



**UNIVERSIDAD DE CHILE
FACULTAD DE ECONOMÍA Y NEGOCIOS
ESCUELA DE ECONOMÍA Y ADMINISTRACIÓN**

LA HETEROGENEIDAD DEL EFECTO PAR EN EL RENDIMIENTO ACADÉMICO DE LOS ESTUDIANTES EN CHILE

**Seminario para optar al título de
Ingeniero Comercial, Mención Economía**

Francisca Muñoz Fernandez

Profesor guía:

Juan Pablo Valenzuela B.

**Santiago, Chile
2012**

ÍNDICE

I) INTRODUCCIÓN	3
II) REVISIÓN BIBLIOGRÁFICA	7
III) METODOLOGÍA	13
IV) DESCRIPCIÓN, ANÁLISIS DE DATOS Y MÉTODO	16
Descripción y análisis de datos	16
Método	22
V) RESULTADOS	24
Resultados nivel individual	24
Resultados nivel colegio	26
VI) CONCLUSIONES	28
VII) BIBLIOGRAFÍA	30
VIII) ANEXO	32

La heterogeneidad del efecto par en el rendimiento académico de los estudiantes en Chile

Francisca Muñoz Fernández

Diciembre, 2012

Profesor guía:

Juan Pablo Valenzuela B.

Resumen ejecutivo

El principal objetivo de esta investigación es analizar la heterogeneidad del efecto par en Chile a través de la interacción entre este efecto y el rendimiento de los alumnos por tipo de establecimiento al cuál pertenece. Haciendo uso del SIMCE 2010 de los 4tos básicos y utilizando un modelo MCO con corrección del sesgo de selección, mostramos que el efecto par de la composición socioeconómica y cultural de las familias no es lineal, sino convexa para los estudiantes que asisten a colegios municipales y particulares subvencionados, siendo mayor en el caso de los estudiantes que asisten a colegios municipales, mientras que para los que asisten a colegios particulares pagados este efecto es lineal. También se observa que aunque mucho más homogénea la composición social de los estudiantes al interior de los colegios particulares subvencionados, existe un efecto par cuando la composición social en este tipo de colegios es más diversa. Una vez que el análisis se lleva a cabo a nivel colegio, los resultados permanecen consistentes pero esta vez encontrando que el efecto par es heterogéneo al interior de las clases de los colegios municipales y particulares subvencionados, lo cual implica que este efecto es mayor cuando los estudiantes que componen el colegio presentan mayor diversidad socioeconómica, al contrario de lo que sucede en los colegios particulares pagados, donde los colegios más heterogéneos en su composición tienen un efecto menor que los homogéneos¹.

¹Esta investigación utilizó como fuente de información las bases de datos del SIMCE del Ministerio de Educación de Chile. El autor agradece al Ministerio de Educación el acceso a la información. Todos los resultados del estudio son de responsabilidad del autor y en nada comprometen a dicha Institución

I) INTRODUCCIÓN

La base del crecimiento de un país está soportada sobre la educación y la calidad de esta, siendo además el motor para disminuir la desigualdad y para encontrar estabilidad social al permitir acercarnos a un set de oportunidades más similar y justo para todos.

Justamente es en esta materia en la que Chile tiene muchas falencias: aún existen importantes brechas en los resultados de país respecto a los estándares internacionales, ello aunque denota mejoramiento a través del tiempo, ha quedado al descubierto con el pobre rendimiento que hemos alcanzado en las pruebas tanto internacionales como nacionales. Por su parte, aunque las brechas del desempeño educativo entre los niños de diferente nivel socioeconómico se han reducido a través del tiempo, éstas aún permanecen en niveles muy elevados.

En suma, nuestro sistema educacional se caracteriza por ofrecer desiguales oportunidades educativas y por lo tanto de logros académicos, caracterizándose tanto por su naturaleza segregadora como por su aún alto déficit de calidad, especialmente para los estudiantes más vulnerables.

La investigación comparada ha permitido identificar consistentemente cuáles son los componentes necesarios para ofrecer un buen sistema educacional. Desde diversas perspectivas se ha intentado encontrar la relación más fehaciente entre inputs y outputs y así tener una mirada más clara de donde poner los esfuerzos para mejorar los resultados. Si bien muchas de estas estimaciones han llegado a resultados consistentes sobre algunos de estos componentes y el peso de cada uno de ellos, creciente atención se ha puesto al rol que juegan los factores sociales, específicamente el ambiente en el cuál se

vinculan e interactúan los estudiantes. Mientras a nivel internacional existe una creciente literatura sobre el posible efecto de la composición del grupo de pares en el rendimiento individual, a nivel nacional aún es un ámbito de estudio poco desarrollado.

La importancia del recién aludido efecto composicional en explicar el rendimiento individual, se ha investigado internacionalmente desde el Reporte de Coleman (1966) para Estados Unidos y Coleman (1968). Asimismo se han desarrollado crecientes estudios teóricos y diversas investigaciones empíricas, las cuales no necesariamente han llegado a conclusiones consistentes.

En Chile, un estudio realizado por Mizala y Torche en el año 2012, concluyó que el nivel promedio de los pares (medido a través de la situación socioeconómica de los padres) explica con mayor fuerza que el propio estatus socioeconómico el rendimiento académico individual. De esta forma, la importancia del efecto par es mayor de la que solemos pensar. Investigar su peso y su naturaleza es clave para poder entender de mejor manera la función de producción escolar y el peso relativo de éste para con los otros inputs tales como background familiar o características individuales.

Siguiendo esta línea, el objetivo que se busca alcanzar a lo largo de esta investigación es determinar el efecto par en Chile; testeando la forma funcional que éste alcanza (ya sea cóncavo o convexo o simplemente lineal) como investigar la heterogeneidad de este efecto al interior del establecimiento, también analizaremos si dichos efectos son similares o diferentes de acuerdo a la dependencia institucional de los establecimientos escolares chileno.

Para esto, utilizaremos la metodología llevada a cabo por Raitano y Vona (2010) utilizando datos entregados por el SIMCE 2010 en el curso de 4to básico para el subsector de matemática, explicada en más detalle en la sección 3.

El punto clave de esta estrategia está en que da espacio para que las variables pares puedan influir de manera diferente en cada sector al que pertenecen los colegios y que se encarga de explotar las posibles no linealidades en el efecto permitiendo identificar si éste crece a tasas crecientes, decrecientes o constantes. Cabe destacar que además podremos dilucidar cuánto del efecto obtenido se debe al grado de heterogeneidad del ambiente (si el tener pares disimiles afecta positiva o negativamente a los resultados).

Basándonos en que si bien el utilizar un modelo lineal en medias para determinar el efecto par nos puede entregar de una manera muy clara información importante acerca de la magnitud y significancia de éste, al momento de crear políticas de reasignación de los alumnos el efecto neto a nivel agregado sería cero, pues solo existiría un efecto redistributivo.

Nuestras estimaciones, que permiten diferenciación en el efecto y no linealidad, dan cuenta que existe un mayor efecto par lineal y una convexidad mayor en el grupo de niños más vulnerables (sector municipal) versus aquellos algo más aventajados (sector subvencionado). Resultados como éstos permiten obtener conclusiones que se asocian a políticas educacionales sobre segregación lo que las hace más importantes aún en un país como Chile, catalogado como el segundo con mayor segregación socioeconómica escolar².

