



Prueba SIMCE 4° Básico 2002

Factores que Inciden en el

Rendimiento de los Alumnos

Nota Técnica

Agosto de 2003

Departamento de Estudios y Estadísticas

Prueba SIMCE 4° Básico 2002

Factores que Inciden en el Rendimiento de los Alumnos

Este documento ha sido elaborado por el Departamento de Estudios y Estadísticas de la División de Planificación y Presupuesto del Ministerio de Educación. El documento fue desarrollado por los profesionales Paula Darville, Rodrigo Díaz y Fernanda Melis bajo la dirección de Vivian Heyl, Jefa del Departamento de Estudios y Estadísticas. Adicionalmente, se agradece el apoyo prestado por el Sistema Nacional de Medición de la Calidad de la Educación (SIMCE) de la Unidad de Curriculum y Evaluación, y en especial, las sugerencias, comentarios y apoyo técnico de Claudia Matus, Coordinadora de Informática y Estadísticas del SIMCE.

Ministerio de Educación
República de Chile

Alameda 1371, Santiago

Agosto de 2003

I. INTRODUCCIÓN

El objetivo de esta nota técnica es indagar sobre los factores que inciden en el rendimiento de los alumnos que rindieron la Prueba SIMCE 4° Básico 2002. Para esto, se utiliza el enfoque tradicional de *función de producción* que analiza la relación entre insumos y productos del proceso educacional. El producto se asocia al resultado de una prueba estandarizada (en este caso el SIMCE de 4° Básico de 2002) y dentro de los insumos se incluyen características de los establecimientos, los alumnos y sus familias¹.

En este trabajo se utilizan distintas metodologías para la estimación de funciones de producción educacional con el objetivo de lograr un mayor entendimiento del proceso educativo y de los factores que están detrás de logro de los estudiantes. En primer lugar se estiman funciones de producción por Mínimos Cuadrados Ordinarios (MCO). Luego, se estiman regresiones de cuantiles, que permiten analizar cómo afectan los insumos educativos en los distintos puntos de la distribución de puntajes. Finalmente, se estiman Modelos Lineales Jerárquicos (Hierarchical Linear Models - HLM), que permiten trabajar con datos multinivel (alumnos y establecimientos). Todas estas regresiones tienen como variable dependiente los resultados a nivel de alumnos de la prueba SIMCE 2002, por lo que estiman el efecto acumulado hasta el momento en que se tomó la prueba, de los insumos educativos sobre el logro de los estudiantes.

De manera adicional, se estima por MCO una función a nivel de establecimientos, en que la variable dependiente es la diferencia de puntaje entre 2002 y 1999. Esta estimación permite explorar cuáles son los factores que inciden en el aumento o disminución del puntaje promedio de los establecimientos en el período de tiempo (1999-2002).

Además de esta introducción, el documento está organizado en tres secciones. En la primera, se presenta el marco teórico de funciones de producción educacional. Luego, en la segunda sección, se discuten aspectos importantes de las distintas metodologías de estimación y se presentan los resultados de las estimaciones para la prueba SIMCE de 4° Básico 2002 en Lenguaje y Matemáticas. Finalmente, en la tercera sección, se presentan las conclusiones.

¹ Los estudios más recientes de estimaciones de *funciones de producción* educacional para Chile son: Mizala y Romaguera (2000, 2002); Bravo, Contreras y Sanhueza (1999); y Dresdner y Valdebenito (2002).

II. MARCO TEÓRICO

1. Función de Producción Educativa

El modelo tradicional con que se representa simplificada el proceso educativo es la *función de producción educativa* que busca explicar el *logro escolar* como función de las características del alumno, la familia, el establecimiento y otras variables de recursos escolares o del ambiente en que se desenvuelve el estudiante.

El proceso educativo tiene muchos objetivos, por ejemplo, entregar determinados conocimientos académicos, lograr un desarrollo integral, físico e intelectual del alumno, la inserción laboral o la integración dentro de la sociedad. Luego, los resultados esperados de la educación son múltiples, y por lo tanto, una primera dificultad que surge es *cómo* medir el logro escolar. En general, los estudios utilizan como medida fundamental el rendimiento de los alumnos, sin embargo, hay que tener claro que esta medida no incorpora todos los resultados esperados del proceso educativo². La manera más usual para medir el rendimiento de los alumnos ha sido a través de resultados de una prueba estandarizada, que en el caso chileno, en general ha correspondido al test estandarizado SIMCE.

En Mizala y Romaguera (2000) se discuten otras dificultades adicionales en relación a la estimación de funciones de producción. Por una parte, se plantea que en general sólo se consideran algunas variables de entrada (insumos) y no la totalidad de posibles factores determinantes del rendimiento escolar. Esto, ya que prácticamente es imposible identificar y cuantificar todas las variables que afectan la calidad de la educación. Además, existen dificultades importantes en los datos disponibles, usando muchas veces datos imperfectos como *proxies* de las verdaderas variables. Por otra parte, se supone que las variables consideradas pueden ser modificadas, cuando en la práctica no lo son (algunos insumos no pueden ser afectados).

A pesar de lo anterior y considerando que las estimaciones de funciones de producción educativa son una representación simplificada del proceso educativo³, éstas constituyen un instrumento útil

² Podrían utilizarse otras medidas como inserción en el mercado laboral, tasas de retención o deserción, etc.

³ Obviamente la *realidad* es mucho más compleja de lo que se puede plantear en un modelo teórico estadístico que explique el logro educativo.

para examinar la importancia relativa de distintos factores que afectan el logro escolar, ya que a partir de ellas se identifican variables que inciden significativamente en el rendimiento educacional.

Siguiendo la metodología tradicional, se plantea el siguiente modelo de función de producción educacional:

$$Y_i^t = f(A_i^t, F_i^t, E_i^t, C_i^t) + \varepsilon \quad [1]$$

Donde:

Y_i^t = Logro escolar del alumno i en el período t.

A_i^t = Características del alumno i en el período t.

F_i^t = Características de la familia del alumno i en el período t.

E_i^t = Características del establecimiento y los profesores del alumno i en el período t.

C_i^t = Características del entorno del alumno i en el período t.

ε = Error aleatorio.

Las características del alumno que afectan su logro escolar se refieren a habilidades innatas, potencial de aprendizaje, motivación y personalidad que difícilmente pueden ser medidas. Adicionalmente se incorporan otras variables relacionadas con el estudiante, como experiencia escolar previa⁴ y si el alumno es hombre o mujer.

Con respecto a la influencia de la familia, dos parecen ser las características con mayor incidencia en el rendimiento del alumno. El nivel económico de la familia, que se traduce en una mayor probabilidad de disponer de medios que favorecen el aprendizaje (recursos materiales, etc.), y el nivel educativo de los padres, que los involucra de manera más directa en el proceso educativo. Ambas variables se relacionan directamente con el nivel socioeconómico de las familias.

Dentro de las características del establecimiento se incluyen condiciones físicas y estructurales, tales como la existencia de bibliotecas, laboratorios, etc, y otras variables como tamaño del establecimiento y relación alumno-profesor. Adicionalmente, se consideran características de los docentes, como años de experiencia y posesión de título de profesor.

⁴ Por ejemplo, si el alumno ha asistido o no a educación preescolar.

Finalmente, la influencia del entorno se puede considerar a través de variables como el nivel socioeconómico o académico del resto de los estudiantes del establecimiento.

Como ya se ha planteado, diversos factores no pueden ser adecuadamente medidos por lo que al estimar el modelo existen variables que deben ser omitidas. El término error (ε) captura el efecto de todas las variables que no puedan ser incluidas en el modelo. Si las variables omitidas no están correlacionadas con las variables explicativas incluidas, los resultados de las regresiones son ineficientes pero insesgados⁵.

Resulta importante destacar que en las estimaciones de ecuaciones como [1] se usan datos de corte transversal. Es información puntual de un momento y por lo tanto no dan al proceso educacional características de continuidad a través del tiempo. Es decir, se basan en un análisis de *valor bruto* del logro del estudiante. Sin embargo, se ha discutido en la literatura que una mejor estimación es aquella que se basa en el *valor agregado* del proceso educativo, la cual requiere contar con observaciones en dos momentos del tiempo para un mismo grupo de estudiantes⁶.

