

El efecto de los factores medioambientales sobre los rendimientos escolares

Rodrigo Pérez Silva¹, Ignacio C. Fernández² y Claudia Sanhueza¹

Resumen

El presente estudio tiene como objetivo evaluar el impacto de diversas variables medioambientales, tales como temperatura superficial, la presencia de áreas verdes dentro y fuera de las escuelas, y un set de medidas de contaminación atmosférica, sobre los rendimientos académicos de los estudiantes del Gran Santiago. Para ello combina información socioeconómica y de pruebas estandarizadas de aprendizaje (SIMCE) con el modelamiento y análisis de información territorial y espacial de factores medioambientales para 1208 escuelas de 4to básico en el período 2010-2018. Modelos parciales, en que cada variable medioambiental se ingresa por separado, indican que factores como la presencia de áreas verdes tanto dentro como fuera de la escuela y la contaminación atmosférica a la hora de la prueba tendrían un impacto estadísticamente significativo sobre los rendimientos académicos promedio de las escuelas del Gran Santiago. En nuestros resultados principales, en los que se estima un modelo de efectos fijos al nivel de la escuela y se consideran todos los factores medioambientales de forma simultánea, variaciones puntuales en la concentración de material particulado fino (PM 2.5) durante el horario de la prueba, provoca una caída significativa en los puntajes obtenidos en la prueba de matemática, pero no así en lectura. Por el contrario, la vegetación al interior de las escuelas tiene un efecto positivo sobre los rendimientos académicos tanto en matemáticas como en lectura, el que persiste aún cuando se controla por diferencias de vegetación al exterior de las mismas. La temperatura en tanto afectaría negativamente los rendimientos obtenidos en matemáticas.

1. Introducción

Según varios informes que miden la calidad del aire (World Air Quality Report, 2019; Greenpace, 2019), las ciudades de Chile y en particular la capital Santiago se encuentran entre las más contaminadas del mundo. Aunque estos niveles de contaminación están presentes desde la colonia, el crecimiento de la ciudad de los últimos 50 años ha agravado la situación (Memoria Chilena, 2020). Estudios previos han demostrado que la contaminación atmosférica puede tener efectos importantes sobre la productividad (Chang et al. 2016; Hanna y Oliva, 2015), el desarrollo de los niños y afectar de forma significativa sus rendimientos escolares (Ebenstein et al. 2016; Miller y Vela, 2013; Mohai et al. 2011). De igual forma, la presencia de áreas verdes ha sido vinculada a menores niveles de estrés y mejores estados de salud (Beyer et al. 2014; Dadvand et al. 2012; Nielsen y Hansen, 2007), pero también a mayores desarrollos cognitivos y mejores rendimientos escolares (Hodson y Sander, 2017; Dadvand et al. 2015; Bharadwaj et al. 2017). Finalmente, algo similar se ha encontrado para las condiciones de ruido (van Kempen et al. 2010; Clark et al. 2005).

A pesar de la importante evidencia relativa a los efectos de los factores medioambientales, no se han encontrado estudios que traten de evaluar de forma combinada el efecto de cada uno de estos

¹ Centro de Economía y Políticas Sociales, Universidad Mayor, Chile.

rodrigo.perez@umayor.cl, claudia.sanhueza@umayor.cl

² Centro de Modelación y Monitoreo de Ecosistemas, Universidad Mayor, Chile.

ignacio.fernandez@umayor.cl

elementos. Más aún, no se cuenta con evidencia de que algún esfuerzo de este tipo se haya realizado para el caso chileno, con la excepción del estudio de Miller y Vela (2013) que estima los efectos específicos de la contaminación acumulada (anual) sobre los resultados en educación.

El objetivo principal de este estudio es determinar el efecto de las características medioambientales sobre los rendimientos académicos de los estudiantes de 4to básico del Gran Santiago para el período 2010-2018. Para este fin, se utiliza información de dos fuentes principales. Por una parte, los puntajes obtenidos en las pruebas de matemáticas y lectura provenientes del Sistema de Medición de la Calidad de la Educación (SIMCE) a nivel de escuelas, y por otra, los datos de contexto medioambiental, principalmente de contaminación ambiental por material particulado 2.5 y 10 (PM2.5 y PM10, respectivamente) presente en el entorno de las escuelas. De forma adicional se considera información de cobertura vegetal al interior de las escuelas y fuera de ellas, así también como información de temperatura superficial en el área de la escuela.

Los efectos de la desigualdad sobre los rendimientos escolares han sido ampliamente discutidos tanto en la literatura académica como a nivel de discusión política (Heckman, 2011; Blanden, 2020; Hanushek et al. 2016). Brechas en ingresos, oportunidades y educación de los padres (Hanushek et al. 2016; Martins y Veiga, 2010; Dubow et al. 2009; Heckman, 2006), así también como la calidad de las escuelas, de sus profesores y directivos, han sido destacados como elementos fundamentales para entender la disparidad en logros educativos entre estudiantes, comunas, regiones y segmentos socioeconómicos (Hanushek et al. 2016; Card y Krueger, 1992; Rivkin et al. 2005; Glewwe y Muralidharan, 2016). La literatura coincide en que, independientemente de la motivación o el esfuerzo que ejerza el estudiante y de su capacidad académica innata, existen ciertas condiciones exógenas al estudiante que modelan los potenciales resultados que el estudiante puede alcanzar en la escuela y que más aún, determinan sus resultados al momento de tratar de ingresar a la universidad y luego en el mercado laboral. Sin embargo, es importante señalar que tales condiciones y la efectividad de las mismas en mejorar los rendimientos académicos varían según el contexto en el que se desarrolla la actividad escolar (Glewwe y Muralidharan, 2016).

Usualmente, hijos de padres con mayor educación e ingresos, tienden a obtener mejores resultados académicos y mejores perspectivas laborales. Adicionalmente, los hijos de estos padres asisten a escuelas que cuentan con más y mejores recursos educativos, lo que refuerza las diferencias de origen (del hogar). Igualmente, y dado que las características de los hogares y de las escuelas no se encuentran aleatoriamente distribuidas en el espacio, existen factores comúnmente asociados a la comuna de residencia y/o del entorno escolar que pueden afectar de manera importante los rendimientos escolares, los que pueden tener implicancias de política significativa y que no han sido suficientemente estudiados. En la literatura económica este tipo de problemas se resuelve regularmente a través del uso de efectos fijos al nivel de la escuela y/o de la comuna (ver, por ejemplo, Card y Krueger, 1992), los que capturan todos aquellos elementos propios de los territorios o escuelas y que son persistentes en el tiempo, pero que sin embargo no permiten la desagregación de estos elementos del territorio ni el diseño de políticas públicas orientadas a su mejoramiento.

Este estudio contribuye a la literatura de diversas formas. Primero, en adición a la recolección de información socioeconómica, integra diversas aproximaciones metodológicas para la recolección, tratamiento y modelamiento de información medioambiental, lo que es usualmente ignorado en la literatura económica. Segundo, es el primer esfuerzo del que tengamos conocimiento en tratar de desagregar los efectos de múltiples factores medioambientales sobre los rendimientos escolares de

forma simultánea, lo que resulta particularmente cierto para el caso chileno donde no existen estudios de este tipo. Tercero, pone de relevancia otras formas de desigualdad que enfrentan los estudiantes chilenos y entrega valiosas contribuciones al desarrollo de políticas públicas de mejoramiento de la infraestructura urbana al nivel territorial y comunal, lo que nuevamente es relevante para el caso chileno por su alto nivel de segregación y desigualdad. Por último, a diferencia del estudio realizado por Miller y Vela (2013) para el caso chileno, quienes sólo utilizan información anual de contaminantes, este estudio permite identificar los niveles de contaminación para el día y las horas alrededor de las cuales las pruebas del SIMCE son tomadas, explotando así la variación temporal y espacial de los contaminantes como estrategia de identificación econométrica. Esto resulta particularmente importante si se considera que existe una variación temporal en la contaminación en los días y horarios de la prueba, la que difiere de año a año entre comunas y escuelas, y que no guarda relación con los niveles de contaminación presentes durante el año.

Los resultados indican que las variables medioambientales tienen impactos significativos sobre los resultados académicos medidos por las pruebas del SIMCE. En específico, nuestros modelos preferidos (con efectos fijos al nivel de la escuela) sugieren que un incremento en los niveles de contaminación de PM2.5 a la hora de la prueba tiene efectos negativos sobre el puntaje promedio obtenido en matemáticas. Variables como la vegetación al interior de la escuela y fuera de ella, en tanto, tienen efectos positivos importantes sobre los puntajes, sin embargo, es la vegetación dentro de la escuela la que explica la mayor parte de los resultados. Finalmente, la temperatura no parece tener efectos sobre los puntajes, lo que puede explicarse por la poca variabilidad en esta variable entre comunas y escuelas, nuestra unidad de medición, en especial dado por el escaso número de estaciones de monitoreo con las que se cuenta para la estimación. Modelos adicionales que solo consideran las escuelas más cercanas a las estaciones, y sobre las cuales se tiene mayor certeza en los niveles de temperatura, muestran que esta variable estaría de hecho negativamente correlacionada con los puntajes de matemáticas.

El resto del documento se estructura de la siguiente forma. El capítulo 2 presenta brevemente los datos del SIMCE y la estrategia de identificación econométrica utilizada para estimar el efecto de las variables medioambientales sobre los rendimientos académicos. El capítulo 3 se enfoca en describir las fuentes de datos geoespaciales utilizadas, especialmente en lo que dice relación con su recolección y generación para las características medioambientales. El capítulo 4, en tanto, presenta los resultados principales asociados a los modelos descritos en el Capítulo 2, y el Capítulo 5 concluye.