Por ejemplo, bajo el supuesto de que el resultado recién mencionado es correcto, el hecho de que niños más aventajados emigren del sector más vulnerable a uno donde hayan niños con características similares a las de ellos, ya sea por políticas educacionales o por algún incentivo a las familias, provocaría una disminución en el rendimiento promedio total ya que la disminución del rendimiento de los más vulnerables sobrepasaría el aumento en

² Dentro del universo de 57 países que realizan la prueba PISA (Valenzuela, Bellei y De los Ríos, 2008)

el rendimiento de los mas aventajados (quienes emigran) y, además, un aumento en el grado de segregación.

Adicionalmente, los resultados obtenidos nos dicen que para el sector municipal y particular subvencionada existe una fuerte relación no lineal entre los pares y el rendimiento académico. Además, que el nivel de heterogeneidad es relevante solo para el sector particular subvencionado.

La estructura de este estudio es de la siguiente manera; en la sección II se hará una revisión bibliográfica internacional y nacional acerca de los estudios realizados investigando el efecto par, en la sección III se intenta describir en detalle la metodología utilizada para luego poder comprender de una manera más clara los resultados, en la sección IV se encuentra una detallada descripción de datos y análisis de estos junto con la descripción del método llevado a cabo. En la sección V se describen los resultados de las estimaciones y por último en la sección VI se pueden encontrar las conclusiones de este estudio.

II) REVISIÓN BIBLIGRÁFICA

Existe un creciente desarrollo de estudios acerca de la relación entre el rendimiento académico de los alumnos y los determinantes de éste. Si bien muchas de estas estimaciones han llegado a resultados consistentes sobre algunos de estos componentes y el peso de cada uno de ellos, el rol que juegan los pares como factores influyentes en el rendimiento individual se mantiene confuso y no lo suficientemente explotado (Sacerdote, 2011).

Es a esta materia a la que se suma nuestra investigación; indagar sobre el efecto a nivel individual de las variables o características de los pares, concretamente en el escenario chileno y por lo tanto haciendo uso de la base de datos nacional entregada por el SIMCE.

La literatura existente en este tema ha concentrado el debate, por largo tiempo, en dos grandes retos que ha habido que enfrentar. En un principio el gran dilema era el problema de la reflexividad generada al incluir como variable explicativa el rendimiento académico de los pares. Al hacer esto, se abre espacio para un fuerte sesgo de endogeneidad puesto que así como los pares tienen algún grado de influencia en los resultados del individuo i (variable dependiente), el rendimiento de este mismo individuo i produce al mismo tiempo algún grado de influencia en el rendimiento de sus pares (Manski 1993). Es por esta situación y por lo engorroso que resulta solucionarlo, que hoy la mayoría de los estudios en esta materia prefieren definir un conjunto de característica de los pares como la variable a considerar para medir el efecto par. Así, normalmente se utiliza el background familiar como aproximación a una forma

de medida del nivel de los pares; el coeficiente acompañado de esta medida será la magnitud potencial del efecto par.

El segundo reto se vincula a los problemas del sesgo de selección los cuales provienen de diversas situaciones. Por una parte, en el contexto educacional es muy difícil que exista asignación o selección de alumnos realmente aleatoria entre los colegios. Esto provoca inevitablemente conclusiones erradas y coeficientes sesgados si no se controla o no se utiliza alguna técnica para controlarlo.

Estudios como Sacerdote (2001); Hoxby & Weingarth (2005) y Duflo, Dupas, Kremer (2010) han tomado provecho de políticas utilizadas dentro de colegios donde mezclan aleatoriamente a sus alumnos en algún año concreto y por lo tanto no existen problemas con la selección de los alumnos. Sacerdote (2001) realizó su experimento con jóvenes en su primer año universitario utilizando el hecho de que eran asignados de manera aleatoria a algún compañero de cuarto. De esta manera no tuvo problemas con el sesgo de selección en sus estimaciones y pudo encontrar evidencia de la existencia de efecto par sobre los resultados académicos de los alumnos. Duflo, Dupas, Kremer (2010) a través de un estudio realizado en Kenia y llevando a cabo un experimento que agrupaba a los alumnos aleatoriamente en nuevos cursos (unos grupos homogéneos según habilidad y otros no), encontraron un efecto positivo directo de tener pares con habilidades altas y también un efecto positivo indirecto para los menos hábiles.

De similar manera, aprovechando los shocks externos, Iberman, Kugler y Sacerdote (2009) realizaron su estudio en base a la emigración que provocó el huracán Katrina a dos localidades de Texas. Encontraron solamente un pequeño efecto lineal promedio sobre el rendimientos de los colegios que recibieron a los alumnos emigrantes.

Entre las técnicas metodológicas más usuales para mitigar éste sesgo, se encuentra la estimación “within school” la cual permite estudiar el efecto par desde la variación en la composición de la sala de clases dentro de un mismo colegio. Se asume aleatorización dentro de los colegios al momento de formar los cursos (algunos países Europeos en la primaria) y así la selección y autoselección de los alumnos al entrar al colegio deja de ser un problema. Ammermueller y Pischke (2006) utilizando esta técnica e intentan encontrar un efecto lineal sobre el rendimiento de los alumnos. Los autores realizan estimaciones del efecto a través de MCO, identificando a un fuerte efecto par promedio, sin embargo una vez que controlaron por error de medición llegaron a resultados muy similares a los encontrados por las estimaciones “within school” las cuáles eran más pequeñas pero aún significantes. Otros autores, entre los cuales se encuentran los trabajos de Hoxby (2000), Betts y Zau (2004) y Hanushek, Kain, Markman y Rivkin (2003), también utilizan la estrategia “within school”, los cuales encuentran un efecto significativo en el rendimiento individual causado por las características de los pares. Asimismo, Burke y Sass (2008) encuentran poca evidencia de un efecto de las características promedio de los pares sobre el rendimiento promedio de los alumnos utilizando la misma estrategia.

En suma, si bien ambos retos tuvieron respuesta por parte de los autores y se hicieron cargo de ellos, los resultados obtenidos seguían siendo muy dispares y aunque las metodologías fueran similares, las conclusiones sobre tamaño, significancia e implicancias del efecto par distanciaban unas de otras.

Un avance metodológico fue el realizado por Hoxby y Weingarth en Wake County, North Carolina, donde ponen en cuestionamiento el modelo lineal en medias para determinar el efecto par. Hoxby y Weingarth (2005) se encargaron de buscar la metodología que se acomodara más a los datos haciendo uso de la política de reasignación sistemática de los estudiantes en los colegios

públicos del condado de Wake. Con el uso de datos de panel de antes y después de la experiencia de reasignación, pudieron identificar los cambios en la composición de la sala de clases provocados por decisiones exógenas, llegaron a concluir que no era correcto que el modelo lineal en medias y el modelo que permite diferenciación del efecto par según el rendimiento individual de cada alumno se estudiaran de manera independiente ya que por si solos no son capaces de determinar el comportamiento del efecto en la población.

Por otra parte, nuevas especificaciones sobre el efecto par han considerado la heterogeneidad de su efecto entre los distintos alumnos según el grupo con el que se vinculan. La literatura existente ha concluido que es efectivo que alumnos con distintos niveles de habilidad o agrupados según algún tipo de característica similar se ven afectados de diferentes maneras por su grupo de pares; Rangvid (2007), quien hizo un estudio en Copenhague, Dinamarca utilizando PISA y Schneeweis y Winter-Ebmer (2007), quienes analizan colegios de Austria en base a datos de PISA 2000 y 2003, aplicaron esta nueva especificación. Ambos estudios llegaron a resultados muy parecidos: fuerte efecto par sobre los alumnos de menor habilidad y un casi nulo impacto asociado a la heterogeneidad dentro de la sala de clases. Sin embargo, Burke y Sass (2008) e Iberman, Kugler y Sacerdote (2009), nombrados anteriormente, una vez que permitieron que el efecto par se diferenciara por grupos, pudieron encontrar evidencia consistente respecto de la existencia de heterogeneidad en el efecto, no así lo hicieron con el efecto lineal. Finalmente, el estudio hecho por Raitano y Vona (2010), a través de interactuar la variable característica de los pares con una dummy que identificaba qué tipo de políticas de selección utilizaba el colegio y a través de la inclusión de la variable pares al cuadrado, permitieron encontrar un efecto diferenciado según grupo y al mismo tiempo cómo se mueve este efecto dentro de cada grupo, de tal forma que también lograron determinar la relevancia del grado de heterogeneidad del grupo en el total del efecto par estimado.

