



INFORMES TÉCNICOS MIDE UC / TECHNICAL REPORTS MIDE UC
Centro de Medición MIDE UC / Measurement Center MIDE UC

IT1201

El efecto de las características de los pares en el aprendizaje
de estudiantes chilenos de enseñanza media

SANDY TAUT & JORGE ESCOBAR

20 de junio de 2012



PONTIFICIA UNIVERSIDAD
CATÓLICA DE CHILE
ESCUELA DE PSICOLOGÍA



Abstract

La presente investigación se enfoca en el efecto de la diversidad social y de rendimiento de los pares en el aprendizaje de estudiantes chilenos de enseñanza media. La metodología de análisis considera una serie de modelos de regresión multinivel de tres niveles (estudiante, curso y escuela). Los datos utilizados corresponden a la prueba SIMCE para los años 2004 y 2006, cuando los mismos estudiantes de 8vo básico fueron examinados por segunda vez en segundo año medio. Se trabajó con datos longitudinales de 112,591 estudiantes distribuidos en 5,830 cursos en 1,792 establecimientos. Los resultados muestran que sobre todo el efecto de la diversidad social a nivel sala de clase tiene un efecto positivo sobre el aprendizaje. Este resultado es particularmente robusto en matemática, especialmente en establecimientos de dependencia municipal y particular subvencionado. El paper analiza interacciones interesantes entre las variables composicionales a nivel curso y variables a nivel individual.

La demanda por una educación equitativa, de igual calidad para todos, aun suena fuerte en Chile después de las recientes protestas estudiantiles durante el año 2011. Esa demanda parece especialmente entendible y urgente en un país en el cual el sistema escolar es altamente estratificado según el origen sociocultural de los alumnos y entre colegios municipales, particulares subvencionados y particulares pagados (Mizala & Romaguera, 2000). Chile es uno de los países donde un porcentaje muy alto de la varianza en los rendimientos se explica por diferencias entre escuelas y el origen sociocultural del alumno (Grupo Iberoamericano de PISA, 2010). Al mismo tiempo, dada la alta estratificación social, geográfica y educacional, además de la selectividad de los colegios particulares subvencionados y particulares pagados, la mayoría de los establecimientos educacionales chilenos se caracterizan por una alta homogeneidad interna respecto del nivel socioeconómico de su alumnado. Sin embargo, para lograr una mayor equidad en el sistema educacional, cada establecimiento y sala de clase, en principio debería aproximarse a los patrones socioeconómicos y de rendimiento que se encuentra en la población. Pero, ¿cuáles son los efectos de la composición de la sala o escuela en el aprendizaje, una vez descontado los factores individuales que inciden en el aprendizaje? ¿Cuáles son las características de los pares que inciden más o menos en el aprendizaje? ¿Es más importante el nivel escuela o el nivel curso?

Este paper se dedica a investigar los efectos de las características de los pares (el efecto llamado “composicional”) en el aprendizaje en la enseñanza media en Chile. El estudio indaga en los efectos del promedio y de la desviación estándar (d.e.) de tres características importantes de los pares para el contexto chileno: (1) nivel socioeconómico, (2) rendimiento previo, y (3) género (ver Figura 1).

Figura 1. Conceptualización del efecto par en el rendimiento.



Como variables dependientes se usan los aprendizajes de los estudiantes entre 1° y 2° medio, usando datos longitudinales de la aplicación censal de SIMCE 2004 (8vo básico) y SIMCE 2006 (2° medio), para Matemática y Lenguaje. Utilizamos una metodología cuantitativa aplicando métodos estadísticos avanzados de regresión múltiple multinivel de tres niveles: nivel individual, curso y escuela. También seguimos la lógica de modelos de valor agregado al controlar por el rendimiento previo a nivel individual (SIMCE 2004) (Lissitz, 2005, 2006; Bryk & Raudenbush, 2002; Raudenbush, 2004).

Antecedentes nacionales e internacionales

Primero, a nivel nacional, pocos estudios han investigado el efecto composicional en el aprendizaje. Segundo, aun menos han utilizado una metodología multinivel, sobre todo tomando en cuenta el nivel individual, *curso y escuela*. Tercero, de igual modo, hay pocos estudios que utilicen datos longitudinales de los estudiantes que pueden dar cuenta de su progreso en el aprendizaje entre dos puntos en el tiempo. No conocemos ningún estudio que abarque los tres puntos. Nuestro trabajo intenta llenar ese vacío.

En primer lugar, es interesante analizar la descomposición de la varianza del rendimiento escolar en base del SIMCE. Ramírez (2007) y Manzi et al. (2008) hicieron análisis en esa línea. Manzi et al. (2008) analizaron los datos SIMCE de varios años (2000-2006) y reportan en general porcentajes de varianza entre escuelas más altas en Matemática que en Lenguaje: en Matemática entre 26% y 45%, y en Lenguaje entre 23% y 34%. Luego, encontraron porcentajes de varianza entre escuelas crecientes de 4°

básico a 8vo básico, y de 8vo básico a 2° medio. Por ejemplo, en Matemática, para 4° básico la varianza entre escuelas variaba entre 26 y 28%, para 8vo entre 30 y 31%, y para 2° medio era de 45%. Ramírez (2007) ofrece una descomposición de la varianza de los resultados SIMCE 1999 4° básico en seis niveles (regiones, provincias, comunas, escuelas, salas, individuos) y encuentra que 21% está a nivel escuela, 4% a nivel curso y 69% entre individuos (los otros tres niveles combinados explican el restante 6%). Lo que llama la atención de su trabajo, es que la variabilidad de rendimiento dentro de un curso promedio es de 161 puntos SIMCE.

Torche y Mizala (2012) investigan el efecto contextual del nivel socioeconómico a nivel escuela en el rendimiento de los estudiantes en SIMCE (4° básico 2002 y 8vo básico 2004), con un interés particular en las diferencias de ese efecto entre el sector municipal versus particular subvencionado. En primer lugar, reportan que la varianza entre escuelas en términos del *nivel socioeconómico* de sus estudiantes es muy alta en Chile: 62%. Luego, analizando los datos en cada sector por separado, encuentran que la varianza entre escuelas es 24% en el sector municipal, pero 47% en el sector particular subvencionado. Interpretan que mientras el *sector* particular subvencionado tiene alta diversidad socioeconómica, *las escuelas* dentro de ese sector se caracterizan por la homogeneidad de su alumnado. La descomposición de la varianza del *rendimiento* tiene resultados similares. La varianza entre escuelas en el sector municipal generalmente es la mitad de la varianza entre escuelas del sector particular subvencionado. Por ejemplo, la varianza entre escuelas usando los resultados de Matemática de 8vo básico (SIMCE 2004) es de 15% en el sector municipal y de 28% en el sector particular subvencionado.

Torche y Mizala (2012) identifican una fuerte influencia del nivel socioeconómico promedio de la escuela en el rendimiento de sus alumnos, particularmente en el sector particular subvencionado. Ese efecto es más fuerte que el efecto que tiene el estatus socioeconómico a nivel individual. Su paper concluye con un llamado a explorar por qué el efecto contextual del NSE es tan fuerte en el sector particular subvencionado, y mencionan como posibles variables relacionados las expectativas y la calidad docente, implementación del currículo, recursos de la escuela, entre otros, que son modificables. Sin embargo, si estas variables no logran explicar o absorber gran parte del efecto del NSE, entonces es la estratificación social misma la que se tiene que reducir para lograr oportunidades de aprendizaje más equitativas. Al mismo tiempo, encuentran un efecto significativo y positivo de la heterogeneidad en el

rendimiento de estudiantes en colegios municipales, no así en colegios particulares subvencionados.

El estudio de McEwan (2003) tiene un propósito similar al nuestro al analizar datos SIMCE de 1997 en búsqueda de efectos de pares en el rendimiento de estudiantes de 8vo básico (en lenguaje y matemática). Sin embargo, McEwan no tuvo disponible datos longitudinales de los alumnos y no consideró la estructura anidada de los datos (individuo, sala de clase, escuela). Usando regresión múltiple OLS sin y con efectos fijos de la escuela, incluyó las siguientes variables de los alumnos: género, raza, nivel educacional de la madre y del padre, ingreso familiar, número de libros en el hogar. Para modelar el efecto de los pares, incluyó el promedio del curso del nivel educacional de la madre y del padre, raza e ingreso. Los resultados no varían mucho entre los modelos con y sin efectos fijos de la escuela, indicando que no parece importante controlar por efectos de asignación no aleatoria *entre escuelas* al modelar el efecto de los pares -- un resultado que parece intuitivamente dudoso en el contexto chileno con un sistema educacional altamente estratificado y tomando en cuenta la selección de alumnos de parte de los colegios. McEwan (2003) encuentra un efecto promedio a nivel sala de clase de la educación de la madre en el rendimiento en Lenguaje: un aumento de una d.e. en el nivel promedio de educación de la madre resulta en un aumento de 0.27 d.e. en el rendimiento en Lenguaje.

A nivel internacional, Hattie (2003) entrega un meta-análisis del efecto composicional en el aprendizaje y concluye que no hay suficiente evidencia que el hecho de agrupar estudiantes según su nivel de rendimiento (o sexo) les beneficia en términos de su aprendizaje. Su revisión de la literatura lo lleva a constatar que los efectos composicionales tienen un impacto más fuerte en actitudes individuales y la equidad del sistema, que en el aprendizaje propiamente tal. Hattie concluye que la discusión de efectos composicionales distrae la atención de lo que realmente importa: la calidad de las prácticas docente en aula, sin importar los niveles de diversidad al interior de cada sala de clase. Es decir, los docentes deberían poder hacerse cargo de la homogeneidad o heterogeneidad de sus estudiantes y entregarles a todos los cursos por igual su enseñanza utilizando las prácticas más eficaces.

Opendakker & Van Damme (2001) reportan un estudio multinivel de tres niveles (individuo, curso, escuela) de la relación entre la composición de la escuela y los

procesos escolares, y sus efectos en el aprendizaje en Matemática de estudiantes belgas de enseñanza media (pero sin usar datos longitudinales de aprendizaje). Como variables a nivel individual utilizaron inteligencia numérica medido al comienzo del año escolar (como indicador similar al rendimiento previo), idioma hablado en casa, género, y nivel educacional del padre. A nivel escuela utilizaron las siguientes variables composicionales: el promedio de inteligencia numérica, la d.e. de inteligencia numérica, la proporción de estudiantes que hablan holandés en sus casas, la proporción de mujeres, y el promedio del nivel educacional del padre. Los autores encuentran un efecto no significativo de la d.e. de la inteligencia numérica a nivel escuela y un efecto muy significativo del promedio de esa variable a nivel escuela en el rendimiento individual en Matemática, una vez descontado el efecto de NSE e inteligencia numérica a nivel individual. El efecto de variables de procesos escolares disminuye al incluir efectos composicionales. También examinan efectos diferenciales por grupo de estudiantes: alto rendimiento, rendimiento medio, y bajo rendimiento. Los autores concluyen, “stimulating the selectivity of schools enhances the relationship between characteristics of school process and school composition and reduces the potential effectiveness of malleable conditions of schooling” (p. 423).

Es interesante notar que los autores argumentan que un aumento de heterogeneidad en el rendimiento promedio de las escuelas tendría efectos positivos para estudiantes de bajo rendimiento y no tendría efecto para estudiantes de rendimiento medio, pero que estudiantes de escuelas con un alto rendimiento promedio no se verían beneficiados por un aumento en la heterogeneidad. Sus resultados también destacan que especialmente niños de alto rendimiento pero de un bajo nivel sociocultural se ven beneficiados por una composición favorable de su escuela. El paper concluye con un llamado a explorar los efectos de variables docente para explicar los efectos conjuntos de variables de composición y de procesos escolares.

Un estudio inglés (Gibbons & Telhaj, 2006) también es muy similar al nuestro en utilizar datos longitudinales antes y después de la transición de la enseñanza básica a la enseñanza media (dos mediciones en el tiempo en la misma cohorte, a los 11 y 15 años de edad). Los autores argumentan que estos datos permiten estimaciones más válidas del efecto de pares “nuevos” ya que reflejan cambios en la composición de los pares al llegar a las escuelas de enseñanza media. Encuentran que el efecto de los pares “nuevos” es más fuerte pero en comparación con otras variables, todavía se considera

algo débil: un mejoramiento del rendimiento de estos pares en una d.e. resulta en el aumento en rendimiento individual de 0.05-0.08 d.e. después de un periodo de tres años. Otro resultado interesante es que los efectos de la diversidad social de los pares (medidos a través de % promedio de niños que reciben almuerzo escolar gratuito) solo tiene un efecto hasta introducir el rendimiento previo promedio de los pares en el modelo.

