicio y el contexto de la escuela (ICSC)	Puntaje que busca aproximarse, sobre la base de características físicas de la institución educativa y características de su entorno, a la calidad de las condiciones en las cuales la escuela opera y su correspondiente potencial para brindar un servicio educativo de calidad orientado a la mejora de los aprendizajes (Ministerio de Educación, 2016c).
Percepción sobre el conocimiento pedagógico de los docentes de Comunicación	Puntaje factorial que refleja la percepción del director sobre los conocimientos y estrategias de los docentes acerca de organización de la información, comprensión lectora y producción de textos (ver Anexo E1).
Percepción sobre la existencia de problemas que afectan el ambiente escolar	Puntaje factorial que recoge la percepción del director sobre problemas que se encuentran en la escuela (ej. indisciplina de los estudiantes, inasistencia de docentes, intimidación entre estudiantes, etc.) (ver Anexo E1)
Percepción sobre la existencia de problemas familiares de los estudiantes	Puntaje factorial que refleja la percepción del director sobre situaciones familiares desfavorables que podrían afectar el desempeño escolar de los estudiantes (ej. escasos recursos económicos, padres con bajo nivel educativo, escaso involucramiento familiar en la escuela, etc.) (ver Anexo E1).

### 6.2.3. Estrategia analítica

Para cumplir con los objetivos previamente mencionados, se estimaron diferentes modelos de regresión multinivel, siendo la variable criterio el rendimiento en Lectura en la ECE 2016. Estos modelos consideraron el ISE y el ISEP como principales variables predictoras<sup>27</sup>, así como otras variables de control y la interacción de estas con tales indicadores socioeconómicos. El modelamiento del efecto del estatus socioeconómico en el rendimiento en Lectura se realizó mediante los siguientes pasos:

- I Se estimó un modelo nulo de rendimiento y se utilizó el coeficiente de correlación intraclase (CCI) para observar la distribución de la varianza entre niveles. En el modelo nulo se descompuso la varianza del rendimiento en Lectura entre dos partes: varianza a nivel del estudiante y varianza a nivel de la escuela. Sobre la base de este modelo nulo, se especificó primero el modelo a nivel del estudiante y luego el modelo a nivel de la escuela.
- II Dado que el objetivo fue analizar el comportamiento del estatus socioeconómico, el ISE fue la primera variable que se introdujo en el modelo.
- III Se especificó un modelo que considerase la mayor cantidad de variables posibles sin incluir efectos de segundo orden, como interacciones o efectos no lineales (cuadrático), siguiendo para ello las recomendaciones de Hox (2010). Se utilizó un procedimiento por pasos hacia adelante que parte desde el modelo nulo, luego se especificó el modelo para el nivel del estudiante y, finalmente, el modelo para el nivel de la escuela. Para especificar el modelo a nivel del estudiante se incluyeron las variables de este nivel una por una, en el siguiente orden: ISE, sexo, lengua materna, repetición, asistencia a educación inicial, no violencia en la escuela, exposición a diferentes tipos de pregunta sobre comprensión de textos y expectativas de máximo nivel educativo. Luego de ingresar todos los predictores de este nivel, se realizó el mismo procedimiento con las variables del nivel de escuela, incluyendo en orden las siguientes variables: ISEP, gestión, ICSC, percepción del director sobre el conocimiento pedagógico del docente, percepción sobre el conocimiento pedagógico de los docentes de Comunicación, percepción sobre la existencia de problemas que afectan el ambiente escolar, percepción de problemas familiares de los estudiantes, y tamaño de la escuela.
- IV Se calcularon los efectos fijos y aleatorios del ISE. Finalmente, se calcularon modelos que especifican interacciones entre el ISE o el ISEP con otras variables.

<sup>27</sup> Ambas variables fueron centradas a la gran media.

### 6.3. Resultados

El CCI del modelo nulo estimado permite obtener la proporción de varianza del nivel 2 (escuela) del rendimiento en Lectura. En este caso, el 41 % de la varianza del rendimiento se encuentra entre las escuelas. Es decir, los estudiantes dentro de una misma escuela tendrían un rendimiento académico homogéneo. Al mismo tiempo, diferencias en éste podrían ser más heterogéneas entre escuelas.

La tabla 6.3 reporta la comparación de bondades de ajuste entre los modelos analizados para determinar cuál de ellos representa mejor los datos. Los resultados muestran que el modelo 11 es el mejor, pues cuenta con valores inferiores de  $\chi^2$  cuadrado, AIC y BIC.

**Tabla 6.3. Comparación de modelos de rendimiento en Lectura**

Modelo	$-2 \times \mathcal{L}$	Modelos comparados	DIF Modelos $-2 \times \mathcal{L}$	AIC	BIC
Modelo nulo	5 482 741,28	-	-	5 482 741,28	5 482 741,28
Modelo 1	5 324 233,64	M0 - M1	-158 507,64	5 324 235,64	5 324 243,10
Modelo 2	2 035 165,61	M1 - M2	-3 289 068,03	2 035 187,61	2 035 255,28
Modelo 3	2 032 656,79	M2 - M3	-2508,82	2 032 680,79	2 032 754,61
Modelo 4	1 811 964,71	M3 - M4	-220 692,07	1 811 996,71	1 812 092,95
Modelo 5	1 811 941,94	M4 - M5	-22,77	1 811 979,94	1 812 094,22
Modelo 6	1 811 751,50	M5 - M6	-190,44	1 811 789,50	1 811 903,78
Modelo 7	1 810 097,55	M6 - M7	-1653,95	1 810 135,55	1 810 249,83
Modelo 8	1 810 048,00	M7 - M8	-49,55	1 810 086,00	1 810 200,28
Modelo 9	1 809 961,78	M8 - M9	-86,23	1 810 001,78	1 810 122,07
Modelo 10	1 809 919,32	M9 - M10	-42,45	1 809 963,32	1 810 095,65
Modelo 11	1 809 911,51	M10 - M11	-7,82	1 809 959,51	1 810 103,86

La tabla 6.4 muestra los parámetros estimados para el modelo 11. Los resultados muestran que existe un efecto estadísticamente significativo del ISE sobre el rendimiento. Por cada desviación estándar del ISE, los puntajes del rendimiento en Lectura tienden a incrementar en 6 puntos controlando por todas las demás variables. Asimismo, el ISEP también tiene una relación estadísticamente significativa con el rendimiento. Por cada punto extra en el ISEP, el puntaje del rendimiento en la prueba de Lectura se incrementaría en 30,58 puntos. De este modo, el ISE y el ISEP se asocian a un mejor rendimiento de los estudiantes independientemente de otras características individuales y escolares.

**Tabla 6.4. Modelo final de rendimiento en Lectura**

	Variable	B	$\beta$	Significancia
<b>Nivel 1</b>	ISE	6,016	0,086	<,001
	Mujer	-1,789	-0,013	<,001
	Castellano	23,500	0,104	<,001
	Repetición	-16,650	-0,095	<,001
	Asistencia a educación inicial	7,597	0,031	<,001
	No violencia en la escuela	2,399	0,020	<,001
	Exposición a tipos de pregunta	8,877	0,078	<,001
	Terminará la secundaria	-1,983	-0,008	,31
	Terminará una carrera técnica	17,039	0,077	<,05
	Terminará una carrera universitaria	25,581	0,179	<,05
Terminará una maestría o doctorado	45,447	0,318	<,05	
<b>Nivel 2</b>	ISEP	30,581	0,357	<,001
	IE estatal	18,958	0,117	<,001
	ICSC	5,082	0,102	<,001
	Conocimiento pedagógico del docente	1,831	0,019	<,01
	Problemas en el ambiente escolar	-1,955	-0,014	<,05
	Problemas familiares	-3,410	-0,029	<,01
	IE mediana	-4,252	-0,026	<,01
	IE pequeña	0,453	0,001	,91
	ISEP $\times$ IE Estatal	-9,519	-0,085	<,001
	ISEP $\times$ problemas familiares	-4,184	-0,028	<,001
	ISEP $\times$ IE mediana	8,309	0,029	<,001
ISEP $\times$ IE pequeña	0,000	0,000	,99	
ISE $\times$ ICSC	1,283	0,031	<,001	
<b>Intercepto</b>		498,846	-	<,001

Para cumplir con el primer objetivo del estudio se calculó un efecto aleatorio del ISE en el rendimiento. En este caso, se analizó la relación del ISE con el rendimiento entre las escuelas y entre los estudiantes dentro de la escuela. Los hallazgos evidencian que el efecto del ISE en el rendimiento varía entre diferentes escuelas.

Asimismo, se identifican múltiples interacciones estadísticamente significativas entre el ISEP y las variables de gestión, problemas familiares y tamaño de la escuela. Dicho efecto es mayor en escuelas no estatales ( $0,357 + 0,085 = 0,442$ ) que en escuelas estatales ( $0,357 - 0,085 = 0,272$ ). El efecto del ISEP en el rendimiento también varía según el tamaño de la escuela. En escuelas grandes el efecto del ISEP es menor ( $0,357 - 0,029 = 0,328$ ) que en escuelas de tamaño mediano ( $0,357 + 0,029 = 0,386$ ). Además, el efecto del ISEP podría ser menor en escuelas con estudiantes que tienen mayores problemas familiares ( $0,357 - 0,028 = 0,329$ ) que en aquellos donde estos problemas se presentan en menor medida ( $0,357 + 0,027 = 0,384$ ).

Los resultados también muestran una interacción entre el nivel del estudiante y el de la escuela. Se encuentra que los estudiantes de mayor ISE y que asisten a escuelas con mejores instalaciones y servicios, tienen mayor rendimiento ( $0,086 + 0,031 = 0,117$ ) que aquellos que asisten a escuelas con condiciones menos favorables ( $0,086 - 0,031 = 0,055$ ). Finalmente, el efecto del estatus socioeconómico también varía según las



diferentes características de la escuela, tales como gestión, tamaño de la escuela y condiciones en las que opera.

#### **6.4. Conclusiones**

Los resultados confirman que en el sistema educativo peruano existen diferencias en el rendimiento en Lectura por escuelas relacionadas al ESE (Ministerio de Educación, 2016a, 2016b, 2017b). Aunque esta relación ya es conocida, los hallazgos muestran que no se presenta de la misma forma en todas las escuelas. El efecto del estatus socioeconómico de los estudiantes es diferente entre las escuelas.

Estos resultados permiten reflexionar sobre las posibles implicancias que puedan surgir para modelos de valor agregado. En otras palabras, son importantes para aproximarse al conocimiento de cuánto de las diferencias en el rendimiento se atribuyen exclusivamente a aspectos que la escuela puede manejar (Raudenbush, 2004).