III. ESTIMACIÓN DE FUNCIONES DE PRODUCCIÓN EDUCACIONAL

1. Estimación de Mínimos Cuadrados Ordinarios y de Cuantilas

1.1 Metodología de Estimación

La estimación por mínimos cuadrados ordinarios (MCO) permite conocer el **efecto promedio** de los distintos insumos educativos (características del alumno y de su establecimiento) sobre el logro escolar. Esta metodología tiene algunos supuestos implícitos importantes. Supone que el modelo de regresión es lineal en los parámetros, los valores de las variables dependientes son fijos en muestreo repetido, el valor medio del error aleatorio es cero, homocedasticidad o igual varianza en el componente aleatorio, no autocorrelación entre los residuos, las covarianzas entre el error y las variables explicativas son nulas, el número de observaciones de la muestra debe superar

⁵ Es decir, las variables consideradas no explican completamente el logro escolar (R^2 bajos), pero los valores de los coeficientes estimados corresponden a los verdaderos valores de los parámetros.

⁶ La Prueba SIMCE no permite realizar este tipo de estudios, dado que se aplica una vez a los alumnos de un cierto nivel.

ampliamente el número de parámetros a estimar, variabilidad en los valores de las variables explicativas, correcta especificación del modelo de regresión, no multicolinealidad perfecta. Estos supuestos son *fuertes*, si se considera que la mayoría de las variables explicativas son del establecimiento, y por lo tanto son exactamente iguales para todos los alumnos que asisten a dicho establecimiento.

Sin embargo, MCO es la metodología más utilizada, por lo que permite comparar los resultados con otros estudios y además analizar cómo difieren las estimaciones de otras metodologías. La ecuación que se estima, basado en la función de producción educacional estándar [1], toma la siguiente forma:

$$Y_i = \beta_0 + \sum_j \beta_j X_{ij} + \varepsilon_{ij} \quad [2]$$

Donde:

Y_i = Logro del alumno i.

β_0 = Constante a estimar.

β_j = Son los parámetros a estimar.

X_{ij} = Característica j del alumno i.

ε_{ij} = Error aleatorio, $\varepsilon_{ij} \sim N(0, \sigma^2)$

Dentro de las características X_j hay características del alumno y también características del establecimiento.

Una de las principales críticas a las estimaciones MCO es que no consideran un posible efecto diferenciado de las características de los alumnos y establecimientos en el logro educacional según la posición en la distribución de puntajes del estudiante. Es decir, no considera que algunos insumos educativos afecten de manera diferente a estudiantes con altos puntajes que a estudiantes con bajos puntajes. Con el fin de capturar estos efectos heterogéneos, además de las estimaciones MCO, se estiman regresiones como [2] usando el Método de Cuantiles. Este método analiza la relación entre la variable dependiente (logro del alumno) y las variables explicativas en distintos puntos de la distribución de puntajes.

1.2 Especificación del Modelo

Se estima un modelo como [2] que relaciona el logro educacional del estudiante, con características de los alumnos y los establecimientos a los que asisten. La variable dependiente es el resultado a nivel de alumnos de la prueba SIMCE 4° Básico 2002. Dentro de las variables explicativas (X_i) se incorpora un conjunto de variables relacionadas con las características de los alumnos y establecimientos.

A nivel de alumnos se considera el nivel socioeconómico medido a través del índice socioeconómico (ISE), que se construye para cada alumno en base a la información de educación de los padres y del ingreso familiar⁷. También se incluye el índice socioeconómico al cuadrado para analizar si existe algún grado de no linealidad en la relación de esta variable y los resultados del SIMCE⁸. Las otras variables que se incluyen son: género, incluida como una dummy que toma el valor uno cuando el alumno es mujer y cero cuando es hombre, y asistencia a preescolar, también incluida como una dummy que toma el valor uno cuando el alumno asistió a kinder y/o pre-kinder y cero en caso contrario.

A nivel de establecimientos se incluye la experiencia docente, medida como el promedio de años de experiencia de los profesores del establecimiento; el tamaño del establecimiento, medido como el número total de alumnos en enseñanza básica; la tasa alumno-profesor, correspondiente al cociente entre el número de alumnos en enseñanza básica y el número de profesores que imparte clases en enseñanza básica en cada establecimiento⁹; la dependencia, incorporada a través de variables dummies para los establecimientos particulares pagados y subvencionados; el área geográfica en que está el establecimiento, medida a través de una dummy que toma valor 1 para establecimientos rurales y 0 para los establecimientos urbanos; y una variable dummy de financiamiento padres, que toma valor 1 para los establecimientos particulares subvencionados con financiamiento compartido y para los establecimientos particulares pagados, y toma valor cero para los establecimientos municipales y particulares subvencionados sin financiamiento compartido.

⁷ Este índice es calculado con la herramienta de análisis multivariado de componentes principales. Ver Mizala y Romaguera (2000).

⁸ Un signo negativo en esta variable implica que aumentos del nivel socioeconómico más allá de un cierto punto no contribuyen a aumentar los resultados de la prueba SIMCE.

⁹ Estos datos fueron corregidos según jornadas equivalentes. Cada jornada equivalente corresponde a un docente que trabaja 30 o más horas en el establecimiento.

Adicionalmente, para medir el efecto de los programas del Ministerio en el logro escolar de los alumnos, se incluyen variables dummies para la JEC, Programa Rural, P900 y Escuelas Críticas¹⁰. Estas variables toman el valor uno si el establecimiento pertenece al programa y cero en caso contrario. En el caso del P900, los establecimientos se separaron en tres grupos: los que *Ingresan* al programa, los que *Salen* y los que *Permanecen*. El primer grupo corresponde a los establecimientos que en 2000 no están en el programa y en 2001 ingresan al programa, el segundo grupo corresponde a los establecimientos que en 2001 y 2002 no están en el programa, pero que en los años anteriores pertenecieron al P900, y el último grupo corresponde a los establecimientos que en 2000 pertenecen al programa y en 2001 siguen perteneciendo¹¹.

Como novedad en relación a otros estudios de función de producción, se incorporan dos variables relacionadas con los docentes. Un índice de confianza de los docentes en su preparación para enseñar cada sector de aprendizaje y un índice de cobertura curricular de cada materia. Ambas variables fueron construidas en base a información proveniente de los cuestionarios realizados a los docentes que enseñan a los alumnos que rinden la prueba. Para mayor detalle ver Mineduc (2003).

La descripción de las variables incluidas en el modelo y las fuentes a partir de las cuales fueron obtenidas se muestran en el Cuadro 1.

En las estimaciones se utilizan sólo los casos que tienen información válida para todas las variables incorporadas en el modelo. Esto significa que un porcentaje importante de los alumnos queda excluido de las estimaciones econométricas por no contar con información para alguna de las variables. La principal motivo por el cual se excluyen alumnos es por que no se puede construir su

¹⁰ Las Escuelas Críticas corresponden a las 66 escuelas de la región metropolitana con peores resultados en la prueba SIMCE 4° Básico 1999. Éstas cuentan con apoyo de instituciones externas (ONGs, universidades y fundaciones) con el objetivo de mejorar el resultado en el SIMCE para el año 2005.

¹¹ El momento del quiebre para construir los grupos P900 (2000-2001), se escoge debido a que a partir de los resultados de la prueba SIMCE 4° Básico 1999, a comienzos el año 2001 se selecciona a los establecimientos que participarían en el programa por los próximos tres años. Algunos de los establecimientos que se seleccionaron ya participaban en el P900 (*Permanecen*) mientras que otros se incorporan al programa (*Ingresaron*). La selección se realiza en base al último resultado SIMCE de 4° Básico.

índice socioeconómico, ya que no cuentan con la información del cuestionario de padres y apoderados¹².

Cuadro 1

Variables	Descripción	Fuente de Información
Preesc	Dummy. 1 si el alumno asistió a Kinder o Prekinder y 0 en caso contrario	Cuestionario a Padres y Apoderados SIMCE
Mujer	Dummy que toma valor 1 si es mujer	
ISE	Índice socioeconómico del alumno	
ISE ²	Índice socioeconómico del alumno al cuadrado	
IP900	Dummy de los que ingresaron a P900 el 2001	Bases de datos de los programas. Información 2002.
PP900	Dummy de los que permanecen en P900 el 2001	
SP900	Dummy de los que salen del P900	
PRUR	Dummy que toma valor 1 si el establecimiento participa en el Programa Rural	
JEC	Dummy JEC que toma valor 1 si el establecimiento tiene JEC	
CRIT	Dummy de las Escuelas Críticas	Bases de datos matrícula y docentes año 2001 MINEDUC
EXP	Años de experiencia promedio de los docentes	
MAT	Matrícula enseñanza básica	
MAT ²	Matrícula enseñanza básica al cuadrado	
CAP	Cuociente Alumno-Profesor en enseñanza básica	
PSUB	Dummy Particular Subvencionado	
PPAG	Dummy Particular pagado	
FPAD	Dummy Financiamiento Padres	
RUR	Dummy que toma valor 1 si el establecimiento es rural.	Cuestionario a Docentes SIMCE
CONF	Índice de confianza de los docentes en su preparación en el sector de aprendizaje	
COB	Cobertura curricular en el sector de aprendizaje	Cuestionario a Padres y Apoderados SIMCE
ISEP	Nivel socioeconómico promedio de los alumnos del establecimiento	
ISEP ²	Nivel socioeconómico promedio de los alumnos del establecimiento al cuadrado	

De un total de 251.408 alumnos que rinde la prueba de Lenguaje, sólo hay 85.641 casos con toda la información válida para las regresiones por MCO (aproximadamente el 34%). En la prueba de Matemáticas la situación es similar. De los 251.182 alumnos que rinde la prueba, sólo 81.612 (32%) tiene toda la información necesaria¹³.