Un estudio destacable viene de Rumberger & Palardy (2003) que van en la misma dirección de Opdenakker & Van Damme (2001) al investigar posibles explicaciones de los efectos composicionales a nivel escuela. Los autores utilizan un modelo multinivel de tres niveles y datos longitudinales con tres mediciones en el tiempo, donde su primer nivel es un modelo de crecimiento intraindividual, el nivel dos es el nivel individual, y el nivel tres el nivel escuela. En un primer paso a nivel contextual, los autores incluyen variables composicionales tales como promedio de NSE, raza, rendimiento, entre otros, y encuentran que el efecto promedio de NSE tiene un efecto casi igual de importante en el crecimiento de los aprendizajes que el efecto de NSE a nivel individual. Estos efectos son pequeños pero sustanciales (un aumento en una d.e. en el NSE resulta en un aumento en los aprendizaje de 0.11 d.e. – en el caso de NSE escuela - o 0.12 d.e. – en el caso de NSE individual, en promedio entre las distintas asignaturas) y afectan a todos los estudiantes, de alto y bajo rendimiento.

Adicionalmente, después de haber controlado por variables individuales y de composición a nivel escuela, las variables estructurales y de recursos no fueron significativas. Sin embargo, procesos y prácticas escolares (como las expectativas de los docentes y el clima escolar) sí fueron relevantes, ya que al incluirlas en el modelo, el efecto composicional del NSE promedio desapareció.

Hay algunos estudios que investigan de manera comparativa el efecto composicional en el aprendizaje. Por ejemplo, Dronkers (2010) utiliza datos PISA con una metodología multinivel y distinguiendo el efecto promedio del efecto de diversidad/variabilidad. Su trabajo llega a la conclusión de que en los sistemas de alta estratificación social existe un efecto positivo muy alto de la heterogeneidad social dentro de las salas de clases. En sus análisis multinivel de 17 países de la OECD, Vandenberghe (2002) concluye que el nivel socioeconómico promedio de los pares

tiene un efecto más fuerte en el rendimiento individual que el NSE individual, y que estudiantes de bajo rendimiento son los más sensibles a los efectos composicionales.

Preguntas de investigación e hipótesis

Nuestras preguntas de investigación son las siguientes:

1. ¿Cuál es el efecto composicional (entendido como las características de los pares a nivel sala de clases y escuela) en el aprendizaje de alumnos de enseñanza media en Chile?
2. ¿Cuáles variables composicionales tienen menor o mayor incidencia?
3. ¿Tienen alumnos de cursos de alta diversidad (socioeconómica, de rendimiento y de género) mejores o peores aprendizajes que alumnos de cursos con baja diversidad?
4. ¿Hay efectos diferenciales de las variables composicionales para alumnos de distinto nivel de aprendizaje u origen sociocultural?
5. ¿Cómo se comparan los resultados por dependencia?

Nuestras hipótesis son:

Hay un efecto composicional social, de rendimiento previo y de género *por sobre* el efecto de características individuales en los aprendizajes de alumnos de enseñanza media en Chile (Hoxby, 2002; McEwan, 2003; Hanushek et al., 2003; Gibbons & Telhaj, 2006; Rumberger & Palardy, 2005).

A nivel general en Chile, estudiantes de cursos con alta diversidad muestran mejores aprendizajes que estudiantes de cursos con baja diversidad (Dronkers, 2010).

Estudiantes de bajo desempeño y de nivel socioeconómico bajo se ven más beneficiados por esta diversidad que estudiantes de alto rendimiento y de nivel socioeconómico alto (Schneeweis & Winter-Ebmer, 2006; Vandenberghe, 2002).

Debido a que los estudiantes de las escuelas municipales y particulares subvencionados obtienen resultados de aprendizaje más bajos y tienen un nivel sociocultural más bajo, deberían beneficiarse más de la diversidad que estudiantes de colegios particulares pagados.

Datos y variables

Bases de datos

El Sistema de Medición de la Calidad de la Educación (SIMCE) es el sistema nacional chileno de evaluación del rendimiento escolar del Ministerio de Educación. Esta prueba se aplica a todos los estudiantes chilenos (aplicación censal). A partir del año 2005 la aplicación de la prueba se alternó de 4° básico, 8° básico y 2° medio. A partir del año 2006 se evalúa todos los años a 4° básico y se alternan 8° básico y 2° medio (ver www.simce.cl). Son pruebas de conocimientos curriculares en varios subsectores en la cual no existe una escala estandarizada, sin embargo, para aspectos generales se considera una escala de un mínimo de 100 y un máximo de 400 puntos, con un promedio de 250 y una desviación estándar de 50 puntos.

Además de la prueba de conocimientos curriculares, SIMCE recoge información de contexto a través de un cuestionario para los padres de los estudiantes testeados. Este cuestionario tiene como fin indagar sobre el nivel educacional de los padres, el ingreso del hogar y el nivel de satisfacción con el establecimiento, entre otros.

Para este estudio se consideraron las pruebas SIMCE 2004 y SIMCE 2006 a alumnos de 8° básico y 2° medio, en Lenguaje y Matemática, respectivamente. El total de alumnos que rindió la prueba SIMCE el año 2006 fueron 244,594, sin embargo para una cantidad de alumnos que rindió esta prueba no se tabuló su respectivo género, con lo cual se trabajó con una cantidad de 243,486 alumnos. De los cuales 118,429 son estudiantes hombres y 125,057 son estudiantes mujeres. De los 243,486 alumnos considerados estos pertenecían a 7,530 cursos en 2,454 colegios, divididos en colegios municipales, particulares subvencionados y particulares pagados.

Por otro lado, del cuestionario de los padres se obtuvo la respuesta de 231,320 de estos. Sin embargo muchas de las respuestas de los padres se perdieron o algunos ítems tenían dos respuestas. Para simplicidad de cálculos y considerando solo las variables de interés en el estudio, se consideraron solamente los casos en los que había una sola respuesta por ítem.

Finalmente, después de haber descartado los casos perdidos en ambas pruebas tanto para los estudiantes como para el cuestionario de los padres, se está trabajando con un universo de 112,591 estudiantes.

En la Tabla 1 se muestra la base de datos con la cual se está trabajando separado por dependencia educacional.

Tabla 1: Cantidad de estudiantes, cursos y escuelas por dependencia escolar

	Estudiantes	Cursos	Escuelas
Municipal	50,476	2,720	589
Subvencionado	52,611	2,570	969
Pagado	9,504	540	234
Total	112,591	5,830	1,792

Variables

Las variables dependientes en este estudio son los resultados de los alumnos en el SIMCE 2006 en lenguaje y matemática. Las variables independientes se diferencian en variables individuales (las que son propiamente tal del alumno) y contextuales (las que pertenecen al curso y escuela). Las variables independientes de tipo individual son: los resultados 2004 en las pruebas de lenguaje y matemática, el género (hombre o mujer), la repitencia (si el estudiante ha repetido alguna vez o no), si la madre pertenece a algún tipo de etnia indígena y el ingreso familiar del estudiante. Por otro lado, las variables independientes contextuales son: la desviación estándar del ingreso familiar (diversidad social del curso en 2004), la desviación estándar de la prueba SIMCE 2004 (diversidad de rendimiento del curso en 2004), ingreso familiar promedio del curso, promedio del curso en las pruebas SIMCE 2004, porcentaje de mujeres, porcentaje de alumnos repitentes, porcentaje de alumnos cuya madre pertenece a algún tipo de etnia indígena, ruralidad (si el colegio es rural o urbano) y la dependencia de la escuela.

De las variables contextuales las únicas que existían previamente eran las de ruralidad y dependencia. Las otras variables fueron construidas a partir de variables individuales. La descripción de las variables independientes se presenta en la Tabla 2 y Tabla 3.

Tabla 2: Variables independientes a nivel individual

Variable	Ítem	Respuesta
Rendimiento 2004	Resultado en la prueba SIMCE 2004 de lenguaje o matemática	
Mujer		1. Hombre 2. Mujer
Repite	El alumno, ¿ha repetido de curso?	1. Nunca 2. Una o más veces
Indígena madre	¿Pertenece la madre a algún pueblo indígena?	1. No 2. Si
Ingreso familiar bruto	En un mes normal, la suma de los ingresos de todas las personas que viven en el hogar del alumno está:	1. Bajo \$ 100,000 2. Entre \$ 100,000 y \$ 200,000 3. Entre \$ 200,001 y \$ 300,000 4. Entre \$ 300,001 y \$ 400,000 5. Entre \$ 400,001 y \$ 500,000 6. Entre \$ 500,001 y \$ 600,000 7. Entre \$ 600,001 y \$ 800,000 8. Entre \$ 800,001 y \$ 1,000,000 9. Entre \$ 1,000,001 y \$ 1,200,000 10. Entre \$ 1,200,001 y \$ 1,400,000 11. Entre \$ 1,400,001 y \$ 1,600,000 12. Entre \$ 1,600,001 y \$ 1,800,000 13. Sobre \$ 1,800,000

Tabla 3: Variables independientes a nivel contextual

Variable	Ítem
Heterogeneidad social	Desviación estándar del Ingreso familiar bruto del curso en 2004
Heterogeneidad rendimiento	Desviación estándar del resultado SIMCE 2004 lenguaje o matemática del curso
Promedio del ingreso familiar bruto	Promedio del ingreso familiar bruto del curso en 2004
Promedio de rendimiento	Promedio en la prueba SIMCE 2004 de lenguaje o matemática del curso
Porcentaje de mujeres	Porcentaje de mujeres por curso
Porcentaje de repitentes	Porcentaje de alumnos repitentes por curso
Porcentaje de madres indígenas	Porcentaje de alumnos cuya madre pertenezca a una etnia indígena por curso
Ruralidad del colegio	1. Urbano 2. Rural
Dependencia del colegio	1. Municipal 2. Particular subvencionado 3. Particular pagado

Análisis y metodología

La metodología utilizada para hacer las estimaciones son modelos multinivel (también conocidos como modelos lineales jerárquicos) de tres niveles, donde se utilizará como primer nivel el estudiante, segundo nivel el curso y tercer nivel la escuela (Bryk & Raudenbush, 2002; Hox, 2010). Tal como Rumberger y Palardy (2003), primero analizamos un modelo solo con variables a nivel individual, para controlar los efectos del origen individual en el aprendizaje. Con esto tratamos de hacer desaparecer las diferencias que existen entre estudiantes al empezar la enseñanza media en sus respectivas escuelas y cursos. Segundo, incluimos variables a nivel curso y escuela (contextual). En un primer paso, queremos ver si la varianza entre cursos o escuelas se puede atribuir a variables composicionales incluidos en el modelo (ver Figura 1). En un segundo paso, incluimos variables contextuales que se escapan del control de escuelas o docentes (ruralidad y dependencia). Finalmente, también testeamos interacciones entre niveles, entre variables composicionales del curso e individuales (ingreso familiar, rendimiento previo, género).

El modelo completo multinivel (sin interacciones) que se utilizará para hacer las estimaciones de las variables antes descritas está dada por la siguiente ecuación matemática:

$$Y_{ijk} = \beta_{0jk} + \beta_{1jk}X_{ijk} + e_{ijk} \longrightarrow \text{Nivel 1}$$

$$\beta_{0jk} = \pi_{00k} + \pi_{0jk}Z_{jk} + r_{0jk} \longrightarrow \text{Nivel 2}$$

$$\beta_{1jk} = \pi_{ijk}$$

$$\pi_{00k} = \gamma_{000} + \gamma_{00k}W_k + \mu_{00k} \longrightarrow \text{Nivel 3}$$

$$\pi_{0jk} = \gamma_{0jk}$$

$$\pi_{ijk} = \gamma_{ijk}$$

$$X_{ijk} = \text{Contiene variables a nivel estudiante (nivel 1)}$$

$$Z_{jk} = \text{Contiene variables a nivel curso (nivel 2)}$$

$$W_k = \text{Contiene variables a nivel escuela (nivel 3)}$$

γ_{000} = Promedio general del rendimiento de todos los estudiantes (gran intercepto)

γ_{ijk} = Pendientes nivel individual

γ_{0jk} = Pendientes nivel curso

γ_{00k} = Pendientes nivel escuela

Donde:

Y_{ijk} = Rendimiento SIMCE 2006

X_{1jk} = Rendimiento SIMCE 2004

X_{2jk} = Si el estudiante es mujer o no

X_{3jk} = Repite

X_{4jk} = Indígena madre

X_{5jk} = Ingreso familiar bruto

Z_{1k} = Heterogeneidad social

Z_{2k} = Heterogeneidad rendimiento

Z_{3k} = Promedio de ingreso familiar bruto

Z_{4k} = Promedio de rendimiento

Z_{5k} = Porcentaje de mujeres

Z_{6k} = Porcentaje de repitentes

Z_{7k} = Porcentaje de madres indígenas

W_1 = Ruralidad

W_2 = Dependencia

μ_{00k} = Efecto aleatorio a nivel escuela

τ_{0jk} = Efecto aleatorio a nivel curso

e_{ijk} = Efecto aleatorio a nivel individual

Finalmente la expresión matemática en su forma expandida es:

$$Y_{ijk} = \gamma_{000} + \gamma_{ijk}X_{ijk} + \gamma_{0jk}Z_{jk} + \gamma_{00k}W_k + \mu_{00k} + r_{0jk} + e_{ijk}$$

El nivel de referencia para las estimaciones de la dependencia escolar son las escuelas municipales.