Es importante notar que los resultados de este estudio podrían deberse a múltiples variables escolares que a su vez tienen un efecto directo o indirecto en el rendimiento. En esta línea, es posible que las escuelas urbanas con estudiantes de estatus socioeconómico muy bajo o muy alto sean predominantemente no estatales. Además, dado que las escuelas estatales no cobran pensión, es probable que estas alberguen estudiantes con condiciones socioeconómicas más homogéneas que las escuelas no estatales, lo cual podría llevar a que también tengan rendimiento más homogéneo. Por otro lado, existe una amplia gama de tipos de escuelas no estatales, que incluyen diferentes costos, niveles de infraestructura y equipamiento (Balarin, 2015; Marcos, 2017), elementos que podrían generar una gran variabilidad en el efecto del ISE en el rendimiento.

Si bien en este estudio gran parte de las diferencias en el rendimiento se atribuye al estatus socioeconómico, estas también pueden deberse a otras variables contextuales del lugar en que se encuentra la escuela (Bronfenbrenner & Morris, 1998). En relación a esto, Putnam (1995) señala que el lugar de residencia no sólo provee el acceso a recursos materiales, sino también permite relacionarse con otras personas de la comunidad y la generación de capital social que impacte en el rendimiento. Si bien los resultados presentados informan sobre las diferencias del efecto del ISE en el rendimiento, no detallan qué mecanismos están ocurriendo para que ocurran estas diferencias.

Estudios posteriores podrían enfocarse en qué otras variables del contexto tienen un efecto en el estatus socioeconómico y en el rendimiento. Incluso, se podría profundizar en cómo el efecto de pares influye en la relación entre el ESE y los logros de aprendizaje.

# Comentarios finales

Capítulo 7

La construcción de un índice que represente de forma adecuada la realidad socioeconómica de las familias de los estudiantes ha representado un importante desafío para la Oficina de Medición de la Calidad de los Aprendizajes, debido a que era importante determinar la aproximación teórica y práctica adecuada para su implementación. Esto sumado al constante análisis de la relación entre el logro académico y el estatus socioeconómico en diferentes evaluaciones, han generado otras reflexiones sobre su medición y la complejidad de su relación con el rendimiento académico. En este marco, la presente publicación se ha centrado en dos líneas: la construcción y medición adecuada del ESE mediante un índice socioeconómico, y la capacidad explicativa del desempeño académico a través de este indicador. Específicamente, se analizó la estabilidad del ISE entre grados y años de una misma escuela, cómo poder calcularlo considerando diferencias contextuales, la existencia de un efecto contextual de la composición socioeconómica de la escuela, y la variabilidad del efecto encontrado sobre el rendimiento. En esta sección se hace un recuento de los principales hallazgos, así como de las oportunidades y los nuevos desafíos que implican estos resultados.

Con respecto a la construcción y medición del ISE, en el capítulo 3 se muestra que el índice socioeconómico promedio de la escuela (ISEP) sería estable entre grados de una misma escuela. Se evidencia entonces que la información socioeconómica de un grado puede ser representativa del resto de la escuela, además de que el ISEP no variaría de forma significativa en el corto plazo. Esto arroja evidencias sobre la segregación escolar y la poca diversidad socioeconómica al interior de las escuelas. La estabilidad del ISE tiene implicancias prácticas para el recojo de la información socioeconómica en las evaluaciones de rendimiento, pues permite tener evidencias sobre la realidad de los estudiantes, recolectando así una menor cantidad de información que a su vez representa un ahorro de costos al erario.

En el capítulo 4 se aborda la problemática que representa la heterogeneidad social al momento de construir un índice socioeconómico. La existencia de contextos con dificultades de acceso y conexión, diferentes niveles de desarrollo económico y limitaciones geográficas de traslado para obtener bienes y servicios, favorecen la aparición de subpoblaciones con condiciones diferenciadas entre sí. Esta realidad sugiere el uso de un método alternativo para la construcción del ISE y de un ACPC que considere las diferentes relaciones entre las variables de este indicador existentes al interior de cada estrato. Los resultados indican que al considerar grupos diferentes, el ISE estimado por ACPC se comporta de manera diferente al ISE estimado de forma tradicional, sobre todo para el área rural. Sin embargo, los resultados nacionales no varían de forma significativa. Si bien no se disipan todas las

dudas que puedan surgir a partir del uso de esta técnica, es importante profundizar en ella para determinar la viabilidad de la construcción de un ISE que considere de forma adecuada el valor de bienes y servicios al interior de cada estrato.

Como se ha explicado, el poder predictivo del ISE sobre el rendimiento es considerable; por lo que, es preciso analizar cómo se presenta este efecto y si puede variar entre las escuelas. En el capítulo 5 se exploró la existencia de un efecto contextual de la composición socioeconómica de la escuela mediante el análisis del ISEP. El estudio de aleatorización y simulación arroja evidencias de que el efecto que se suele encontrar del ISEP no sería un subproducto de la agregación del ISE, sino que representaría el efecto del contexto socioeconómico del estudiante sobre su rendimiento. También muestra la importancia de identificar e interpretar correctamente el centrado de las variables utilizadas en un modelo de regresión multinivel. Todo esto confirma las evidencias previamente encontradas en otros análisis hechos por la UMC, y permite tener un panorama más claro sobre el efecto de la composición socioeconómica al interior de las escuelas.

Por último, en el capítulo 6 se encuentra que el efecto del ESE no es idéntico en todas las escuelas. Los resultados muestran que esta relación puede variar por una serie de condiciones, como son la propia composición socioeconómica de la escuela, su gestión, el tamaño de ésta y las condiciones de su infraestructura. Sin embargo, como también resalta el estudio, existen otras variables individuales y escolares que tienen una importante capacidad predictiva del rendimiento de los estudiantes, por lo que resultan ser variables que deben ser exploradas con mayor profundidad.

Tal como muestran estos hallazgos, el índice socioeconómico construido por la UMC es un indicador sólido y presenta un comportamiento consistente con lo encontrado en otras investigaciones. Los resultados reportados proporcionan evidencias empíricas que sustentan el beneficio de construir un indicador de estatus socioeconómico sin la utilización de dos componentes tradicionales: prestigio ocupacional e ingreso. El estudio permite, además, tener un panorama más claro sobre la naturaleza y funcionamiento del ESE y la composición socioeconómica de la escuela, puesto que se focaliza en la construcción de este índice y sus efectos diferenciados sobre el rendimiento.

A partir de este estudio se podrían plantear nuevas preguntas e inquietudes que exploren con mayor profundidad los temas tratados. Por ejemplo, sería importante continuar recolectando evidencia que documente la estabilidad del ISE e ISEP en el tiempo y entre grados en una misma escuela, a mediano plazo y entre niveles de la escolaridad (primaria y secundaria). Por otro lado, se podría ahondar en la heterogeneidad social de las múltiples subpoblaciones presentes en el país, comparando cómo varían los efectos del ISE e ISEP en función al tipo de análisis de componentes principales utilizado en la construcción del ISE. Asimismo, se podría

examinar con mayor profundidad el efecto contextual del estatus socioeconómico y observar cómo se relaciona con otras variables o indicadores.

Finalmente, es importante resaltar que esta investigación y los estudios que la componen han buscado, a través de metodologías más sofisticadas, aportar al diálogo sobre el ESE y los factores asociados al rendimiento académico.

# Referencias

---

## Referencias

- Alkan, B. B., Atakam, C. & Alkan, N. (2015). A comparison of different procedures for principal component analysis in the presence of outliers. *Applied Statistics*, 1-7.
- Asociación Peruana de Empresas de Investigación de Mercados. (2016). *Niveles socioeconómicos 2016*. Lima: Autor.
- Asociación Peruana de Empresas de Investigación de Mercados. (2017). *Niveles socioeconómicos 2017*. Lima: Autor.
- Balarin, M. (2015). *Las múltiples formas y efectos de la participación del sector privado en la educación*. Lima: Proyecto FORGE.
- Balarin, M. (2016). El contexto importa: reflexiones acerca de cómo los contextos y la composición escolar afectan el rendimiento y la experiencia educativa de los estudiantes. En S. Cueto (Ed.), *Investigación para el desarrollo en el Perú. Once balances* (p. 27-53). Lima: Grupo de Análisis para el Desarrollo.
- Banerjee, A. & Duflo, E. (2012). *Repensar la pobreza. Un giro radical en la lucha contra la desigualdad global*. México: Taurus.
- Bellei, C. (2013). El estudio de la segregación socioeconómica y académica de la educación chilena. *Estudios Pedagógicos*, 34(1), 325-345.
- Bronfenbrenner, U. & Morris, P. (1998). The ecology of developmental processes. En W. Damo & R. M. Lerner (Eds.), *Handbook of child psychology: Theoretical models of human development* (p. 993-1028). Hoboken: John Wiley.
- Cairns, R. B., Cairns, B. D. & Neckerman, H. J. (1989). Early School Dropout: Configurations and Determinants. *Child Development*, 60(6), 1437-1452.
- Caro, D. (2002). *Estimación del nivel socioeconómico de las familias: propuesta metodológica para la Evaluación Nacional de Rendimiento del 2001*. Lima: Unidad de Medición de la Calidad Educativa.
- Caro, D. & Cortés, D. (2012). Measuring family socioeconomic status: An illustration using data from PIRLS 2006. *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*, 5.
- Castellano, K., Rabe-Hesketh, S. & Skrondal, A. (2014). Composition, Context, and Endogeneity in School and Teacher Comparisons. *Journal of Educational and Behavioral Statistics*, 39(5), 333-367.
- Centro de Investigación y Desarrollo del Instituto Nacional de Estadística e Informática. (2007). *Perú: características socio económicas de los hogares*. Lima: Instituto Nacional de Estadística e Informática.
- Coleman, J. (1988). Social Capital in the Creation of Human Capital. *American Journal of Sociology*, 94, S95-S120.
- Coleman, J., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D. & York, R. L. (1966). *Equality of educational opportunity*. Washington, D.C.: US Department of Health, Education and Welfare.
- Crane, J. (1991). The Epidemic Theory of Ghettos and Neighbourhood Effects on Dropping Out and Teenage Childbearing. *The American Journal of Sociology*, 96, 1226-1259.