2. Metodología

El modelo básico para estimar el efecto de los indicadores medioambientales sobre el rendimiento académico de los estudiantes es un modelo de regresión lineal que utiliza el puntaje promedio de matemáticas o lectura en la escuela como variable dependiente y el indicador ambiental de interés como variable independiente, más controles. Se estiman modelos parciales para determinar el efecto de cada una de las variables ambientales por separado sobre los puntajes en las pruebas SIMCE. Para el caso de la contaminación atmosférica, por ejemplo, se estima el siguiente modelo. Sea Y_{ijt} el puntaje obtenido por la escuela i en j (matemáticas, lectura) en el año t , se estima

$$Y_{ijt} = \beta CP_{it} + CA\delta + X\theta + T_t + \varepsilon_{ijt} \quad (1)$$

Donde CP_{it} es la contaminación promedio al momento de la prueba (9-11am) el día en que se rinde la misma, CA es un vector de variables que identifican la contaminación acumulada que registra la escuela, tal como la contaminación del día de la prueba en PM10 y la contaminación registrada en promedio entre los meses de mayo y agosto. Este vector captura el efecto acumulado de la contaminación sobre los rendimientos, lo que permite que β identifique de forma precisa el efecto de la variación en la contaminación durante el horario de la prueba sobre el rendimiento académico. El vector X contiene información sobre el grupo socioeconómico de la escuela, la dependencia, la comuna donde se ubica, y el número máximo de alumnos presentes para rendir cualquiera de las pruebas SIMCE, entre matemáticas y lectura. Esta última variable es un proxy del tamaño regular de la sala. T_t es dummy para cada año considerado, que captura variaciones estructurales en los niveles de contaminación y en los puntajes de las pruebas, que es común para todas las escuelas.

El problema que surge al estimar el modelo presentado en la ecuación (1) es que es posible que los estudiantes (o sus padres) elijan escuelas en función no solo de variables educativas (búsqueda de mejores colegios, mejores profesores, etc.) sino que también en función de variables medioambientales. De esta forma, si los padres eligen matricular a sus hijos en escuelas ubicadas en sectores de la ciudad que tengan mejor calidad del aire o más áreas verdes, por ejemplo, entonces la estimación del efecto de las variables medioambientales podría encontrarse sesgada. Esto es especialmente cierto si los padres que tienen más información y recursos para elegir, también tienen más altos ingresos, son más educados y poseen un mayor capital cultural al interior del hogar. Si bien esto coincide con el grupo socioeconómico al que pertenece la escuela, la comuna en la que se encuentra y con la dependencia de la misma, todo lo cual está considerado en la regresión, si la selección ocurre al interior de grupos socioeconómicos, dependencias y comunas, entonces los resultados siguen estando sesgados.

Más aún y pese a que la estimación se basa en el efecto que tiene la contaminación a la hora de la prueba sobre los rendimientos académicos de los estudiantes, si esta variación en la contaminación horaria está correlacionada con la contaminación acumulada observada y capturada por el vector CA , entonces la inclusión de dicha variable no corrige el sesgo originalmente encontrado.

Una alternativa para resolver este problema es utilizar efectos fijos al nivel de la escuela. La utilización de efectos fijos produce que solo variaciones en los niveles de contaminación, y por ende en los puntajes, al nivel de la escuela, en especial durante el horario en que se toma la prueba, tengan impacto. Esta situación es efectivamente exógena a la decisión de selección de escuela por parte de los padres. Esto es, si bien los padres pueden elegir escuelas ubicadas en zonas con menores niveles de contaminación, es poco probable que tomen dicha decisión en función de movimientos horarios en la contaminación para un día específico del año. Más aún, como la prueba se toma a la misma hora para todas las escuelas del país, es también exógeno para los tomadores de decisión, quienes no tienen información que les permita anticipar las variaciones horarias en los niveles de contaminación para cada escuela. De esta forma, el modelo de efectos fijos se estima tal como

$$Y_{ijt} = \beta CP_{it} + CA\delta + \phi_i + T_t + \varepsilon_{ijt} \quad (2)$$

Es importante notar que en el modelo presentado en la ecuación (2) se excluye el vector X que contiene variables que no varían en el tiempo, tal como la comuna en la que se encuentra la escuela, pero que también incluye variables que en principio pueden variar, tales como la dependencia o el grupo socioeconómico. Estas últimas variables se excluyen principalmente porque la proporción de

escuelas para las que estas variables varían en el tiempo es menor y produce que la identificación se realice sólo para aquellas escuelas donde sí hay variación. De cualquier forma, modelos donde sí se incluyen las variables de dependencia y grupo socioeconómico se estiman, sin variaciones significativas en los resultados.

La identificación causal del efecto de la contaminación sobre los rendimientos académicos, entonces, proviene del hecho de que la variación puntual en la contaminación atmosférica es exógena y aleatoria. Si bien los padres pueden anticipar las condiciones ambientales (y de contaminación atmosférica especialmente) que resultan estables en el tiempo (razón por la cual se controla por la contaminación acumulada, por ejemplo) ellos no pueden anticipar la variación puntual de la contaminación al momento de la prueba SIMCE en un día determinado. Más aún, como se muestra en la Figura 2, la contaminación acumulada y la puntual no se encuentran correlacionadas, observándose incluso y contrario a lo que ocurre con la contaminación acumulada, mejores condiciones ambientales en el sector poniente del Gran Santiago a las 9-11 am los días de toma de la Prueba SIMCE. Es particularmente interesante notar también que no existe una correlación exacta tampoco entre la contaminación del día 1 y la del día 2 de toma de prueba, reforzando nuestra hipótesis de aleatoriedad en la contaminación atmosférica durante el horario de las pruebas SIMCE.

En resumen, dado lo exógeno y aleatorio de los movimientos del material particulado durante el horario de la prueba, el hecho de que se controla por la contaminación acumulada, y la no correlación entre la contaminación acumulada y la puntual, consideramos que nuestro modelo reúne las condiciones de un cuasi-experimento que permite identificar el efecto causal de la contaminación sobre los rendimientos académicos de los estudiantes del Gran Santiago. Desafortunadamente no podemos hacer esta misma aseveración para el caso de las otras variables medioambientales consideradas, para las cuales no contamos con variaciones exógenas en su stock en momentos dados.

Un detalle adicional que debe ser considerado en los modelos previos es la naturaleza de los datos de puntaje del SIMCE. De acuerdo a lo informado por la Agencia de la Calidad de la Educación, los puntajes obtenidos en una prueba dada sólo son comparables con los obtenidos en la prueba previa o siguiente, pero las pruebas previa y siguiente no son comparables entre sí. Es decir, por ejemplo, los resultados obtenidos en una prueba el año 2015 son solo comparables con los puntajes de la misma prueba el año 2014 o bien con los del año 2016, pero los puntajes del año 2014 no son comparables con los del 2016 por sí mismos. Esto puede tener efectos negativos importantes en las estimaciones de efectos fijos, por cuanto para que esta estrategia tenga validez se requiere que las variaciones interanuales de los puntajes de las pruebas del SIMCE capturen efectivamente variaciones en los rendimientos académicos de los estudiantes.

Es importante notar que, en el Gran Santiago se cuentan con solo 10 estaciones de monitoreo, las que en ocasiones pueden encontrarse muy lejos de las escuelas para las cuales se estiman niveles de contaminación. Además, los valores de los indicadores de contaminación atmosférica provienen de modelos geostatísticos cuya precisión depende del número de muestras, por lo que sus estimaciones tienden a ser menos precisas en aquellos lugares donde existe una densidad menor de monitores. Por esta razón, se estiman modelos adicionales considerando sólo aquellas escuelas que se encuentran localizadas en un radio de 500, 1000, 1500, 2000 y 3000 metros alrededor de la estación de monitoreo de contaminación. Por parsimonia se presentan sólo los resultados para el buffer más cercano a la estación de monitoreo.

Como ha sido notado previamente, el modelo presentado en la ecuación (2) incorpora tanto el efecto acumulado de la contaminación, bajo los supuestos fuertes de que los y las estudiantes no cambian de comuna o colegio a lo largo del tiempo y de que viven en la misma comuna del colegio al que asisten, como el efecto puntual en el momento de la prueba. Como los supuestos mencionados son difíciles de testear a menos que se cuente con datos de panel al nivel del estudiante por un período considerable, la principal variable de interés es la contaminación a la hora de la prueba, pero sin embargo utilizamos controles por contaminación anual.

Sea $t = 0$ definido como el primer período de vida de un estudiante. Como el aprendizaje es un proceso acumulativo, en el siguiente período ($t = 0$), el aprendizaje de ese período dependerá del aprendizaje en $t = 0$, que está condicionada por la contaminación a la que fue expuesto el estudiante en dicho período, y también de la contaminación en $t = 1$. Así, en período t , el aprendizaje dependerá de la exposición a la contaminación acumulada y la contaminación en el período t . Esto es,

$$L_t = L(L_{t-1}, p_t) = L(p_0, \dots, p_{t-1}, p_t) = L(S_p, p_t) \quad (3)$$

Donde L_t es el aprendizaje total en el tiempo t , p_t es la contaminación en el tiempo t y S_p es la contaminación acumulada desde $t = 0$ hasta $t - 1$, y es fija para cada individuo.

Dado que la contaminación en t probablemente se correlaciona positivamente con la contaminación acumulada en $t - 1$, el modelo empírico (1) de corte transversal que incluye solo contaminación en t estará sobreestimando el efecto de la contaminación sobre el aprendizaje. Sin embargo, el modelo de datos de panel (2) asumimos que el efecto fijo de la escuela considera, entre otros factores, la exposición acumulada de los estudiantes de la escuela a la contaminación, y, por lo tanto, el modelo representado por la ecuación (2) constituye nuestro modelo preferido.