En la investigación nacional, no son numerosos los estudios realizados acerca del efecto composicional. Sapelli y Vial (2005) encuentran que el efecto par no es lo suficientemente importante como para explicar las diferencias de rendimiento entre una escuela municipal y una particular subvencionada. Por su parte, McEwan (2003), utilizando una metodología “within school” encuentra un efecto positivo de los años de educación de la madre promedio de los pares sobre los resultados de la prueba SIMCE de lenguaje del año 1997 para los octavos básicos. Mizala y Torche (2012), que si bien buscan analizar la estratificación socioeconómica del rendimiento de los alumnos y la implicancia del copago, encuentran que el sector particular subvencionado a pesar de ser el más heterogéneo en cuanto al nivel socioeconómico de los alumnos, es el que cuenta con escuelas más homogéneas dando cuenta de un posible efecto de esta homogeneidad en los resultados obtenidos. Por último, Rodríguez (2010) realiza una investigación, enfocado a similares objetivos que la nuestra, utilizando la base de datos SIMCE 2005 para 4tos básicos y, a través de metodologías semiparamétricas y MCO corregido por sesgo de selección y con interacciones, encuentra, a modo general, evidencia de no linealidades y que el efecto par actúa con mayor fuerza entre los estudiantes más vulnerables.

Bajo este contexto, nuestra investigación utilizará un modelo que permite estudiar el efecto par de una manera más amplia diferenciando entre grupos y permitiendo la identificación de efectos no lineales. A la vez, analizamos el efecto de la heterogeneidad del grupo de pares y la dirección que éste toma. Nuestra especificación intenta mitigar el sesgo de selección dado que la base a utilizar cuenta con datos acerca de la forma de selección de parte de los colegios y además, con la ayuda de estadística del INE, podremos hacernos cargo de la autoselección de los padres ante diferentes colegios utilizando la estrategia aplicada por Mac Ewan (2001).

Así, utilizando la metodología que será explicada en más detalle en la siguiente sección y llevándola concretamente al caso de Chile al analizar los resultados diferenciados por dependencia de los establecimientos y además haciéndonos cargo del problema que trae el sesgo de selección, intentaremos encontrar un resultado coherente para el parámetro en cuestión teniendo como objetivo determinar la magnitud y la forma que este efecto sigue. Averiguar si éste es creciente, decreciente o constante a medida que nos encontramos con grupo más o menos hábil (y así llegar a un resultado agregado o promedio) y cuánto de esto se debe a la heterogeneidad del grupo.

III) METODOLOGÍA

Tal como se explicó en las secciones anteriores, la modalidad a seguir será muy similar a la llevada a cabo por Raitano y Vona (2011) la cuál cuenta con dos ventajas concretas: i) al incluir término que represente posibles no linealidades nos permitirá investigar la forma funcional que sigue el efecto y ii) al hacer interactuar las variables que captura el efecto par con cada tipo de establecimiento, podremos estudiar el efecto diferenciado en la población educacional.

Antes que todo es importante aclarar que nuestro estudio se basa en la influencia de los pares en el rendimiento escolar desde una perspectiva académica, es decir, es una investigación mucho más estadística que teórica llevada a cabo a través de datos recopilados por el Sistema de Medición de la Calidad de la Educación 2010 (SIMCE 2010) de los 4tos básicos.

Con el objetivo de no caer en los problemas típicos al enfrentar una estimación de este tipo, como son los de reflexividad y sesgo de selección, se proponen estrategias para resolverlas. El problema de reflexividad se evita siguiendo a Sacerdote (2011) pues no se considerará en ningún caso el rendimiento de los pares como variable explicativa, sino que la medida a utilizar para investigar sobre el efecto par será un nuevo índice que incluya las características familiares de los compañeros (background familiar) y la cantidad de libros en casa como aproximación al capital cultural.

Por su lado, el problema de sesgo de selección será mitigado de dos maneras; i) En pos de controlar por la selección no aleatoria de los alumnos por parte de

los colegios, se utilizaron variables de control tales como si los colegios seleccionan o no por habilidad o por ciertas características como certificado de matrimonio de los padres o certificado de remuneraciones. Se utilizará la metodología utilizada por Contreras, Sepúlveda y Bustos (2010), la cual considera que un colegio selecciona cuando más del 50% de los apoderados admite que el colegio realiza alguna forma de selección (ya sea a través de exámenes o certificados) y ii) para mitigar el sesgo de autoselección de los alumnos al colegio, se llevó a cabo el método desarrollado por Lee Heckman (1983), como corrección al método de dos etapas, para casos donde la elección se mueve entre más de dos alternativas como es el caso de municipales, particulares subvencionados y particulares pagados.

Específicamente, para nuestro modelo utilizamos la función de producción escolar normalmente utilizada en el ámbito de la educación, agregándole variables de interacción y una especificación que haga cuenta del efecto par y las características de éste.

$$PS_{i,c} = \alpha + \beta_1 * X_i + \beta_2 * X_c + \beta_3 * \overline{X_c} + \beta_4 * X_{bf} + (\delta + \varphi * D_{ps=1} + \phi * D_{pp=1}) f(\overline{BF}) + \mu_i \quad (1)$$

En la ecuación (1), la variable dependiente será el puntaje SIMCE obtenido por el alumno i en el colegio c ($PS_{i,c}$) y será el resultante de diversos vectores; X_i que incluirá las variables del individuo i, X_c con las variables a nivel del establecimiento que no dependen del colegio en sí, $\overline{X_c}$ con las variable específicas de cada establecimiento como porcentaje de mujeres, porcentaje de repitentes y cantidad de alumnos y X_{bf} que incluirá variables representando las características o background familiar propio. Lo novedoso e interesante es la inclusión de la función $f(\overline{BF})$ en pos de capturar la influencia par medida a través del background familiar promedio de los compañeros neto del individual.

Esta forma funcional, además de incluir las características promedio de los pares para calcular el efecto par lineal, incluye otra especificación para capturar el efecto no lineal; el promedio de los pares al cuadrado y la desviación estándar de los pares. Lo primero servirá para determinar la forma que sigue el efecto y lo segundo para determinar cuánto de esta concavidad se debe a la heterogeneidad dentro de la sala de clases.

Notar que esta especificación interactúa con un coeficiente por sí solo (δ) para obtener datos a nivel general y que también lo hace con dos variables dummies; $\delta * f(\overline{BF}) + \varphi * D_{ps=1} * f(\overline{BF}) + \phi * D_{pp=1} * f(\overline{BF})$, la primera, $D_{ps} = 1$ si el establecimiento es particular subvencionado y la segunda $D_{pp} = 1$ si es particular pagado.

Adicionalmente, se hicieron estimaciones para cada sector por separado.

$$PS_{i,c} = \alpha + \beta_1 * X_i + \beta_2 * X_c + \beta_3 * \overline{X_c} + \beta_4 * X_{bf} + \delta * f(\overline{BF}) + \mu_i \quad (1.2)$$

La cual cuenta con la misma descripción de variables anterior y se estimará la misma para cada sector. Así, la expresión $\delta * f(\overline{BF})$ será la que nos hablará sobre el efecto par lineal y no lineal de cada tipo de establecimiento.