¹² Además del alto porcentaje de los padres que no contesta la encuesta, en esta aplicación del cuestionario la pregunta relacionada con nivel educacional es complicada de responder por lo que muchos padres la omitieron.

¹³ En el anexo 1 se presentan las estadísticas descriptivas para la población total y para los datos utilizados en la regresión.

Las estadísticas descriptivas muestran que los datos utilizados en la regresión no parecen tener diferencias sustantivas con los datos totales (en ambas pruebas). Sin embargo, la distribución del total de puntajes de la prueba de lenguaje es distinta a la distribución de los casos utilizados en la regresión.¹⁴

En el anexo 2 se muestran las estadísticas descriptivas de los datos usados en cada una de las regresiones.

1.3 Resultados

A partir de la especificación señalada en la sección anterior, se estima el modelo utilizando los resultados de las pruebas de Matemáticas y Lenguaje por separado. Dentro de las variables explicativas del modelo podría existir algún grado de multicolinealidad, lo que no cumpliría uno de los supuestos básicos de MCO. Para analizar la existencia de multicolinealidad entre las variables, se realizaron las pruebas pertinentes, concluyéndose que no existen indicios de multicolinealidad (ver anexo 3).

En primer lugar se estima el modelo utilizando MCO y luego para las cuantiles 0.10, 0.25, 0.50, 0.75 y 0.90. En el Cuadro 2 se muestran los resultados para Lenguaje y en el Cuadro 3 se muestran los resultados para Matemáticas.

La regresión para la cuantila x estima la relación entre puntaje e insumos educativos en el percentil x de puntajes. Es decir, la regresión para la cuantila 0.10 estima los resultados en el 10% de puntajes más bajos, mientras que en la cuantila 0.90 estima los resultados en el 10% de mejores puntajes. Las cuantiles 0.25 0.50 0.75 representan situaciones intermedias.

Las estimaciones por MCO muestran que el nivel socioeconómico de los alumnos tiene una fuerte incidencia en los resultados de las pruebas de Lenguaje y Matemáticas. Al observar las estimaciones para las diferentes cuantiles de resultados se observa un impacto menor del nivel socioeconómico en el cuantil 0.75 y en el cuantil 0.90, es decir, los correspondientes a los estudiantes de mayor rendimiento.

¹⁴ Ver anexo 1 para mayor detalle.

Con respecto a la variable de género, se observa que en la prueba de Lenguaje, las niñas obtienen en promedio cerca de 7 puntos más que los niños, mientras que en Matemáticas los niños obtienen en promedio alrededor de 4 puntos más que las niñas. Sin embargo, este efecto es diferenciado a través de la distribución de puntajes. En la prueba de Lenguaje las mayores diferencias entre niñas y niños se observan en los estudiantes de menor rendimiento (cuantil 0.10), las que van disminuyendo entre los estudiantes de mejores desempeños. Por otro lado, en la prueba de Matemáticas sucede lo contrario, ya que las mayores diferencias entre niños y niñas se producen en los alumnos de mejores rendimientos (cuantil 0.90) y las menores diferencias se observan en el cuantil 0.10.

También se observa que el tamaño del establecimiento y la experiencia docente tienen un efecto significativo y positivo sobre el logro escolar. Del análisis de cuantiles, no se observa que este efecto sea diferenciado entre estudiantes de alto y bajo rendimiento.

En Lenguaje, los alumnos en establecimientos rurales tienen mejores resultados que los alumnos en establecimientos urbanos de similares características. En Matemáticas la dummy rural no es significativa, por lo que no existe una diferencia en los resultados de los alumnos en establecimientos rurales y urbanos.

Los establecimientos que están en Jornada Escolar Completa (JEC) y los que cuentan con financiamiento de los padres tienen mejores resultados en la prueba SIMCE en comparación a establecimientos de similares características, respectivamente. En Lenguaje los alumnos en JEC tienen en promedio cerca de 3 puntos más que aquellos que no están, y además se observa que este impacto es similar en todas las cuantiles. Por otra parte, en Matemáticas el efecto de la JEC es menor, ya que los alumnos en JEC sólo obtienen cerca de 1 punto más que aquellos de similares características que no están. Incluso para las cuantiles 0.25, 0.50 y 0.90 este efecto no es estadísticamente significativo. Con respecto al financiamiento de padres se observa que el efecto es más importante en los alumnos del cuantil 0.25, en comparación con el resto de los estudiantes según distribución de puntajes.

El efecto de la dependencia del establecimiento sobre el logro muestra que los establecimientos particulares pagados tienen mejores resultados que los particulares subvencionados, y estos a su vez, tienen mejores resultados que los municipales, tanto en la prueba de Lenguaje como en Matemáticas. Si bien estas diferencias se observan en toda la distribución de puntajes, el efecto de la dependencia disminuye en los tramos altos de la distribución de puntajes (cuantiles 0.75 y 0.90).

Cuadro 2

Regresiones por MCO y Cuantilas. Prueba SIMCE 4º básico 2002

	MCO		q10		q25		q50		q75		p90	
	Coef.	t	Coef.	t	Coef.	t	Coef.	t	Coef.	t	Coef.	t
Preesc	2,11*	2,91	1,01	0,74	0,27	0,24	2,19*	2,39	3,04	3,53	3,73*	3,58
Mujer	7,90*	23,66	14,61*	23,31	11,41*	21,66	6,85*	16,14	4,43	11,11	4,36*	9,19
IP900	-14,75*	-29,04	-13,28*	-13,54	-16,05*	-19,70	-16,24*	-25,16	-14,83	-24,69	-13,93*	-19,58
PP900	-20,44*	-27,87	-19,71*	-14,03	-23,13*	-19,79	-23,58*	-25,30	-19,31	-22,20	-17,50*	-16,90
SP900	-8,65*	-16,43	-7,31*	-7,31	-9,60*	-11,48	-9,59*	-14,34	-8,34	-13,34	-8,00*	-10,77
PRUR	-7,44*	-6,68	-8,37*	-4,01	-8,42*	-4,79	-9,31*	-6,57	-8,58	-6,46	-4,93*	-3,09
JEC	2,54*	6,94	3,24*	4,71	3,14*	5,43	2,06*	4,43	2,67	6,13	2,33*	4,46
CRIT	-28,29*	-17,78	-26,01*	-8,70	-30,36*	-12,08	-31,70*	-15,69	-26,92	-14,20	-24,38*	-10,81
EXP	0,41*	13,12	0,54*	9,19	0,51*	10,25	0,44*	11,00	0,34	9,13	0,30*	6,69
MAT	0,01*	8,60	0,01*	5,33	0,01*	7,33	0,01*	6,95	0,01	6,29	0,00*	3,64
MAT ²	0,00*	-6,68	0,00*	-4,71	0,00*	-6,07	0,00*	-5,31	0,00	-4,69	0,00*	-2,80
CAP	0,01*	3,24	0,01	1,77	0,02*	4,38	0,01*	2,57	0,00	1,58	0,00	0,25
PSUB	4,73*	7,93	7,92*	7,07	6,82*	7,25	4,87*	6,43	3,24	4,55	3,10*	3,68
PPAG	16,86*	13,04	26,60*	11,24	24,53*	12,36	17,17*	10,46	14,05	8,88	10,23*	5,18
FPAD	12,16*	20,39	12,28*	11,06	15,56*	16,61	13,41*	17,70	10,71	14,92	9,56*	11,12
RUR	3,23*	4,76	5,56*	4,25	5,52*	5,10	3,72*	4,31	1,16	1,44	1,12	1,16
ISE	8,86*	37,73	9,83*	22,10	10,40*	28,22	9,42*	31,57	7,82	27,51	7,08*	20,31
ISE ²	0,27*	2,26	1,12*	4,84	0,05	0,28	-0,26	-1,68	0,01	0,05	-0,08	-0,47
CONF	1,58*	4,34	1,24	1,81	1,56*	2,71	1,69*	3,65	1,57	3,61	0,78	1,49
COB	5,19*	13,26	6,69*	9,05	6,15*	9,91	5,18*	10,43	4,18	9,02	4,56*	8,29
Constante	210,50*	121,09	130,71*	40,07	169,81*	61,85	214,82*	97,24	252,63	121,62	280,80*	113,92
N		85641		85641		85641		85641		85641		85641
R ² Ajustado		0,146		0,073		0,085		0,087		0,079		0,072

Nota: (*) Coeficiente significativo al 5%.