Cabe notar, sin embargo, que si bien en términos teóricos tanto la contaminación acumulada como la puntual debieran producir efectos negativos sobre los aprendizajes, como la contaminación acumulada es observable por los padres y niños, y como los datos con los que se cuenta se encuentra solo al nivel de la escuela, es posible que su efecto no sea posible de capturar en nuestros resultados.

Por último, un modelo equivalente a (2) se estima usando cobertura de vegetación dentro de la escuela, fuera de la escuela, ambas, y temperatura superficial. Adicionalmente, se los resultados principales consideran todas las variables medioambientales de forma conjunta, para evaluar así su efecto adicional sobre los rendimientos académicos.

3. Datos

3.1. Georeferenciación y Poligonización Establecimientos Educativos

Se utilizó la base de datos “Establecimientos de Educación Escolar”, disponibles en el portal de Información Territorial del Gobierno de Chile (www.ide.cl) para obtener los datos georeferenciados de los establecimientos de educación Escolar existentes en Chile. Estos datos fueron luego incorporados a una plataforma de información geográfica (QGIS 3.12) y filtrados para obtener únicamente los datos georeferenciados de los colegios pertenecientes al área urbana de las 34 comunas que forman el conglomerado urbano conocido como el Gran Santiago. Estos datos fueron luego sujeto de un segundo filtro con el fin de dejar en la base de datos únicamente aquellos

establecimientos educacionales que contarán con reportes de los resultados del SIMCE de matemáticas y lenguaje para el período cubierto por el estudio (2010-2018).

Una vez filtrados los datos georreferenciados de los establecimientos educacionales se procedió a hacer una inspección visual en Google Earth Pro (V. 7.3.3) para cerciorarse de que cada una de las coordenadas geográficas reportadas efectivamente representarán los establecimientos educacionales detallados en la base de datos. En caso de haber errores en las coordenadas geográficas para un establecimiento educacional, se procedió a buscar en la base de datos o en la ficha de cada colegio la dirección exacta, y con ello realizar una nueva georreferenciación de dichos establecimientos a sus coordenadas correctas. Luego de este proceso se obtuvo una base de datos georreferenciada con 1,208 establecimientos educacionales para el Gran Santiago (Figura 1).

Para la poligonización de los establecimientos educacionales se procedió a generar una capa en formato shape (.shp) que contiene 1,208 polígonos, correspondientes a cada uno de los establecimientos educacionales identificados anteriormente. Para ello se hizo una fotointerpretación de imágenes satelitales en Google Earth Pro, de manera de generar un polígono del perímetro de cada establecimiento educacional georreferenciado. En caso de existir dudas respecto al perímetro de algún establecimiento, se complementó el trabajo de fotointerpretación con imágenes “Google Street View” disponibles en los servidores de mapas e imágenes de Google. Los polígonos fueron generados en Google Earth en formato .kml y luego transformados en formato .shp en QGIS.

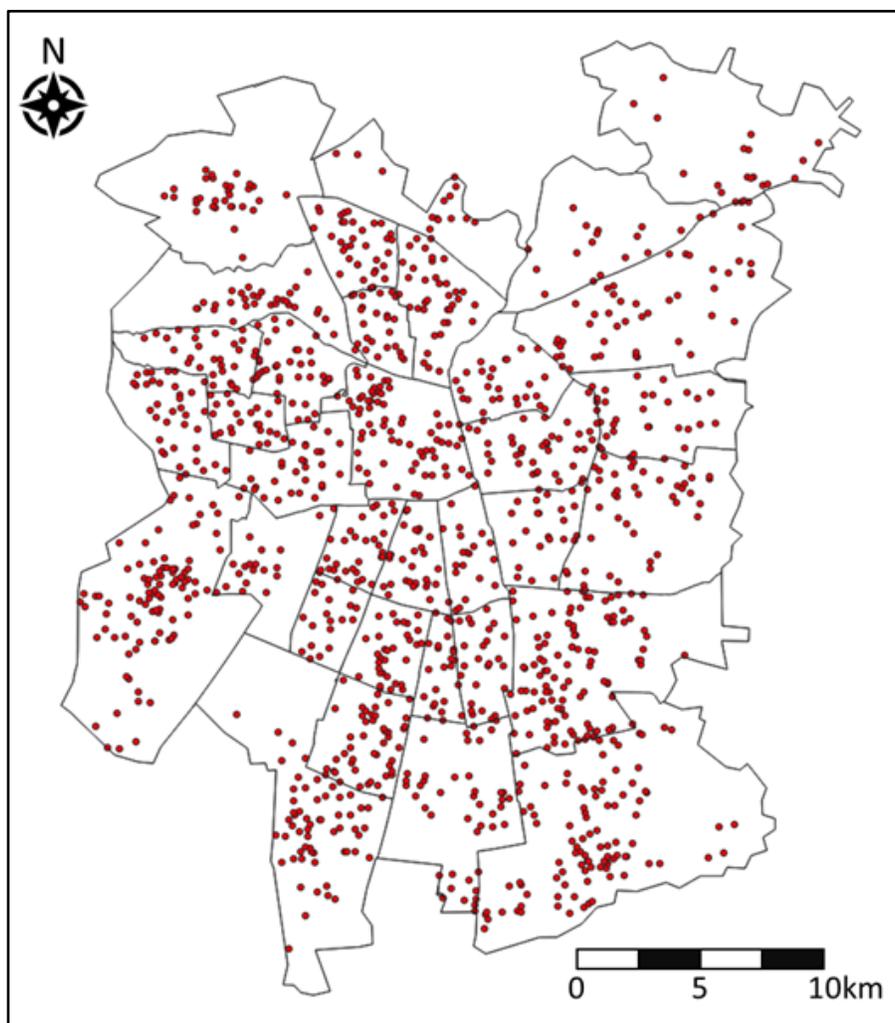


Figura 1: Distribución de los 1208 establecimientos educacionales utilizados para el estudio. Las líneas grises denotan el área urbana de las 34 municipalidades que conforman el Gran Santiago.

3.2. Estimaciones de Material Particulado en Establecimientos Educacionales

Los datos de material particulado para la ciudad de Santiago se obtuvieron del Sistema de Información Nacional de Calidad del Aire (SINCA) del Ministerio de Medio Ambiente del Gobierno de Chile, disponible en línea a través del enlace <https://sinca.mma.gob.cl/index.php>. Esta base de datos cuenta con una serie de contaminantes atmosféricos para 10 estaciones de monitoreo que se encuentran distribuidos en la ciudad de Santiago. De esta base de datos se obtuvieron los datos de registro horario para el material particulado grueso (<10 micrones de diámetro, PM10) y para el material particulado fino (<2.5 micrones de diámetro, PM2.5) para las 10 estaciones de monitoreo antes señaladas. En base a la información validada disponible en el SINCA, se calculó para cada estación de monitoreo los siguientes indicadores de contaminación atmosférica: (1) Promedio de PM2.5 entre 9-11am (horario de la prueba SIMCE) para los días en que se tomó la prueba, (2) promedio diario de PM10 para los días de la prueba, (3) promedio de PM10 para la temporada de episodios críticos (01 mayo al 31 de agosto), (4) promedio de PM10 anual, y (5) porcentaje de horas anuales en que las estaciones

superaron la concentración horaria de $100 \mu\text{g}/\text{m}^3$ de PM10. Debido a la restringida disponibilidad de lecturas validadas para PM2.5, el cálculo de los datos de material particulado se limitó al período 2010-2013 y 2015-2018.

Una vez calculados los indicadores de contaminación atmosférica descritos, se procedió a generar los modelos de distribución espacial de estos indicadores para toda el área urbana de Santiago. Para decidir el mejor modelo a utilizar para realizar la interpolación de los datos, se procedió a generar para cada indicador/año el respectivo semivariograma (Olea, 1999) de manera de evaluar la estructura espacial de los datos y determinar la mejor aproximación geoestadística para generar los modelos. Junto con ello, también se evaluó el nivel de correlación espacial con otras cuatro variables ambientales (i.e. temperatura superficial, índice NDBI de superficie construida, índice NDVI de vegetación, elevación/altura) para identificar potenciales variables explicativas que podrían ser incorporadas en los modelos.

En aquellos casos en que se encontraron indicios de autocorrelación espacial (i.e. PM10 promedio temporada episodios críticos, PM10 promedio anual, PM10 porcentaje hora sobre $100\mu\text{g}/\text{m}^3$) se utilizó el procedimiento de “Regression Kriging” (Hengl et al, 2007) utilizando un modelo de elevación digital como variable regresora. Mientras que en aquellos casos en que no se encontró una estructura de autocorrelación espacial (i.e. PM2.5 entre 9-11 am y PM10 promedio diario para los días de la prueba) se procedió a utilizar una interpolación “B-spline” multinivel (Lee et al, 1997). En ambos casos los modelos geoestadísticos se construyeron en base a los algoritmos del paquete SAGA disponibles en QGIS. Todos los modelos fueron generados con una resolución de 100 metros/pixel. En la figura 2 se muestra un ejemplo gráfico de los modelos de contaminación generados por estas técnicas. Una vez construidos los modelos de distribución espacial de contaminantes atmosféricos para cada uno de los indicadores/año, se procedió a asignar dichos valores a cada uno de los polígonos de las escuelas generados en el paso anterior utilizando para ello el comando “zonal statistics” (opción mean) del paquete GDAL disponible en QGIS.