Por último, una nueva especificación a nivel colegio fue estimada a modo de obtener resultados más genéricos por sector y no tener que lidiar con los sesgos y problemas que traen las estimaciones a nivel individual. Es importante tener en cuenta que estimaciones a nivel-escuela si bien no cuentan con las complicaciones de estimaciones individuales, tendrán un sesgo de agregación que se espera esté representado en coeficientes de mayores magnitudes.

$$PS_c = \alpha + \beta_1 * X_c + \beta_2 * \overline{X_c} + (\delta + \varphi * D_{m=1} + \phi * D_{ps=1}) * f(\overline{BF}) + \mu_c \quad (2)$$

Se espera que a nivel colegio los resultados sean consistentes con aquellos encontrados a nivel individual y así agregarle robustez a nuestras estimaciones.

IV) DESCRIPCIÓN, ANÁLISIS DE DATOS Y MÉTODO

Descripción y análisis de datos

La base de datos utilizada en esta investigación es una fusión de cuatro bases de datos, todas proporcionadas por el Ministerio de Educación.

La primera corresponde a los resultados de la prueba estandarizada SIMCE (Sistema Nacional de medición de la calidad de la Educación) en Matemáticas³.

Específicamente los datos utilizados pertenecen a los alumnos de 4to básico en el año 2010. La segunda es un cuestionario realizado a los padres de los alumnos la cual provee información variada acerca de la situación familiar y del alumno; características familiares como ingresos y educación de los padres, compromiso de ellos para con el estudio de sus hijos, características del alumno, etc. La tercera fuente es un cuestionario realizado a los profesores que en ese entonces dictaban matemática al curso que dio la prueba SIMCE. Ésta incluye tanto características del profesor como características a nivel colegio. Y por último, la cuarta fuente son datos a nivel de establecimiento que se utilizó para obtener características específicas de cada colegio y poder controlar por ellas. Las cuatro bases son recogidas por el Ministerio e incluyen la totalidad de los colegios públicos, subvencionados y particulares pagados en Chile cuando participaron del SIMCE los 4tos básicos el año 2010.

³ Se consideró la prueba de Matemáticas ya que el año 2010 el cuestionario a los profesores fue hecho a quienes dictaban esta materia.

Para estas bases de datos del año 2010 se tienen 7.998 colegios y 255.662 alumnos, de los cuáles aproximadamente el 76% (167.939 alumnos) fueron utilizados en esta investigación. Las observaciones eliminadas fueron aquellos alumnos que no contaban con los datos necesarios para las estimaciones hechas y aquellos colegios que contaban con una matrícula total en educación básica menor a 100 alumnos para evitar errores sistemáticos provocados por contar con colegios representados por muy poco alumnos (Kane y Staiger 2002).. De este modo, la base utilizada contiene datos a nivel individual del alumno, de los padres, de los profesores y del establecimiento en el cuál estudia. En las Tablas 1.1 y 1.2 del anexo se encuentra un resumen de las variables utilizadas en este trabajo (descripción y estadísticas).

Algunas de las variables a nivel individual fueron promediadas a través del “análisis de componentes principales” con el fin de obtener un índice que represente el background familiar (estado socioeconómico) y cultural, éste incluye años de educación del padre, años de educación de la madre, ingresos familiares per cápita y cantidad de libros en el hogar. Lo llamaremos “esec” por sus siglas “estatus socioeconómico y cultural”.

El objetivo de este índice, además de medir el nivel del alumno en cuanto a su background familiar y cultural y analizar la distribución de los alumnos en los distintos tipos de colegios, es tener una medida con la cual calcular el nivel promedio de los compañeros con los que estudia el alumno. De esta manera, calculando el promedio por curso de este índice, neto del individual, podremos obtener el nivel del grupo de pares. Intentando rescatar no linealidades, se definió también como una nueva variable la desviación estándar de este índice por curso, el que capturará el nivel de heterogeneidad dentro de la sala de clases y una variable igual al promedio del índice al cuadrado para obtener un efecto diferenciado.

La siguiente tabla resume las estadísticas de los resultados en la prueba de matemática a modo general y según tipo de colegio:

Tabla 1: Estadística descriptiva puntaje obtenido en SIMCE por sector y total. Análisis a nivel individual y a nivel colegio (4° básico (2010))

	Nivel individual		Nivel colegio	
	Media	Desviación estándar	Media de desviación estándar colegios	Desviación estándar promedio colegios
Municipal	239,47	51,7	46,27	22,46
Subvencionados	261,34	50,48	44,02	22,64
Particular pagado	301,89	43,26	40,07	16,33
Total	256,26	53,17	44,64	28,75
N° observaciones	167,939 alumnos		4116 colegios	

Fuente: Elaboración propia utilizando datos SIMCE 2010.

Se observa que el puntaje promedio individual es aproximadamente 22 puntos inferior para los colegios municipales versus colegios subvencionados y más de 60 puntos menor con respecto a los establecimientos particulares pagados. Al observar la desviación estándar por sector, se concluye que el sector con menos diferencia entre sus alumnos es el particular pagado, luego los subvencionados y por último, por una diferencia muy baja, los municipales que cuentan con aproximadamente 7 puntos más de desviación con respecto a la media en comparación con los particular pagado. Cabe destacar que la diferencia entre la dispersión de los puntajes dentro del sector subvencionado y municipal es menos de 1 punto lo que podría entregar luces de que si bien hay grandes diferencias en el promedio obtenido en ambos sectores, entre los alumnos de cada sector existe casi la misma brecha.

En suma a esto, podemos observar que el sector municipal y el sector subvencionado cuentan con una brecha muy similar entre sus colegios (22,5 y 22,6 respectivamente) y un tanto mayor versus el sector particular pagado con una dispersión de 16,3; si bien los alumnos pertenecientes al sector subvencionado obtienen puntajes más homogéneos entre ellos que aquellos pertenecientes al sector municipal, cuando analizamos a nivel colegio encontramos que en este último sector las diferencias entre el rendimiento de los colegios se hacen prácticamente iguales a las del sector particular subvencionado.

La siguiente tabla muestra estadísticas de características propias de cada sector con el fin de vincularlas con lo visto anteriormente:

Tabla 2: Estadística descriptiva datos sector y total. Análisis a nivel individual (4° básico (2010)).

	Selección por habilidad	Porcentaje alumnos repitentes	Expectativas	Media esec
Municipal	55%	15%	80%	-0,839
Particular Subvencionado	80%	7%	92%	0,308
Particular Pagado	90%	2%	100%	2,754
Total	71%	9%	88%	0,14
N° observaciones				167.939

Fuente: Elaboración propia a partir de datos SIMCE 2010

Al comparar los datos entregados por la tabla 2 con la tabla anterior, podemos decir que justamente el sector que más selecciona y con mayor “esec” es aquel que mejor resultados SIMCE obtiene. Esta relación se mantiene para los otros dos sectores; los subvencionados (80% de los colegios), que seleccionan más que los municipales (55% de los colegios), obtienen mayores puntajes que estos ultimo y a la vez menores puntajes que el sector más seleccionador (particular pagado: 90% de los colegios lo hace de alguna manera, ya sea

exigiendo certificado de notas del colegio anterior, el tener que rendir un examen de ingreso o exigiendo evaluación preescolar).