- Datcher, L. (1982). Effects of Community and Family Background on Achievement. *The Review of Economics and Statistics*, 64(1), 32-41.
- Dickinson, E. R. & Adelson, J. L. (2014). Exploring the Limitations of Measures of Students' Socioeconomic Status (SES). *Practical Assessment, Research & Evaluation*, 19(1), 1-14.
- Diez Roux, A. (2017). A glossary for multilevel analysis. *Journal of Epidemiology and Community Health*, 56(8), 588-594.
- Duncan, C., Jones, K. & Moon, G. (1998). Context, Composition and Heterogeneity: Using Multilevel Models in Health Research. *Social Science & Medicine*, 46(1), 97-117.
- Duncan, O. D., Featherman, D. L. & Duncan, B. (1972). *Socioeconomic background and achievement*. New York: Seminar Press.
- Enders, C. & Tofighi, D. (2007). Centering Predictor Variables in Cross-Sectional Multilevel Models: A New Look at an Old Issue. *Psychological Methods*, 12(2), 121-138.
- Ensminger, M. E. & Slusarcick, A. L. (1992). Paths to High School Graduation or Dropout: A Longitudinal Study of a First-Grade Cohort. *Sociology of Education*, 65(2), 95-113.
- Escobal, J. & Torero, M. (2000). *¿Cómo enfrentar una geografía adversa?: el rol de los activos públicos y privados*. Lima: Grupo de Análisis para el Desarrollo.
- Feaster, D., Brincks, A., Robbins, M. & Szapocznik, J. (2011). Multilevel Models to Identify Contextual Effects on Individual Group Member Outcomes: A Family Example. *Family Process*, 50(2), 167-183.
- Finn, J. D. & Rock, D. A. (1997). Academic Success Among Students at Risk for School Failure. *Journal of Applied Psychology*, 82(2), 221-234.
- Flury, B. N. (1984). Principal Components in k Groups. *Journal of the American Statistical Association*, 79(388), 892-898.
- Gottfried, M. & Ream, R. (2014). Socioeconomic Status and Education. En D. J. Brewer & L. O. Picus (Eds.), *Encyclopedia of Education Economics and Finance* (p. 687-690). Los Angeles: SAGE Publications.
- Harker, R. & Tymms, P. (2004). The Effects of Student Composition on School Outcomes. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, 15(2), 177-199.
- Hertzman, C. (1994). The Lifelong Impact of Childhood Experiences: A Population Health Perspective. *Daedalus*, 123(4), 167-180.
- Hertzman, C. & Weins, M. (1996). Child development and long-term outcomes: A population health perspective and summary of successful interventions. *Social Science Medicine*, 43(7), 1083-1095.
- Hofmann, D. A. & Gavin, M. B. (1998). Centering Decisions in Hierarchical Linear Models: Implications for Research in Organizations. *Journal of Management*, 24(5), 623-641.
- Hox, J. (2010). *Multilevel Analysis. Techniques and applications*. New York: Routledge.



- Instituto Nacional de Estadística e Informática. (2016). *Perú: síntesis estadística 2016*. Lima: Autor.
- Janosz, M., LeBlanc, M., Boulerice, B. & Tremblay, R. E. (1997). Disentangling the Weight of School Dropout Predictors: A Test on Two Longitudinal Samples. *Journal of Youth and Adolescence*, 26(6), 733-762.
- Jennrich, R. (1970). An Asymptotic  $\chi^2$  Test for the equality of two correlation matrices. *Journal of American Statistical Association*, 65(330), 904-912.
- Johnson, M. K., Crosnoe, R. & Elder, G. H. (2001). Students' Attachment and Academic Engagement: The Role of Race and Ethnicity. *Sociology of Education*, 74, 318-340.
- Juneja, J. (2012). Common factors, principal components analysis, and the term structure of interest rates. *International Review of Financial Analysis*, 24, 48-56.
- Lareau, A. (1987). Social Class Differences in Family-School Relationships: The Importance of Cultural Capital. *Sociology of Education*, 60, 73-85.
- León, J. & Collahua, Y. (2016). El efecto del nivel socioeconómico en el rendimiento de los estudiantes peruanos: un balance de los últimos 15 años. En S. Cueto (Ed.), *Investigación para el desarrollo en el Perú. Once balances* (p. 109-162). Lima: Grupo de Análisis para el Desarrollo.
- Marcos, M. (2017). Colegios privados de bajo costo ¿Mejor educación a un precio razonable? Aproximación desde los resultados en Matemática en la Evaluación Censal de Estudiantes 2016. *Pensamiento Crítico*, 22(2), 31-48.
- Martínez, D. (2010). El aporte de Adolfo Figueroa al análisis de la relación entre equidad, productividad y competitividad. En F. Jiménez (Ed.), *Teoría económica y desarrollo social: exclusión, desigualdad y democracia. Homenaje a Adolfo Figueroa* (p. 25-47). Lima: Pontificia Universidad Católica del Perú.
- Ministerio de Educación. (2016a). *¿Cuánto aprenden nuestros estudiantes al término de la educación primaria? Informe de logros de aprendizaje y sus factores asociados en la Evaluación Muestral 2013*. Lima: Oficina de Medición de Calidad de los Aprendizajes.
- Ministerio de Educación. (2016b). *Regiones en perspectiva: la influencia de los factores asociados al aprendizaje al término de la educación primaria. Estudios Breves*, 2.
- Ministerio de Educación. (2016c). *Reporte técnico del índice de calidad del servicio y el contexto de la escuela*. Lima: Oficina de Medición de Calidad de los Aprendizajes.
- Ministerio de Educación. (2017a). *¿Cómo se relaciona la infraestructura de la escuela con los aprendizajes de los estudiantes? Zoom Educativo*, 3.
- Ministerio de Educación. (2017b). *El Perú en PISA 2015. Informe nacional de resultados*. Lima: Oficina de Medición de Calidad de los Aprendizajes.
- Ministerio de Educación. (2017c). *¿Qué sucede con los aprendizajes en la transición de primaria a secundaria? Una mirada desde la equidad y la*

- eficacia escolar diferenciada*. Lima: Oficina de Medición de Calidad de los Aprendizajes.
- Ministerio de Educación. (2017d). *Reporte técnico de la Evaluación Censal de Estudiantes (ECE 2016). Segundo y cuarto de primaria, segundo de secundaria*. Lima: Oficina de Medición de Calidad de los Aprendizajes.
- Ministerio de Educación. (2017e). *Reporte técnico del estatus socioeconómico en la Evaluación Censal de Estudiantes 2016*. Lima: Oficina de Medición de Calidad de los Aprendizajes.
- Mueller, C. W. & Parcel, T. L. (1981). Measures of Socioeconomic Status: Alternatives and Recommendations. *Child Development*, 52(1), 13-30.
- Murillo, F. (2016). Midiendo la segregación escolar en América Latina. Un análisis metodológico utilizando el TERCE. *Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 14(4), 33-60.
- Nash, R. (2003). Is the School Composition Effect Real?: A Discussion With Evidence From the UK PISA Data. *School Effectiveness and School Improvement*, 14(4), 441-457.
- National Center for Education Statistics. (2012). *Improving the Measurement of Socioeconomic Status for the National Assessment of Educational Progress: A Theoretical Foundation*. Washington, D.C.: Consorcio de Investigación Económica y Social.
- Paccagnella, O. (2006). Centering or not Centering in Multilevel Models? The Role of the Group Mean and the Assessment of Group Effects. *Evaluation Review*, 30(1), 66-85.
- Piketty, T. (2014). *El capital en el siglo XXI*. Madrid: Fondo de Cultura Económica.
- Pokropek, A. (2015). Phantom Effects in Multilevel Compositional Analysis: Problems and Solutions. *Sociological Methods & Research*, 44(4), 677-705.
- Putnam, R. D. (1995). Bowling Alone: America's Declining Social Capital. *Journal of democracy*, 6(1), 65-78.
- R Core Team. (2017). R: A Language and Environment for Statistical Computing [Manual de software informático]. Vienna, Austria.
- Raudenbush, S. W. (2004). What Are Value-Added Models Estimating and What Does This Imply for Statistical Practice? *Journal of Educational and Behavioral Statistics*, 29(1), 121-129.
- Raudenbush, S. W. & Kasim, R. M. (1998). Cognitive Skill and Economic Inequality: Findings from the National Adult Literacy Survey. *Harvard Educational Review*, 68(1), 33-79.
- Raudenbush, S. W. & Willms, J. D. (1995). The Estimation of School Effects. *Journal of Educational and Behavioral Statistics*, 20(4), 307-335.
- Rivas, A. (2015). *América Latina después de PISA. Lecciones aprendidas de la educación en siete países (2000-2015)*. Buenos Aires: Fundación CIPPEC.
- Rumberger, R. W. (1995). Dropping out of Middle School: A Multilevel Analysis of Students and Schools. *American Educational Research Journal*, 32, 583-625.

- Sirin, S. R. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research*, 75(3), 417-453.
- Snijders, T. & Bosker, J. (1999). *Multilevel Analysis. An introduction to basic and advanced multilevel modeling*. London: Sage Publications.
- Televantou, I., Marsh, H. W., Kyriakides, L., Nagengast, B., Fletcher, J. & Malmberg, L.-E. (2015). Phantom effects in school composition research: consequences of failure to control biases due to measurement error in traditional multilevel models. *School Effectiveness and School Improvement*, 26(1), 75-101.
- Therborn, G. (2015). *Los campos de exterminio de la desigualdad*. Buenos Aires: Fondo de Cultura Económica.
- Trendafilov, N. T. (2003). Stepwise estimation of common principal components. *Computational Statistics and Data Analysis*, 54, 3446-3457.
- van Buuren, S. & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3), 1-67.
- van Ewijk, R. & Slegers, P. (2010). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review*, 5, 134-150.
- Vazquez, E. (2012). *Segregación escolar por nivel socioeconómico midiendo el fenómeno y explorando sus determinantes* (Tesis de Master no publicada). Universidad Nacional de La Plata.
- Véliz, C. (2016). *Análisis Multivariante. Métodos estadísticos multivariantes para la investigación*. México: Cengage.
- Voelkl, K. E. (1995). School Warmth, Student Participation, and Achievement. *Journal of Experimental Education*, 63(2), 127-138.
- White, K. R. (1982). The Relation Between Socioeconomic Status and Academic Achievement. *Psychological Bulletin*, 91(3), 461-481.
- Yang, Y. & Gustafsson, J.-E. (2004). Measuring Socioeconomic Status at Individual and Collective Levels. *Educational Research and Evaluation*, 10(3), 259-288.
- Ziyatdinov, A., Kanaan-Izquierdo, S., Trendafilov, N. T. & Perera-Lluna, A. (2014). *cpca: Methods to perform Common Principal Component Analysis (CPCA)* [Manual de software informático]. (R package version 0.1.2)

# Anexos

---

## Anexos

### A1. Preguntas usadas en la elaboración del índice socioeconómico

**Tabla A1.1. Preguntas usadas en la elaboración del índice socioeconómico para 2.º y 4.º grado de primaria**

Indicador	Pregunta
Educación de los padres	¿Cuál es el nivel de estudio más alto que la madre o apoderada del estudiante ha alcanzado? ¿Cuál es el nivel de estudio más alto que el padre o apoderado del estudiante ha alcanzado?
Material de construcción de la vivienda	¿Cuál es el material predominante de las PAREDES de la casa del estudiante? ¿Cuál es el material predominante de los PISOS de la casa del estudiante? ¿Cuál es el material predominante de los TECHOS de la casa del estudiante?
Servicios básicos en el hogar	En la casa del estudiante, ¿cuál es el tipo de alumbrado? ¿De dónde proviene el agua que usan en la casa del estudiante? ¿Cómo es el baño en la casa del estudiante?
Activos y otros servicios en el hogar	La casa del estudiante tiene: radio. La casa del estudiante tiene: licuadora. La casa del estudiante tiene: plancha eléctrica. La casa del estudiante tiene: televisión. La casa del estudiante tiene: celular. La casa del estudiante tiene: reproductor de video. La casa del estudiante tiene: computadora de escritorio. La casa del estudiante tiene: laptop o computadora portátil. La casa del estudiante tiene: tablet. La casa del estudiante tiene: equipo de sonido. La casa del estudiante tiene: consola de videojuegos (Xbox, Wii, etc.). La casa del estudiante tiene: horno microondas. La casa del estudiante tiene: refrigeradora. La casa del estudiante tiene: lavadora. La casa del estudiante tiene: moto. La casa del estudiante tiene: auto y/o camioneta. La casa del estudiante tiene: teléfono fijo. La casa del estudiante tiene: plan de datos o internet móvil. La casa del estudiante tiene: conexión a internet. La casa del estudiante tiene: servicio de televisión por cable.