3.3. Estimaciones de Cobertura Vegetacional en Establecimientos Educativos

Para estimar la cobertura vegetacional se utilizaron imágenes satelitales Landsat para calcular el “Soil Adjusted Vegetation Index” (SAVI) (Huete, 1988), el cual es un indicador de vegetación considerado útil para estimar la cobertura vegetacional en áreas con superficies contrastantes, como el área urbana de Santiago (Fernández et al, 2018). Para ello se buscaron imágenes provenientes de la familia de satélites Landsat, que poseían 0% de cobertura de nubes para la ciudad de Santiago, y cuya fecha de captura fuera lo más cercana a las fechas en que se tomó la prueba SIMCE cada año (Tabla 1). Una vez identificadas y obtenidas las imágenes, se procedió a calcular el SAVI para toda el área de Santiago, y luego esta información asignada a cada uno de los establecimientos a través del mismo proceso usado para asignar los valores de contaminación atmosférica. En base a los mismos datos de vegetación, se procedió a calcular el SAVI para el entorno de las escuelas, tomando para ello buffers de 100, 200 y 500 metros de distancia en torno a cada escuela.

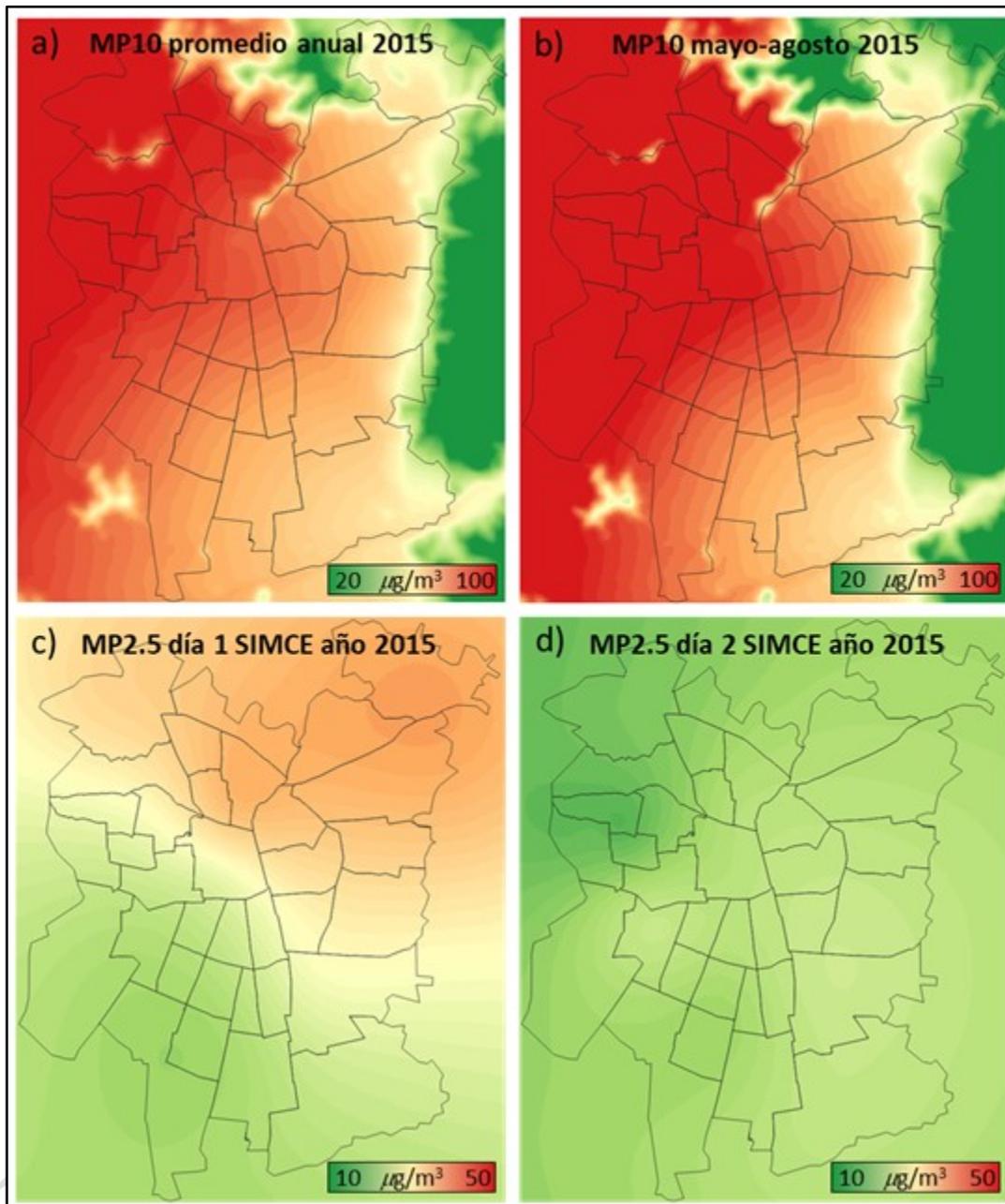


Figura 2. Ejemplo de los modelos de contaminación del aire por material particulado generados para este estudio. a) y b) Modelos de contaminación promedio anual y promedio período mayo-agosto por MP10, respectivamente. Ambos modelos construidos en base a la técnica “regression kriging”. c) y d) Modelos de contaminación promedio por MP2.5 para el horario en que se tomaron las pruebas SIMCE de matemáticas y lectura durante el año 2015, respectivamente. Ambos modelos construidos en base a la técnica “multilevel b-spline interpolation”.

3.4 Estimaciones de Temperatura Superficial en Establecimientos Educativos

Los modelos de distribución de temperatura para el horario de la prueba se construyeron integrando los datos de temperatura reportados por las 10 estaciones meteorológicas asociadas a las estaciones de contaminación atmosférica del SINCA, con la temperatura superficial obtenida de la banda térmica de las imágenes satelitales utilizadas para estimar la cobertura vegetal. Debido a que la fecha de las imágenes satelitales no coincide con la fecha en la que se tomaron las pruebas (Tabla 1) y que las imágenes fueron tomadas alrededor de las 2pm hora local, no es posible asignar directamente los valores de temperatura superficial obtenidos por el satélite a los establecimientos educativos durante el horario de la prueba. Para resolver este problema, se decidió utilizar la información de temperatura obtenida de las imágenes satelitales como la variable que diera cuenta de la distribución espacial de la temperatura, utilizando los datos de temperatura proveniente de las estaciones meteorológicas para calibrar el valor de las temperaturas en el día y horario de la prueba. De esta forma, para cada año se generó un modelo de temperatura en base a una regresión espacial entre los datos de punto de temperatura de las estaciones meteorológicas y la capa de temperatura superficial, usando esta última como variable predictora. Con ello se generó una nueva capa de temperatura superficial, pero ajustada a los valores de temperatura estimada para el día y horario de la prueba. En la figura 3 se muestra un ejemplo gráfico de modelos de temperatura construidos para el día 1 y 2 de la prueba SIMCE del año 2016.

Tabla 1. Fechas prueba SIMCE y de imágenes satelitales para cálculo de SAVI

Año	SIMCE		Fecha Imágenes	Satélite
	Día 1 (Matemáticas)	Día 2 (Lectura)		
2010	19 octubre	20 octubre	14 noviembre	Landsat 5
2011	12 octubre	13 octubre	17 noviembre	Landsat 5
2012	17 octubre	18 octubre	10 octubre	Landsat 7
2013	08 octubre	09 octubre	05 octubre	Landsat 8
2015	06 octubre	07 octubre	11 octubre	Landsat 8
2016	15 noviembre	16 noviembre	14 noviembre	Landsat 8
2017	07 noviembre	08 noviembre	17 noviembre	Landsat 8
2018	06 noviembre	07 noviembre	04 noviembre	Landsat 8

Nota: 2018 no está considerada en los modelos. Fuente: elaboración propia en base a información del SIMCE.

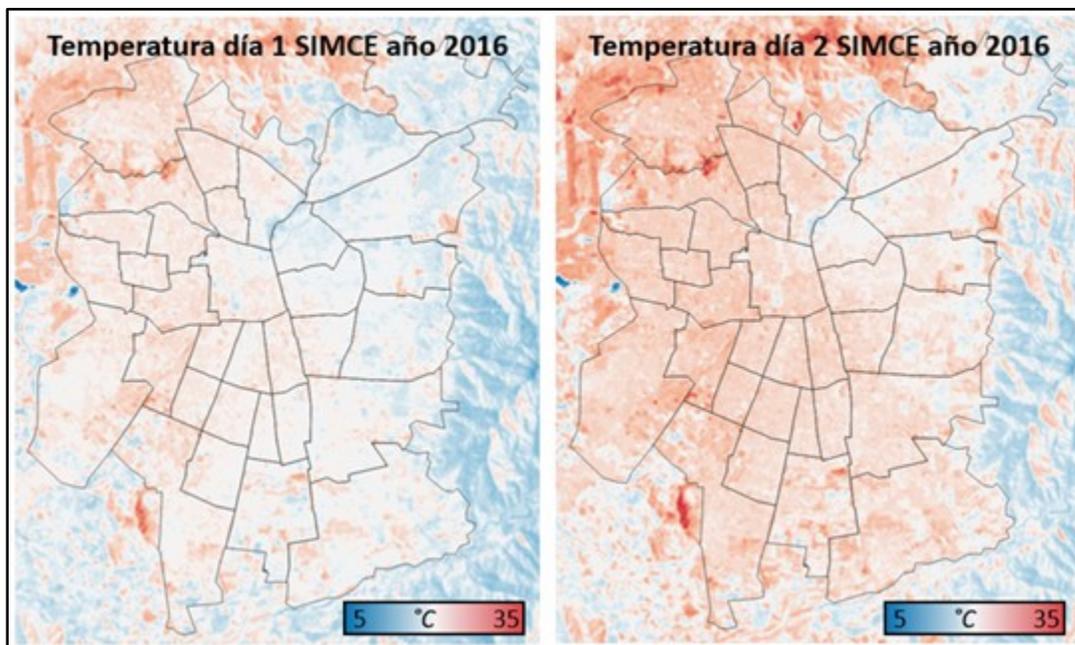


Figura 3. Ejemplo de los modelos de temperatura generados para el estudio. Los modelos muestran la distribución espacial de la temperatura promedio estimada para el horario en que se rindió la prueba SIMCE durante el año 2016..