Por otro lado, si observamos el porcentaje de alumnos repitentes por sector, podemos deducir que aparte de existir una relación positiva entre mejores resultados y menos repitentes, también puede darse una causalidad contraria; es decir, si se tienen muchos repitentes en un mismo sector y por lo tanto mayores probabilidades de contar con cursos con repitentes, y además este sector cuenta con bajos resultados, puede haber una influencia negativa en el rendimiento individual por tener mayor cantidad de compañeros repitentes (tal como concluye Bellei (2007))

En cuanto a las expectativas, quisimos incluirlas para contar con variables motivacionales y ver la relación de éstas para con el rendimiento de los alumnos. A simple vista se puede deducir una relación positiva entre las expectativas de los padres de que sus hijos alcancen educación superior y los resultados en el colegio.

Por último, el índice “esec”, cuenta con una media muy disímil entre los tres sectores; -0,84 para el sector municipal, 0,31 para el sector subvencionado y 2,75 para el sector particular pagado. Esta situación nos da claras luces de la distribución entre sectores donde es fácil concluir que aquellos alumnos con un background familiar y cultural más bajo se concentran en el sector municipal reafirmando lo expuesto en la primera sección al hablar de un sistema educacional segregado.

La siguiente tabla nos muestra estadísticas sobre el índice “esec” por dependencia.

Tabla 2: Estadística descriptiva índice “esec” por dependencia. Análisis a nivel individual y a nivel colegio (4° básico (2010)).

	Nivel individual individual		Nivel colegios	
	(1) Media esec	(2) Desviación estándar esec	(3) Media de desviación estándar colegios	(4) Desviación estándar del esec promedio colegio
Municipal	-0,839	1,392	1,178	0,716
Particular	0,308	1,457	0,665	0,933
Subvencionados Particular pagado	2,754	0,748	1,115	0,337
Total	0,02	1,03	1,107	1,246
N° observaciones	167,939 alumnos		4116 colegios	

Fuente: Fuente: Elaboración propia a partir de datos SIMCE 2010

Se observa que la dispersión del “esec” (columna 2) es mayor en el sector subvencionado que en los otros dos sectores, sin embargo la diferencia con el sector municipal es muy pequeña (1,46 versus 1,39). Esto significaría que el sector con mayor heterogeneidad es el subvencionado, seguido por el municipal y luego por el particular pagado que cuenta con una dispersión mucho menor (0,75).

Si analizamos la dispersión por sector pero ahora respecto a las medias de los colegios (columna 3), observamos que en los tres sectores ésta disminuye; hay mayor homogeneidad entre un colegio y otro que entre todos los alumnos del sector. Ahora, si observamos el promedio de la desviación estándar por colegio, es decir la media de la dispersión arrojada dentro de cada colegio, se puede concluir que el sector subvencionado es el que cuenta con una mayor grado de homogeneidad entre alumnos que pertenecen a un mismo colegio; si bien es el

sector con mayor heterogeneidad en la totalidad de sus alumnos, es el que cuenta con mayor homogeneidad dentro de cada colegio, tal como

Método

Lo primero que se hizo fue unir las bases de datos para contar con las variables explicativas y de control que se utilizaron en la investigación. La decisión de qué variables integrar y cuáles no, fue tomada en base a diversa literatura que se ha encargado de investigar cuáles son los componentes que explican el rendimiento de los alumnos tanto en el escenario internacional como nacional (Hanushek (2007), Mizala y Torche (2010)).

En pos de depurar la base de datos, se eliminaron las observaciones con algunas variables sin información y aquellos valores que salían fuera de rango en cada variable⁴ con ello nos aseguramos de eliminar posibles problemas que puede traer el error de medición.

El siguiente paso, fue la construcción del índice compuesto que combine distintas dimensiones para representar la composición par: “esec” (condición socioeconómico cultural)⁵, el cual intenta rescatar el background familiar (estado socioeconómico) y cultural; incluye años de educación del padre, años de educación de la madre, ingresos familiares y cantidad de libros en el hogar como intento de medida del capital cultural familiar.

Por último, basándonos en la metodología de Lee-Heckman (1983), como corrección por sesgo de selección cuando se cuenta con más de dos

⁴ No hubo mayores problemas ya que los valores perdidos estaban distribuidos aleatoriamente en su mayoría, situación que fue comprobado al testearlo a través de diferentes metodologías y los resultados siempre fueron consistentes.

⁵ Para más detalles de la creación de este tipo de índice véase “Análisis factorial”; Morales, Pedro (2011) y “Guía metodológica Diseño de indicadores compuestos de desarrollo sostenible”; Schuschny, y Soto (2009)

alternativas de elección, obtendremos el inverso del ratio de mills y podremos controlar por el sesgo de selección⁶ asociado a que los padres no escogen aleatoriamente el colegio para sus hijos. La elección del modelo para esta estimación se hizo siguiendo la metodología llevada a cabo por McEwan (2001) donde se incluyen como variables explicativas la cantidad de colegios municipales, particulares subvencionados y particulares pagados por kilómetro cuadrado en la comuna donde el alumno estudia y así se satisface la obligatoriedad de incluir al menos una variable no correlacionada con el modelo de rendimiento escolar definido anteriormente. Con la ayuda del informe “División político administrativa y censal, 2007” del INE se pudo construir la variable que resumiera densidad de colegios de cada sector por comuna.

De esta forma, a través de un multinomial logit se estimó la ecuación de elección de colegio teniendo como variable dependiente cada uno de los sectores para luego en una segunda etapa, y en base a los resultados obtenidos, estimar la ecuación de rendimiento incluyendo ahora los lambdas (inverso del ratio de mills) que absorbe el sesgo mencionado (para detalles de la estimación ver tabla 2 del anexo.).

La primera estimación que se hizo fue una especificación básica de la ecuación (1), sección III, sin incluir el efecto par (estimación típica en la mayoría de los estudios de esta literatura). Luego, se incluyeron las variables que capturaban las características de la composición par y finalmente se analizó la interacción entre el efecto par y la dependencia del establecimiento, primero haciendo interactuar las variables pares con las dummies de dependencia y luego haciendo estimaciones por separado para cada sector. Por último, se llevaron a cabo estimaciones a nivel colegio.

⁶ El no controlar por el sesgo de selección o no hacerse cargo de este problema nos llevaría a problemas de identificación y resultados sesgados ya que bien es sabido que además de que los colegios seleccionen a sus alumnos, el alumno también selecciona su colegio dependiendo en el gusto o intereses de los padres.

V) RESULTADOS

Estimación a nivel alumno

Las estimaciones hechas a nivel de los alumnos se hicieron mediante la técnica de Mínimos Cuadrados Ordinarios (MCO) con corrección de sesgo de selección, ajustando la estimación de los errores por clústers, reconociendo a los colegios como la unidad básica de la muestra y una cierta asociación de no observables entre los estudiantes de un mismo curso. Una de las características importantes de destacar esta forma de estimación es que los errores estándares arrojados son robustos.

En la tabla 3 del anexo se pueden ver los resultados de los coeficientes estimados para cada modelo, junto a su error estándar y su grado de significancia⁷.

Los seis modelos estiman una regresión general que incluye los tres tipos de colegio. La importancia de esta comparación es que cada modelo varía al incluir variables relacionada a los pares.

La especificación (1) es el modelo considerado como estándar por la literatura pues no incluye variables pares.

Los resultados obtenidos muestran la robustez de nuestras estimaciones al

⁷ La cantidad de asteriscos denotará el p-value y por lo tanto el nivel de significancia: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

permanecer consistentes en las distintas especificaciones. A modo general, un aumento del promedio del “esec” de los compañeros está relacionado con un mejor rendimiento individual. Se observa que el tamaño de este efecto lineal es significativo a más del 1% para cada una de las regresiones.