Resultados

La presentación de los resultados se organiza en tres sub-secciones. La primera de ella se enfoca en los resultados descriptivos que se obtuvieron de las variables. La segunda parte muestra los resultados inferenciales, esto es, cómo y cuánto predicen las covariables (variables independientes) sobre las dependientes (resultados de la prueba SIMCE 2006 de lenguaje y matemática). Las estimaciones se realizan utilizando toda la muestra, así como también diferenciado por dependencia escolar. En la tercera parte se analizan posibles interacciones entre variables para ver si existe algún efecto diferencial de alguna de las variables independientes sobre el rendimiento escolar.

Descriptivos

En la Tabla 4 y Tabla 5 se presenta un resumen descriptivo de las variables continuas, tanto a nivel individual como contextual utilizando toda la muestra.

Tabla 4: Resumen descriptivos de las variables continuas a nivel individual

Variable	Mínimo	Media	Máximo	DE
Lenguaje 2006	120.43	263.84	398.16	51.71
Matemática 2006	93.46	264.78	426.58	64.85
Lenguaje 2004	94.20	265.44	391.67	49.48
Matemática 2004	118.79	266.18	406.09	49.51
Ingreso familiar	1.00	3.85	13.00	3.05

Tabla 5: Resumen descriptivos de las variables continuas a nivel contextual

Variable	Mínimo	Media	Máximo	DE
Heterogeneidad social	0.00	1.65	4.29	0.74
Heterogeneidad lenguaje 2004	11.23	39.79	82.86	7.63
Heterogeneidad matemática 2004	11.11	37.87	75.66	7.18
Promedio ingreso familiar	1.00	3.85	13.00	2.49
Lenguaje 2004 promedio	164.20	265.40	353.90	29.89
Matemática 2004 promedio	183.20	266.20	367.40	32.30
Porcentaje de mujeres	0.00	0.54	1.00	0.24
Porcentaje de repitentes	0.00	0.09	1.00	0.11
Porcentaje de madre indígena	0.00	0.06	1.00	0.10

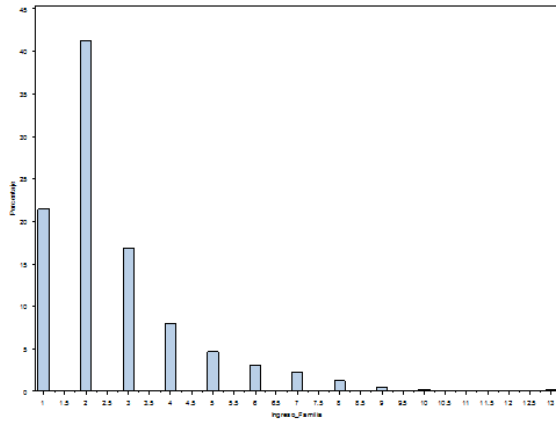
Finalmente en la Tabla 6 se puede observar un resumen descriptivo de las variables discretas tanto a nivel individual como contextual.

Tabla 6: Resumen descriptivo de las variables discretas

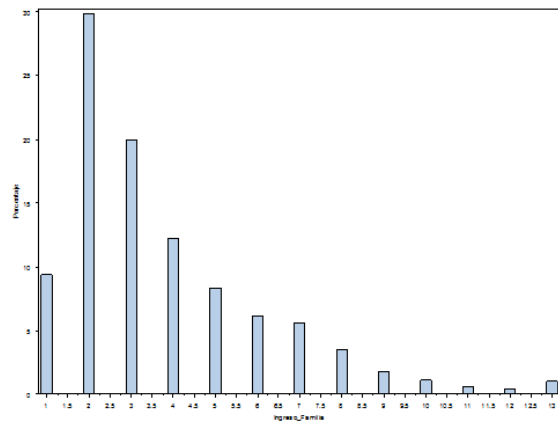
Variable	Porcentaje
Estudiantes mujeres	54.39%
Estudiantes repitentes	8.49%
Estudiantes con madre indígena	5.69%
Escuelas rurales	5.25%
Escuelas municipales	32.87%
Escuelas particulares subvencionados	54.07%
Escuelas particulares pagados	13.06%

En los siguientes histogramas se observa la dispersión del ingreso familiar por dependencia escolar:

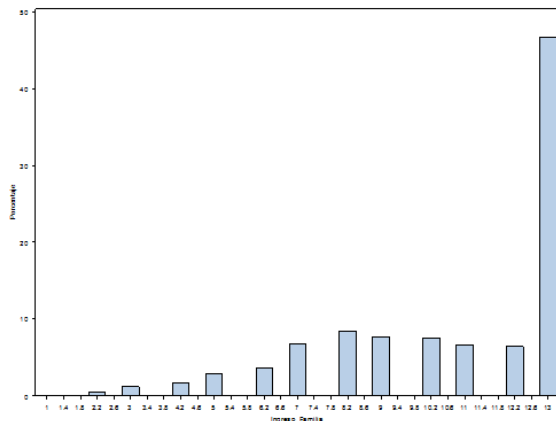
Municipales



Subvencionados



Pagados



Se observa que la dispersión del ingreso de las familias en las escuelas municipales y subvencionadas es similar. Por el contrario en los pagados el ingreso se concentra en el último nivel (sobre \$ 1.800.000 mensuales).

Resultados de análisis inferenciales

Muestra completa

Previamente a realizar las estimaciones utilizando las variables independientes mencionadas anteriormente, se hizo la estimación de los modelos nulos. Estos modelos no incluyen variables predictoras, sino que solo incluyen los efectos aleatorios. En la

Tabla 7 se observa cuanta varianza es atribuida a los tres niveles del estudio (estudiante, curso y escuela).

Tabla 7: Descomposición de la varianza en los tres niveles

	Lenguaje	%	Matemática	%
<i>Efectos aleatorios</i>				
Estudiante	1632.39	58.73	2059.91	46.64
Curso	219.16	7.88	427.43	9.68
Escuela	928.22	33.39	1929.37	43.68

Se puede apreciar que el mayor nivel de varianza se encuentra a nivel estudiante, donde el porcentaje de varianza es más grande en la prueba de lenguaje que en la prueba de matemática. En matemática existe un mayor porcentaje de variabilidad entre escuelas y entre cursos, en comparación a lo que sucede en lenguaje. En general, el porcentaje de varianza a nivel escuela es grande, en comparación internacional, con una varianza no despreciable a nivel curso. En la Tabla 8 se muestra el porcentaje de variabilidad atribuida al estudiante, curso y escuela por dependencia educacional.

Tabla 8: Porcentaje de descomposición de la varianza por dependencia escolar

<i>Efectos</i>	MUNICIPALES		SUBVENCIONADOS		PAGADOS	
	Lenguaje	Matemática	Lenguaje	Matemática	Lenguaje	Matemática
<i>Aleatorios</i>	%	%	%	%	%	%
Estudiante	67.07	56.75	68.36	56.52	89.57	80.82
Curso	13.08	16.89	5.69	7.01	1.58	3.19
Escuela	19.85	26.36	25.95	36.47	8.85	15.99

Se puede observar que hay diferencias interesantes respecto de la descomposición de la varianza entre las distintas dependencias, y estas diferencias demandan el análisis por cada dependencia por separado (ya que dentro de la muestra total no existe homogeneidad de varianza). Por ejemplo, en los colegios particulares pagados la variabilidad que hay a nivel curso es prácticamente nula (y la del nivel escuela es bastante menor) cuando se compara con los colegios municipales y particulares subvencionados. Siguiendo en la misma línea se observa que en los colegios municipales la variabilidad a nivel curso es sustancial, aunque algo menor que la atribuida a la escuela, situación que cambia en los colegios particulares subvencionados donde parte de la varianza del curso la capta el nivel escuela.

Estos resultados nos llevarían a pensar que debido a la alta variabilidad que posee la escuela nuestro foco de atención debería centrarse ahí. Sin embargo, nuestra intención es evaluar el efecto de los pares en el aprendizaje y a pesar de que los resultados indican que la mayor variabilidad en el rendimiento se encuentra a nivel escuela, la literatura demuestra que los pares a nivel de curso inciden de manera más directa y significativa en el aprendizaje que los compañeros de escuela. Por lo tanto, las variables que evalúan el efecto de los pares en nuestro estudio (variables composicionales) las trabajaremos solo a nivel curso¹

Para las estimaciones multinivel estimamos cinco modelos. El primer modelo solo incluye variables a nivel individual que inciden en el aprendizaje. En el segundo modelo se incluyen las variables contextuales de diversidad (social y rendimiento) a nivel curso. En el tercer modelo se incluyen las variables de promedio del ingreso y rendimiento a nivel curso. En el cuarto modelo se consideran otras variables composicionales como porcentaje de mujeres, repitentes y estudiantes con madre indígena por curso. Finalmente en el modelo 5 se incluyen las variables a nivel escuela, la ruralidad del establecimiento y su respectiva dependencia educacional. La intención de utilizar esta metodología (agregando nuevas variables paulatinamente en distintos modelos) es poder observar qué variables en un principio son significativas y cuáles al ir agregando nuevas variables van perdiendo significancia.

Cabe mencionar que aun existiendo cierto grado de asociación entre ambas variables composicionales de heterogeneidad (la correlación entre heterogeneidad social y de rendimiento en matemática es de 0.38), al realizar las estimaciones estas no presentan problemas de autocorrelación. Por lo tanto, no existe razón para realizar las estimaciones utilizando de forma separada ambas medidas de heterogeneidad al interior de los cursos.

En la Tabla 9 y Tabla 10 se muestran los resultados de las estimaciones utilizando toda la muestra para la prueba de lenguaje y matemática.

¹ Además de esta justificación teórica, nuestro propio estudio entrega una justificación empírica. Analizamos también el efecto composicional a nivel escuela y observamos que el tamaño del efecto a nivel escuela era menor que el del nivel curso, indicando que la incidencia del efecto de los pares en el aprendizaje efectivamente es mayor al interior de los cursos que al interior de las escuelas.

Tabla 9: Estimación prueba de lenguaje

	Modelo1	Modelo2	Modelo3	Modelo4	Modelo5
<i>Nivel individual</i>					
Lenguaje 2004 (centrado gran media)	0.619** (280.40)	0.616** (278.75)	0.594** (259.51)	0.594** (259.52)	0.594** (259.53)
Mujer	0.009** (4.10)	0.009** (4.14)	0.007** (3.31)	0.006** (2.98)	0.006** (2.98)
Repite	-0.044** (-22.90)	-0.044** (-22.70)	-0.040** (-20.80)	-0.037** (-18.93)	-0.037** (-18.93)
Indígena madre	-0.001 (-0.60)	-0.001 (-0.54)	0.001 (0.26)	0.000 (0.19)	0.000 (0.19)
Log Ingreso familiar (Intercepto)	0.054** (20.80)	0.048** (18.24)	0.027** (9.83)	0.027** (9.84)	0.027** (9.82)
	259.139** (620.70)	250.535** (237.69)	169.205** (77.02)	175.187** (70.56)	175.697** (70.02)
<i>Nivel curso</i>					
Heterogeneidad social		0.077** (18.52)	-0.004 (-0.85)	-0.005 (-1.07)	0.001 (0.34)
Heterogeneidad lenguaje 2004		-0.000 (-0.11)	0.015** (5.19)	0.014** (5.02)	0.014** (5.08)
Log Ingreso familiar (promedio)			0.051** (8.03)	0.053** (8.21)	0.032** (3.56)
Lenguaje 2004 (promedio)			0.180** (36.71)	0.168** (31.36)	0.169** (31.39)
% de mujeres				0.002 (0.45)	0.002 (0.46)
% de repitentes				-0.018** (-5.63)	-0.019** (-5.85)
% de madres indígenas				0.001 (0.30)	0.000 (0.03)
<i>Nivel escuela</i>					
Ruralidad					0.007* (2.01)
Subvencionado					0.012* (2.29)
Pagado					0.024** (3.67)
Devianza	1,102,445	1,102,130	1,100,291	1,100,259	1,100,244
R² Nivel 1	39.67	39.69	39.76	39.76	39.76
R² Nivel 2	73.71	73.06	81.18	81.34	81.37
R² Nivel 3	84.98	88.37	94.94	95.00	95.05

Estimación máxima verosimilitud, coeficientes estandarizados. * p<0.05, ** p<0.01. En paréntesis se muestra el valor t.