3.5 Set de Datos para Realizar los Análisis

El SIMCE es una prueba tomada en el mes de octubre o noviembre de cada año y evalúa los conocimientos que los estudiantes tienen respecto a los contenidos curriculares mínimos establecidos por el Ministerio de Educación de Chile para cada nivel. En adición a los resultados de las pruebas de matemáticas y lectura, la base de datos del SIMCE entrega información de caracterización de las escuelas para cada escuela, variables que permiten controlar por aspectos socioeconómicos, la dependencia (municipal, particular subvencionado, particular pagado), de localización (comuna, región, provincia), el tamaño de la sala, entre otros.

Si bien la información del SIMCE está disponible y de acceso libre para todos los años entre 1988 y 2019, los datos de contaminación diaria por PM2.5 para los días de la prueba, en los cuales nos centramos, no estuvieron disponibles para todas las estaciones de monitoreo el año 2014, por lo que tuvimos que excluir este año del análisis. Por otra parte, el SIMCE de 2019 fue solo aplicado a 8vo básico, razón por la cual no es posible de utilizar en nuestro estudio que solo considera información para 4tos básicos. Adicionalmente, no fue posible incorporar información de temperaturas para el año 2017, ya que también hubo estaciones de monitoreo que no presentaron información para los días de la prueba. De esta forma, para el estudio se conforma un panel balanceado de escuelas con información completa de contaminación por PM2.5, PM10, vegetación, temperaturas, y resultados del SIMCE para los años 2010 a 2018, pero excluyendo los años 2014 y 2017. Así, la base final consta con información para 1,208 escuelas observadas en 7 años, es decir un total de 8,456 observaciones.

3. Resultados

Se presentan varios sets de resultados separados. Para todas ellas la variable dependiente es (separadamente) el puntaje obtenido en las pruebas de matemáticas y de lectura, considerando modelos OLS y Efectos Fijos (EF). La Tabla 2 presenta los resultados para el efecto de la contaminación durante el horario de la prueba sobre los rendimientos académicos de los estudiantes, controlando por contaminación horaria, diaria y acumulada, además de variables socioeconómicas.

Tabla 2. Efecto de la contaminación sobre el puntaje del SIMCE

	Matemáticas		Lectura	
	OLS	EF	OLS	EF
PM2.5 9-11am	-0.154*** (0.0449)	-0.0811* (0.0427)	0.0228 (0.0413)	-0.0236 (0.0383)
PM10 Día de la prueba	0.106*** (0.0319)	0.0396 (0.0324)	-0.00443 (0.0283)	-0.00383 (0.0232)
PM10 Mayo-Agosto	0.00337 (0.0335)	0.0461* (0.0242)	-0.0158 (0.0279)	0.0227 (0.0181)
Número de alumnos	0.188*** (0.0153)	-0.114*** (0.0141)	0.132*** (0.0118)	-0.0765*** (0.0128)
Corporación Municipal	-2.264 (1.844)		-1.304 (1.424)	
Particular Subvencionado	4.394*** (1.343)		5.049*** (1.027)	
Particular Pagado	4.613 (3.555)		5.246* (3.021)	
GSE. Medio Bajo	11.52*** (1.404)		12.29*** (1.182)	
GSE. Medio	26.36*** (1.665)		27.18*** (1.401)	
GSE. Medio Alto	43.57*** (2.067)		43.01*** (1.709)	
GSE. Alto	58.23*** (3.308)		55.25*** (2.809)	
Intercepto	206.7*** (4.293)	255.1*** (2.432)	232.6*** (3.542)	272.5*** (1.671)
Observaciones	8,456	8,456	8,456	8,456
R2	0.549	0.039	0.572	0.057
Número de escuelas		1,208		1,208

Nota: incluye controles por año y comuna (OLS). Errores estándar clusterizados al nivel de la escuela (OLS) y de la comuna (EF) se reportan en paréntesis debajo de cada coeficiente. *** p<0.01, ** p<0.05, * p<0.1

Considerando los modelos OLS, los resultados presentados en la Tabla 2 muestran que un incremento en la contaminación a la hora de la prueba tiene un impacto negativo y significativo sobre el puntaje

de las pruebas de matemáticas. En particular, cada incremento de 1 microgramo de PM2.5 por metro cúbico genera una reducción de 0.154 en el puntaje obtenido en matemáticas y no tiene efectos para lectura. Si se considera que en promedio el PM2.5 presenta una variación de más de 20 gramos por metro cúbico en el período observado, lo anterior equivale a una diferencia de más de 2 puntos en el puntaje obtenido en un año cualquiera. En el año 2012 o 2013, sin embargo, la diferencia en la concentración de PM2.5 a la hora de la prueba fue de 34 puntos, lo que supone que las escuelas con mayores niveles de contaminación tuvieron una diferencia de casi 4 puntos en su puntaje respecto a aquellas ubicadas en sectores menos contaminados, los que serían exclusivamente atribuidos a variaciones puntuales en los niveles de contaminación a la hora de la prueba.

Parte del efecto anterior se pierde cuando se consideran los resultados utilizando efectos fijos. Si bien en ambos casos, tanto para matemáticas como para lectura, el coeficiente es negativo, este nuevamente solo resulta significativo para matemáticas. Como es de esperar, los resultados estimados bajo el modelo de efectos fijos reducen el tamaño del coeficiente, pasando, en valor absoluto, de 0.154 a 0.0811 en el caso de matemáticas, y de 0.0228 a 0.0236 con cambio de signo en lectura.

Resulta importante notar que las medidas de contaminación acumulada resultan significativas de forma casi exclusiva para el caso de matemáticas, pero de forma interesante estas estarían positivamente correlacionadas con los rendimientos. Es importante notar, sin embargo, que, al tratarse de datos acumulados, estas pueden estar reflejando el hecho de que comunas de altos ingresos, como Vitacura, Ñuñoa, Providencia, Las Condes y Lo Barnechea, entre otras, concentran niveles relativamente altos de contaminación acumulada anual, muy por sobre los niveles de Cerrillos, Cerro Navia, Lo Prado y otras comunas donde se concentran niveles socioeconómicos más bajos.³

Esto es relevante por cuanto indicaría que las variables socioeconómicas incluidas, tales como grupo socioeconómico, dependencia y comuna entre otras, efectivamente capturan la disparidad que existe entre los niveles de ingreso y las condiciones medioambientales de largo plazo. En cambio, ellas no logran capturar variaciones puntuales y acotadas en los niveles de contaminación, sugiriendo que la estrategia de identificación es correcta.

Finalmente, hay pocas diferencias en los puntajes en función de la dependencia, con excepción de la diferencia observada entre las escuelas particulares subvencionadas en la prueba de lectura (OLS). Probablemente la razón para la poca diferencia encontrada en esta variable se deba a que la mayor parte de la variación es capturada por el grupo socioeconómico. Para esta última variable las diferencias son importantes. Cada grupo socioeconómico tiene mejores resultados que el GSE que lo precede, con diferencias significativas y positivas en todos los grupos. A modo de ejemplo, las escuelas que pertenecen al grupo socioeconómico alto obtienen 55 puntos más en promedio por sobre aquellas del grupo socioeconómico bajo. Esta diferencia es de casi 58 puntos para el caso de la prueba de matemáticas.

Tabla 3 muestra los resultados para el efecto de la vegetación al interior de las escuelas sobre los rendimientos académicos. La vegetación, medida por el índice SAVI, tiene un efecto positivo y

³ Modelos adicionales que solo incluyen la contaminación acumulada, sin la variación puntual al momento de la prueba, no presentan resultados significativos. Esto es esperable por cuanto el efecto acumulado, toda vez que es persistente en el tiempo, debiera ser capturado por el uso de efectos fijos ya sea al nivel de la comuna o de la misma escuela. De igual forma, utilizar otras medidas de contaminación acumulada, como el porcentaje de horas al año que se supera la norma, o la contaminación anual total entregan los mismos resultados que la medida de la contaminación en los meses de mayo a agosto. Se prefirió esta última medida, por cuando captura el fenómeno en los meses de mayor contaminación.

significativo para ambas pruebas y en los dos tipos de modelos considerados. De forma interesante y a diferencia de lo que se observa para contaminación, cambios en los índices vegetacionales al interior de las escuelas tienen efectos aún más pronunciados cuando se consideran las especificidades de las escuelas (modelo de efectos fijos) que lo que predeciría el modelo OLS. Por ejemplo, un incremento de una unidad en el índice SAVI está asociado con un incremento de 14.5 puntos en el puntaje promedio obtenido en matemáticas bajo el modelo de efectos fijos. Es decir, cada unidad de variación en el índice está asociado con un puntaje promedio en matemáticas 14.5 puntos más alto, pese a controlar por todos aquellos aspectos fijos de la escuela (calidad del alumnado, de los profesores, infraestructura, ingreso de los padres, localización, etc.). Para el caso de lectura este efecto llega a casi 11 puntos en el modelo de efectos fijos, más de 2 puntos más que lo que se estima en el modelo OLS.