El modelo (2) muestra que al incluir el nivel promedio de los pares, el efecto relacionado a un aumento de éste es positivo y significativo. Esto se mantiene para las siguientes dos especificaciones ((3) y (4)), sin embargo, en el modelo (3) se puede observar que el efecto de la heterogeneidad al no ser significativa demuestra no ser relevante para los estudiantes de 4to básico del año 2010. El modelo (4) muestra significancia en el término que analiza la no linealidad, lo que nos dice que el efecto par sería convexo y por lo tanto cuadrático.

Las especificaciones cuadráticas ((5) y (6)) nos muestran la evidencia de no linealidades en el efecto par con alto grado de significancia. En cuanto a la heterogeneidad, se observa que ésta no es relevante en ninguno de los casos. Estos resultados indican que para 4to básico en la asignatura de matemática para el año 2010, los alumnos no serían perjudicados en términos académicos por tener un grupo de pares más heterogéneo, situación que está en línea con la literatura internacional donde la heterogeneidad del grupo normalmente no es relevante en el rendimiento de los alumnos (Rangvid, 2007).

Ahora, si se considera que cada tipo de colegios tiene una función de producción diferenciada, para lo cual las estimaciones a nivel individual se realizan en forma separada para colegios municipales, particulares subvencionados y particulares pagados (ecuación (1.2) de la sección III), obtenemos resultados parecidos a los que nos llevó el modelo (6).

En la tabla 4 del anexo se aprecia que la forma funcional del efecto par y de heterogeneidad dependerá del tipo de establecimiento al que asistan los

estudiantes. A partir de especificaciones diferenciadas, se puede concluir que el efecto par es convexo tanto en el sector municipal como particular subvencionado, aunque de mayor magnitud para el sector municipal.

Es muy interesante observar también que para el sector particular subvencionado existe un efecto positivo y significativo de grupos más heterogéneos, lo cual implica que una composición más diversa del colegio se ve más beneficiada de las mejores condiciones socioeconómicas promedio que presente este tipo de establecimientos.

En cuánto al sector particular pagado, no hay evidencia de efecto de heterogeneidad y no linealidad, de tal forma que la especificación de linealidad en medias es la que representa adecuadamente este tipo de establecimientos.

Finalmente, a modo de conclusión, para similares niveles de escolaridad, son los colegios municipales los que presentan un mayor efecto par para sus estudiantes, contando también con presencia de no linealidades. Por otro lado, el único sector que se ve afectado por la heterogeneidad de la sala de clases es el particular subvencionado el cuál, al igual que el municipal, cuenta con efecto no lineal convexo. Por último, el sector particular pagado solo cuenta con efecto par lineal de menor magnitud al encontrado para el municipal y de mayor magnitud al del particular subvencionado.

Si asumimos (en base a los datos que tenemos que se describieron en la sección anterior) que es en el sector particular pagado donde se concentran los alumnos con mejores condiciones socioeconómicas, en el sector subvencionado aquellos de niveles medios y en el sector municipal se concentran los estudiantes vulnerables, los resultados observados a nivel general está también en línea con la literatura internacional donde se ha observado que tanto alumnos de mayor NSE como de menos NSE son los más afectados positivamente por contar con pares de mejor NSE en comparación

con los grupos medios (Hoxby y Weingarth (2005)).

Estimaciones a nivel colegio

La misma regresión general que se utilizó para estimar los coeficientes a nivel de alumno, fue utilizada a nivel colegio (ecuación 2 de la sección III). Las diferencias entre ambos modelos es que ahora para el caso de los colegios la variable dependiente será el promedio del puntaje en la prueba Simce de matemática obtenido por el colegio.

Los resultados de las estimaciones se encuentran en la tabla 5 del anexo en la cual se observa que los resultados obtenidos son muy similares a cuando se hace a nivel de alumnos. . Por su parte, al realizar las estimaciones en forma diferenciada de acuerdo a la dependencia institucional de los establecimientos, se concluye que nuevamente es mayor el efecto entre los colegios municipales que particulares subvencionados y pagados, aunque ello proviene de efectos diferenciados entre los tres el efecto lineal es mayor para el sector particular pagado aunque cuenta con efectos negativos y altamente significativos de la heterogeneidad par.

Una vez que se estima la misma especificación para cada sector ((1.1c), (1.2c) y (1.3c)), el coeficiente de la heterogeneidad pasa a ser positivo para el sector municipal y particular subvencionado siendo mayor para este último (mismos resultados que los obtenidos por las estimaciones a nivel alumno), mientras que para el sector particular pagado es negativo.

Finalmente, hay evidencia de que para el sector municipal y particular subvencionado el efecto sería convexo, resultado similar al observado a nivel individual.

VI) CONCLUSIONES

Esta investigación analiza la influencia de la composición de la sala de clases en el rendimiento individual para los diferentes tipos de colegios en Chile. Es muy poca la literatura existente en esta materia a nivel nacional y a diferencia de los estudios ya realizados, este estudio se centra exclusivamente en el efecto composicional, no solamente desde la perspectiva de identificar no linealidades en su efecto, sino que también busca identificar la existencia de un efecto diferenciado por sector y por la diversidad en la composición socioeconómica de los alumnos de cada curso teniendo como objetivo determinar la existencia de la heterogeneidad del efecto par.

Las principales conclusiones que pudimos obtener a nivel individual es que para los alumnos de 4to básico que rindieron el SIMCE 2010 de matemática existe un significativo y positivo efecto par para los tres sectores educacionales, el cual tiene una función convexa para los estudiantes que asisten a colegios municipales y particulares subvencionados, siendo mayor para el municipal. Por su parte, en el sector particular pagado no encontramos evidencia de no linealidades.

Por otro lado, se encuentra que si bien la heterogeneidad en la composición de los estudiantes al interior de los colegios no es relevante para los sectores municipal y particular pagado, si lo es para el particular subvencionado,

afectando positivamente a los alumnos pertenecientes a este sector.

En general, se demuestra que el efecto par tiene una influencia mayor en los alumnos que pertenecen al sector municipal, es decir en los más vulnerables, resultados que pueden tener muchas implicancias al momento de crear políticas educacionales que influyan en la distribución de los alumnos a través de los tres sectores.

A nivel colegio, los resultados son bastante consistentes con los obtenidos a nivel individual, el efecto par es significativo para entender la heterogeneidad de los resultados académicos entre colegios y además se observa que, al igual a los obtenidos a nivel individual, hay evidencia de no linealidad para el sector municipal y particular subvencionado, de mayor magnitud en el primero, mientras que entre los colegios particulares pagados sólo se observa una relación lineal.

Finalmente, se concluye que la heterogeneidad es relevante para los tres sectores. Sin embargo, mientras los colegios pertenecientes al sector municipal y particular subvencionados se ven afectados positivamente, este efecto es considerablemente mayor entre éstos últimos, para los particulares pagados este efecto es negativo, es decir, colegios pertenecientes a este sector con mayor diversidad socioeconómica tendrán menores resultados promedio en el Simce de matemática en 4º básico.

VII) BIBLIOGRAFÍA

Ammermueller, A. y Pischke, J.S., (2006). Peer effects in European primary schools: Evidence from PIRLS. NBER Working Paper No. 12180.

Base de Datos SIMCE [años]. Santiago, Chile: SIMCE, Ministerio de Educación.

Betts, J.R. (2011). The Economics of Tracking in Education, Handbook of the Economics of Education, 3(4), 341 – 381.

Betts, J.R. y Zau, A. (2004). Peer groups and academic achievement: Panel evidence from administrative data. Unpublished manuscript.