Tabla 10: Estimación prueba de matemática

	Modelo1	Modelo2	Modelo3	Modelo4	Modelo5
<i>Nivel individual</i>					
Matemática 2004 (centrado gran media)	0.604** (296.80)	0.603** (296.20)	0.585** (280.82)	0.584** (280.51)	0.584** (280.51)
Mujer	-0.037** (-20.00)	-0.037** (-20.03)	-0.038** (-20.63)	-0.042** (-22.23)	-0.042** (-22.23)
Repite	-0.047** (-28.00)	-0.047** (-27.85)	-0.044** (-26.57)	-0.042** (-24.72)	-0.042** (-24.72)
Indígena madre	-0.002 (-1.00)	-0.002 (-0.91)	-0.001 (-0.48)	-0.000 (-0.24)	-0.000 (-0.24)
Log Ingreso familiar (Intercepto)	0.027** (11.60) 264.284** (432.00)	0.023** (9.92) 244.505** (167.58)	0.011** (4.44) 127.292** (44.62)	0.010** (4.39) 131.012** (40.07)	0.010** (4.38) 130.997** (39.20)
<i>Nivel curso</i>					
Heterogeneidad social		0.096** (20.10)	0.014** (2.77)	0.012* (2.37)	0.012* (2.23)
Heterogeneidad matemática 2004		0.016** (4.79)	0.014** (4.62)	0.015** (4.90)	0.014** (4.77)
Log Ingreso familiar (promedio) Matemática 2004 (promedio)			0.036** (4.65)	0.036** (4.75)	0.031** (2.98)
			0.237** (40.08)	0.224** (35.81)	0.224** (35.66)
% de mujeres				0.038** (8.71)	0.038** (8.67)
% de repitentes				-0.023** (-6.77)	-0.023** (-6.67)
% de madres indígenas				-0.004 (-1.04)	-0.004 (-1.01)
<i>Nivel escuela</i>					
Ruralidad					-0.000 (-0.06)
Subvencionado Pagado					0.015* (2.34) 0.008 (1.06)
Devianza	1,123,135	1,122,733	1,120,659	1,120,520	1,120,514
R² Nivel 1	44.36	44.37	44.40	44.40	44.40
R² Nivel 2	63.37	63.76	73.54	74.15	74.17
R³ Nivel 3	78.22	82.93	92.49	92.92	92.92

Estimación máxima verosimilitud, coeficientes estandarizados. * p<0.05, ** p<0.01. En paréntesis se muestra el valor t.

De los resultados se puede observar que la inclusión de variables contextuales no altera la significancia de las estimaciones en las variables individuales. Como se podía esperar, el resultado de los alumnos en la prueba SIMCE 2004 (el rendimiento previo) es la variable más relevante respecto de la predicción del rendimiento SIMCE 2006. Sin embargo llama la atención la relevancia del promedio del curso en el rendimiento SIMCE 2004, incluso por sobre el efecto del rendimiento previo individual. Las variables de *repetencia* a nivel individual y el *porcentaje de repitentes a nivel curso* tienen un fuerte efecto negativo en ambas asignaturas, lo que significa que hacer repetir a un alumno le afecta negativamente y que los alumnos que interactúan con compañeros que hayan repetido al menos una vez los perjudica en su aprendizaje personal. Estos resultados se ven tanto para lenguaje como para matemática.

Respecto a la variable género (en nuestro caso, ser mujer), esta tiene un efecto positivo y pequeño en lenguaje pero negativo y más grande en matemática, lo que implica que a las estudiantes mujeres les va significativamente peor en matemática que a los hombres. Un resultado interesante se relaciona con el *porcentaje de mujeres* en las salas de clases en la prueba de matemática, ya que se observa que tiene un efecto positivo en el aprendizaje, pero como se mencionó anteriormente las mujeres obtienen peores resultados que los hombres. Es decir, mientras más mujeres hay en un curso mejor será el rendimiento individual de los alumnos, en general.

Las variables que corresponden a la diversidad social y académica del aula presentan efectos positivos en ambas pruebas, siendo más claro en matemática, ya que ahí se observa que las estimaciones son todas positivas y significativas en los cinco modelos. Llama la atención la relevancia de la heterogeneidad social, en comparación con la heterogeneidad del rendimiento, teniendo un efecto positivo en el aprendizaje. Sin embargo, a partir del modelo 3 que introduce variables promediadas a nivel curso tanto de rendimiento previo como de ingreso familiar se pierde la relevancia de esta variable. Es muy probable que esta pérdida se deba al ingresar la variable de promedio del ingreso familiar, debido a que son variables que poseen una alta correlación (0.72). Por lo tanto, parte de la variabilidad atribuida a la *Heterogeneidad social* la captura el ingreso familiar promedio del curso.

A nivel escuela, respecto de los efectos de la dependencia escolar se observa que utilizando las escuelas municipales como nivel de referencia, tanto los particulares

subvencionados como los particulares pagados tienen mejores resultados que los municipales en lenguaje (en parte porque los resultados en lenguaje se relacionan más fuertemente con el nivel socioeconómico de las familias). Sin embargo, en matemática esta tendencia no se mantiene, donde se observa que los particulares subvencionados tienen mejor rendimiento que los municipales pero los particulares pagados no se diferencian de estos últimos, siempre controlando por variables individuales y composicionales. Los efectos de la dependencia escolar son pequeños y no logran explicar un porcentaje significativo de la varianza que queda por explicar después de haber controlado por variables individuales y contextuales a nivel curso.

Finalmente, observando la devianza y los R^2 se puede concluir que los modelos finales (modelo 5) logran explicar casi en su totalidad la variabilidad atribuida inicialmente al nivel escuela, y también una gran parte de la varianza a nivel curso. No así en el caso individual donde aun existe más de la mitad de la variabilidad que no se logra explicar por nuestros modelos.

Dependencia escolar

El interés ahora es poder observar que ocurre en detalle cuando se analizan los establecimientos de forma separada según su dependencia. Para este caso se analizó exclusivamente el modelo completo (antes llamado Modelo 5), el que contiene todas las variables (individuales y contextuales). La Tabla 11 muestra las estimaciones diferenciados por dependencia educacional.²

² Para ver el resumen descriptivo por dependencia, puede consultar el anexo al final del paper.

Tabla 11: Estimaciones para las pruebas de lenguaje y matemática diferenciado por dependencia escolar

	<i>MUNICIPALES</i>		<i>SUBVENCIONADOS</i>		<i>PAGADOS</i>	
	Leng.	Mat.	Leng.	Mat.	Leng.	Mat.
<i>Nivel individual</i>						
Lenguaje 2004 (centrado gran media)	0.594** (171.50)	-	0.607** (181.47)	-	0.590** (70.06)	-
Matemática 2004 (centrado gran media)	-	0.562** (174.97)	-	0.602** (199.82)	-	0.682** (93.18)
Mujer	-0.003 (-0.76)	-0.053** (-17.85)	0.016** (4.78)	-0.038** (-12.97)	0.004 (0.38)	-0.043** (-5.43)
Repite	-0.042** (-13.86)	-0.046** (-17.33)	-0.036** (-12.01)	-0.044** (-17.02)	-0.044** (-5.33)	-0.037** (-5.29)
Indígena madre	0.003 (0.91)	-0.001 (-0.22)	-0.002 (-0.70)	-0.000 (-0.16)	0.001 (0.17)	0.003 (0.46)
Log Ingreso familiar (Intercepto)	0.024** (7.09)	0.006* (2.34)	0.024** (6.78)	0.011** (3.42)	0.010 (1.12)	0.017* (2.23)
	170.146** (51.43)	114.221** (24.17)	181.242** (43.91)	142.064** (26.92)	223.230** (13.13)	264.586** (14.27)
<i>Nivel curso</i>						
Heterogeneidad social	-0.007 (-0.91)	0.007 (0.79)	0.013 (1.57)	0.020* (2.14)	0.005 (0.27)	-0.019 (-1.08)
Heterogeneidad rendimiento^a	0.012** (2.82)	0.015** (3.24)	0.012** (2.63)	0.016** (3.48)	0.021 (1.67)	-0.023 (-1.78)
Log Ingreso familiar (promedio)	0.019 (1.93)	0.014 (1.12)	0.017 (1.63)	0.024 (1.95)	0.040* (2.14)	0.001 (0.03)
Lenguaje 2004 (promedio)	0.187** (25.95)	-	0.142** (18.56)	-	0.033* (2.22)	-
Matemática 2004 (promedio)	-	0.250** (28.55)	-	0.179** (21.10)	-	0.038* (2.34)
% de mujeres	0.006 (1.05)	0.051** (7.08)	0.000 (0.08)	0.036** (5.44)	-0.002 (-0.11)	0.005 (0.31)
% de repitentes	-0.020** (-4.04)	-0.024** (-4.48)	-0.022** (-4.17)	-0.021** (-3.80)	0.005 (0.45)	-0.022 (-1.84)
% de madres indígenas	-0.001 (-0.20)	-0.001 (-0.11)	0.002 (0.41)	-0.008 (-1.24)	-0.021 (-1.89)	-0.013 (-1.15)
<i>Nivel escuela</i>						
Ruralidad	0.009 (1.74)	0.002 (0.35)	0.008 (1.27)	-0.002 (-0.24)	0.000 (0.02)	0.005 (0.36)
Devianza	491,394	503,082	515,497	524,808	93,204	92,124
R² Nivel 1	39.16	41.54	40.82	46.06	36.66	50.74
R² Nivel 2	86.41	79.41	71.52	62.88	-9.18	7.72
R³ Nivel 3	92.43	86.00	91.61	88.64	69.16	75.84

Estimación máxima verosimilitud, coeficientes estandarizados. * p<0.05, ** p<0.01.

a. Se refiere a la heterogeneidad de las pruebas lenguaje 2004 y matemática 2004 respectivamente.

En paréntesis se muestra el valor t.

Al observar la tabla de resultados se puede concluir que las variables individuales no varían prácticamente entre dependencias. Constatamos que el ingreso familiar incide mucho más en el aprendizaje en los colegios municipales y particulares

subvencionados que en los particulares pagados (donde no sale significativo en lenguaje). El tema género también se matiza: la ventaja de las niñas en lenguaje se concentra en los colegios particulares subvencionados, mientras que en todos los tipos de colegios se encuentra la ventaja de los niños en matemática (aunque lo más fuerte ocurre en los colegios municipales, y lo menos en los particulares subvencionados).

A nivel contextual se observa que en los colegios particulares pagados solo las variables de ingreso familiar promedio y el promedio de las pruebas SIMCE 2004 son levemente significativas. Otras variables composicionales, tales como porcentaje de mujeres, repitentes y madres indígenas, así como también las variables de diversidad social y diversidad del rendimiento, no son estadísticamente significativas.

Por el contrario, en los establecimientos municipales como particulares subvencionados algunas variables composicionales, como diversidad del rendimiento previo, son significativas. Este último resultado refleja que el efecto de los pares tiene una incidencia sobre todo en los estratos medios – bajos en el aprendizaje de los estudiantes.

Llama la atención la fuerte incidencia del promedio del rendimiento previo del curso en el aprendizaje de los estudiantes en todas las dependencias. Aun controlando por el rendimiento previo a nivel individual, todavía sale significativo el promedio. Por lo tanto, esa característica composicional tiene un efecto por sobre el efecto del rendimiento previo a nivel individual, en todas las dependencias.

Interacción entre variables

En nuestro estudio consideramos interacciones para ver si existe algún efecto diferencial de los efectos composicionales en el aprendizaje de los estudiantes. De los resultados anteriores se ha decidido incluir interacciones en los modelos 2 y 3. En las tablas a continuación se indican modelos 2a, 2b, 2c y 2d, donde estos modelos contienen las mismas variables que se utilizaron en el modelo 2 de las tablas 9 y 10, y los modelos 3a y 3b que corresponde al modelo 3 de las mismas tablas 9 y 10.

El modelo 2a y 2b incluye la interacción del ingreso familiar del estudiante con la heterogeneidad social del curso por un lado, y la del ingreso familiar del estudiante con la heterogeneidad de rendimiento del curso por otro lado. Los modelos 2c y 2d incluyen la interacción del género del estudiante con la heterogeneidad social del curso, e

interacción del género con la heterogeneidad de rendimiento del curso, respectivamente. Finalmente, en el modelo 3a incluimos la interacción del rendimiento previo del estudiante con la heterogeneidad social del curso, y el 3b incluye la interacción del rendimiento previo del estudiante con la heterogeneidad de rendimiento del curso. En la Tabla 12 y Tabla 13 se muestran los resultados de estimación de los modelos 2 y 3 para la prueba de lenguaje, mientras que en la Tabla 14 y Tabla 15 los resultados de estimación en matemática.