Tabla 3. Efecto de las áreas verdes al interior de las escuelas sobre el puntaje del SIMCE

	Matemáticas		Lectura	
	OLS	EF	OLS	EF
Vegetación (SAVI)	13.56** (5.595)	14.50*** (4.634)	8.220* (4.577)	10.62* (5.265)
Número de alumnos	0.189*** (0.0155)	-0.112*** (0.0140)	0.133*** (0.0120)	-0.0742*** (0.0125)
Corporación Municipal	-2.267 (1.844)		-1.317 (1.424)	
Particular Subvencionado	4.556*** (1.348)		5.129*** (1.032)	
Particular Pagado	4.398 (3.550)		5.134* (3.020)	
GSE. Medio Bajo	11.52*** (1.407)		12.31*** (1.185)	
GSE. Medio	26.29*** (1.668)		27.18*** (1.404)	
GSE. Medio Alto	43.34*** (2.082)		42.91*** (1.720)	
GSE. Alto	57.66*** (3.283)		54.94*** (2.798)	
Intercepto	209.7*** (2.918)	258.1*** (1.018)	231.1*** (2.428)	272.8*** (0.901)
Observaciones	8,456	8,456	8,456	8,456
R2	0.550	0.040	0.572	0.057
Número de escuelas		1,208		1,208

Nota: incluye controles por año y comuna (OLS). Errores estándar clusterizados al nivel de la escuela (OLS) y de la comuna (EF) se reportan en paréntesis debajo de cada coeficiente. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

El resto de los coeficientes, asociados a los grupos socioeconómicos, la dependencia, el puntaje en la otra prueba y el número de alumnos por sala se mantienen en los mismos niveles previamente

reportados en la Tabla 2. Resulta interesante notar que el número de alumnos en la sala resulta estar negativamente correlacionado con los puntajes obtenidos en las pruebas, lo que es altamente consistente con la literatura.

Tabla 4. Efecto de las áreas verdes en el barrio de las escuelas (buffer 500m) sobre el puntaje del SIMCE

	Matemáticas		Lectura	
	OLS	EF	OLS	EF
Vegetación (SAVI)	19.35*** (6.734)	13.93** (5.174)	16.16*** (5.444)	3.055 (5.148)
Número de alumnos	0.190*** (0.0155)	-0.113*** (0.0138)	0.134*** (0.0120)	-0.0758*** (0.0124)
Corporación Municipal	-2.061 (1.842)		-1.150 (1.423)	
Particular Subvencionado	4.746*** (1.341)		5.320*** (1.025)	
Particular Pagado	4.864 (3.531)		5.473* (3.016)	
GSE. Medio Bajo	11.19*** (1.420)		12.03*** (1.186)	
GSE. Medio	25.69*** (1.701)		26.67*** (1.418)	
GSE. Medio Alto	42.58*** (2.121)		42.23*** (1.739)	
GSE. Alto	56.50*** (3.350)		53.85*** (2.841)	
Intercepto	209.5*** (2.912)	257.6*** (1.042)	230.8*** (2.431)	273.4*** (0.879)
Observaciones	8,456	8,456	8,456	8,456
R2	0.550	0.039	0.573	0.057
Número de escuelas		1,208		1,208

Nota: incluye controles por año y comuna (OLS). Errores estándar clusterizados al nivel de la escuela (OLS) y de la comuna (EF) se reportan en paréntesis debajo de cada coeficiente. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Modelos adicionales consideran la vegetación fuera de la escuela en buffers de 100, 200 y 500 metros alrededor de la escuela (excluyendo a la misma) a fin de evaluar el efecto que las áreas verdes circundantes pueden tener sobre los rendimientos académicos de los estudiantes de 4to básico. En línea con lo esperado, la vegetación fuera de las escuelas (en el barrio) está positivamente correlacionada con los rendimientos en la prueba SIMCE, especialmente en matemáticas. A modo de ejemplo, un incremento de una unidad en el índice SAVI en un área de 500m alrededor de la escuela está asociada con casi 14 puntos adicionales de puntaje en la prueba de matemáticas, pero no tiene efectos sobre los resultados de la prueba de lectura, al menos bajo el modelo de efectos fijos. Los modelos OLS presentan resultados aún mayores (19.35 para matemática y 16.16 para lectura) y en

ambos casos altamente significativos. Los resultados asociados a vegetación fuera de la escuela están presentados en la Tabla 4.⁴

Tabla 5. Efecto de la temperatura superficial en las escuelas sobre el puntaje del SIMCE

	Matemáticas		Lectura	
	OLS	EF	OLS	EF
Temperatura el día de la prueba	0.238 (0.435)	-0.0992 (0.309)	0.270 (0.388)	0.207 (0.322)
Número de alumnos	0.187*** (0.0154)	-0.115*** (0.0140)	0.131*** (0.0119)	-0.0764*** (0.0127)
Corporación Municipal	-2.247 (1.844)		-1.306 (1.424)	
Particular Subvencionado	4.397*** (1.342)		5.025*** (1.028)	
Particular Pagado	4.533 (3.567)		5.180* (3.029)	
GSE. Medio Bajo	11.52*** (1.402)		12.31*** (1.181)	
GSE. Medio	26.32*** (1.663)		27.19*** (1.400)	
GSE. Medio Alto	43.52*** (2.066)		43.01*** (1.708)	
GSE. Alto	58.23*** (3.308)		55.30*** (2.806)	
Intercepto	205.0*** (9.845)	261.6*** (6.819)	226.9*** (6.862)	270.5*** (5.268)
Observaciones	8,456	8,456	8,456	8,456
R2	0.549	0.038	0.572	0.057
Número de escuelas		1,208		1,208

Nota: incluye controles por año y comuna (OLS). Errores estándar clusterizados al nivel de la escuela (OLS) y de la comuna (EF) se reportan en paréntesis debajo de cada coeficiente. *** p<0.01, ** p<0.05, * p<0.1

La Tabla 5 muestra el efecto de la temperatura sobre los rendimientos académicos. A diferencia de lo observado para contaminación y para vegetación, los primeros resultados asociados a la temperatura parecieran no tener un efecto significativo sobre los puntajes obtenidos en el SIMCE. Si bien en matemáticas, un incremento en la temperatura se encuentra asociado con disminuciones en los puntajes promedios del SIMCE, este efecto es estimado de forma imprecisa y su efecto es no

⁴ Modelos para otras distancias fueron también estimados, con resultados similares a los presentados acá. De igual forma, se corrieron modelos adicionales utilizando la variable NDVI en vez de SAVI para estimar el efecto de la vegetación dentro y fuera de la escuela sobre los puntajes del SIMCE. Adicionalmente, se consideró el uso de un indicador de porcentaje del suelo que se encuentra construido, como una proxy para la falta de áreas verdes. Todos estos resultados son consistentes con aquellos mostrados por el índice SAVI y no presentan mayores variaciones relevantes ni en la magnitud ni dirección de los efectos estimados. Por parsimonia, estos resultados no son presentados acá, pero están disponibles previa solicitud a los autores.

significativo. Al igual como sucede de forma previa y como resulta esperable, colegios de grupos socioeconómicos más altos presentan mejores resultados y el número de alumnos en la sala de clases afecta negativamente los resultados.

Un elemento importante a considerar cuando se analizan los resultados de la Tabla 5 para temperatura (y de alguna manera también para el caso de la contaminación en la Tabla 2) es que, para asignar información medioambiental a las escuelas, solo se cuenta con 10 estaciones de monitoreo en el Gran Santiago. Esto lógicamente afecta la variabilidad de las estimaciones entre escuelas, pudiendo aminorar el real efecto de la temperatura (y de la contaminación) sobre los puntajes. Para evaluar si este efecto está presente y afectando los resultados presentados, la Tabla 7 repite la información final donde se incorporan todas las variables medioambientales de forma simultánea, pero utilizando la información solo de aquellas escuelas que se encuentran en un radio de 500 metros de distancia de la estación de monitoreo más cercana. Esto permite tener más confianza en los datos y por tanto, si bien reduce el número de escuelas posibles de analizar, permitiría evaluar de mejor forma el efecto tanto de la temperatura como potencialmente de la contaminación.

Por último, la Tabla 6 presenta el resumen de todas las variables medioambientales de forma conjunta. Esto es, se busca conocer cuál es el efecto marginal de cada variable medioambiental por separado sobre los puntajes del SIMCE, toda vez que se toman en cuenta todas las otras variables del tipo. La lógica de realizar este modelo es evaluar si algunos de los efectos previamente observados se eliminan al agregar controles medioambientales, que pudieran estar correlacionados entre sí, tal como es el caso con la vegetación dentro y fuera de la escuela, por ejemplo.

De acuerdo a los resultados de la Tabla 6, los efectos se mantienen mayoritariamente inalterados para contaminación. A modo de ejemplo, si en los resultados de la Tabla 2, donde no se controla por otros aspectos medioambientales, el coeficiente asociado al modelo de efectos fijos para matemáticas era -0.0811, ahora este llega a -0.0810 y se mantiene significativo. Esto resulta interesante, por cuanto sugiere que la estrategia de identificación utilizada es correcta, dado que el coeficiente estimado no se modifica al agregar controles. Para el caso de lectura, existe una pequeña caída en el estimado puntual, pero este sigue siendo estadísticamente cero, también reflejando invariabilidad respecto al estimado de la Tabla 2.

Es importante notar que los efectos tanto de la contaminación como de la vegetación son relevantes si se comparan por ejemplo con los efectos del número de estudiantes reportado en el anexo. De hecho, el efecto positivo de la vegetación significativamente más relevante.

El efecto heterogéneo en matemáticas y lenguaje se ha observado en otros estudios que evalúan otros factores en los resultados de aprendizaje, tales como intervenciones tempranas, inteligencia, autoestima, género (Lauermann, et al, 2020). En este caso en particular, como los conceptos numéricos tienen un origen ontogenético y una base neuronal que son distintos e independientes del lenguaje (Gelman, R y Butterworth, 2005), puede ser posible que la contaminación esté afectando de manera heterogénea la corteza cerebral. Sin embargo, esto último es una hipótesis que debe ser estudiada en el futuro, probablemente en el contexto de un estudio interdisciplinario con neurocientíficos.