Cuadro 3

Regresiones por MCO y Cuantilas. Prueba SIMCE 4º básico 2002												
Matemáticas												
	<i>MCO</i>		<i>q10</i>		<i>q25</i>		<i>q50</i>		<i>q75</i>		<i>p90</i>	
	<i>Coef.</i>	<i>t</i>	<i>Coef.</i>	<i>t</i>	<i>Coef.</i>	<i>t</i>	<i>Coef.</i>	<i>t</i>	<i>Coef.</i>	<i>t</i>	<i>Coef.</i>	<i>t</i>
Preesc	2,02*	2,68	2,21	1,67	2,44*	2,14	2,00*	2,06	1,22	1,27	1,37	1,20
Mujer	-4,12*	-11,91	-1,52*	-2,52	-3,46*	-6,63	-4,79*	-10,74	-4,99*	-11,39	-5,21*	-10,05
IP900	-13,22*	-25,19	-13,33*	-14,21	-13,91*	-17,36	-14,02*	-20,70	-13,27*	-20,17	-12,26*	-15,89
PP900	-19,06*	-24,95	-17,40*	-12,78	-20,13*	-17,32	-20,18*	-20,47	-21,77*	-22,65	-19,65*	-17,39
SP900	-7,26*	-13,29	-7,90*	-8,21	-8,60*	-10,41	-7,22*	-10,25	-6,57*	-9,53	-5,87*	-7,22
PRUR	-6,08*	-5,18	-7,70*	-3,74	-7,97*	-4,49	-6,38*	-4,21	-5,28*	-3,56	-4,32*	-2,46
JEC	0,89*	2,36	1,42*	2,13	0,71	1,23	0,33	0,67	1,30*	2,70	0,65	1,15
CRIT	-27,80*	-16,82	-21,36*	-7,38	-24,72*	-9,91	-30,21*	-14,17	-32,48*	-15,53	-26,86*	-10,84
EXP	0,40*	12,40	0,44*	7,81	0,44*	8,94	0,40*	9,58	0,41*	10,04	0,38*	7,74
MAT	0,01*	10,29	0,01*	5,70	0,01*	7,53	0,01*	9,07	0,01*	8,81	0,01*	6,22
MAT ²	0,00*	-7,09	0,00*	-3,70	0,00*	-5,33	0,00*	-6,25	0,00*	-6,53	0,00*	-4,93
CAP	0,00*	1,67	0,01*	2,32	0,01*	2,48	0,00	0,76	0,00	0,16	0,00	-0,42
PSUB	4,63*	7,48	5,94*	5,48	5,33*	5,70	5,26*	6,59	4,72*	6,02	3,01*	3,27
PPAG	20,81*	15,62	25,97*	11,20	24,06*	12,12	22,61*	13,15	18,01*	10,54	17,16*	8,28
FPAD	10,82*	17,53	12,00*	11,16	12,84*	13,87	11,56*	14,51	9,63*	12,21	8,05*	8,59
RUR	0,51	0,72	1,81	1,44	0,42	0,39	0,46	0,51	0,55	0,63	1,01	0,97
ISE	8,37*	34,38	9,49*	22,18	9,71*	26,63	9,01*	28,69	7,75*	24,85	6,48*	17,27
ISE ²	0,31*	2,55	1,02*	4,56	0,42*	2,23	-0,09	-0,58	-0,01	-0,05	-0,18	-0,96
CONF	3,78*	6,16	3,95*	3,66	3,52*	3,79	4,29*	5,42	3,64*	4,69	4,08*	4,38
COB	6,07*	13,92	6,41*	8,33	6,51*	9,81	6,67*	11,83	6,39*	11,62	4,93*	7,55
Constante	203,22*	100,11	131,39*	36,99	165,99*	54,31	201,41*	76,88	239,37*	93,07	273,09*	89,45
N		81612		81612		81612		81612		81612		81612
R ² Ajustado		0,141		0,071		0,081		0,083		0,076		0,066

Nota: (*) Coeficiente significativo al 5%.

Con respecto a los programas P900, Rural y al conjunto de Escuelas Críticas, se observa que los estudiantes en establecimientos que participan en dichos programas tienen puntajes significativamente menores que los alumnos que estudian en el conjunto de establecimientos comparables, respectivamente. Si bien esto implica que estos establecimientos tienen, en promedio, peores puntajes, no significa que hayan disminuido su promedio entre 1999 y 2002¹⁵. Además, una razón que podría explicar el efecto negativo de estos programas en el logro, es que estas variables estén jugando el rol de proxies de pobreza, ya que como se sabe, estos establecimientos están dentro de los más pobres del país.

El índice de confianza en la preparación para enseñar de los docentes tiene un importante impacto al momento de explicar los resultados, especialmente en Matemáticas. En efecto, en esta prueba aquellos niños que tienen profesores con un alto índice de confianza en su preparación tienen alrededor de 6 puntos más que aquellos que tienen un índice bajo. Sin embargo, en la prueba de Lenguaje esta variable no tiene impacto en los resultados de los alumnos.

En relación a la cobertura curricular, esta variable muestra un impacto positivo en el logro de los alumnos, tanto en Lenguaje como en Matemáticas. Los alumnos cuyos profesores tienen un índice alto de cobertura curricular pueden llegar a tener, en promedio, 15 puntos más que aquellos que tienen un índice bajo.

2. Estimación de Modelos Lineales Jerárquicos (HLM)

2.1 Metodología de Estimación

En esta sección se estima una función de producción educacional con Modelos Lineales Jerárquicos (HLM). La principal ventaja de estos modelos es que permiten trabajar con información de distintos niveles para explicar los logros educacionales.

El modelo general de función de producción (ecuación [1]) se desagrega en dos niveles. El primero detalla las características de cada alumno y su familia (nivel 1: alumno) y el segundo nivel incorpora las características del establecimiento y sus profesores (nivel 2: establecimiento) para estimar los coeficientes del primer nivel. Esta característica de los Modelos Lineales Jerárquicos es

¹⁵ Un análisis de este tipo se desarrolla en la sección 3 de este documento.

clave, ya que reconoce que existe una dependencia del alumno con el establecimiento al que asiste. El modelo general se presenta a continuación:

$$\text{Nivel 1:} \quad Y_{ij} = \beta_{0j} + \sum_{k=1}^n \beta_{kj} X_{kij} + r_{ij} \quad [3]$$

$$\text{Nivel 2:} \quad \beta_{kj} = \gamma_{k0} + \sum_l \gamma_{kl} W_{lj} + \mu_{kj} \quad k = 0, \dots, n \quad [4]$$

Donde:

Y_{ij} = Logro escolar del alumno i que pertenece al establecimiento j.

X_{kij} = Característica k del alumno i que pertenece al establecimiento j.

β_{0j} = Logro promedio del establecimiento j. Coeficiente a estimar.

β_{kj} = Coeficientes del primer nivel que son modelados a partir de las variables del segundo nivel.

Efecto que tiene la característica k sobre un alumno del establecimiento j.

r_{ij} = Efecto aleatorio del nivel 1, $r_{ij} \sim N(0, \sigma^2)$.

W_{lj} = Característica l del establecimiento j.

γ_{k0} = Efectos fijos. Es el promedio general sobre los colegios de la característica k.

γ_{kl} = Coeficientes del segundo nivel (efectos fijos).

μ_{ij} = Efectos aleatorios del nivel 2, $\mu_{ij} \sim N_{multi\ var\ iada}(0, \tau_{ij})$.

La manera tradicional de construir un Modelo Lineal Jerárquico es a través de un análisis sistemático, en que se parte del modelo más simple (que sólo incluye las constantes y los términos aleatorios) para luego ir agregando variables explicativas en ambos niveles (alumno y establecimiento). En general, el criterio para incorporar una variable en el modelo es que su inclusión reduzca la varianza total (σ^2), y en el caso de ser una variable de segundo nivel que reduzca la varianza del componente que intenta explicar.