Tabla 12: Estimación prueba de lenguaje con interacciones en el modelo 2

	Modelo2a	Modelo2b	Modelo2c	Modelo2d
<i>Nivel individual</i>				
Lenguaje 2004 (centrado gran media)	0.617** (278.71)	0.617** (278.78)	0.617** (278.82)	0.617** (278.84)
Mujer	0.009** (4.16)	0.009** (4.12)	-0.002 (-0.29)	0.036** (3.16)
Repite	-0.044** (-22.64)	-0.044** (-22.73)	-0.044** (-22.75)	-0.044** (-22.71)
Indígena madre	-0.001 (-0.51)	-0.001 (-0.55)	-0.001 (-0.54)	-0.001 (-0.53)
Log Ingreso familiar	0.066** (10.25)	0.071** (5.80)	0.048** (18.30)	0.048** (18.29)
(Intercepto)	249.114** (220.32)	248.585** (177.09)	251.209** (230.15)	249.087** (202.84)
<i>Nivel curso</i>				
Heterogeneidad social	0.090** (15.84)	0.078** (18.78)	0.072** (14.96)	0.077** (18.68)
Heterogeneidad lenguaje 2004	-0.001 (-0.20)	0.006 (1.32)	-0.001 (-0.27)	0.005 (1.20)
<i>Interacción</i>				
Log Ingreso familiar* Heterogeneidad social	-0.028** (-3.17)	-	-	-
Log Ingreso familiar* Heterogeneidad lenguaje 2004	-	-0.024 (-1.95)	-	-
Mujer* Heterogeneidad social	-	-	0.013* (2.19)	-
Mujer* Heterogeneidad lenguaje 2004	-	-	-	-0.027* (-2.41)
Devianza	1,102,114	1,102,120	1,102,083	1,102,082
R² Nivel 1	39.53	39.53	39.74	39.74
R² Nivel 2	78.15	78.11	72.72	72.64
R² Nivel 3	88.51	88.55	88.36	88.39

Estimación máxima verosimilitud, coeficientes estandarizados. * p<0.05, ** p<0.01. En paréntesis se muestra el valor t.

Tabla 13: Estimación prueba de lenguaje con interacciones en el modelo 3

	Modelo3a	Modelo3b
<i>Nivel individual</i>		
Lenguaje 2004 (centrado gran media)	0.582** (101.83)	0.689** (55.88)
Mujer	0.007** (3.37)	0.007** (3.42)
Repite	-0.040** (-20.88)	-0.040** (-20.94)
Indígena madre	-0.001 (0.30)	0.001 (0.30)
Log Ingreso familiar	0.027** (9.73)	0.027** (9.73)
(Intercepto)	167.900** (75.64)	171.800** (76.10)
<i>Nivel curso</i>		
Heterogeneidad social	-0.004 (-1.03)	-0.004 (-0.95)
Heterogeneidad lenguaje 2004	0.016** (5.61)	0.012** (4.33)
Log Ingreso familiar (promedio)	0.052** (8.03)	0.052** (8.02)
Lenguaje 2004 (promedio)	0.182** (36.65)	0.175** (34.99)
<i>Interacción</i>		
Lenguaje 2004 (centrado gran media)*Heterogeneidad social	0.014* (2.39)	-
Lenguaje 2004 (centrado gran media)*Heterogeneidad lenguaje 2004	-	-0.094** (-7.87)
Devianza	1,100,151	1,100,095
R² Nivel 1	40.17	40.16
R² Nivel 2	85.31	85.30
R² Nivel 3	94.77	94.74

Estimación máxima verosimilitud, coeficientes estandarizados. * p<0.05, ** p<0.01. En paréntesis se muestra el valor t.

Tabla 14: Estimación prueba de matemática con interacciones en el modelo 2

	Modelo2a	Modelo2b	Modelo2c	Modelo2d
<i>Nivel individual</i>				
Matemática 2004 (centrado gran media)	0.617** (296.21)	0.603** (296.22)	0.602** (296.17)	0.602** (296.14)
Mujer	0.009** (-20.01)	-0.037** (-20.04)	-0.047** (-9.69)	-0.044** (-4.19)
Repite	-0.044** (-27.74)	-0.046** (-27.81)	-0.047** (-27.87)	-0.047** (-27.85)
Indígena madre	-0.001 (-0.86)	-0.002 (-0.88)	-0.002 (-0.93)	-0.002 (-0.93)
Log Ingreso familiar	0.069** (6.88)	0.055** (4.89)	0.023** (9.91)	0.023** (9.93)
(Intercepto)	243.356** (158.67)	241.620** (131.83)	245.106** (163.91)	244.886** (150.16)
<i>Nivel curso</i>				
Heterogeneidad social	0.077** (17.67)	0.095** (19.87)	0.091** (17.44)	0.097** (20.24)
Heterogeneidad matemática 2004	0.006** (4.68)	0.025** (5.38)	0.027** (4.80)	0.015** (3.72)
<i>Interacción</i>				
Log Ingreso familiar* Heterogeneidad social	-0.022** (-3.11)	-	-	-
Log Ingreso familiar* Heterogeneidad matemática 2004	-	-0.033** (-2.86)	-	-
Mujer*Heterogeneidad social	-	-	0.012* (2.33)	-
Mujer*Heterogeneidad matemática 2004	-	-	-	0.008 (0.72)
Devianza	1,122,718	1,122,719	1,122,652	1,122,657
R² Nivel 1	44.14	44.14	44.55	44.55
R² Nivel 2	65.06	65.11	64.76	64.73
R² Nivel 3	82.62	82.68	82.77	82.78

Estimación máxima verosimilitud, coeficientes estandarizados. * p<0.05, ** p<0.01. En paréntesis se muestra el valor t.

Tabla 15: Estimación prueba de matemática con interacciones en el modelo 3

	Modelo3a	Modelo3b
<i>Nivel individual</i>		
Matemática 2004 (centrado gran media)	0.542** (99.57)	0.605** (50.47)
Mujer	-0.037** (-20.49)	-0.037** (-20.44)
Repite	-0.045** (-26.89)	-0.044** (-26.65)
Indígena madre	-0.001 (-0.43)	-0.001 (-0.37)
Log Ingreso familiar	0.010** (4.34)	0.010** (4.43)
(Intercepto)	124.346** (42.74)	126.000** (42.56)
<i>Nivel curso</i>		
Heterogeneidad social	0.015** (2.79)	0.014* (2.56)
Heterogeneidad matemática 2004	0.016** (5.11)	0.014** (4.50)
Log Ingreso familiar (promedio)	0.035** (4.51)	0.034** (4.38)
Matemática 2004 (promedio)	0.241** (39.77)	0.240** (39.28)
<i>Interacción</i>		
Matemática 2004 (centrado gran media)*Heterogeneidad social	0.047** (8.53)	- -
Matemática 2004 (centrado gran media)*Heterogeneidad matemática 2004	- -	-0.021 (-1.82)
Devianza	1,120,358	1,120,427
R² Nivel 1	45.06	45.04
R³ Nivel 2	76.71	73.86
R² Nivel 3	92.19	92.19

Estimación máxima verosimilitud, coeficientes estandarizados. * p<0.05, ** p<0.01. En paréntesis se muestra el valor t.

A nivel individual se observa que los efectos no varían mucho en comparación a los análisis hechos anteriormente para ambas pruebas. A nivel contextual en el modelo 2 (a, b, c y d) se observa que en la prueba de lenguaje el efecto de la diversidad de rendimiento no es significativo, sin embargo en el modelo 3a y 3b el efecto es significativo. Posiblemente el efecto del rendimiento previo promedio a nivel curso logre agregar mayor variabilidad a la variable *Heterogeneidad lenguaje 2004*, logrando así que esta sea significativa (esto en estadística es lo que se conoce como el efecto supresión entre variables). Por otro lado, se observa que la *Heterogeneidad social* sigue teniendo un efecto positivo en el aprendizaje en lenguaje, sin embargo esto solo ocurre en el modelo 2, en los modelos 3 (a y b) el efecto se pierde.

Para la prueba de matemática todos los efectos a nivel contextual son significativos (con efecto positivo en todos ellos), manteniendo la tendencia mostrada en los resultados anteriores. Las variables composicionales de diversidad (social y de rendimiento) siguen teniendo un efecto positivo en el aprendizaje, al igual que el promedio del ingreso familiar y el promedio de rendimiento previo.

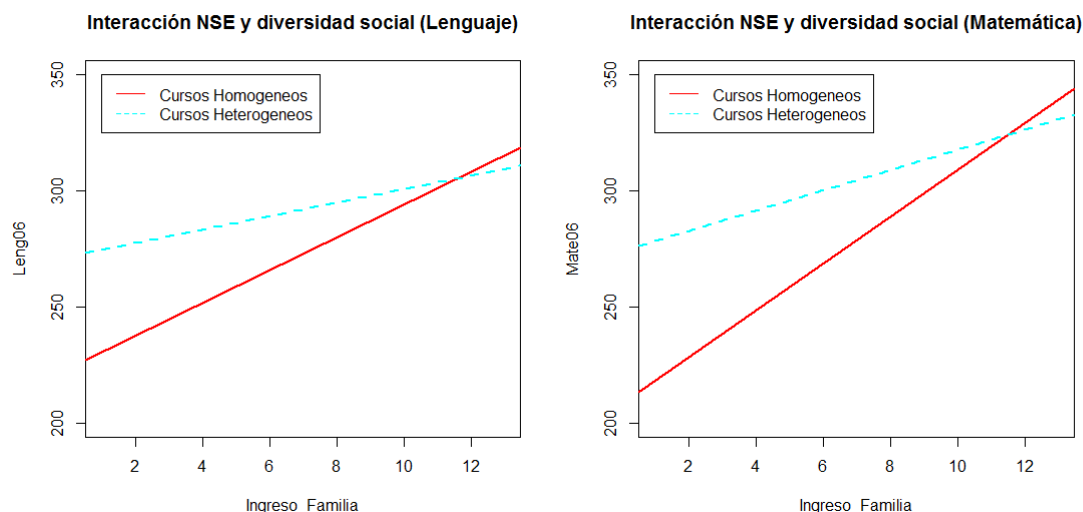
Para el caso de la prueba de lenguaje se observa que prácticamente todas las interacciones son significativas (salvo el caso de la interacción entre el ingreso familiar individual y la heterogeneidad de rendimiento del curso). En matemática la situación es algo similar ya que muchas interacciones son significativas (salvo el caso de la interacción del género con heterogeneidad de rendimiento del curso, y rendimiento previo con heterogeneidad de rendimiento del curso).

En lenguaje la interacción entre el ingreso familiar y la heterogeneidad social es negativa, lo que quiere decir que el efecto del ingreso familiar de los estudiantes en el aprendizaje en lenguaje es menor en los cursos de alta diversidad socioeconómica cuando son comparados con los cursos de baja diversidad. *Esto quiere decir que los cursos de mayor diversidad social logran amortiguar el efecto del ingreso familiar en el aprendizaje en lenguaje para los niños de bajo ingreso familiar; y en la medida en que aumenta el ingreso familiar el efecto positivo de la diversidad social del curso se ve más débil.* Por otro lado, observamos que el efecto de la interacción entre rendimiento previo con heterogeneidad de rendimiento en lenguaje es negativo. *El efecto positivo de tener un buen rendimiento previo en la prueba de lenguaje es mayor en cursos más homogéneos que en los heterogéneos. Los buenos alumnos se ven más favorecidos si están insertos en cursos con poca diversidad de rendimiento. Al mismo tiempo, los alumnos con bajo rendimiento previo se ven beneficiados por la heterogeneidad de rendimiento previo de sus pares.*

En matemática, la interacción del ingreso familiar con la heterogeneidad social, e ingreso familiar con heterogeneidad de rendimiento son ambas negativas. Tal como se explicó para lenguaje, *esto indica que los alumnos de bajos ingresos familiares se benefician por un ambiente de alta heterogeneidad social y de rendimiento.* Finalmente, la interacción del rendimiento previo con heterogeneidad social es significativa y positiva, *lo que indica que estudiantes con buen rendimiento previo se ven más*

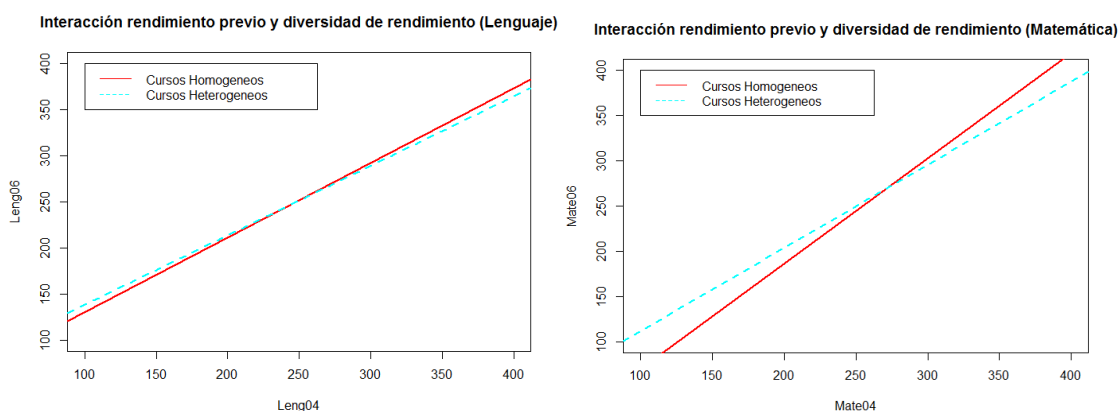
beneficiados en cursos más heterogéneos socialmente que los que asisten a cursos más homogéneos.