Contreras, D. Sepúlveda, P. y Bustos, S. (2010). When Schools Are the Ones that Choose: The Effects of Screening in Chile. *Social science quarterly* 91 (5):1349-1368.

Duflo, E., Dupas, P. y Kremer, M. (2008). Peer effects, Teacher Incentives and the impact of tracking: Evidence from a randomized evaluation in Kenya, NBER Working Paper No. 14475, Noviembre 2008.

Dupriez, V. (2010). Fundamentals of educational planning; Methods of grouping learners at school. Paris, France: UNESCO.

Elacqua, G, Schneider, M. Y Buckley, J. (2006). “School Choice in Chile: Is it Class or the Classroom?” *Journal of Policy Analysis and Management* 25(3):577–601.

Hanushek, E. (2007). Education Production Functions: Developed Countries Evidence. In Dominic J. Brewer and Patrick J. McEwan (Ed.), *Economics of Education* (Amsterdam: Elsevier, 2010), 2010, pp. 132-136.

Hoxby, C. y Weingarth, G. (2005). Taking race out of the equation: School reassignment and the structure of peer effects. Mimeograph.

Manski, C. F. (1993). "Identification of endogenous social effects: The reflection problem." *Review of Economic Studies*, 60(3), 531–542.

McEwan, P. (2003). "Peer effects on student achievement: evidence from Chile", *Economics of Education Review* Vol. 22 (2): 131-141.

McEwan, P. J. (2001). The effectiveness of public, Catholic, and non-religious private schooling in Chile's voucher system. *Education Economics*, 9(2), 103–128.

Mizala, A. y Torche, F. (2012). Bringing the schools back in: the stratification of educational achievement in the Chilean voucher system. *International Journal of Educational Development*, 32, 132-144.

Morales, P. (2011). El análisis factorial en la construcción e interpretación de tests, escalas y cuestionarios. Universidad Pontificia Comillas, Madrid. Facultad de Ciencias Humanas y Sociales.

Raitano, M. y Vona, F. (2011). Peer heterogeneity, school tracking and student's performances: Evidence from PISA 2006. Working Paper Dipartimento di Economia Pubblica, n° 143.

Rangvid-Schindler B. (2007). "School composition effects in Denmark: quantile regression evidence from PISA 2000", *Empirical Economics* 32: 359-386.

Rodríguez, J. (2010). La no linealidad del efecto par educacional: Evidencia para Chile. Recuperado desde <http://www.econ.uchile.cl/uploads/addon/archivo/post/5716/Rodriguezpeereffect2010122010.pdf>

Sacerdote, B. (2011). Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far? "Handbook of the Economics of Education" 3(4), 249 – 277.

Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *Q.J. Econ.* 116, 681-704.

Sapelli, C. y Vial, B., (2005). Private vs. public voucher schools in Chile: new evidence on efficiency and peer effects. Working Paper N 289 Instituto de Economía, Universidad Católica de Chile.

Schuschny, y Soto (2009). Guía metodológica Diseño de indicadores compuestos de desarrollo sostenible. (LC/W.255), Documento de proyectos n°225, Comisión económica para América Latina (CEPAL)/GTZ mayo.

Valenzuela, J.P., Bellei, C. y De Los Ríos, D. (2011). "Selección e Integración social, ¿Asignatura pendiente?". Segregación escolar en Chile. Martinic, Sergio

y Gregory Elacqua (Eds.) (2010) Fin de ciclo? Cambios en la gobernanza del sistema educativo. UNESCO – PUC, Santiago, Chile. (8), 209 – 229.

VIII) ANEXO

Tabla 1
Descripción de variables y estadística descriptiva, 4to básico 2010.

Variable	Descripción	Media	Desv. Est.
Características individuales			
Puntaje matemática		254	53.34
Sexo	1 si es hombre	0.5	0.5
Kinder	1 si asistió a preescolar	0.88	0.32
Repitencia	1 si repitió alguna vez	0.09	0.29
Asistencia	1 si cuenta con buena asistencia	0.82	0.39
Características familiares			
Esec	Índice de Estado socioeconómico y cultural; Incluye escolaridad de padre y madre, ingresos familiares y cantidad de libros en casa	0.02	1.03
Cantidad personas	Número personas que vive con alumno	4.50	1.76
Expectativas	1 si apoderado espera estudios superiores	0.88	0.33
Características colegio/curso			
Dependencia (3 dummies)	1 si Municipal	0.38	0.49
	1 si subvencionado	0.54	0.5
	1 si particular pagado	0.08	0.27
Esec curso	"Esec" promedio del curso, neto del individual	0.14	1.26
D.E. Esec	Desviación estándar "esec"	1.11	0.22
Ruralidad	1 si establecimiento se encuentra en sector rural	0.94	0.24
Alumnos colegio	Cantidad alumnos que rindieron la prueba	71.2	45.9
Selección habilidad	1 si colegio selecciona por habilidad	0.72	0.45
Sel. por Características familiares	1 si colegio selecciona en base a características familiares	0.18	0.38
Porcentaje repitentes	Porcentaje alumnos repitentes por curso	0.09	0.09
Características profesores			
Experiencia profesor	Años de experiencia del profesor	16.6	12.4
Postgrado	1 si profesor tiene postgrado	0.55	0.5
Expectativas profesor	1 si profesor espera que alumnos vayan a ed. Superior	0.66	0.47
Preocupación director	1 capturando si director se preocupa de alumnos	0.36	0.48

N° observaciones

167939

Tabla 1.2

Estadística descriptiva para el conjunto de estudiantes y por dependencia institucional 4to básico 2010.

Variables	Total		Municipal		Subvencionado		Particular pagado	
	Media	Desv. Est.	Media	Desv. Est.	Media	Desv. Est.	Media	Desv. Est.
Puntaje matemáticas	254	53.34	239.47	51.7	261.34	50.48	301.89	43.26
<i>Variables Individuales/Familiares</i>								
Sexo (1 si es hombre)	0.5	0.5	0.5	0.5	0.5	0.5	0.49	0.5
Asistió a kínder	0.88	0.32	0.85	0.35	0.9	0.3	0.95	0.21
Ha Repetido	0.09	0.29	0.15	0.35	0.07	0.25	0.02	0.15
Asistencia (1 si se ausentó menos de 10 días al año)	0.82	0.39	0.79	0.41	0.83	0.38	0.89	0.32
Esec	0.15	1.67	-0.74	1.39	0.38	1.46	2.79	0.75
No persona en la familia	4.5	1.76	4.6	1.91	4.42	1.67	4.4	1.57
Expectativas	0.88	0.33	0.8	0.4	0.92	0.27	1	0.06
<i>Variables Curso/Colegio</i>								
Esec curso	0.14	1.26	-0.75	0.74	0.37	0.95	2.78	0.36
Desv. estándar Esec	1.11	0.22	1.2	0.17	1.12	0.18	0.66	0.16
Urbanidad	0.94	0.24	0.88	0.32	0.98	0.16	0.98	0.14
Alumnos colegio (4º básico)	71.2	45.9	62.08	36.45	77.14	50.99	73.9	42.37
Selección habilidad	0.72	0.45	0.55	0.5	0.8	0.4	0.9	0.3
Sel. por características económicas	0.18	0.38	0.06	0.24	0.22	0.42	0.42	0.49
Porcentaje repitentes	0.09	0.09	0.15	0.1	0.07	0.08	0.02	0.03
Años de experiencia profesor	16.6	12.4	20.77	13.13	13.71	11.11	16.59	11.79
Profesores con postgrado o postítulo	0.55	0.5	0.6	0.49	0.52	0.5	0.52	0.5
Expectativas profesor	0.66	0.47	0.52	0.5	0.71	0.45	0.98	0.15
Preocupación director	0.36	0.48	0.39	0.49	0.35	0.48	0.26	0.44
Nº observaciones	167939		63574		91099		13266	

Tabla 2

Multinomial Logit de modelo elección colegio: determinantes de atender a municipal, subvencionado o particular pagado. 4to básico 2010.