Para el caso de vegetación dentro de la escuela, el efecto cae ligeramente para matemáticas (de 14.5 a 11.7) y se incrementa levemente para lectura (de 10.6 a 11.4), sin embargo, debe tomarse en cuenta

que los resultados presentados en la Tabla 3 no consideraban la vegetación fuera de la escuela, variable que puede estar altamente correlacionada con la vegetación al interior de las escuelas, y los presentados acá si lo hacen. De manera interesante, la vegetación al interior de las escuelas es positiva y significativa para explicar los rendimientos del SIMCE, pero la vegetación en las afueras de las escuelas no lo es (al menos en los modelos de efectos fijos), lo que permite deducir importantes herramientas de política que son resorte de los sostenedores y que serán mencionadas más adelante.

Por último, la temperatura, tal como fue descrito anteriormente, pareciera no tener impactos sobre los rendimientos de los estudiantes del Gran Santiago. El resto de los indicadores (no presentados acá para ahorrar espacio), en tanto, se mantiene en niveles similares a los informados, con efectos positivos para los grupos socioeconómicos más altos y con mejores resultados para aquellas escuelas con un número de alumnos por sala es más bajo.

Tabla 6. Efecto de las variables medioambientales sobre el puntaje del SIMCE

	Matemáticas		Lectura	
	OLS	EF	OLS	EF
PM2.5 9-11am	-0.148*** (0.0448)	-0.0810* (0.0418)	0.0354 (0.0415)	-0.0191 (0.0389)
PM10 Día de la prueba	0.100*** (0.0317)	0.0365 (0.0315)	-0.0172 (0.0285)	-0.00663 (0.0250)
PM10 Mayo-Agosto	0.0148 (0.0337)	0.0439* (0.0235)	-0.00824 (0.0275)	0.0227 (0.0189)
Vegetación (SAVI)	8.603 (6.012)	11.71** (5.533)	3.676 (4.959)	11.37* (5.692)
Vegetación (SAVI) Buffer 500m	16.75** (7.308)	8.038 (6.607)	15.64*** (5.935)	-2.163 (6.344)
Temperatura el día de la prueba	0.498 (0.443)	0.133 (0.331)	0.488 (0.388)	0.218 (0.355)
Intercepto	194.6*** (11.24)	250.4*** (8.565)	223.8*** (7.848)	268.3*** (6.975)
Observaciones	8,456	8,456	8,456	8,456
R2	0.551	0.041	0.573	0.058
Número de escuelas		1,208		1,208

Nota: incluye controles por GSE, dependencia, número de alumnos, año y comuna (OLS). Errores estándar clusterizados al nivel de la escuela (OLS) y de la comuna (EF) se reportan en paréntesis debajo de cada coeficiente. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Dado que no se cuenta con información precisa de contaminación al nivel de la escuela, esta debió ser construida a través del método ya descrito de forma previa. Lo anterior implica necesariamente que existen errores en la contaminación estimada para cada escuela, el que podría acrecentarse con la distancia de la estación de monitoreo. Para evaluar si nuestros resultados eran, a su vez, el resultado de eventuales errores en la estimación inicial de la contaminación, rehicimos todos nuestros modelos, pero considerando únicamente a aquellas escuelas que se encontraban a una distancia máxima de la

estación de monitoreo. Así, se consideraron modelos adicionales para aquellas escuelas que estaban en un radio de 500, 1000, 1500, 2000 y 3000 metros de la estación de monitoreo.

Lógicamente el acotar el buffer a niveles bajos (ej. 500 metros), reduce de forma importante el tamaño de la muestra y dificulta la estimación, especialmente cuando se utilizan efectos fijos, pero permite mayor variabilidad en la información entre escuelas y estimar con mayor precisión los efectos, tanto para temperatura como para contaminación. Así, por ejemplo, cuando se utiliza el buffer de 500 metros de distancia entre la escuela y la estación de monitoreo, el efecto de la contaminación resulta negativo y significativo nuevamente solo para matemática, sin embargo, su coeficiente se incrementa sustancialmente, llegando a -0.473 (el modelo comparable de las Tablas 2 y 6 solo reporta un coeficiente de -0.081). Para el caso de lectura, este también se incrementa fuertemente, llegando a -0.36, pero se mantiene estadísticamente no significativo.

Tabla 7. Efecto de las variables medioambientales sobre el puntaje del SIMCE. Escuelas en un buffer de 500 metros de la estación de monitoreo

	Matemáticas		Lectura	
	OLS	EF	OLS	EF
PM2.5 9-11am	-0.287 (0.336)	-0.473* (0.262)	-0.302 (0.243)	-0.360 (0.233)
PM10 Día de la prueba	0.0883 (0.156)	0.135 (0.152)	0.110 (0.145)	0.240** (0.102)
PM10 Mayo-Agosto	0.282** (0.130)	0.310** (0.133)	0.189* (0.106)	0.260* (0.135)
Vegetación (SAVI)	58.81 (42.26)	-6.244 (27.40)	41.84 (38.11)	-40.61 (34.12)
Vegetación (SAVI) Buffer 500m	2.105 (35.21)	-5.181 (27.45)	-6.142 (42.14)	-34.58 (40.30)
Temperatura el día de la prueba	-3.352 (2.640)	-3.777* (2.061)	-2.637* (1.415)	-2.113 (1.713)
Intercepto	236.8*** (58.32)	326.0*** (46.49)	237.1*** (31.06)	287.1*** (36.28)
Observaciones	168	168	168	168
R2	0.810	0.141	0.782	0.161
Número de escuelas		24		24

Nota: incluye controles por GSE, dependencia, número de alumnos, año y comuna (OLS). Errores estándar clusterizados al nivel de la escuela (OLS) y de la comuna (EF) se reportan en paréntesis debajo de cada coeficiente. *** p<0.01, ** p<0.05, * p<0.1

Quizás el resultado más interesante de la Tabla 7, además del incremento en el efecto de la contaminación puntual, es el hecho de que la temperatura en las escuelas resulta en todos los casos negativamente asociada a los puntajes obtenidos tanto en matemáticas como en lectura. Más aún, en el caso de matemáticas esto resulta significativo con un estimado puntual de -3.8. Para el caso de lectura, en tanto, sólo resulta ser significativo en el modelo OLS, pero no así en el de efectos fijos.

Finalmente, se realizó un test placebo para comprobar que los resultados señalados como causales y asociados específicamente a la contaminación atmosférica estuvieran realmente identificando cambios en rendimiento académico asociado a variaciones puntuales en la contaminación al momento de la prueba. El ejercicio realizado consistió en repetir las regresiones mostradas en la Tabla 2 pero utilizando la contaminación PM 2.5 del día de la otra prueba. Es decir, para el caso de la prueba de matemáticas se utilizó la contaminación PM 2.5 de las 9-11am del día en que se rindió la prueba de lectura y viceversa.

Tabla 8. Test Placebo para evaluar el efecto de la contaminación sobre el puntaje del SIMCE

	Matemáticas		Lectura	
	OLS	EF	OLS	EF
PM2.5 9-11am (Día cambiado)	0.0444 (0.0377)	-0.000487 (0.0336)	-0.0121 (0.0315)	0.00195 (0.0339)
PM10 Día de la prueba	0.0286 (0.0249)	0.00225 (0.0246)	0.00431 (0.0236)	-0.0123 (0.0196)
PM10 Mayo-Agosto	0.00233 (0.0347)	0.0485* (0.0240)	-0.0139 (0.0276)	0.0206 (0.0186)
Número de alumnos	0.188*** (0.0153)	-0.115*** (0.0141)	0.132*** (0.0118)	-0.0763*** (0.0127)
Corporación Municipal	-2.257 (1.844)		-1.303 (1.424)	
Particular Subvencionado	4.413*** (1.342)		5.048*** (1.027)	
Particular Pagado	4.597 (3.562)		5.254* (3.020)	
GSE. Medio Bajo	11.50*** (1.402)		12.30*** (1.183)	
GSE. Medio	26.31*** (1.663)		27.19*** (1.401)	
GSE. Medio Alto	43.51*** (2.067)		43.01*** (1.709)	
GSE. Alto	58.21*** (3.313)		55.25*** (2.809)	
Intercepto	207.5*** (4.320)	255.5*** (2.422)	232.6*** (3.480)	272.8*** (1.807)
Observaciones	8,456	8,456	8,456	8,456
R2	0.549	0.039	0.572	0.057
Número de escuelas		1,208		1,208

Nota: incluye controles por GSE, dependencia, número de alumnos, año y comuna (OLS). Errores estándar clusterizados al nivel de la escuela (OLS) y de la comuna (EF) se reportan en paréntesis debajo de cada coeficiente. *** p<0.01, ** p<0.05, * p<0.1

Si la variación puntual en los contaminantes está capturando una relación espuria y no causal (se encuentra por ejemplo serialmente correlacionada en el tiempo y es por ende predecible y posible de anticipar por los padres y madres de los estudiantes), entonces deberíamos seguir encontrando efectos negativos de la contaminación puntual sobre los puntajes. Si por el contrario es la contaminación del día de la prueba la que tiene efectos y no la del día anterior o siguiente, entonces el estimador asociado al PM 2.5 debería ser (estadísticamente) cero. Los estimados obtenidos se presentan en la Tabla 8.