Los siguientes gráficos nos muestran visualmente la interacción entre ingreso familiar del estudiante y diversidad social del curso:



En estos gráficos se observa que para los estudiantes que tienen el ingreso familiar más bajo (bajo \$100.000) hay una gran diferencia de alrededor de una d.e., de 44 y 61 puntos en lenguaje y matemática, respectivamente, a favor de los cursos con alta diversidad social. Mientras que para los estudiantes con el ingreso familiar más alto (superior a \$1.800.000) hay una pequeña diferencia de 6 y 8 puntos en lenguaje y matemática, respectivamente, a favor de cursos homogéneos. Es importante notar que las rectas se dan vuelta (de una ventaja de cursos heterogéneos a una ventaja de cursos homogéneos) solo a partir de un ingreso familiar del estudiante muy alto (casi en el máximo de la escala de ingresos del cuestionario SIMCE) y esta ventaja es muy pequeña en comparación con la ventaja de cursos heterogéneos para ingresos bajos y medios.

Los siguientes gráficos muestran la interacción entre rendimiento previo del estudiante y diversidad de rendimiento del curso:



En estos gráficos se observa que para los estudiantes que tienen un puntaje SIMCE previo de 100 puntos hay una diferencia de 8 y 41 puntos en lenguaje y matemática, respectivamente, a favor de cursos heterogéneos. Mientras que para los estudiantes que tienen un puntaje SIMCE previo de 400 puntos hay una diferencia de 9 y 31 puntos en lenguaje y matemática, respectivamente, a favor de cursos homogéneos.

Discusión y conclusiones

En primer lugar, en base a los análisis de descomposición de la varianza en un modelo multinivel de tres niveles, nuestro estudio demuestra que el nivel de la sala de clase incide en el aprendizaje de los estudiantes de enseñanza media (8% en lenguaje y 10% en matemática para la muestra que incluye todas las dependencias), adicionalmente a nivel individual (59% en lenguaje y 47% en matemática) y de escuela (33% en lenguaje, 44% en matemática). Ese resultado después se matiza al descomponer la varianza por dependencia, donde se muestra particularmente grande el efecto del nivel curso para establecimientos municipales, ya algo más pequeño para particulares subvencionados (donde el nivel escuela incide más comparado con los colegios municipales), y desapareciendo casi en los particulares pagados (2% lenguaje, 3% matemática). El nivel escuela y sala de clase tienen un mayor impacto en el aprendizaje en matemática que en lenguaje.

Si se acepta que variables a nivel individual, de aula, y de escuela inciden en el aprendizaje, entonces otra pregunta es cuáles son las variables relevantes que explican la varianza en el aprendizaje entre alumnos. Nuestro estudio encuentra que el

rendimiento en lenguaje está predicho a nivel individual sobre todo por el rendimiento previo del alumno (positivamente), si repitió alguna vez (negativamente), y, en menor medida, su ingreso familiar (ingreso familiar y rendimiento previo están altamente correlacionados). En matemática, las variables individuales más relevantes son nuevamente el rendimiento previo del alumno (positivamente), si repitió alguna vez (negativamente), y si es mujer (negativamente).

Sin embargo, nuestro principal interés en este paper fueron las variables composicionales a nivel aula. En términos de la heterogeneidad social, encontramos que es bastante relevante si se incluye como primera variable a nivel de curso, tanto en lenguaje como en matemática. A más alta diversidad social, mejores son los rendimientos de los alumnos, aun controlando por otras variables a nivel individual. En nuestro estudio, contrastando con resultados de Torche y Mizala (2012), hay un efecto positivo similar de la heterogeneidad de rendimiento en matemática pero no así en lenguaje.

Al incluir el promedio del rendimiento previo y del ingreso familiar a nivel curso, el efecto de las variables de heterogeneidad tiende a perderse (ver también Gibbons & Telhaj, 2006). El promedio del ingreso familiar es mucho menos relevante (aunque significativo) que el promedio del rendimiento previo. En todo caso y como se menciono anteriormente, esas dos variables (rendimiento previo e ingreso) están altamente correlacionados (0.72 y 0.75, en lenguaje y matemática, respectivamente). Llama la atención la fuerte relevancia del promedio del curso por sobre el rendimiento previo individual ya incluido en el modelo (ver también Opdenakker & Van Damme, 2001). También llama la atención que el efecto del promedio del ingreso familiar en el aprendizaje es tanto o más relevante que el ingreso familiar a nivel individual (también ver Torche & Mizala, 2012; y Rumberger & Palardy, 2003). Otras características del curso no se acercan a tener la misma relevancia. Las variables a nivel escuela, por sobre las del nivel individual y curso, no mejoran significativamente el modelo.

Cuando se analizan los datos por dependencia, en general para los establecimientos municipales y particulares subvencionados los resultados anteriormente descritos se mantienen, pero el patrón es distinto en el caso de los particulares pagados. A nivel individual, todos los efectos de las variables son más pequeños y las variables composicionales a nivel curso pierden relevancia. En los

colegios municipales y particulares subvencionados, sigue siendo el promedio del rendimiento previo del curso la variable composicional más relevante.

En términos del valor explicativo de los modelos, constatamos que nuestros modelos, aunque utilizan datos de panel y controlan por el rendimiento previo de los alumnos, no logran explicar más del 44% de la varianza atribuida al nivel individual/estudiante. Al mismo tiempo los modelos explican casi la totalidad de la varianza atribuida al nivel escuela (al menos 93%), y gran parte de la varianza atribuida al nivel aula (al menos 74%). Debido a la poca varianza atribuible al nivel curso en los establecimientos particulares pagados, los resultados del fit de los modelos son poco confiables en ese caso.

Por último, de las interacciones entre variables de heterogeneidad del curso y variables relevantes del nivel individual, destacamos que alumnos de ingresos familiares bajos y medios se ven claramente beneficiados por la diversidad social en el aula, con una diferencia de alrededor de una d.e. (de entre 44 y 61 puntos) para los alumnos de ingresos familiares más bajos. En comparación, el beneficio de cursos homogéneos para alumnos de ingresos familiares altos (y eso además solo se da en el caso de ingresos muy altos) es muy pequeño (de entre 6 y 8 puntos). En términos de la diversidad de rendimiento previo del curso y su interacción con el rendimiento previo de los estudiantes, los resultados van en la dirección esperada según la literatura (para resultados similares, ver Oppendakker & Van Damme, 2001).

Nuestro trabajo es relevante no solo por sus resultados, sino también por las preguntas que pone a discusión respecto de la importancia de efectos composicionales para disminuir la inequidad del sistema educacional chileno. Según nuestro estudio, en Chile aulas más diversas socialmente tendrían claros efectos positivos en el aprendizaje de alumnos de ingresos familiares bajo y medios, y no perjudicarían en forma significativa a los alumnos de alto ingreso familiar (ver Dronkers, 2010). En 2009 se promulgó una ley (ver Ley General de Educación, ley no. 20370, art. 12) que prohíbe la selección de alumnos hasta 6° básico de parte de colegios subvencionados por el estado chileno, en base de su rendimiento o características socioeconómicas. Según este estudio, parece importante asegurar una adecuada implementación de esa ley para lograr algo más de diversidad en las aulas chilenas. Después de haber investigado el efecto

composicional en el aprendizaje, ahora falta examinar sus efectos en términos de indicadores socioemocionales y motivacionales de los alumnos chilenos.

En el futuro también queremos entender mejor como afectan variables a nivel aula el rendimiento de los alumnos. Para estos efectos podemos usar tanto variables disponibles del cuestionario SIMCE para profesores (recursos a nivel aula, prácticas y creencias docente), como información disponible de la Evaluación Docente (solo para docentes de establecimientos municipales). Ya hemos explorado en una muestra de escuelas municipales el efecto de la calidad docente, medida a través de la Evaluación Docente, en nuestros modelos. El puntaje final del docente en base de los instrumentos de la ED incide en el aprendizaje de sus alumnos en matemática (se mantiene a través de distintos modelos), pero no así en lenguaje. Esos resultados confirman que el impacto del aula (y del profesor) en matemática es superior comparado con lenguaje, donde parece ser más difícil contrapesar el origen sociocultural de los alumnos. Es interesante notar que la heterogeneidad social sigue siendo relevante aun controlando por la calidad docente (pero antes de incluir promedios a nivel curso). Eso indica que la diversidad social tiene un efecto por sobre el efecto de la calidad docente, al menos en colegios municipales de enseñanza media. Nos interesa sobre todo explorar con más detalle la interacción entre efectos composicionales del aula y variables atribuibles al docente.

Referencias

- Bryk, A. S., & Raudenbush, S. W. (2002). *Hierarchical linear models: applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Dronkers, J. (2010, September). *Positive but also negative effects of ethnic diversity in schools on educational performance? An empirical test using cross-national PISA data*. Paper presented at the conference Integration and Inequality in Educational Institutions, Teerhof, Germany. Paper retrieved from <http://mpira.ub.uni-muenchen.de/25598>
- Gibbons, S. & Telhaj, S. (2006). Peer effects and pupil attainment: Evidence from secondary school transition. Retrieved from http://eprints.lse.ac.uk/19421/1/Peer_Effects_and_Pupil_Attainment_Evidence_from_Secondary_School_Transition.pdf
- Grupo Iberoamericano en PISA (2010). Iberoamérica en PISA 2006. Paris: OECD.
- Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics*, 18, 527-544. doi: 10.1002/jae.741
- Hattie, J. (2003). Classroom composition and peer effects. *International Journal of Educational Development*, 37, 449-481. doi: 10.1016/S0883-0.355(03)0001 5-6
- Hox, J. J. (2010). *Multilevel Analysis: Techniques and Applications* (2nd ed.). New York: Routledge.
- Hoxby, C. (2002). The power of peers: How does the makeup of a classroom influence achievement? *Education Next*. Retrieved from http://media.hoover.org/sites/default/files/documents/ednext20022_56.pdf
- Lissitz, R. (2005). Value-added models in education. Maple Grove: JAM Press.
- Lissitz, R. (2006). Longitudinal and Value-added Models of Student Performance. Maple Grove: JAM Press.
- McEwan, P. J. (2003). Peer effects on student achievement: evidence from Chile. *Economics of education review*, 22, 131-141. doi: 10.1016/S0272-7757(02)00005-5
- Manzi, J., Strasser, K., San Martín, E., & Contreras, D. (2008). The Quality of Education in Chile: Final report for the Inter-American Development Bank. Washington, D.C.: BID. Retrieved from <http://www.iadb.org/res/laresnetwork/files/pr300finaldraft.pdf>

- Mizala, A. & 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. doi: 10.1016/j.ijedudev.2010.09.004
- Mizala, A., & Romaguera, P. (2000). School Performance and Choice. The Chilean Experience. *The Journal of Human Resources*, 35(2), 392-417.
- Opdenakker, M.C., & Van Damme, J. (2001). Relationship between School Composition and Characteristics of School Process and their Effect on Mathematics Achievement. *British Educational Research Journal*, 27(4), 407-432. doi: 10.1080/0141192012007143 4
- Raudenbush, S. W. (2004). What are value-added models estimating and what does this imply for statistical practice? *Journal of Educational and Behavioral Statistics*, 29(1), 121-129.
- Ramírez, M.J. (2007). Diferencias dentro de las salas de clases distribución del rendimiento en matemáticas. *Estudios Públicos*, 106.
- Rumberger, R. W., & Palardy, G. J. (2005). Does Segregation Still Matter? The impact of student Composition on Academic Achievement in High School. *Teachers College Record*, 107(9), 1999-2045.
- Schneeweis, N., & Winter-Ebmer, R. (2006). Peer effects in Austrian schools. *Empirical economics*, 32, 387-409. doi: 10.1007/s00181-006-0091-4
- Vandenberghe, V. (2002). Evaluating the magnitude and the stakes of peer effects analyzing science and math achievement across OECD. *Applied Economics*, 34(10), 1283-1290.