	Municipal		Subv.	
	Coef.	SE	Coef.	SE
"Esec"	-2.450***	(-0.0179)	-1.971***	(-0.0175)
Sexo	0.00814	(-0.0272)	0.000499	(-0.0257)
Expectativas	-0.915***	(-0.177)	-0.551***	(-0.177)
Kinder	-0.0437	(-0.0589)	-0.0221	(-0.0574)
Repitencia	0.237***	(-0.0787)	-0.118	(-0.0776)
Asistencia	-0.424***	(-0.0396)	-0.293***	(-0.0381)
Cantidad personas	0.00816**	(-0.00333)	-0.00179	(-0.00327)
N colegios municipales por km ² en la municipalidad donde estudiante vive	0.569***	(-0.1)	-0.366***	(-0.0962)
N colegios part. subvencionados por km ² en la municipalidad donde estudiante vive	-0.226***	(-0.0514)	0.843***	(-0.048)
N colegios part. pagados por km ² en la municipalidad donde estudiante vive	-1.064***	(-0.0458)	-2.000***	(-0.0432)
Constant	6.508***	(-0.192)	6.437***	(-0.191)
LR chi2 (20)				78524.55
Prob > chi2				0.0000
Pseudo R2				0.2598
Observaciones				167,939

Tabla 3 : Regresión MCO corregido por sesgo de selección. Rendimiento estudiantes en puntaje Matemática Simce 2010, 4tos básico⁸⁹.

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
"Esec" promedio		6.014***	5.929***	5.669***	5.709***	7.153***
		(-0.473)	(-0.481)	(-0.478)	(-0.481)	(1.011)
Dev. Estándar "esec"			-1.495		1.452	1.832
			(-1.781)		(-1.976)	(2.771)
Cuadrado "esec" promedio				0.799***	0.897***	2.013***
				(-0.247)	(-0.276)	(0.671)
Esec promedio * Subv						-1.537
						(1.124)
Dev. Estándar "esec" *						0.728
						(3.997)
Cuadrado "esec" * Subv.						-1.092
						(0.796)
Esec promedio * PP						7.621
						(5.032)
Dev. Estándar "esec" * PP						-10.12
						(7.313)
Cuadrado "esec" * PP						-
						(1.105)
Variables de control						
Características individuales	√	√	√	√	√	√
Características familiares	√	√	√	√	√	√
Características colegio	√	√	√	√	√	√
Características profesores	√	√	√	√	√	√
Constant	215.2***	220.9***	222.6***	220.4***	218.7***	219.9***
	-1.97	-2.003	-2.881	-2.007	-3.19	-3.903
Observations	166,919	166,919	166,919	166,919	166,919	166,919
R-squared	0.228	0.233	0.233	0.233	0.233	0.233

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

⁸ Para cada variable la primera línea muestra el coeficiente estimado y la segunda el error estándar robusto, ajustado por cluster a nivel colegio.

⁹ Fuente: Creación propia a partir del SIMCE 2010.

Tabla 4: Regresión MCO. Rendimiento estudiantes en puntaje Matemática Simce 2010, 4tos básico¹⁰¹¹.

VARIABLES	(1) Municipal	(1) Subv.	(1) Part. Pag	(2) Municipal	(2) Subv	(2) Part. Pag	(3) Municipal	(3) Subv.	(3) Part. Pag	(4) Municipal	(4) Subv	(4) Part. Pag
"Esec" promedio	7.878***	5.843***	9.795***	7.957***	5.772***	7.599**	10.06***	4.236***	7.352	10.09***	4.353***	4.932
	(-0.92)	(-0.597)	(-2.302)	(-0.927)	(-0.628)	(-3.077)	(-1.247)	(-0.682)	(-13.72)	(-1.254)	(-0.682)	(-14.68)
Desv. Estándar "esec"				-2.413	-0.808	-7.677				0.963	5.203*	-7.689
				(-2.73)	(-2.659)	(-6.339)				(-2.816)	(-2.93)	(-6.355)
Cuadrado "esec" promedio							2.261***	1.651***	0.47	2.328***	1.994***	0.513
							(-0.715)	(-0.388)	(-2.731)	(-0.743)	(-0.435)	(-2.836)
Variables de control												
Características individuales	√	√	√	√	√	√	√	√	√	√	√	√
Características familiares	√	√	√	√	√	√	√	√	√	√	√	√
Características colegio	√	√	√	√	√	√	√	√	√	√	√	√
Características profesores	√	√	√	√	√	√	√	√	√	√	√	√
Constante	230.5***	230.2***	213.0***	233.4***	231.1***	224.6***	229.2***	227.5***	216.0***	228.0***	221.3***	227.9***
	(-2.352)	(-3.062)	(-10.23)	(-4.111)	(-4.198)	(-14.31)	(-2.287)	(-3.051)	(-19.88)	(-4.132)	(-4.595)	(-23.91)
Observaciones	63,060	90,594	13,265	63,060	90,594	13,265	63,060	90,594	13,265	63,060	90,594	13,265
R-squared	0.142	0.158	0.091	0.142	0.158	0.092	0.143	0.159	0.091	0.143	0.159	0.092

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

¹⁰ Para cada variable la primera línea muestra el coeficiente estimado y la segunda el error estándar robusto, ajustado por cluster a nivel colegio.

¹¹ Fuente: Creación propia a partir del SIMCE 2010.

Tabla 5. : Regresión MCO corregido por sesgo de selección. Promedio colegio en puntaje Matemática Simce 2010, 4tos básico¹²¹³.

VARIABLES	(1c)	(2c)	(3c)	(1.1c) Municipal	(1.2c) Subv	(1.3c) Part. Pag.
"Esec" promedio	12.40***	11.74***	12.49***	14.31***	11.00***	17.80***
	-0.0697	-0.071	(1.144)	-0.158	-0.102	-0.962
Desv. Estándar "esec"	-1.632***	4.381***	4.040	2.233***	7.468***	-5.041***
	-0.264	-0.298	(2.768)	-0.458	-0.426	-1.013
Cuadrado "esec" promedio		1.751***	2.534***	2.132***	2.096***	0.0413
		-0.0408	(0.618)	-0.0933	-0.0626	-0.158
Esec promedio * Subv			-0.544			
			(1.205)			
Desv. Estándar "esec" * Subv			1.992			
			(3.954)			
Cuadrado "esec" promedio *			-0.868			
			(0.734)			
Esec promedio * PP			4.980			
			(4.016)			
Desv. Estándar "esec" * PP			-9.420			
			(7.063)			
Cuadrado "esec" promedio *			-2.547***			
			(0.752)			
Variables de control						
Características individuales						
Características familiares	√	√	√	√	√	√
Características colegio	√	√	√	√	√	√
Características profesores	√	√	√	√	√	√
Constant	250.8***	240.6***	242.1***	243.8***	237.8***	233.1***
	-0.392	-0.457	-0.6	-0.688	-0.687	-2.561
Observations	4116	4116	4116	1564	2223	329
R-squared	0.622	0.626	0.626	0.378	0.477	0.389

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

¹² Para cada variable la primera línea muestra el coeficiente estimado y la segunda el error estándar robusto.

¹³ Fuente: Creación propia a partir del SIMCE 2010.