**Tabla A1.2. Preguntas usadas en la elaboración del índice socioeconómico para 2.º grado de secundaria**

Indicador	Pregunta
Educación de los padres	¿Cuál es el nivel de estudio más alto que la madre o apoderada del estudiante ha alcanzado?
	¿Cuál es el nivel de estudio más alto que el padre o apoderado del estudiante ha alcanzado?
Material de construcción de la vivienda	¿Cuál es el material predominante de las PAREDES de tu casa?
	¿Cuál es el material predominante de los PISOS de tu casa?
	¿Cuál es el material predominante de los TECHOS de tu casa?
Servicios básicos en el hogar	En tu casa, ¿cuál es el tipo de alumbrado?
	¿De dónde proviene el agua de tu casa?
	¿Cómo es el baño de tu casa?
Activos y otros servicios en el hogar	En tu casa tienes: radio.
	En tu casa tienes: licuadora.
	En tu casa tienes: plancha eléctrica.
	En tu casa tienes: televisión.
	En tu casa tienes: celular.
	En tu casa tienes: reproductor de video.
	En tu casa tienes: computadora de escritorio.
	En tu casa tienes: laptop o computadora portátil.
	En tu casa tienes: tablet.
	En tu casa tienes: equipo de sonido.
	En tu casa tienes: consola de videojuegos (Xbox, Wii, etc.).
	En tu casa tienes: horno microondas.
	En tu casa tienes: refrigeradora.
	En tu casa tienes: lavadora.
	En tu casa tienes: moto.
En tu casa tienes: auto y/o camioneta.	
En tu casa tienes: teléfono fijo.	
En tu casa tienes: plan de datos o internet móvil.	
En tu casa tienes: conexión a internet.	
En tu casa tienes: servicio de televisión por cable.	

## A2. Densidad del ISE en 2.º grado de secundaria

Figura A2.1. Densidad del ISE de 2.º grado de secundaria según área de la escuela

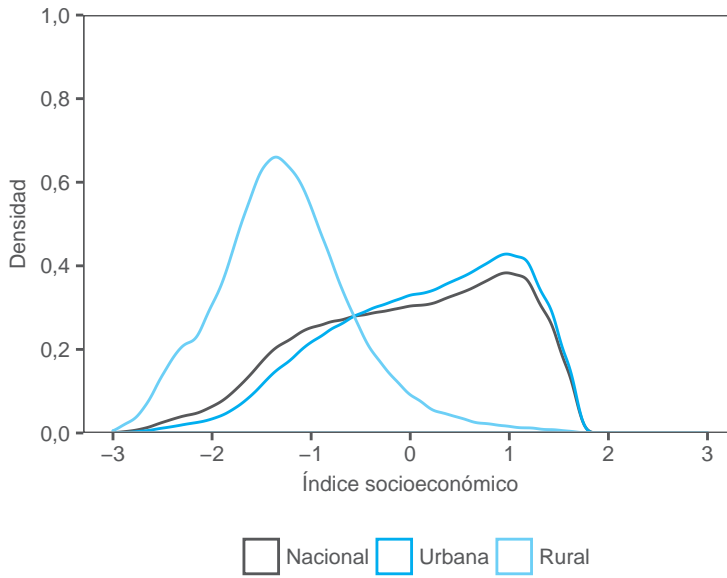
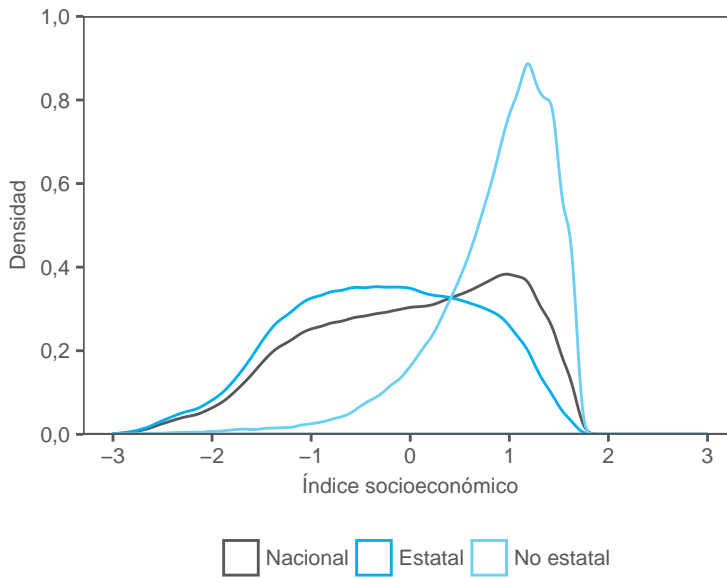
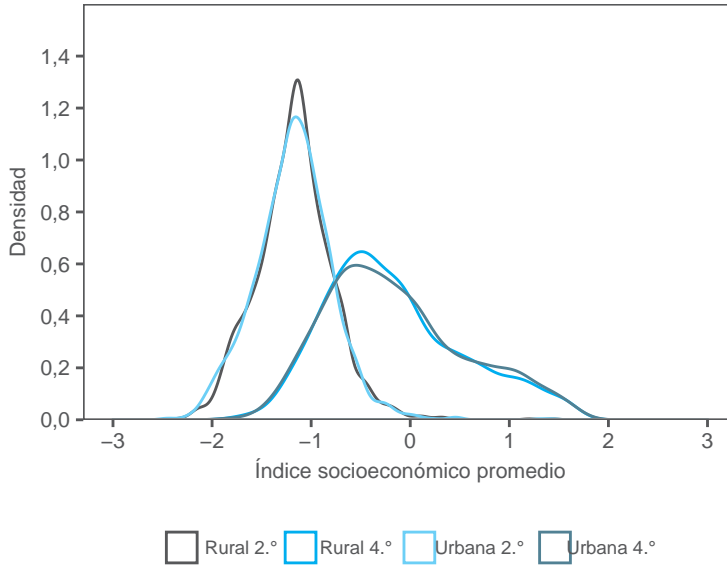
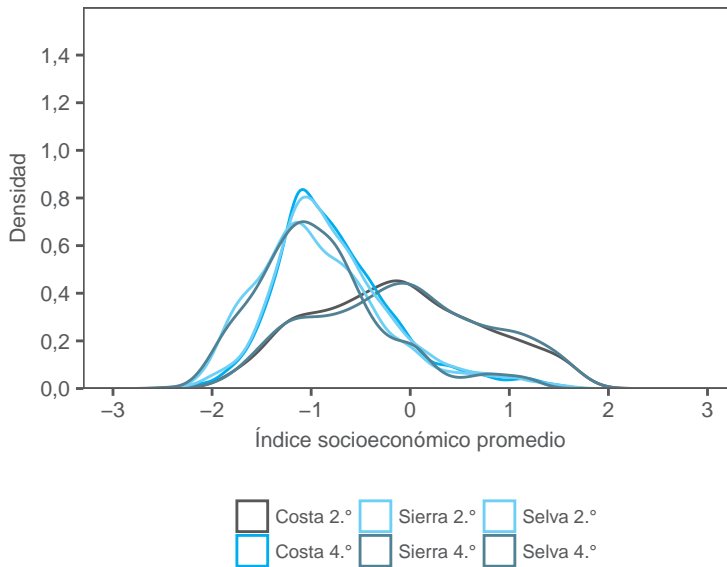


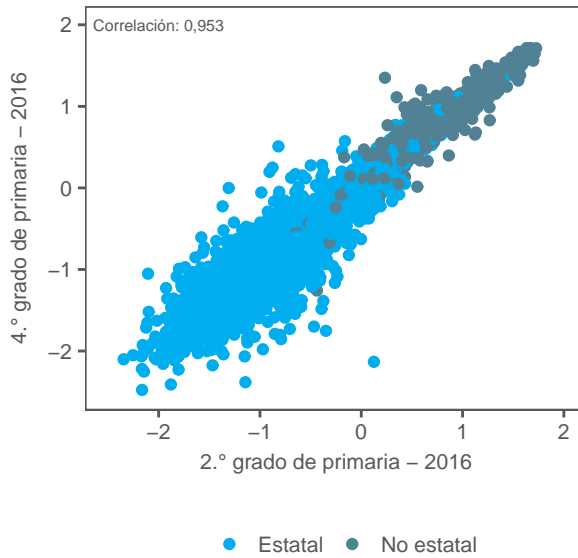
Figura A2.2. Densidad del ISE de 2.º grado de secundaria según gestión de la escuela



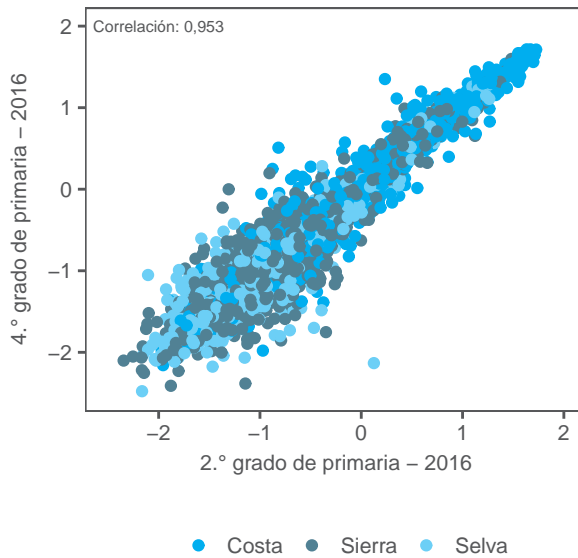
**B1. Densidad y correlación del ISEP entre años y entre grados****Figura B1.1. Densidad del ISEP de 2.º y 4.º grado de primaria según área de la escuela - ECE 2016****Figura B1.2. Densidad del ISEP de 2.º y 4.º grado de primaria según región natural - ECE 2016**