Anexo

Resumen descriptivo por dependencia escolar

MUNICIPALES:

Tabla 16: Resumen descriptivo variables continuas a nivel individual

Variable	Mínimo	Media	Máximo	DE
Lenguaje 2006	120.71	251.86	398.16	49.97
Matemática 2006	93.52	248.16	426.58	62.30
Lenguaje 2004	97.51	255.15	391.67	48.53
Matemática 2004	121.72	255.04	406.09	47.52
Ingreso familiar	1.00	2.67	13.00	1.75

Tabla 17: Resumen descriptivo variables continuas a nivel contextual

Variable	Mínimo	Media	Máximo	DE
Heterogeneidad social	0.00	1.37	3.92	0.61
Heterogeneidad lenguaje 2004	11.23	39.22	82.86	7.62
Heterogeneidad matemática 2004	11.11	37.05	71.65	7.10
Promedio ingreso familiar	1.00	2.67	8.46	0.97
Lenguaje 2004 promedio	164.22	255.15	346.25	29.11
Matemática 2004 promedio	183.17	255.04	367.39	30.20
Porcentaje mujeres	0.00	0.55	1.00	0.27
Porcentaje repitencia	0.00	0.11	1.00	0.12
Porcentaje madre indígena	0.00	0.07	0.85	0.11

Tabla 18: Resumen descriptivo variables discretas

Variable	Porcentaje
Estudiantes Mujeres	54.73%
Estudiantes repitentes	10.69%
Estudiantes con madre indígena	7.16%
Escuelas rurales	6.79%

SUBVENCIONADOS:

Tabla 19: Resumen descriptivo variables continuas a nivel individual

Variable	Mínimo	Media	Máximo	DE
Lenguaje 2006	120.43	266.77	398.16	49.51
Matemática 2006	93.46	268.32	426.58	61.30
Lenguaje 2004	94.20	267.95	391.67	47.74
Matemática 2004	118.79	267.74	406.09	46.91
Ingreso familiar	1.00	3.76	13.00	2.42

Tabla 20: Resumen descriptivo variables continuas a nivel contextual

Variable	Mínimo	Media	Máximo	DE
Heterogeneidad social	0.30	1.81	4.29	0.71
Heterogeneidad lenguaje 2004	16.05	40.69	69.93	7.36
Heterogeneidad matemática 2004	12.40	38.79	75.66	6.96
Promedio ingreso familiar	1.09	3.76	10.60	1.49
Lenguaje 2004 promedio	172.95	267.96	330.24	25.58
Matemática 2004 promedio	194.83	267.74	342.80	26.90
Porcentaje mujeres	0.00	0.55	1.00	0.27
Porcentaje repitencia	0.00	0.07	0.83	0.09
Porcentaje madre indígena	0.00	0.05	1.00	0.09

Tabla 21: Resumen descriptivo variables discretas

Variable	Porcentaje
Estudiantes Mujeres	54.66%
Estudiantes repitentes	7.27%
Estudiantes con madre indígena	5.23%
Escuelas rurales	5.26%

PAGADOS:

Tabla 22: Resumen descriptivo variables continuas a nivel individual

Variable	Mínimo	Media	Máximo	DE
Lenguaje 2006	120.79	311.27	398.16	42.21
Matemática 2006	93.65	333.43	426.58	46.49
Lenguaje 2004	107.19	306.19	391.67	40.22
Matemática 2004	123.96	316.71	406.09	40.57
Ingreso familiar	1.00	10.62	13.00	2.84

Tabla 23: Resumen descriptivo variables continuas a nivel contextual

Variable	Mínimo	Media	Máximo	DE
Heterogeneidad social	0.00	2.26	4.10	0.81
Heterogeneidad lenguaje 2004	15.04	37.83	75.08	8.49
Heterogeneidad matemática 2004	14.83	37.10	63.80	8.13
Promedio ingreso familiar	5.73	10.62	13.00	1.61
Lenguaje 2004 promedio	235.64	306.19	353.92	14.17
Matemática 2004 promedio	253.60	316.71	362.94	16.91
Porcentaje mujeres	0.00	0.51	1.00	0.28
Porcentaje repitencia	0.00	0.04	0.50	0.05
Porcentaje madre indígena	0.00	0.00	0.17	0.02

Tabla 24: Resumen descriptivo variables discretas

Variable	Porcentaje
Estudiantes Mujeres	51.10%
Estudiantes repitentes	3.56%
Estudiantes con madre indígena	0.47%
Escuelas rurales	1.28%

Tabla 25: Matriz de correlaciones

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
1. Lenguaje 2006	-												
2. Matemática 2006	0.72	-											
3. Lenguaje 2004*	0.75	0.65	-										
4. Matemática 2004*	0.66	0.78	0.70	-									
5. Ingreso familiar**	0.37	0.41	0.34	0.39	-								
6. Lenguaje 2004 promedio	0.61	0.66	0.60	0.61	0.53	-							
7. Matemática 2004 promedio	0.59	0.68	0.56	0.65	0.56	0.93	-						
8. % mujeres por curso	0.05	-0.03	0.06	-0.05	-0.03	0.09	-0.08	-					
9. % repitencia	-0.37	0.40	-0.34	-0.35	-0.27	-0.57	-0.53	-0.09	-				
10. % madre indígena	-0.16	-0.19	-0.16	-0.17	-0.23	-0.26	-0.27	-0.01	0.14	-			
11. Ingreso familiar** promedio	0.46	0.52	0.43	0.49	0.75	0.72	0.75	-0.04	-0.37	-0.30	-		
12. Heterogeneidad social	0.36	0.40	0.34	0.37	0.50	0.56	0.56	-0.02	-0.32	-0.20	0.72	-	
13. Heterogeneidad lenguaje 2004	-0.14	-0.15	-0.15	-0.15	-0.07	-0.25	-0.22	-0.10	0.11	0.00	-0.09	-0.07	-
14. Heterogeneidad matemática 2004	-0.04	-0.04	-0.04	-0.05	-0.02	-0.07	-0.08	-0.04	0.03	0.00	-0.02	0.00	0.38
15. Género	-0.06 ^b	0.09 ^b	-0.08 ^b	0.11 ^b	0.05 ^b	-0.05 ^b	0.05 ^b	-0.54 ^b	0.05 ^b	0.00 ^b	0.02 ^b	0.01 ^b	0.06 ^b
16. Repite	-0.40 ^b	-0.41 ^b	-0.38 ^b	-0.36 ^b	-0.21 ^b	-0.39 ^b	-0.37 ^b	-0.07 ^b	0.69 ^b	0.09 ^b	-0.26 ^b	-0.22 ^b	0.08 ^b
17. Indígena madre	-0.15 ^b	-0.17 ^b	-0.15 ^b	-0.16 ^b	-0.28 ^b	-0.22 ^b	-0.23 ^b	-0.01 ^b	0.11 ^b	0.85 ^b	-0.25 ^b	-0.17 ^b	0.00 ^b
18. Ruralidad	-0.18 ^b	-0.21 ^b	-0.20 ^b	-0.19 ^b	-0.21 ^b	-0.33 ^b	-0.29 ^b	-0.04 ^b	0.23 ^b	0.16 ^b	-0.29 ^b	-0.24 ^b	-0.01 ^b
19. Municipal	-0.26 ^b	-0.29 ^b	-0.24 ^b	-0.25 ^b	-0.45 ^b	-0.39 ^b	-0.39 ^b	0.01 ^b	0.23 ^b	0.17 ^b	-0.60 ^b	-0.44 ^b	-0.08 ^b
20. Part. Subvencionado	0.07 ^b	0.06 ^b	0.06 ^b	0.04 ^b	0.08 ^b	0.10 ^b	0.06 ^b	0.01 ^b	-0.13 ^b	-0.06 ^b	0.12 ^b	0.26 ^b	0.14 ^b
21. Part. pagado	0.50 ^b	0.58 ^b	0.45 ^b	0.56 ^b	0.95 ^b	0.74 ^b	0.85 ^b	-0.07 ^b	-0.25 ^b	-0.29 ^b	0.99 ^b	0.45 ^b	-0.14 ^b

b. Correlaciones biserials

* Variables centradas en torno a la gran media

** Variables ingresadas como logaritmo en las estimaciones

	14.	15.	16.	17.	18.	19.	20.	21.
1. Lenguaje 2006								
2. Matemática 2006								
3. Lenguaje 2004*								
4. Matemática 2004*								
5. Ingreso familiar**								
6. Lenguaje 2004 promedio								
7. Matemática 2004 promedio								
8. % mujeres por curso								
9. % repitencia								
10. % madre indígena								
11. Ingreso familiar** promedio								
12. Heterogeneidad social								
13. Heterogeneidad lenguaje 2004								
14. Heterogeneidad matemática 2004	-							
15. Género	0.02 ^b	-						
16. Repite	0.02 ^b	-0.11 ^a	-					
17. Indígena madre	0.00 ^b	0.01 ^a	0.09 ^a	-				
18. Ruralidad	-0.01 ^b	-0.02 ^a	0.14 ^a	0.12 ^a	-			
19. Municipal	-0.13 ^b	0.01 ^a	0.16 ^a	0.17 ^a	-0.03 ^a	-		
20. Part. Subvencionado	0.15 ^b	0.01 ^a	-0.06 ^a	0.01 ^a	0.15 ^a	-0.51 ^a	-	
21. Part. pagado	-0.06 ^b	-0.03 ^a	-0.17 ^a	-0.32 ^a	-0.23 ^a	-0.49 ^a	-0.48 ^a	-

a. Correlaciones tetracóricas b. Correlaciones biserials

* Variables centradas en torno a la gran media

** Variables ingresadas como logaritmo en las estimaciones

Tabla 26: Estimación prueba de lenguaje

	Modelo1	Modelo2	Modelo3	Modelo4	Modelo5
<i>Nivel individual</i>					
Lenguaje 2004 (centrado gran media)	0.646** (280.40)	0.644** (278.75)	0.621** (259.51)	0.621** (259.52)	0.621** (259.53)
Mujer	0.900** (4.10)	0.900** (4.14)	0.710** (3.31)	0.670** (2.98)	0.669** (2.98)
Repite	-8.192** (-22.90)	-8.105** (-22.70)	-7.399** (-20.80)	-6.922** (-18.93)	-6.922** (-18.93)
Indígena madre	-0.279 (-0.60)	-0.238 (-0.54)	0.114 (0.26)	0.085 (0.19)	0.084 (0.19)
Log Ingreso familiar (Intercepto)	3.951** (20.80)	3.506** (18.24)	1.993** (9.83)	1.994** (9.84)	1.991** (9.82)
	259.139** (620.70)	250.535** (237.69)	169.205** (77.02)	175.187** (70.56)	175.697** (70.02)
<i>Nivel curso</i>					
Heterogeneidad social		5.377** (18.52)	-0.267 (-0.85)	-0.336 (-1.07)	0.116 (0.34)
Heterogeneidad lenguaje 2004		-0.002 (-0.11)	0.099** (5.19)	0.096** (5.02)	0.098** (5.08)
Log Ingreso familiar (promedio)			5.084** (8.03)	5.287** (8.21)	3.157** (3.56)
Lenguaje 2004 (promedio)			0.311** (36.71)	0.291** (31.36)	0.292** (31.39)
% de mujeres				0.335 (0.45)	0.337 (0.46)
% de repitentes				-8.863** (-5.63)	-9.237** (-5.85)
% de madres indígenas				0.557 (0.30)	0.059 (0.03)
<i>Nivel escuela</i>					
Ruralidad					2.162* (2.01)
Subvencionado					1.198* (2.29)
Pagado					4.389** (3.67)
Devianza	1,102,445	1,102,130	1,100,291	1,100,259	1,100,244
R² Nivel 1	39.67	39.69	39.76	39.76	39.76
R² Nivel 2	73.71	73.06	81.18	81.34	81.37
R³ Nivel 3	84.98	88.37	94.94	95.00	95.05

Estimación máxima verosimilitud, coeficientes NO estandarizados. * p<0.05, ** p<0.01. En paréntesis se muestra el valor t.