Las principales ventajas de trabajar con los modelos HLM, en relación a las estimaciones a nivel de alumnos con MCO de la sección anterior, son que:

- Considera la heterogeneidad que existe a través de los establecimientos. Es decir, considera que la relación entre las características de los alumnos y los logros educativos varía a

través de los establecimientos. Esto es importante ya que reconoce la existencia de un **efecto establecimiento**. La ecuación [4] muestra cómo las características del establecimiento (W) permiten estimar los coeficientes a nivel de alumno.

- HLM resuelve el problema de una inadecuada estimación de los errores estándar debido a la dependencia de las variables a nivel de alumno con el establecimiento al que asisten. Este problema se resuelve incorporando un término aleatorio único para cada establecimiento.
- Corrige el problema que se produce al agregar información de los alumnos al nivel de establecimientos (sesgo de agregación). Esto se soluciona, pues la relación de las variables se descompone al nivel que correspondan.

2.2 Especificación del Modelo

Tal como se indica, se parte con el ajuste de un modelo básico al que se le van agregando variables explicativas en ambos niveles, de manera de explicar la varianza total y la varianza de cada nivel considerado. La primera estimación (modelo 1) sólo incluye las constantes y los términos aleatorios de cada nivel. Luego, el modelo 2 amplía las variables en el primer nivel, considerando las características del alumno y su familia (nivel socioeconómico, nivel socioeconómico al cuadrado, asistencia a educación preescolar y género). Finalmente se estima un modelo 3 que incluye variables a nivel de establecimiento. Estas variables son las mismas que se incorporaron en las estimaciones por MCO y cuantilas. La novedad es que también se incluye una variables del nivel socioeconómico promedio de los alumnos del establecimiento (además de la variable ISE en el primer nivel). Esto es interesante ya que separa el efecto de la condición socioeconómica en un efecto a nivel individual y un efecto a nivel de establecimiento (nivel socioeconómico del conjunto de alumnos)

A partir de las ecuaciones [3] y [4] los modelos a estimar se describen de la siguiente manera (ver Cuadro 1 anterior para la nomenclatura de las variables):

Modelo 1:

$$\text{Nivel 1: } Y_{ij} = \beta_{0j} + r_{ij}$$

$$\text{Nivel 2: } \beta_{0j} = \gamma_{00} + \mu_{0j}$$

Modelo 2:

$$\text{Nivel 1: } Y_{ij} = \beta_{0j} + \beta_{1j}(\text{Preesc}) + \beta_{2j}(\text{Mujer}) + \beta_{3j}(\text{ISE}) + \beta_{4j}(\text{ISE}^2) + r_{ij}$$

$$\text{Nivel 2: } \beta_{0j} = \gamma_{00} + \mu_{0j}$$

$$\beta_{3j} = \gamma_{30} + \mu_{3j}$$

$$\beta_{kj} = \gamma_{k0} \quad k = 1,2,4$$

Modelo 3:

$$\text{Nivel 1: } Y_{ij} = \beta_{0j} + \beta_{1j}(\text{Preesc}) + \beta_{2j}(\text{Mujer}) + \beta_{3j}(\text{ISE}) + \beta_{4j}(\text{ISE}^2) + r_{ij}$$

Nivel 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{IP900}) + \gamma_{02}(\text{PP900}) + \gamma_{03}(\text{SP900}) + \gamma_{04}(\text{PRUR}) + \gamma_{05}(\text{JEC}) + \gamma_{06}(\text{CRIT})$$

$$+ \gamma_{07}(\text{EXP}) + \gamma_{08}(\text{MAT}) + \gamma_{09}(\text{CAP}) + \gamma_{010}(\text{PSUB}) + \gamma_{011}(\text{PPAG}) + \gamma_{012}(\text{FPAD})$$

$$+ \gamma_{013}(\text{RUR}) + \gamma_{014}(\text{CONF}) + \gamma_{015}(\text{COB}) + \gamma_{016}(\text{ISEP}) + \gamma_{017}(\text{ISEP}^2) + \mu_{0j}$$

$$\beta_{3j} = \gamma_{30} + \gamma_{31}(\text{PSUB}) + \gamma_{32}(\text{PPAG}) + \gamma_{33}(\text{RUR}) + \gamma_{34}(\text{ISEP}) + \mu_{3j}$$

$$\beta_{kj} = \gamma_{k0} \quad k = 1,2,4$$

2.3 Resultados

Los Cuadros 4a y 5a resumen los resultados de las estimaciones de los tres Modelos Lineales Jerárquicos para las pruebas de Lenguaje y Matemáticas respectivamente. Luego, en los Cuadros 4b y 5b se muestran los componentes de la varianza y la varianza total. Las estadísticas descriptivas de las variables utilizadas en las estimaciones se presentan en el anexo 4.

El modelo 1 (Cuadros 4a y 5a) muestra que el promedio general estimado es 248 puntos en Lenguaje y 244 puntos en Matemáticas. En la prueba de Lenguaje la varianza estimada a nivel de alumno (varianza al interior de las escuelas) es de 2.124 y la varianza estimada a nivel de logro

promedio de los establecimientos (varianza entre escuelas) es de 711 (Cuadro 4b), lo que indica que existe una mayor variación entre alumnos que entre establecimientos. Esto también se observa en la prueba de Matemáticas.

Al incorporar variables explicativas en los niveles 1 y 2, en general no se observan diferencias muy importantes en relación a los modelos estimados por MCO y cuantiles, y los resultados son consistentes con lo esperado.

Los resultados del modelo 2 muestran la importante incidencia del nivel socioeconómico en los resultados del SIMCE en las pruebas de Lenguaje y Matemáticas. El índice socioeconómico al cuadrado es significativo en ambas pruebas. En cuanto a la variable asistencia a preescolar, el efecto es positivo y significativo, mientras que con la variable género se observa que las mujeres tienen un mejor desempeño en Lenguaje y los hombres un mejor desempeño en Matemáticas.

Al analizar las varianzas (Cuadros 4b y 5b) es claro que la incorporación de las variables en el nivel 1 las reduce. Comparando los modelos 1 y 2 se observa que se reduce la varianza entre escuelas y también disminuye la varianza al interior de las escuelas (r_{ij}) levemente (2.124 a 2.103 en Lenguaje y 2.156 a 2.137 en Matemáticas).

Si bien los resultados del modelo 3 son similares a lo obtenido a través de MCO también se observan diferencias interesantes.

Con respecto a las variables de programas (P900, Rural, Escuelas Críticas) se observa que tienen puntajes significativamente menores, aunque los coeficientes estimados en este caso son bastante más pequeños que con MCO, incluso para el programa Rural el coeficiente estimado no es significativamente distinto de cero.

También se concluye que los establecimientos rurales tienen un mejor desempeño que los urbanos, que los alumnos en establecimientos con JEC tienen mejores resultados que los alumnos en establecimientos sin JEC, que los establecimientos con financiamiento de padres obtienen mejores resultados, y que existe un efecto positivo y significativo de la experiencia de los docentes en el establecimientos. El cociente alumno-profesor sólo es significativo para el caso de Lenguaje, mientras que la confianza de los profesores sólo es significativa en Matemáticas.

Cuadro 4a
Modelo HLM SIMCE 4° Básico. Efectos Fijos
Lenguaje

Variable	Modelo 1		Modelo 2		Modelo 3	
	Coeficiente	t	Coeficiente	t	Coeficiente	t
Para Bo (logro promedio del establecimiento)						
Constante	248.367	526.238	242.290	307.182	242.619	206.146
IP900					-6.919	-6.838
PP900					-12.424	-8.589
SP900					-3.750	-3.761
PRUR					-2.503	-1.473
JEC					3.260	4.755
CRIT					-17.879	-6.572
EXP					0.205	3.791
MAT					0.002	2.751
CAP					0.008	3.611
PSUB					-0.938	-0.883
PPAG					-5.148	-1.564
FPAD					6.969	6.072
RUR					7.963	6.031
CONF					0.598	0.752
COB					5.283	6.353
ISEP					23.297	24.078
ISEP2					-1.418	-2.476
Para B1(Asistencia Pre-escolar)						
Constante			3.756	5.106	0.739	0.983
Para B2 (Mujer)						
Constante			7.089	20.581	7.066	20.740
Para B3 (ISE)						
Constante			7.575	29.725	4.193	10.550
PSUB					0.116	0.187
PPAG					-1.497	-0.905
RURAL					-2.811	-2.878
ISEP					-1.497	-2.054
Para B4 (ISE2)						
Constante			0.521	4.440	0.483	2.156

Nota: Los números en negrita corresponden a coeficientes significativos al 5%.