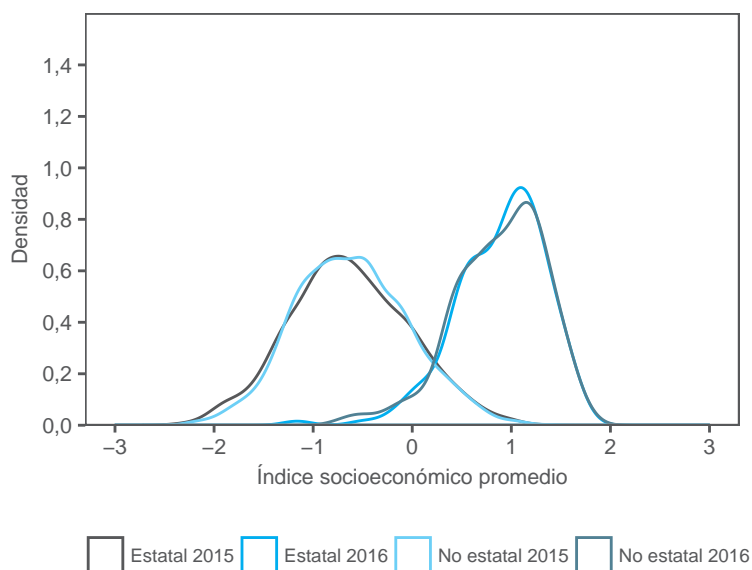
**Figura B1.3. *Dispersión del ISEP de 2.º grado y 4.º grado de primaria según gestión de la escuela - ECE 2016***



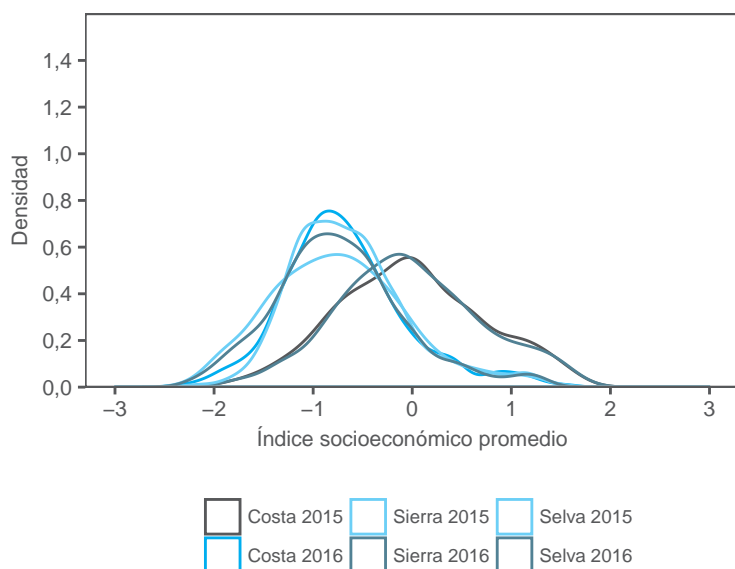
**Figura B1.4. *Dispersión del ISEP de 2.º grado y 4.º grado de primaria según región natural - ECE 2016***



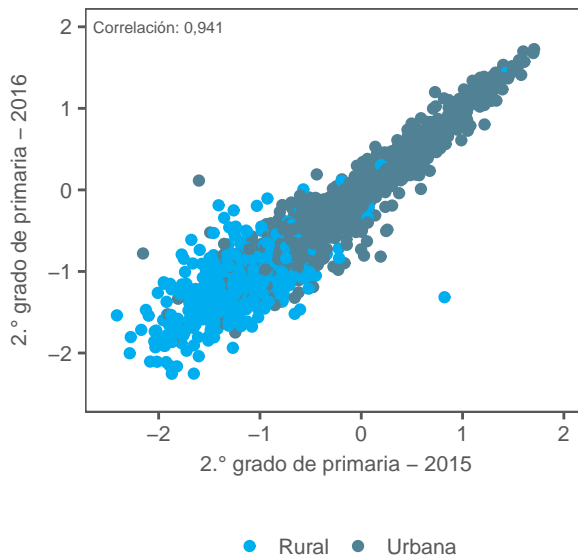
**Figura B1.5. Densidad del ISEP de 2.º grado de primaria según gestión de la escuela - ECE 2015 y 2016**



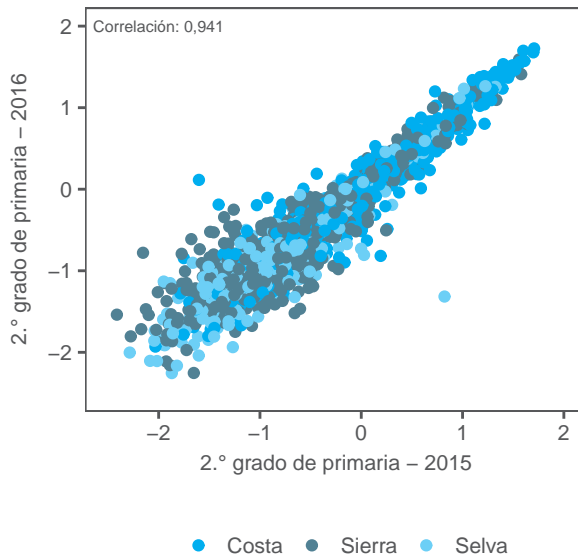
**Figura B1.6. Densidad del ISEP de 2.º grado de primaria según región natural - ECE 2015 y 2016**



**Figura B1.7. Dispersión del ISEP de 2.º grado de primaria según área de la escuela - ECE 2015 y 2016**

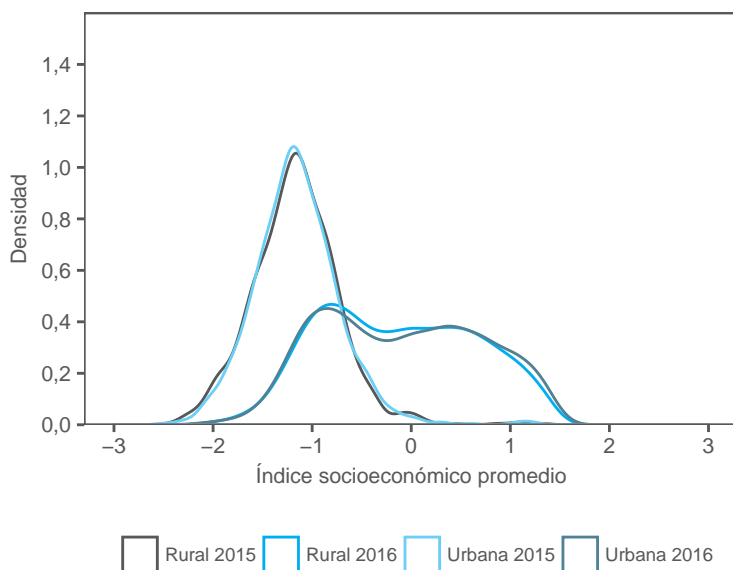


**Figura B1.8. Dispersión del ISEP de 2.º grado de primaria ECE según región natural - ECE 2015 y 2016**

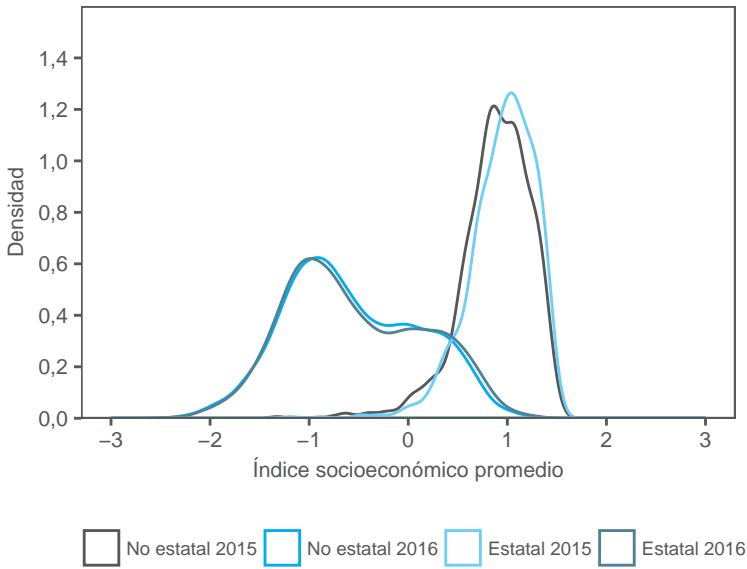


**Tabla B1.1. Distribución de la varianza del ISE de 2.º grado de secundaria - ECE 2015 y 2016**

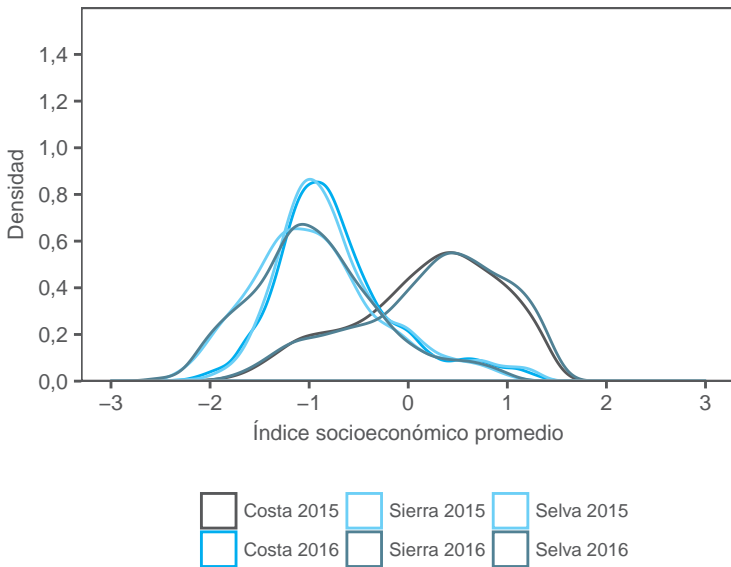
Estrato	Grupo	Estudiante	Escuela
<b>Nacional</b>	ECE 2015	0,356	0,644
	ECE 2016	0,345	0,655
	ECE 2015 - 2016	0,351	0,649
<b>Urbano</b>	ECE 2015	0,387	0,613
	ECE 2016	0,373	0,627
	ECE 2015 - 2016	0,381	0,619
<b>Rural</b>	ECE 2015	0,681	0,319
	ECE 2016	0,674	0,326
	ECE 2015 - 2016	0,688	0,312
<b>Estatal</b>	ECE 2015	0,477	0,523
	ECE 2016	0,463	0,537
	ECE 2015 - 2016	0,473	0,527
<b>No Estatal</b>	ECE 2015	0,594	0,406
	ECE 2016	0,622	0,378
	ECE 2015 - 2016	0,610	0,390
<b>Costa</b>	ECE 2015	0,394	0,606
	ECE 2016	0,382	0,618
	ECE 2015 - 2016	0,388	0,612
<b>Sierra</b>	ECE 2015	0,534	0,466
	ECE 2016	0,522	0,478
	ECE 2015 - 2016	0,533	0,467
<b>Selva</b>	ECE 2015	0,516	0,484
	ECE 2016	0,501	0,499
	ECE 2015 - 2016	0,509	0,491