Como era de esperar, la contaminación del día anterior o siguiente al de la toma de la prueba no tiene efectos sobre los puntajes obtenidos por los y las estudiantes en las escuelas del Gran Santiago. Esto sugiere que es efectivamente la contaminación puntual al momento de la prueba la que causa las disminuciones en los puntajes de la prueba SIMCE y no se trataría de una correlación espuria.

4. Conclusiones

Los resultados presentados previamente sugieren que los factores medioambientales tienen un impacto medible sobre los rendimientos académicos de los estudiantes del Gran Santiago, los que se representan en caídas en puntajes como consecuencia de alzas puntuales en la contaminación, y en incrementos en los rendimientos asociados a mejoras en niveles de áreas verdes en las escuelas y fuera de ellas. De igual forma, escuelas localizadas en áreas de mayores temperaturas, presentan menores resultados en matemáticas que aquellas que gozan de mejores condiciones ambientales.

Uno de los aspectos interesantes del estudio es el hecho de que si bien, muchas de estas variables pueden estar correlacionadas entre sí, como la vegetación fuera y dentro de la escuela, o la misma vegetación y la temperatura, los resultados presentados sugieren que todas ellas pueden ser importantes en mejorar los rendimientos académicos y en reducir las brechas entre las áreas más ricas y más vulnerables del Gran Santiago.

La información presentada se circunscribe exclusivamente a una región del país, la Región Metropolitana, y en particular a las 34 comunas que conforman el Gran Santiago. Esto permite aventurar que estudios que consideren más ciudades y regiones pudieran amplificar los resultados aquí encontrados. Esto, por cuanto las diferencias en niveles de contaminación entre regiones del sur de Chile y aquellas ubicadas al norte, por ejemplo, son aún más pronunciadas que las que se pueden observar al interior de una sola ciudad. Si las variaciones horarias en contaminación son también más altas, entonces nuestros resultados pueden ser considerados sesgados a la baja, pudiendo suceder algo similar con respecto a la temperatura y las áreas verdes tanto dentro como fuera de las escuelas.

De igual forma, el hecho de que la prueba SIMCE se tome en el mes de noviembre, un mes en que la contaminación suele ser más baja que durante los meses de invierno, sugiere que los resultados acá encontrados pueden tratarse nuevamente de estimaciones sesgadas a la baja. Esto es, exámenes similares tomados en otras épocas del año o bien en áreas de mayor contaminación (por ejemplo, el sur de Chile) pudieran tener efectos aún más marcados, si es que las diferencias en los niveles de contaminación son más acentuadas.

Además de los efectos asociados a la contaminación, es llamativo que la inversión que pueden hacer las escuelas en áreas verdes tiene efectos importantes sobre los resultados académicos. Si bien esto no es algo nuevo, dado que se trata de algo directamente observable por los padres y apoderados, uno esperaría que otras variables, tales como ingresos o la comuna de residencia, capturarían dicho efecto.

En adición a los resultados que sustentan la hipótesis de que tanto la contaminación atmosférica como la temperatura superficial tienen efectos negativos sobre los aprendizajes, lo que acá se postula es que incrementar los espacios verdes al interior de la escuela, puede tener impactos positivos sobre los rendimientos, lo que es independiente de lo que suceda fuera de la escuela. Esto es particularmente importante para escuelas localizadas en sectores más vulnerables de la ciudad, usualmente con escasas áreas verdes comunales.

Por último, para futuros estudios resultará importante orientarse a analizar los efectos que la contaminación atmosférica particularmente, pero también otras condicionantes medioambientales, pueden tener sobre otros resultados, como, por ejemplo, la asistencia. Es de esperar que sectores que presenten mayores problemas ambientales, en particular episodios puntuales con altos niveles de contaminación, puedan tener una menor asistencia a clases ya sea por precaución o bien por un mayor desarrollo de enfermedades en los y las niñas. Si bien la información del SIMCE no permite capturar cambios en el número de estudiantes que asiste a dar la prueba respecto del total de estudiantes de cada curso, otras fuentes de información sí permitirían desarrollar esta hipótesis.

Referencias

- Beyer, K.M.M. y otros (2014) Exposure to neighborhood green space and mental health: evidence from the survey of the health of Wisconsin. *International Journal of Environmental Research and Public Health* 11: 3453-3472
- Bharadwaj, P., Gibson, M., Graff Zivin, J., Neilson, C. A. (2017) "Gray matters: Fetal pollution exposure and human capital formation." *Journal of the Association of Environmental and Resource Economists* 4:2 (2017): 505–542.
- Card, D. y Krueger, A.B. (1992) Does school quality matter? Returns to education and the characteristics of public schools in the United States. *Journal of Political Economy* 100 (1): 1-40
- Chang, T., Zivin, J.G., Gross, T. y Neidell, M. (2016) Particulate pollution and the productivity of pear packers. *American Economic Journal: Economic Policy* 8 (3): 141-169
- Chen Y. y Rosenthal, S.S. (2008) Local amenities and life-cycle migration: do people move for jobs or fun. *Journal of Urban Economics* 64: 519-537
- Clark, C. y otros (2005) Exposure-effect relations between aircraft and road traffic noise exposure at school and reading comprehension. *American Journal of Epidemiology* 163 (1): 27-37
- Dadvand, P. y otros (2012) Green space, health inequality and pregnancy. *Environmental International* 40: 110-115
- Dadvand, P. y otros (2015) Green spaces and cognitive development in primary schoolchildren. *PNAS* 112 (26): 7937-7942
- Dubow, E.F., Boxer, P. y Huesmann, L.R. (2009) Long-term effects of parents' education on children's educational and occupational success: mediation by family interactions, child aggression, and teenage aspirations. *Merrill Palmer Quarterly* 55 (3): 224-249
- Ebenstein, A., Lavy, V., y Roth, S. (2016) The long-run economic consequences of high-stakes examinations: evidence from transitory variation in pollution. *American Economic Journal: Applied Economics* 8 (4): 36-65
- Fernández, I.C., Wu, J. y Simonetti, J.A. (2018) The urban matrix matters: Quantifying the effects of surrounding urban vegetation on natural habitat remnants in Santiago de Chile. *Landscape and Urban Planning* 187: 181-190
- Gelman, R y Butterworth, B. (2005) Number and language: How are they related?. *Trends in Cognitive Sciences* 9(1) 6-10. DOI: 10.1016/j.tics.2004.11.004
- Glewwe, P and K. Muralidharan. (2016) Chapter 10 - Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications, Editor(s): Eric A. Hanushek,

- Stephen Machin, Ludger Woessmann, Handbook of the Economics of Education, Elsevier, Volume 5, 2016, Pages 653-743, ISSN 1574-0692, ISBN 9780444634597.
- Hanna, R. y Oliva, P. (2015) The effect of pollution on labor supply: evidence from a natural experiment in Mexico City. *Journal of Public Economics* 122: 68-79
- Hanushek, E., Machin, S. y Woessmann, L. (2016) Handbook of the Economics of Education, Volume 5. 1st Edition
- Heckman, J. J. (2006). Skill Formation and the Economics of Investing in Disadvantaged Children. *Science*, 312(5782), 1900–1902. <https://doi.org/10.1126/science.1128898>
- Heckman, J. J. (2011). The Economics of Inequality: The Value of Early Childhood Education. *American Educator*, 35(1), 31.
- Hengl, T., Heuvelink, G.B.M. y Rossiter, D.G. (2007) About regression-kriging: From equations to case studies. *Computers and Geosciences* 33: 1301–1315
- Hodson, C.B. y Sander, H.A. (2017) Green urban landscapes and school-level academic performance. *Landscape and Urban Planning* 160: 16-27
- Huete, A.R. (1988) A Soil-Adjusted Vegetation Index (SAVI). *Remote Sensing of the Environment* 25: 295–308.
- Jo Blanden (2020) Education and inequality, Editor(s): Steve Bradley, Colin Green, The Economics of Education (Second Edition), Academic Press, 2020, Pages 119-131, ISBN 9780128153918, <https://doi.org/10.1016/B978-0-12-815391-8.00010-0>.
- Lee, S., Wolberg, G. y Shin, S.Y. (1997) Scattered data interpolation with multilevel b-splines. *IEEE Trans Vis Comput Graph* 3: 228–244
- Lauermann, F., Meißner, A., & Steinmayr, R. (2020). Relative importance of intelligence and ability self-concept in predicting test performance and school grades in the math and language arts domains. *Journal of Educational Psychology*, 112(2), 364–383. <https://doi.org/10.1037/edu0000377>
- Martins, L. y Veiga, P. (2010) Do inequalities in parents' education play an important role in PISA students' mathematics achievement test score disparities? *Economics of Education Review* 29 (6): 1016-1033
- Miller, S.J. y Vela, M.A. (2013) The effects of air pollution on educational outcomes: evidence from Chile. IDB Working Paper No. IDB-WP-468.
- Mohai, P., Kweon, B., Lee, S., y Ard, K. (2011) Air pollution around schools is linked to poorer student health and academic performance. *Health Affairs* 30 (5): 852-862
- Nielsen, T.S. y Hansen, K.B. (2007) Do green areas affect health? Results from a Danish survey on the use of green areas and health indicators. *Health & Place* 13: 839-850
- Olea, R.A. (1999) The Semivariogram. In: *Geostatistics for Engineers and Earth Scientists*. Springer US, Boston, MA, pp 67–90
- Rivkin, S.G., Hanushek, E.A., y Kain, J.F. (2005) Teachers, schools, and academic achievement. *Econometrica* 73 (2): 417-458
- Van Kempen, E. y otros (2010) Neurobehavioral effects of transportation noise in primary schoolchildren: a cross-sectional study. *Environmental Health* 9: 1-13
- Zheng, L. (2016) What city amenities matter in attracting smart people? *Papers in Regional Science* 95 (2): 309-328