Cuadro 4b
Modelo HLM SIMCE 4° Básico 2002. Efectos Aleatorios
Lenguaje

	Componente de la Varianza	Grados de Libertad	Chi 2	Valor P
Modelo 1				
Constante β_0	710.760	4,142.000	31,530.781	0.000
Efectos nivel 1 r_{ij}	2,123.993			
Modelo 2				
Constante β_0	457.492	3,979.000	16,187.920	0.000
Pendiente β_3	30.438	3,979.000	4,914.326	0.000
Efectos nivel 1 r_{ij}	2,101.357			
Modelo 3				
Constante β_0	213.230	3,962.000	10,151.578	0.000
Pendiente β_3	17.289	3,975.000	4,600.163	0.000
Efectos nivel 1 r_{ij}	2,102.630			

Cuadro 5a
Modelo HLM SIMCE 4° Básico 2002. Efectos Fijos
Matemáticas

Variable	Modelo 1		Modelo 2		Modelo 3	
	Coefficiente	t	Coefficiente	t	Coefficiente	t
Para Bo (logro promedio del establecimiento)						
Constante	243.699	509.435	242.888	298.205	245.014	203.210
IP900					-6.530	-6.195
PP900					-12.142	-8.725
SP900					-3.125	-2.993
PRUR					-3.293	-1.784
JEC					1.482	2.077
CRIT					-19.515	-8.263
EXP					0.227	4.037
MAT					0.004	4.073
CAP					0.003	1.739
PSUB					-0.795	-0.712
PPAG					-4.251	-1.236
FPAD					6.473	5.304
RUR					4.957	3.680
CONF					3.084	2.315
COB					4.898	5.327
ISEP					20.883	20.492
ISEP2					-0.372	-0.613
Para B1(Asistencia Pre-escolar)						
Constante			4.188	5.497	0.700	0.898
Para B2 (Mujer)						
Constante			-4.791	-14.202	-4.838	-14.494
Para B3 (ISE)						
Constante			7.429	28.302	3.747	9.163
PSUB					0.557	0.879
PPAG					-0.241	-0.144
RURAL					-2.272	-2.368
ISEP					-1.489	-2.021
Para B4 (ISE2)						
Constante			0.365	2.988	0.315	1.430

Nota: Los números en negrita corresponden a coeficientes significativos al 5%.

Cuadro 5b
Modelo HLM SIMCE 4° Básico 2002. Efectos Aleatorios
Matemáticas

	Componente de la Varianza	Grados de Libertad	Chi 2	Valor P
Modelo 1				
Constante β_0	732.197	4,142.000	31,637.805	0.000
Efectos nivel 1 r_{ij}	2,155.529			
Modelo 2				
Constante β_0	475.921	3,979.000	16,512.423	0.000
Pendiente β_3	39.101	3,979.000	5,004.220	0.000
Efectos nivel 1 r_{ij}	2,135.338			
Modelo 3				
Constante β_0	239.234	3,962.000	10,912.791	0.000
Pendiente β_3	21.576	3,975.000	4,709.449	0.000
Efectos nivel 1 r_{ij}	2,136.794			

Con respecto al nivel socioeconómico, en el modelo 3 se incluye el índice socioeconómico promedio del establecimiento (ISEP) que es positivo y significativo. Esto mostraría el importante efecto del nivel socioeconómico del establecimiento en los resultados de la prueba SIMCE. Cabe notar que en este caso el coeficiente estimado para el ISE cae de manera importante (modelo 3 versus modelo 2), lo que significa que el nivel socioeconómico tiene un efecto individual, así como también un efecto a través del establecimiento. Es decir, establecimientos que en promedio tienen alumnos de nivel socioeconómico más alto obtienen mejores resultados.

Un resultado distinto se observa para las dependencias, ya que las dummies de dependencia no son significativas. Este fenómeno se explica por la dificultad que existiría para diferenciar el efecto de la dependencia del efecto del nivel socioeconómico, haciendo que la variable ISEP capture todo el efecto sobre el logro escolar. Tal como discuten Mizala y Romaguera (2002), para poder diferenciar claramente el impacto dependencia y nivel socioeconómico sobre el logro educacional, sería necesario que en cada dependencia existan alumnos de todos los niveles socioeconómicos, lo que permitiría testear si alumnos de similares niveles socioeconómicos tienen resultados distintos por el sólo hecho de estar en establecimientos de distinto tipo.

En los Cuadros 4b y 5b, se observa que las variables incluidas en el nivel 2 explican una proporción importante de la varianza entre escuelas (β_0) y de la pendiente del índice socioeconómico (β_3). En Lenguaje, el componente de la varianza entre escuelas se reduce de 457,49 a 213,23 y el de la pendiente β_3 de 30,43 a 17,28, es decir se reducen en 53,4% y 42,4% respectivamente. En Matemáticas, las reducciones son de la misma magnitud (ver Cuadro 5b).

2.4 Análisis de la Varianza

Tal como se indicó, las variables incorporadas en el modelo 2 explican parte de la varianza total de puntajes (varianza total del modelo nulo). En el caso de la prueba de Lenguaje el modelo 2 explica un 70% de la varianza entre escuelas y sólo un 1% de la varianza al interior de las escuelas (entre los alumnos)¹⁶. Esto significa las variables contribuyen a explicar una parte importante de las diferencias de desempeños entre escuelas, sin embargo, no explican las diferencias al interior de las escuelas, lo que probablemente se debe a que las diferencias de resultados entre alumnos están

¹⁶ Esto se obtiene de comparar como se reduce la varianza entre escuelas y al interior de las escuelas entre el modelo nulo y el modelo con todas las variables explicativas (modelo 2).

determinadas por habilidades, motivación, actitud hacia el estudio y/o otras características personales de los estudiantes que no están siendo consideradas en el modelo. En el caso de la prueba de Matemáticas, el modelo final explica un 67% de la varianza entre escuelas y cerca de un 1% de la varianza al interior de las escuelas.

El cuadro 6 descompone la varianza explicada en la contribución de las distintas variables para las pruebas de Lenguaje y Matemáticas. Claramente, son las diferencias socioeconómicas, tanto a nivel de estudiantes (sus familias) como a nivel de escuelas (nivel socioeconómico promedio de los alumnos de la escuela), las que explican en mayor medida las diferencias entre escuelas. El nivel socioeconómico explica el 63% de la varianza entre escuelas en la prueba de Lenguaje y el 63% de la varianza entre escuelas en la prueba de Matemáticas.

Otras variables que explican en menor medida la varianza entre escuelas son ingresos y permanecen a P900, escuelas críticas, financiamiento de padres, ser rural y cobertura de aprendizaje.

En cuanto a la varianza al interior de las escuelas, son nivel socioeconómico y género, las variables que explican un 1% de la varianza en Lenguaje y Matemáticas.

Cuadro 6
Descomposición de la Varianza en Lectura y Matemáticas

Variable	Lenguaje		Matemáticas		
	Contribución a explicar la varianza entre escuelas	Contribución a explicar la varianza al interior de las escuelas	Contribución a explicar la varianza entre escuelas	Contribución a explicar la varianza al interior de las escuelas	
Variables a nivel individual	ISE	34%	0.6%	35%	0.7%
	MUJER	1%	0.5%	0%	0.3%
	ISE2	0%	0.0%	0%	0.0%
	PREESC	0%	0.0%	0%	0.0%
Variables a nivel de escuela	ISEP	29%	0.0%	28%	0.0%
	PP900	1%	0.0%	1%	0.0%
	RUR	1%	0.0%	0%	0.0%
	FPAD	1%	0.0%	1%	0.0%
	COB	1%	0.0%	1%	0.0%
	CRIT	1%	0.1%	1%	0.0%
	IP900	1%	0.0%	0%	0.0%
	RESTO	1%	0.0%	1%	0.0%
Total	70%	1%	67%	1%	

3. Estimaciones de Diferencias de Puntaje entre 1999-2002

3.1 Especificación del Modelo

Las estimaciones anteriores representan la situación en un momento puntual del tiempo y por lo tanto analizan los desempeños relativos de los distintos alumnos y establecimientos en ese momento, pero no entregan información en cuanto a los avances o retrocesos que los establecimientos puedan tener en relación a la prueba SIMCE. En esta sección, se desea analizar principalmente los efectos que tienen los programas focalizados del Ministerio (P900, Rural y Escuelas Críticas), en cuanto a la variación de puntajes que se obtienen a través de un período de tiempo. Este análisis entrega información importante sobre los programas considerando que no constituye un análisis de los resultados e impacto de los mismos.