**Figura B1.9. Densidad del ISEP de 2.º grado de secundaria según área de la escuela - ECE 2015 y 2016**

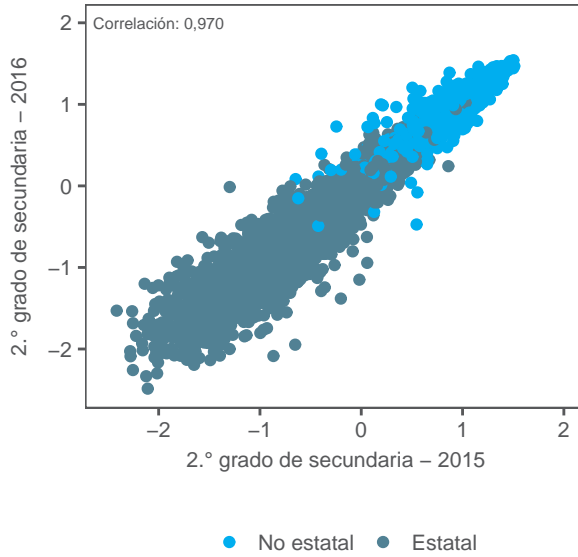
**Figura B1.10. Densidad del ISEP de 2.º grado de secundaria según gestión de la escuela - ECE 2015 y 2016**



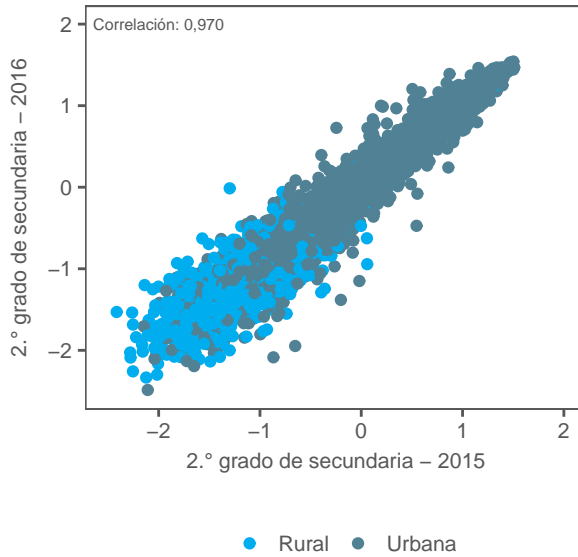
**Figura B1.11. Densidad del ISEP de 2.º grado de secundaria según región natural - ECE 2015 y 2016**



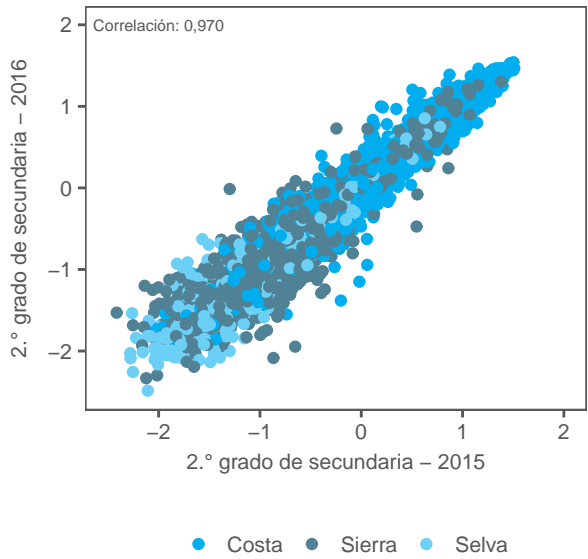
**Figura B1.12. Dispersión del ISEP de 2.º grado de secundaria según gestión de la escuela - ECE 2015 y 2016**



**Figura B1.13. Dispersión del ISEP de 2.º grado de secundaria según área de la escuela - ECE 2015 y 2016**



**Figura B1.14. Dispersión del ISEP de 2.º grado de secundaria según región natural - ECE 2015 y 2016**



**C1. Análisis de componentes principales comunes por área y región natural****Tabla C1.1. Prueba de igualdad de matrices de correlación, subpoblaciones vs. nacional - 2.º grado de secundaria**

Test de igualdad de matrices	$\chi^2$	Significancia
Nacional vs. Urbana	3869,547	<0,001
Nacional vs. Rural	18128,02	<0,001
Nacional vs. Costa	6013,855	<0,001
Nacional vs. Sierra	3650,899	<0,001
Nacional vs. Selva	2186,857	<0,001
Urbana vs. Rural	15347,700	<0,001
Costa vs. Sierra	5316,014	<0,001
Costa vs. Selva	2596,091	<0,001
Sierra vs. Selva	1551,168	<0,001

**Tabla C1.2. Cargas al componente y varianza explicada de los indicadores del índice socioeconómico - 2.º grado de secundaria**

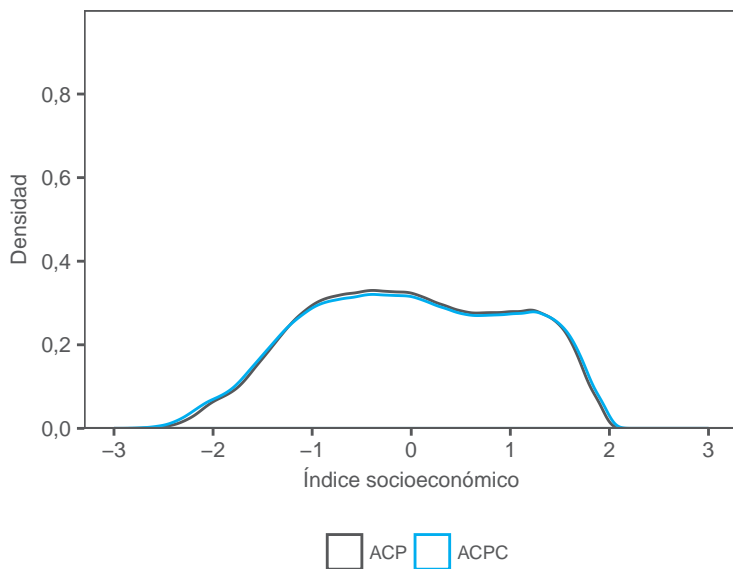
Cargas al componente	Área		Región natural		
	Urbana	Rural	Costa	Sierra	Selva
Máximo nivel educativo	0,682	0,412	0,678	0,692	0,623
Calidad de la vivienda	0,764	0,617	0,750	0,765	0,721
Servicios básicos	0,686	0,666	0,670	0,722	0,736
Otros servicios	0,822	0,707	0,821	0,781	0,805
Activos en el hogar	0,875	0,849	0,863	0,878	0,881
Varianza explicada	0,59	0,44	0,58	0,59	0,58

**Tabla C1.3. Cargas al componente y varianza explicada de los indicadores del índice socioeconómico considerando heterogeneidad social - 2.º grado de secundaria**

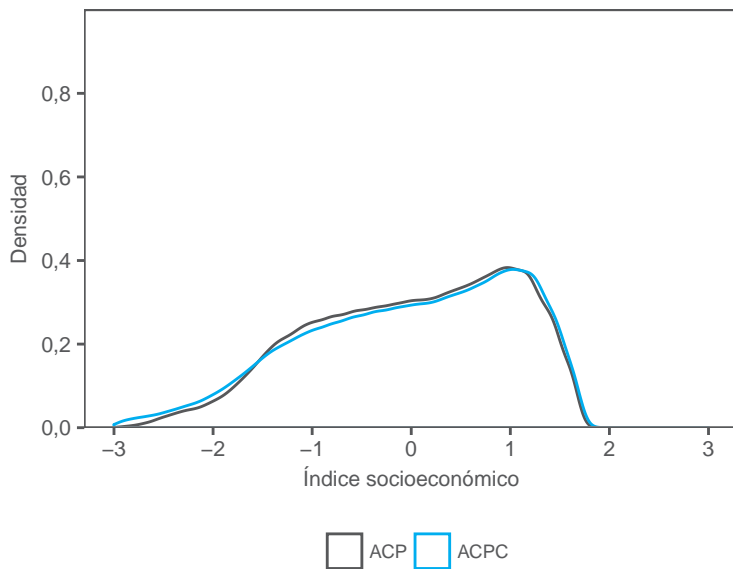
Cargas al componente	Área		Región natural		
	Urbana	Rural	Costa	Sierra	Selva
Máximo nivel educativo	0,594	0,514	0,663	0,672	0,661
Calidad de la vivienda	0,741	0,640	0,743	0,753	0,741
Servicios básicos	0,722	0,624	0,707	0,716	0,705
Otros servicios	0,820	0,709	0,800	0,810	0,797
Activos en el hogar	0,925	0,799	0,871	0,882	0,797
Varianza explicada	0,59	0,44	0,58	0,59	0,57



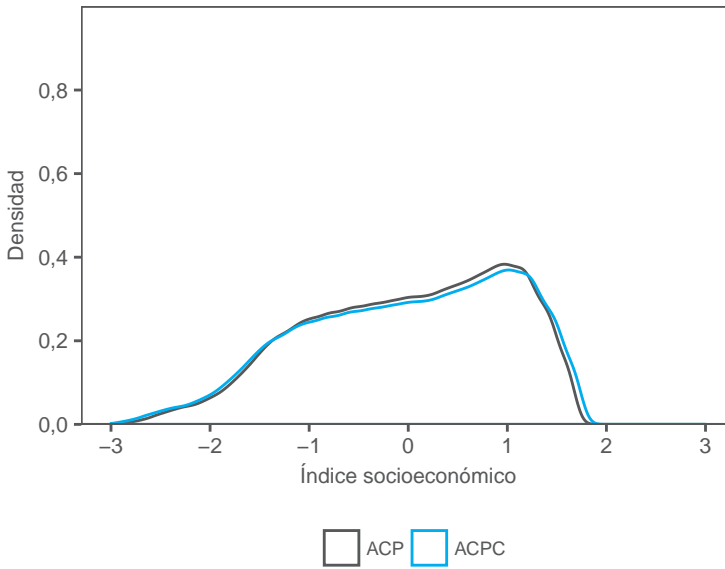
**Figura C1.1. Densidad del ISE calculado mediante ACP y ACPC considerando diferencias por región natural - 4.º grado de primaria**