Tabla 27: Estimación prueba de matemática

	Modelo1	Modelo2	Modelo3	Modelo4	Modelo5
<i>Nivel individual</i>					
Matemática 2004 (centrado gran media)	0.791** (296.80)	0.789** (296.20)	0.766** (280.82)	0.765** (280.51)	0.765** (280.51)
Mujer	-4.780** (-20.00)	-4.792** (-20.03)	-4.894** (-20.63)	-5.424** (-22.23)	-5.424** (-22.23)
Repite	-10.887** (-28.00)	-10.824** (-27.85)	-10.296** (-26.57)	-9.753** (-24.72)	-9.754** (-24.72)
Indígena madre	-0.460 (-1.00)	-0.431 (-0.91)	-0.225 (-0.48)	-0.117 (-0.24)	-0.117 (-0.24)
Log Ingreso familiar (Intercepto)	2.463** (11.60)	2.116** (9.92)	0.973** (4.44)	0.960** (4.39)	0.959** (4.38)
	264.284** (432.00)	244.505** (167.58)	127.292** (44.62)	131.012** (40.07)	130.997** (39.20)
<i>Nivel curso</i>					
Heterogeneidad social		8.465** (20.10)	1.257** (2.77)	1.064* (2.37)	1.085* (2.23)
Heterogeneidad matemática 2004		0.149** (4.79)	0.126** (4.62)	0.133** (4.90)	0.129** (4.77)
Log Ingreso familiar (promedio) Matemática 2004 (promedio)			4.460** (4.65)	4.553** (4.75)	3.834** (2.98)
			0.476** (40.08)	0.450** (35.81)	0.450** (35.66)
% de mujeres				9.035** (8.71)	8.997** (8.67)
% de repitentes				-14.074** (-6.77)	-13.938** (-6.67)
% de madres indígenas				-2.757 (-1.04)	-2.696 (-1.01)
<i>Nivel escuela</i>					
Ruralidad					-0.106 (-0.06)
Subvencionado					1.927* (2.34)
Pagado					1.930 (1.06)
Devianza	1,123,135	1,122,733	1,120,659	1,120,520	1,120,514
R² Nivel 1	44.36	44.37	44.40	44.40	44.40
R² Nivel 2	63.37	63.76	73.54	74.15	74.17
R³ Nivel 3	78.22	82.93	92.49	92.92	92.92

Estimación máxima verosimilitud, coeficientes NO estandarizados. * p<0.05, ** p<0.01. En paréntesis se muestra el valor t.

Tabla 28: Estimación prueba de lenguaje y matemática diferenciado por dependencia escolar

	<i>MUNICIPALES</i>		<i>SUBVENCIONADOS</i>		<i>PAGADOS</i>	
	Leng.	Mat.	Leng.	Mat.	Leng.	Mat.
<i>Nivel individual</i>						
Lenguaje 2004 (centrado gran media)	0.611** (171.50)	-	0.630** (181.47)	-	0.619** (70.06)	-
Matemática 2004 (centrado gran media)	-	0.737** (174.97)	-	0.787** (199.82)	-	0.782** (93.18)
Mujer	-0.253 (-0.76)	-6.593** (-17.85)	1.586** (4.78)	-4.658** (-12.97)	0.303 (0.38)	-4.029** (-5.43)
Repite	-6.774** (-13.86)	-9.368** (-17.33)	-6.874** (-12.01)	-10.499** (-17.02)	-9.917** (-5.33)	-9.174** (-5.29)
Indígena madre	0.534 (0.91)	-0.143 (-0.22)	-0.476 (-0.70)	-0.117 (-0.16)	0.860 (0.17)	2.089 (0.46)
Log Ingreso familiar (Intercepto)	2.071** (7.09)	0.755* (2.34)	1.969** (6.78)	1.074** (3.42)	1.246 (1.12)	2.311* (2.23)
	170.146** (51.43)	114.221** (24.17)	181.242** (43.91)	142.064** (26.92)	223.230** (13.13)	264.586** (14.27)
<i>Nivel curso</i>						
Heterogeneidad social	-0.571 (-0.91)	0.715 (0.79)	0.932 (1.57)	1.745* (2.14)	0.236 (0.27)	-1.116 (-1.08)
Heterogeneidad rendimiento^a	0.076** (2.82)	0.133** (3.24)	0.081** (2.63)	0.142** (3.48)	0.106 (1.67)	-0.129 (-1.78)
Log Ingreso familiar (promedio)	2.962 (1.93)	2.591 (1.12)	2.291 (1.63)	3.873 (1.95)	10.548* (2.14)	0.191 (0.03)
Lenguaje 2004 (promedio)	0.320** (25.95)	-	0.276** (18.56)	-	0.099* (2.22)	-
Matemática 2004 (promedio)	-	0.516** (28.55)	-	0.409** (21.10)	-	0.105* (2.34)
% de mujeres	1.174 (1.05)	11.696** (7.08)	0.088 (0.08)	8.369** (5.44)	-0.232 (-0.11)	0.753 (0.31)
% de repitentes	-7.872** (-4.04)	-11.950** (-4.48)	-11.807** (-4.17)	-14.260** (-3.80)	4.145 (0.45)	-19.397 (-1.84)
% de madres indígenas	-0.478 (-0.20)	-0.378 (-0.11)	1.255 (0.41)	-5.247 (-1.24)	-52.578 (-1.89)	-36.320 (-1.15)
<i>Nivel escuela</i>						
Ruralidad	2.720 (1.74)	0.887 (0.35)	1.925 (1.27)	-0.558 (-0.24)	0.116 (0.02)	2.804 (0.36)
Devianza	491,394	503,082	515,497	524,808	93,204	92,124
R² Nivel 1	39.16	41.54	40.82	46.06	36.66	50.74
R² Nivel 2	86.41	79.41	71.52	62.88	-9.18	7.72
R³ Nivel 3	92.43	86.00	91.61	88.64	69.16	75.84

Estimación máxima verosimilitud, coeficientes NO estandarizados. * p<0.05, ** p<0.01. En paréntesis se muestra el valor t.

Tabla 29: Estimación prueba de lenguaje con interacciones en el modelo 2

	Modelo2a	Modelo2b	Modelo2c	Modelo2d
<i>Nivel individual</i>				
Lenguaje 2004 (centrado gran media)	0.644** (278.71)	0.644** (278.78)	0.644** (278.82)	0.645** (278.84)
Mujer	0.904** (4.16)	0.895** (4.12)	-0.162 (-0.29)	3.727** (3.16)
Repite	-8.085** (-22.64)	-8.115** (-22.73)	-8.125** (-22.75)	-8.112** (-22.71)
Indígena madre	-0.221 (-0.51)	-0.239 (-0.55)	-0.236 (-0.54)	-0.231 (-0.53)
Log Ingreso familiar	4.886** (10.25)	5.216** (5.80)	3.515** (18.30)	3.513** (18.29)
(Intercepto)	249.114** (220.32)	248.585** (177.09)	251.209** (230.15)	249.087** (202.84)
<i>Nivel curso</i>				
Heterogeneidad social	6.307** (15.84)	5.454** (18.78)	5.039** (14.96)	5.411** (18.68)
Heterogeneidad lenguaje 2004	-0.004 (-0.20)	0.042 (1.32)	-0.006 (-0.27)	0.032 (1.20)
<i>Interacción</i>				
Log Ingreso familiar* Heterogeneidad social	-0.772** (-3.17)	-	-	-
Log Ingreso familiar* Heterogeneidad lenguaje 2004	-	-0.043 (-1.95)	-	-
Mujer* Heterogeneidad social	-	-	0.667* (2.19)	-
Mujer* Heterogeneidad lenguaje 2004	-	-	-	-0.069* (-2.41)
Devianza	1,102,114	1,102,120	1,102,083	1,102,082
R² Nivel 1	39.53	39.53	39.74	39.74
R² Nivel 2	78.15	78.11	72.72	72.64
R² Nivel 3	88.51	88.55	88.36	88.39

Estimación máxima verosimilitud, coeficientes NO estandarizados. * p<0.05, ** p<0.01. En paréntesis se muestra el valor t.

Tabla 30: Estimación prueba de lenguaje con interacciones en el modelo 3

		Modelo3a	Modelo3b
<i>Nivel individual</i>			
	Lenguaje 2004 (centrado gran media)	0.608** (101.83)	0.720** (55.88)
	Mujer	0.721** (3.37)	0.732** (3.42)
	Repite	-7.429** (-20.88)	-7.447** (-20.94)
	Indígena madre	0.130 (0.30)	0.129 (0.30)
	Log Ingreso familiar	1.968** (9.73)	1.970** (9.73)
	(Intercepto)	167.900** (75.64)	171.800** (76.10)
<i>Nivel curso</i>			
	Heterogeneidad social	-0.331 (-1.03)	-0.306 (-0.95)
	Heterogeneidad lenguaje 2004	0.108** (5.61)	0.084** (4.33)
	Log Ingreso familiar (promedio)	5.170** (8.03)	5.165** (8.02)
	Lenguaje 2004 (promedio)	0.314** (36.65)	0.303** (34.99)
<i>Interacción</i>			
	Lenguaje 2004 (centrado gran media)*Heterogeneidad social	0.008* (2.39)	-
	Lenguaje 2004 (centrado gran media)*Heterogeneidad lenguaje 2004	-	-0.002** (-7.87)
	Devianza	1,100,151	1,100,095
	R² Nivel 1	40.17	40.16
	R² Nivel 2	85.31	85.30
	R² Nivel 3	94.77	94.74

Estimación máxima verosimilitud, coeficientes NO estandarizados. * p<0.05, ** p<0.01. En paréntesis se muestra el valor t.

Tabla 31: Estimación prueba de matemática con interacciones en el modelo 2

	Modelo2a	Modelo2b	Modelo2c	Modelo2d
<i>Nivel individual</i>				
Matemática 2004 (centrado gran media)	0.789** (296.21)	0.789** (296.22)	0.789** (296.17)	0.789** (296.14)
Mujer	-4.787** (-20.01)	-4.795** (-20.04)	-6.100** (-9.69)	-5.721** (-4.19)
Repite	-10.786** (-27.74)	-10.812** (-27.81)	-10.832** (-27.87)	-10.822** (-27.85)
Indígena madre	-0.407 (-0.86)	-0.421 (-0.88)	-0.443 (-0.93)	-0.444 (-0.93)
Log Ingreso familiar	3.706** (6.88)	5.079** (4.89)	2.113** (9.91)	2.117** (9.93)
(Intercepto)	243.356** (158.67)	241.620** (131.83)	245.106** (163.91)	244.886** (150.16)
<i>Nivel curso</i>				
Heterogeneidad social	9.327** (17.67)	8.346** (19.87)	8.063** (17.44)	8.521** (20.24)
Heterogeneidad matemática 2004	0.145** (4.68)	0.230** (5.38)	0.149** (4.80)	0.135** (3.72)
<i>Interacción</i>				
Log Ingreso familiar* Heterogeneidad social	-0.858** (-3.11)	-	-	-
Log Ingreso familiar* Heterogeneidad matemática 2004	-	-0.077** (-2.86)	-	-
Mujer*Heterogeneidad social	-	-	0.809* (2.33)	-
Mujer*Heterogeneidad matemática 2004	-	-	-	0.025 (0.72)
Devianza	1,122,718	1,122,719	1,122,652	1,122,657
R² Nivel 1	44.14	44.14	44.55	44.55
R² Nivel 2	65.06	65.11	64.76	64.73
R² Nivel 3	82.62	82.68	82.77	82.78

Estimación máxima verosimilitud, coeficientes NO estandarizados. * p<0.05, ** p<0.01. En paréntesis se muestra el valor t.

Tabla 32: Estimación prueba de matemática con interacciones en el modelo 3

	Modelo3a	Modelo3b
<i>Nivel individual</i>		
Matemática 2004 (centrado gran media)	0.710** (99.57)	0.793** (50.47)
Mujer	-4.855** (-20.49)	-4.845** (-20.44)
Repite	-10.409** (-26.89)	-10.310** (-26.65)
Indígena madre	-0.205 (-0.43)	-0.173 (-0.37)
Log Ingreso familiar	0.948** (4.34)	0.967** (4.43)
(Intercepto)	124.346** (42.74)	126.000** (42.56)
<i>Nivel curso</i>		
Heterogeneidad social	1.300** (2.79)	1.195* (2.56)
Heterogeneidad matemática 2004	0.142** (5.11)	0.126** (4.50)
Log Ingreso familiar (promedio)	4.437** (4.51)	4.307** (4.38)
Matemática 2004 (promedio)	0.483** (39.77)	0.482** (39.28)
<i>Interacción</i>		
Matemática 2004 (centrado gran media)*Heterogeneidad social	0.033** (8.53)	-
Matemática 2004 (centrado gran media)*Heterogeneidad matemática 2004	-	-0.001 (-1.82)
Devianza	1,120,358	1,120,427
R² Nivel 1	45.06	45.04
R³ Nivel 2	76.71	73.86
R² Nivel 3	92.19	92.19

Estimación máxima verosimilitud, coeficientes NO estandarizados. * p<0.05, ** p<0.01. En paréntesis se muestra el valor t.