Para lo anterior, se realiza la estimación de una ecuación como [2] por Mínimos Cuadrados Ordinarios, pero con dos diferencias fundamentales. Primero, la regresión es a nivel de establecimientos, y segundo, la variable dependiente (a explicar) es la diferencia del puntaje promedio del establecimiento entre el 2002 y 1999. En definitiva se estima un modelo como:

$$\Delta Y_i = \beta_0 + \sum_j \beta_j X_{ij} + \varepsilon_{ij} \quad [2]$$

Donde:

ΔY_i = Diferencia de puntaje promedio entre 2002 y 1999 del establecimiento i.

β_0 = Constante a estimar.

β_j = Parámetros a estimar.

X_{ij} = Característica j del establecimiento i.

ε_{ij} = Error aleatorio, $\varepsilon_{ij} \sim N(0, \sigma^2)$

Dentro de las variables X_j , sólo se incluyen variables a nivel de establecimiento: experiencia promedio de los docentes, tamaño del establecimiento, tasa alumno-profesor, dependencia, área geográfica (si es rural o urbano), confianza docente, cobertura curricular y pertenencia a los programas (P900, Rural, JEC y Escuelas Críticas). Además se incorpora el índice socioeconómico promedio de los alumnos de cada establecimiento.

3.2 Resultados

Los Cuadros 7 y 8, muestran los resultados de las estimaciones de diferencias. En total se utilizó la información de 4.132 establecimientos para Lenguaje y 3.944 establecimientos para Matemáticas, que son los que tienen resultados en las pruebas SIMCE de 1999 y 2002 y además información para todas las otras variables. Las estadísticas descriptivas están en el anexo 5.

Si bien, anteriormente se constata que los alumnos de los establecimientos que están en el programa P900 tienen puntajes menores que el conjunto de los establecimientos comparables, se observa que este programa contribuye de manera positiva a aumentar en promedio los puntajes entre 1999 y 2002. En efecto, la diferencia de puntaje entre 1999 y 2002 de las escuelas que ingresaron al P900, en relación a sus similares, es en promedio 12 puntos más en Matemáticas y 11 puntos más en Lenguaje.

En cuanto al programa Rural, también se observa que contribuye significativamente a aumentar los puntajes en el período 1999-2002. Las diferencias de puntaje promedio de los establecimientos que pertenecen a este programa son 6 puntos más en Lenguaje y 5 puntos más en Matemáticas, que los establecimientos similares que no pertenecen al programa Rural.

Finalmente, aunque los alumnos que asisten a escuelas que son atendidas por el programa para Escuelas Críticas obtienen puntajes inferiores al conjunto de los establecimientos comparables, este programa contribuye de manera importante a aumentar, en promedio, los puntajes entre 1999 y 2002. En efecto, la diferencia de puntaje entre 1999 y 2002 de estas escuelas, en relación a sus similares, es en promedio 11 puntos más en Matemáticas y 17 puntos más en Lenguaje.

Cuadro 7

<i>Regresiones por MCO para Lenguaje a nivel de establecimientos</i>		
<i>Var Dep: Diferencia de puntaje 2002-1999</i>		
	<i>Coef.</i>	<i>t</i>
IP900	11,365*	11,171
PP900	9,820*	7,381
SP900	0,090	0,091
PRUR	5,648*	4,076
JEC	1,328	1,849
CRIT	16,865*	6,063
EXP	-0,077	-1,383
MAT	0,004*	2,324
MAT ²	0,000	-1,433
CAP	0,001	0,233
PSUB	0,292	0,273
PPAG	-1,918	-0,732
FPAD	0,724	0,590
RUR	8,061*	7,383
CONF	1,474*	1,983
COB	2,771*	3,458
ISEP	4,118*	4,882
ISEP²	-0,899*	-2,412
Constante	-17,961	-5,665
N	4132	
R ²	0,065	

Nota: (*) Coeficiente significativo al 5%.

Cuadro 8

<i>Regresiones por MCO para Matemáticas a nivel de establecimientos</i>		
<i>Var Dep: Diferencia de puntaje 2002-1999</i>		
	<i>Coef.</i>	<i>t</i>
IP900	12,219*	11,134
PP900	8,744*	6,099
SP900	-0,487	-0,451
PRUR	4,674*	3,066
JEC	1,154	1,489
CRIT	10,906*	3,614
EXP	-0,073	-1,214
MAT	0,007*	3,268
MAT²	0,000*	-2,200
CAP	0,000	0,069
PSUB	-1,580	-1,360
PPAG	-1,989	-0,700
FPAD	0,950	0,720
RUR	3,765*	3,171
CONF	1,631	1,289
COB	1,945*	2,108
ISEP	6,746*	7,405
ISEP²	-1,796*	-4,541
Constante	-18,945*	-4,863
N	3944	
R ²	0,059	

Nota: (*) Coeficiente significativo al 5%.

IV. CONCLUSIONES

En general, los resultados aquí obtenidos son consistentes con otros estudios de función de producción educacional para Chile, en que se señala que el nivel socioeconómico, la dependencia, el tamaño del establecimiento, la Jornada Escolar Completa, la experiencia docente y el financiamiento de padres son variables que tienen un impacto positivo en el rendimiento escolar.

Se constata además que existe un desempeño distinto entre mujeres y hombres. Por un lado las mujeres tienen mejores resultados en Lenguaje, mientras que los hombres tienen mejores resultados en la prueba de Matemáticas.

También se muestra que variables relacionadas con los docentes, como la cobertura curricular y la confianza docente para enseñar los sectores de aprendizaje, tienen un impacto positivo y significativo en el desempeño escolar.

Sin embargo, a través del análisis de cuantiles, se muestra que existe un efecto diferenciado de las variables según el rendimiento del alumno. Por ejemplo, se observa que en los tramos de mayor rendimiento el impacto del nivel socioeconómico es menor.

Con respecto a los programas, se muestra que los establecimientos que están participando en P900, Rural y Escuelas Críticas, parecen tener puntajes menores que el resto de los establecimientos. Sin embargo, el análisis de diferencias muestra que estos programas contribuyen de manera positiva en los resultados promedios de los establecimientos.

El modelo HLM constata los resultados obtenidos por otras metodologías, poniendo énfasis en las diferencias que existen entre el efecto a nivel de alumno y el efecto a nivel de establecimiento.

V. REFERENCIAS BIBLIOGRÁFICAS

Bravo D., Contreras D. y Sanhueza C.(1999), “Rendimiento Educacional, Desigualdad y Brecha de Desempeño Privado / Público”, Documento de Trabajo N° 163, Departamento de Economía, U. de Chile.

Dresdner J. y Valdebenito A. (2002) “Estimación de una función de producción educativa para la octava región”. *Mimeo*, Universidad de Concepción.

MINEDUC (2003) “Prueba SIMCE 4° Básico 2002. Análisis de Resultados”. Departamento de Estudios y Estadísticas, Ministerio de Educación.

MINEDUC (2002) “Prueba SIMCE 2° Medio 2001. Factores que Inciden en el Rendimiento de los Alumnos”. Nota Técnica, Departamento de Estudios y Estadísticas, Ministerio de Educación, Diciembre.

Mizala, A. y Romaguera P. (1998) “Desempeño Escolar y Elección de Colegios: La Experiencia Chilena”. Documento de Trabajo 36, Centro de Economía Aplicada, DII, Universidad de Chile, Junio.

Mizala, A. y Romaguera P. (2000) “Determinación de Factores Explicativos de los Resultados Escolares en Educación Media en Chile”. Documento de Trabajo 85, Centro de Economía Aplicada, DII, Universidad de Chile.

Mizala, A. y Romaguera P. (2002) “Equity and Educational Performance”. Documento de Trabajo 136, Centro de Economía Aplicada, DII, Universidad de Chile.

Ostoic, Carolina (2002) “Nuevas Metodologías para el Análisis de los Resultados Educativos: Una Aplicación de los Modelos Lineales Jerárquicos a los Datos de Chile”. Tesis de Magíster en Economía Aplicada, Universidad de Chile.

ANEXO 1: Comparación de Distribuciones

a. Lenguaje

Se compara la distribución de puntajes de toda la población con la distribución de puntajes de los casos utilizados en la regresiones. Tal como se desprende de las estadísticas descriptivas, los datos utilizados para la regresión no parecen tener diferencias sustantivas con los datos totales.