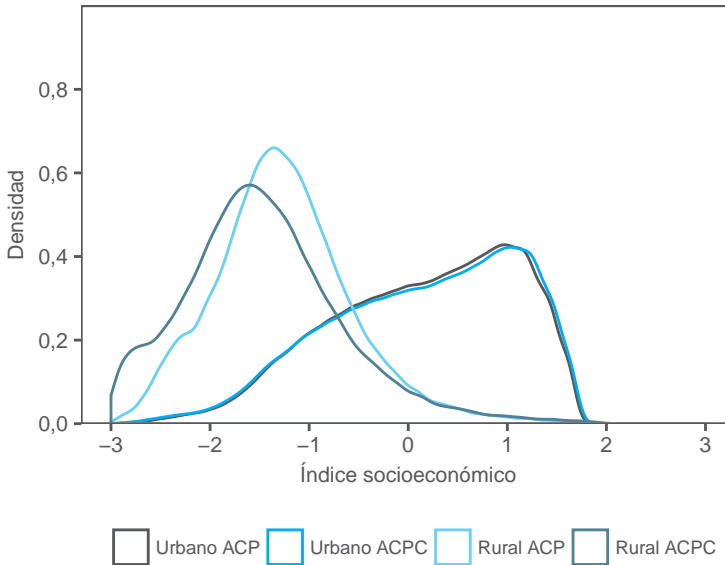
**Figura C1.2. Densidad del ISE calculado mediante ACP y ACPC considerando diferencias por área de la escuela - 2.º grado de secundaria**



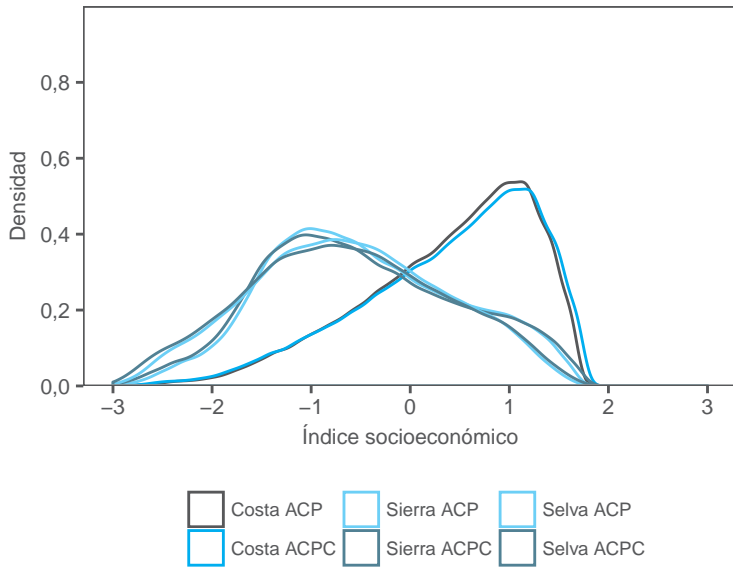
**Figura C1.3. Densidad del ISE calculado mediante ACP y ACPC considerando diferencias por región natural - 2.º grado de secundaria**



**Figura C1.4. Densidad según área del ISE calculado mediante ACP y ACPC considerando diferencias por área de la escuela - 2.º grado de secundaria**



**Figura C1.5. Densidad según región del ISE calculado mediante ACP y ACPC considerando diferencias por región natural - 2.º grado de secundaria**



**D1. Modelos multinivel con datos aleatorizados****Tabla D1.1. Modelos con datos aleatorizados de 2.º grado de primaria - ECE 2015**

Modelo	Datos	Lectura			Matemática		
		Intercepto	ISE	ISEP	Intercepto	ISE	ISEP
Modelo 1:	Aleatorizados	607,10	32,88		600,63	28,33	
ISE	Reales	576,83	11,94		568,22	10,86	
Modelo 2:	Aleatorizados	601,55	32,88		595,83	28,33	
ISE(GM)	Reales	574,81	11,94		566,39	10,86	
Modelo 3:	Aleatorizados	601,54	32,88		595,82	28,34	
ISE(MC)	Reales	569,69	7,90		561,89	7,77	
Modelo 4:	Aleatorizados	607,11		32,90	600,62		28,30
ISEP	Reales	601,21		54,63	589,96		48,84
Modelo 5:	Aleatorizados	601,55		32,90	595,82		28,30
ISEP(GM)	Reales	591,96		54,63	581,67		48,84
Modelo 6:	Aleatorizados	607,11	32,88	0,02	600,62	28,34	-0,03
ISE+ISEP	Reales	601,21	7,90	46,73	589,96	7,77	41,07
Modelo 7:	Aleatorizados	607,10	32,88	0,02	600,63	28,34	-0,03
ISE+ISEP(GM)	Reales	593,29	7,90	46,73	582,99	7,77	41,07
Modelo 8:	Aleatorizados	601,55	32,88	0,02	595,82	28,34	-0,03
ISE(GM)+ISEP	Reales	599,87	7,90	46,73	588,65	7,77	41,07
Modelo 9:	Aleatorizados	601,55	32,88	0,02	595,82	28,34	-0,03
ISE(GM)+ISEP(GM)	Reales	591,95	7,90	46,73	581,67	7,77	41,07
Modelo 10:	Aleatorizados	607,11	32,88	32,90	600,62	28,34	28,30
ISE(MC)+ISEP	Reales	601,21	7,90	54,63	589,96	7,77	48,84
Modelo 11:	Aleatorizados	601,55	32,88	32,90	595,82	28,34	28,30
ISE(MC)+ISEP(GM)	Reales	591,95	7,90	54,63	581,67	7,77	48,84

**Tabla D1.2. Modelos con datos aleatorizados de 2.º grado de primaria - ECE 2016**

Modelo	Datos	Lectura			Matemática		
		Intercepto	ISE	ISEP	Intercepto	ISE	ISEP
Modelo 1:	Aleatorizados	597,67	33,98		625,07	31,93	
ISE	Reales	566,51	14,98		586,15	15,23	
Modelo 2:	Aleatorizados	592,26	33,98		619,96	31,93	
ISE(GM)	Reales	564,13	14,98		583,72	15,23	
Modelo 3:	Aleatorizados	592,28	33,99		619,97	31,92	
ISE(MC)	Reales	556,61	12,33		576,38	12,75	
Modelo 4:	Aleatorizados	597,63		33,75	625,06		31,93
ISEP	Reales	589,18		51,00	607,48		49,29
Modelo 5:	Aleatorizados	592,26		33,75	619,96		31,93
ISEP(GM)	Reales	581,05		51,00	599,62		49,29
Modelo 6:	Aleatorizados	597,64	33,99	-0,23	625,06	31,92	0,01
ISE+ISEP	Reales	589,17	12,33	38,67	607,48	12,75	36,54
Modelo 7:	Aleatorizados	597,68	33,99	-0,23	625,07	31,92	0,01
ISE+ISEP(GM)	Reales	583,01	12,33	38,67	601,65	12,75	36,54
Modelo 8:	Aleatorizados	592,22	33,99	-0,23	619,96	31,92	0,01
ISE(GM)+ISEP	Reales	587,20	12,33	38,67	605,44	12,75	36,54
Modelo 9:	Aleatorizados	592,26	33,99	-0,23	619,96	31,92	0,01
ISE(GM)+ISEP(GM)	Reales	581,04	12,33	38,67	599,61	12,75	36,54
Modelo 10:	Aleatorizados	597,64	33,99	33,76	625,06	31,92	31,93
ISE(MC)+ISEP	Reales	589,17	12,33	51,01	607,48	12,75	49,29
Modelo 11:	Aleatorizados	592,26	33,99	33,76	619,96	31,92	31,93
ISE(MC)+ISEP(GM)	Reales	581,04	12,33	51,01	599,61	12,75	49,29

**Tabla D1.3. Modelos con datos aleatorizados de 4.º grado de primaria - ECE 2016**

Modelo	Datos	Lectura			Matemática		
		Intercepto	ISE	ISEP	Intercepto	ISE	ISEP
Modelo 1:	Aleatorizados	488,77	39,05		476,62	28,87	
ISE	Reales	466,78	20,73		456,08	14,44	
Modelo 2:	Aleatorizados	484,21	39,05		473,25	28,87	
ISE(GM)	Reales	464,36	20,73		454,39	14,44	
Modelo 3:	Aleatorizados	484,21	39,05		473,25	28,87	
ISE(MC)	Reales	454,68	16,15		447,87	11,85	
Modelo 4:	Aleatorizados	488,77		39,01	476,62		28,78
ISEP	Reales	482,12		49,13	467,86		35,87
Modelo 5:	Aleatorizados	484,21		39,01	473,25		28,78
ISEP(GM)	Reales	476,38		49,13	463,67		35,87
Modelo 6:	Aleatorizados	488,77	39,05	-0,04	476,62	28,87	-0,09
ISE+ISEP	Reales	482,10	16,15	32,98	467,85	11,85	24,02
Modelo 7:	Aleatorizados	488,77	39,05	-0,04	476,62	28,87	-0,09
ISE+ISEP(GM)	Reales	478,25	16,15	32,98	465,05	11,85	24,02
Modelo 8:	Aleatorizados	484,21	39,05	-0,04	473,24	28,87	-0,09
ISE(GM)+ISEP	Reales	480,21	16,15	32,98	466,46	11,85	24,02
Modelo 9:	Aleatorizados	484,21	39,05	-0,04	473,25	28,87	-0,09
ISE(GM)+ISEP(GM)	Reales	476,36	16,15	32,98	463,66	11,85	24,02
Modelo 10:	Aleatorizados	488,77	39,05	39,01	476,62	28,87	28,78
ISE(MC)+ISEP	Reales	482,10	16,15	49,13	467,85	11,85	35,87
Modelo 11:	Aleatorizados	484,21	39,05	39,01	473,25	28,87	28,78
ISE(MC)+ISEP(GM)	Reales	476,36	16,15	49,13	463,66	11,85	35,87

**Tabla D1.4. Modelos con datos aleatorizados de 2.º grado de secundaria - ECE 2015**

Modelo	Datos	Lectura			Matemática		
		Intercepto	ISE	ISEP	Intercepto	ISE	ISEP
Modelo 1:	Aleatorizados	565,31	33,48		552,11	26,05	
ISE	Reales	555,49	14,63		543,90	11,32	
Modelo 2:	Aleatorizados	564,16	33,48		551,22	26,05	
ISE(GM)	Reales	554,98	14,63		543,51	11,32	
Modelo 3:	Aleatorizados	564,16	33,48		551,22	26,05	
ISE(MC)	Reales	550,71	12,02		540,29	9,26	
Modelo 4:	Aleatorizados	565,30		33,31	552,10		25,90
ISEP	Reales	564,89		45,91	550,79		34,37
Modelo 5:	Aleatorizados	564,16		33,31	551,22		25,90
ISEP(GM)	Reales	563,30		45,91	549,61		34,37
Modelo 6:	Aleatorizados	565,30	33,48	-0,16	552,10	26,05	-0,15
ISE+ISEP	Reales	564,88	12,02	33,88	550,78	9,26	25,11
Modelo 7:	Aleatorizados	565,31	33,48	-0,16	552,11	26,05	-0,15
ISE+ISEP(GM)	Reales	563,72	12,02	33,88	549,93	9,26	25,11
Modelo 8:	Aleatorizados	564,15	33,48	-0,16	551,21	26,05	-0,15
ISE(GM)+ISEP	Reales	564,46	12,02	33,88	550,46	9,26	25,11
Modelo 9:	Aleatorizados	564,16	33,48	-0,16	551,22	26,05	-0,15
ISE(GM)+ISEP(GM)	Reales	563,30	12,02	33,88	549,60	9,26	25,11
Modelo 10:	Aleatorizados	565,30	33,48	33,32	552,10	26,05	25,90
ISE(MC)+ISEP	Reales	564,88	12,02	45,89	550,78	9,26	34,36
Modelo 11:	Aleatorizados	564,16	33,48	33,32	551,22	26,05	25,90
ISE(MC)+ISEP(GM)	Reales	563,30	12,02	45,89	549,60	9,26	34,36

**Tabla D1.5. Modelos con datos aleatorizados para Historia, Geografía y Economía de 2.º grado de secundaria - ECE 2016**

Modelo	Datos	Intercepto	ISE	ISEP
Modelo 1:	Aleatorizados	503,99	31,19	
	Reales	496,83	14,90	
Modelo 2:	Aleatorizados	504,23	31,19	
	Reales	496,93	14,90	
Modelo 3:	Aleatorizados	504,23	31,19	
	Reales	492,63	11,46	
Modelo 4:	Aleatorizados	503,99		31,25
	Reales	503,42		40,96
Modelo 5:	Aleatorizados	504,23		31,25
	Reales	503,74		40,96
Modelo 6:	Aleatorizados	503,99	31,19	0,05
	Reales	503,42	11,46	29,49
Modelo 7:	Aleatorizados	503,99	31,19	0,05
	Reales	503,66	11,46	29,49
Modelo 8:	Aleatorizados	504,24	31,19	0,05
	Reales	503,50	11,46	29,49
Modelo 9:	Aleatorizados	504,23	31,19	0,05
	Reales	503,74	11,46	29,49
Modelo 10:	Aleatorizados	503,99	31,19	31,25
	Reales	503,42	11,46	40,96
Modelo 11:	Aleatorizados	504,23	31,19	31,25
	Reales	503,74	11,46	40,96

## D2. Estudio de simulación - código en R

```
library(plyr); library(lme4); library(matrixStats); library(truncnorm)

b0<-c(1,2)
b1<-1
b2<-1
sdm<-0.5
sdsd<-0.15
sim<-1000

m<-c(5,10,20,40,80) # Tamaño del clúster
n<-c(20,40,100,1000) # Número de clústeres

mn<-expand.grid(m,n); rm(m,n)
mn1<-cbind(mn,NA,NA,NA,NA,NA,NA)
mn2<-cbind(mn,NA,NA,NA,NA,NA,NA)
colnames(mn1)<-c("n","ncl","me_b0","me_b1","me_b2","de_b0","de_b1","de_b2")
colnames(mn2)<-c("n","ncl","me_b0","me_b1","me_b2","de_b0","de_b1","de_b2")

set.seed(343)

for(i in 1:nrow(mn)){
  co1<-NULL
  co2<-NULL

  for(j in 1:sim){
    ncl<-mn[i,2]
    id<-rep(1:mn[i,2],each=mn[i,1])

    de<-rtruncnorm(ncl,a=0,b=Inf,mean=sdm,sd=sdsd)

    xb<-rnorm(ncl)

    mat<-mapply(rnorm,n=mn[i,1],mean=xb,sd=de)

    x2<-rep(colMeans(mat),each=mn[i,1])
    x1<-c(mat)

    b0j<-rep(runif(mn[i,2],min=b0[1],max=b0[2]),each=mn[i,1])
    eij<-rnorm(mn[i,1]*mn[i,2])
    uij<-rep(rnorm(mn[i,2]),each=mn[i,1])

    yij1<-b0j+b1*x1+b2*x2+uj+eij
    yij2<-b0j+b1*x1+uj+eij

    co1<-rbind(co1,lmer(yij1~1+x1+x2+(1|id))@beta)
    co2<-rbind(co2,lmer(yij2~1+x1+x2+(1|id))@beta)

  }

  mn1[i,3:8]<-c(colMeans(co1),colSds(co1))
  mn2[i,3:8]<-c(colMeans(co2),colSds(co2))
}

mn1;mn2
```

## E1. Indicadores utilizados en el modelo de regresión multinivel

**Tabla E1.1. Tasas de respuestas de los ítems de no violencia en la escuela**

Pregunta	Nunca	Pocas veces	Muchas veces	Siempre	Casos perdidos	Carga factorial
Uno o más PROFESORES(AS) me amenazaron con lastimarme o pegarme	93,6 %	4,8 %	0,8 %	0,8 %	3,3 %	0,752
Uno o más COMPAÑEROS(AS) me agredieron físicamente	77,2 %	18,7 %	2,6 %	1,5 %	3,6 %	0,751
Uno o más COMPAÑEROS(AS) se burlaron de mí o me insultaron	54,9 %	34,4 %	7,4 %	3,2 %	3,8 %	0,692
Uno o más PROFESORES(AS) se burlaron de mí o me insultaron	90,7 %	7,2 %	1,3 %	0,9 %	3,8 %	0,754
Uno o más COMPAÑEROS(AS) me robaron algo	58,4 %	30,3 %	7,7 %	3,6 %	4,0 %	0,565
Uno o más PROFESORES(AS) me agredieron físicamente	93,1 %	4,9 %	1,1 %	0,9 %	3,6 %	0,755
Uno o más COMPAÑEROS(AS) me amenazaron con lastimarme o pegarme	83,8 %	12,5 %	2,3 %	1,4 %	3,5 %	0,800

CFI = 0,953, TLI = 0,949, RMSEA = 0,048

**Tabla E1.2. Tasas de respuestas de los ítems de exposición a diferentes tipos de pregunta sobre comprensión de textos**

Pregunta	Nunca	Pocas veces	Muchas veces	Siempre	Casos perdidos	Carga factorial
Preguntas que piden encontrar información mencionada en el texto	4,1 %	26,6 %	43,8 %	25,5 %	14,8 %	0,734
Preguntas que piden deducir algo que no está mencionado en el texto	3,6 %	23,1 %	45,9 %	27,4 %	14,8 %	0,774
Preguntas que piden dar tu opinión sobre las ideas del texto	2,3 %	19,5 %	45,7 %	32,6 %	14,8 %	0,673
Preguntas que piden dar tu opinión sobre la forma cómo está escrito el texto	3,2 %	22,4 %	43,4 %	31,0 %	14,8 %	0,621

CFI = 1,000, TLI = 1,000, RMSEA = 0,000



**Tabla E1.3. Tasas de respuestas de los ítems de percepción sobre el conocimiento pedagógico de los docentes de Comunicación**

Pregunta	No lo dominan	Lo dominan con bastante dificultad	Lo dominan con algo de dificultad	Lo dominan	Casos perdidos	Carga factorial
Conocimiento sobre los niveles de escritura (pre silábico, silábico, silábico alfabético y alfabético)	0,7 %	3,9 %	35,5 %	59,9 %	4,7 %	0,791
Conocimiento sobre los tipos de comprensión lectora (literal, inferencial y reflexión)	0,4 %	3 %	30,2 %	66,4 %	4,7 %	0,804
Estrategias para la organización de la información (construcción de esquemas, diagramas, mapas conceptuales)	0,9 %	6,6 %	41 %	51,4 %	4,4 %	0,832
Enfoque comunicativo para la enseñanza de la lectura y escritura	0,6 %	4,8 %	36,9 %	57,8 %	4,5 %	0,866
Proceso de la producción escrita (planificación, textualización, edición)	0,7 %	5,7 %	40,4 %	53,2 %	5 %	0,852

CFI = 0,990, TLI = 0,988, RMSEA = 0,057

**Tabla E1.4. Tasas de respuestas de los ítems de percepción sobre la existencia de problemas que afectan el ambiente escolar y de problemas familiares de los estudiantes**

Pregunta	No ocurre	A veces ocurre	Regularmente ocurre	Casi siempre ocurre	Casos perdidos	Carga factorial
Desmotivación de los estudiantes para estudiar	20,3 %	56,5 %	18,2 %	5,0 %	4,9 %	0,624
Ausentismo escolar no justificado	47,2 %	42,5 %	7,7 %	2,7 %	4,2 %	0,659
Indisciplina de los estudiantes en clase	28,4 %	60,9 %	8,6 %	2,1 %	4,6 %	0,623
Estudiantes que trabajan para apoyar económicamente a sus familias	62,2 %	25,3 %	8,5 %	4,0 %	4,5 %	0,332
Situaciones de intimidación o amenaza entre estudiantes	74,2 %	23,6 %	1,6 %	0,6 %	4,9 %	0,712
Inasistencia por parte de los docentes	67,3 %	30,8 %	1,3 %	0,6 %	4,5 %	0,477
Docentes que no preparan adecuadamente sus clases	54,1 %	42,0 %	3,0 %	0,8 %	4,3 %	0,530
Estudiantes que trabajan para apoyar económicamente a sus familias	62,2 %	25,3 %	8,5 %	4,0 %	4,5 %	0,403
Padres de familia con escasos recursos económicos	8,8 %	30,9 %	28,5 %	31,9 %	4,6 %	0,826
Padres de familia con un bajo nivel educativo	6,6 %	29,0 %	32,5 %	31,9 %	4,7 %	0,929
Escaso involucramiento familiar en las tareas escolares	4,5 %	27,8 %	33,0 %	34,6 %	4,9 %	0,752

CFI = 0,963, TLI = 0,948, RMSEA = 0,086