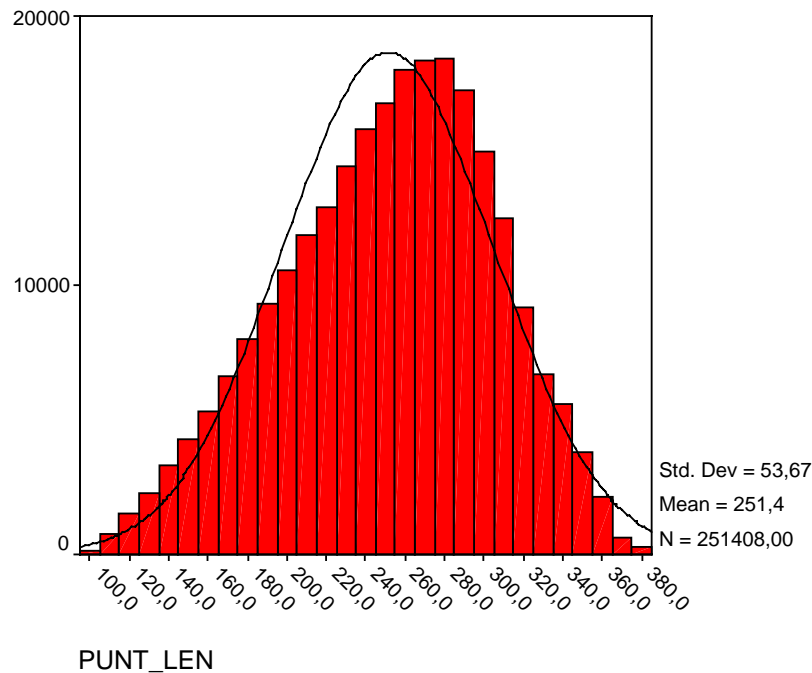
Población Total

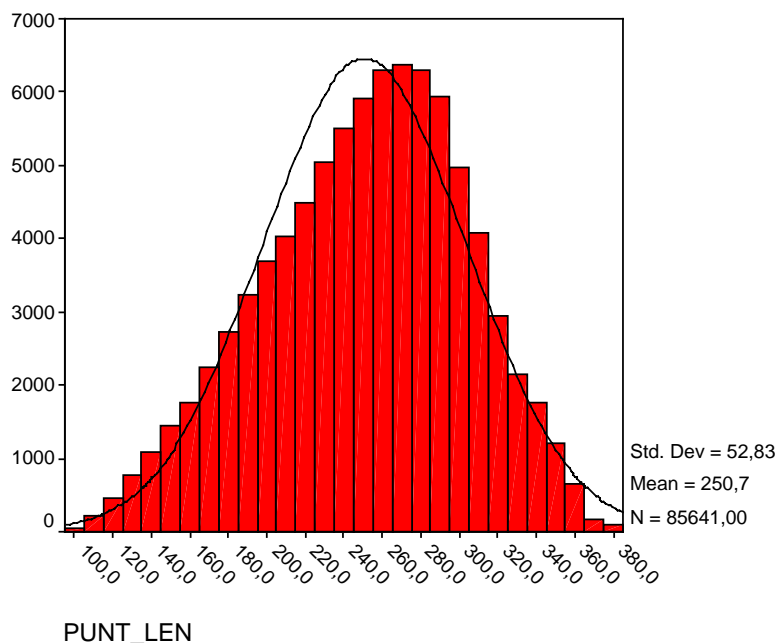
	Estadísticas Descriptivas						
	N	Mínimo	Máximo	Media	DS	Skewness	Kurtosis
Puntaje Lenguaje	251408	100.08	376.35	251.40	53.67	-0.272	-0.46
Preesc	220831	0.00	1.00	0.93	0.25		
Mujer	254389	0.00	1.00	0.49	0.50		
IP900	254520	0.00	1.00	0.18	0.38		
PP900	254520	0.00	1.00	0.06	0.25		
SP900	254520	0.00	1.00	0.14	0.34		
PRUR	254520	0.00	1.00	0.04	0.19		
JEC	254520	0.00	1.00	0.38	0.49		
CRIT	254520	0.00	1.00	0.01	0.11		
EXP	244028	0.00	49.00	17.54	7.01		
MAT	247146	2.00	4777.00	694.26	512.17		
MAT2	247146	4.00	22819729.00	744313.82	1715610.83		
CAP	242876	1.90	2673.75	33.75	73.49		
PSUB	254520	0.00	1.00	0.40	0.49		
PPAG	254520	0.00	1.00	0.07	0.26		
FPAD	254520	0.00	1.00	0.36	0.48		
RURI	254126	0.00	1.00	0.12	0.32		
ISE	117436	-2.30	4.98	0.01	1.00		
ISE2	117436	0.00	24.79	1.00	1.98		
CONF	208017	1.00	3.00	2.62	0.49		
COB	199557	1.24	4.00	3.32	0.46		

Datos Regresión

	Estadísticas Descriptivas						
	N	Mínimo	Máximo	Media	DS	Skewness	Kurtosis
Puntaje Lenguaje	85641	100.27	376.35	250.71	52.83	-0.262	-0.437
Preesc	85641	0.00	1.00	0.93	0.25		
Mujer	85641	0.00	1.00	0.49	0.50		
IP900	85641	0.00	1.00	0.19	0.39		
PP900	85641	0.00	1.00	0.07	0.25		
SP900	85641	0.00	1.00	0.15	0.36		
PRUR	85641	0.00	1.00	0.04	0.19		
JEC	85641	0.00	1.00	0.38	0.49		
CRIT	85641	0.00	1.00	0.01	0.11		
EXP	85641	0.00	45.00	17.86	6.89		
MAT	85641	2.00	4777.00	706.08	497.67		
MAT2	85641	4.00	22819729.00	746220.33	1605574.21		
CAP	85641	0.00	1.00	0.38	0.49		
PSUB	85641	0.00	1.00	0.04	0.21		
PPAG	85641	0.00	1.00	0.31	0.46		
FPAD	85641	0.00	1.00	0.12	0.32		
RURI	85641	-2.30	4.98	-0.02	0.97		
ISE	85641	0.00	24.79	0.94	1.85		
ISE2	85641	1.00	3.00	2.62	0.49		
CONF	85641	1.55	4.00	3.32	0.46		

Se presentan, además, las distribuciones de puntajes para los casos totales y los datos utilizados en la regresión.





Los tests U de Mann-Whitney, Wilcoxon y Kolmogorov-Smirnov, permiten contrastar estadísticamente las formas de las distribuciones. El primero es la más conocida de las pruebas para dos muestras independientes. Contrasta si dos poblaciones muestradas son equivalentes en su posición, el test se basa en los signos (positivos y negativos) de los residuos de la regresión. El segundo test, es más general que los anteriores y detecta las diferencias entre las posiciones y formas de las distribuciones. Se basa en la diferencia máxima absoluta entre las funciones de distribución acumulada observadas para ambas muestras, en definitiva prueba si las dos muestras analizadas provienen de la misma distribución poblacional.

A continuación se presentan los resultados de los tests.

	Puntaje Lenguaje
Mann-Whitney U	10669770752,000
Wilcoxon W	14337003520,000
Z	-3,889
Asymp. Sig. (2-tailed)	,000
Kolmogorov-Smirnov Z	2.811
Asymp. Sig. (2-tailed)	.000

El estadígrafo Z (que es la normalización de U) es idéntico para U y W, y tiene una distribución asintótica normal estándar (N(0.1)). La hipótesis alternativa de esta prueba establece que las

distribuciones no son las mismas, pero implica solamente un desplazamiento en la tendencia central de una con respecto de la otra y no una diferencia de forma o dispersión.

A partir de esto se concluye que ambas distribuciones son distintas. Por una parte la posición de las distribuciones de puntaje es distinta y que, además, la forma de éstas también difiere.

b. Matemáticas

Al igual que con la prueba de lenguaje en este caso se compara la distribución de puntajes de toda la población con la distribución de puntajes de los casos utilizados en la regresiones. Tal como se desprende de las estadísticas descriptivas, los datos utilizados para la regresión no parecen tener diferencias sustantivas con los datos totales.

En particular, las estadísticas de kurtosis (medida de concentración de la distribución en torno a la media) y skewness (medida de simetría o más precisamente de falta de simetría) muestran que no parece existir diferencias importantes entre los datos totales y los utilizados en la regresión.

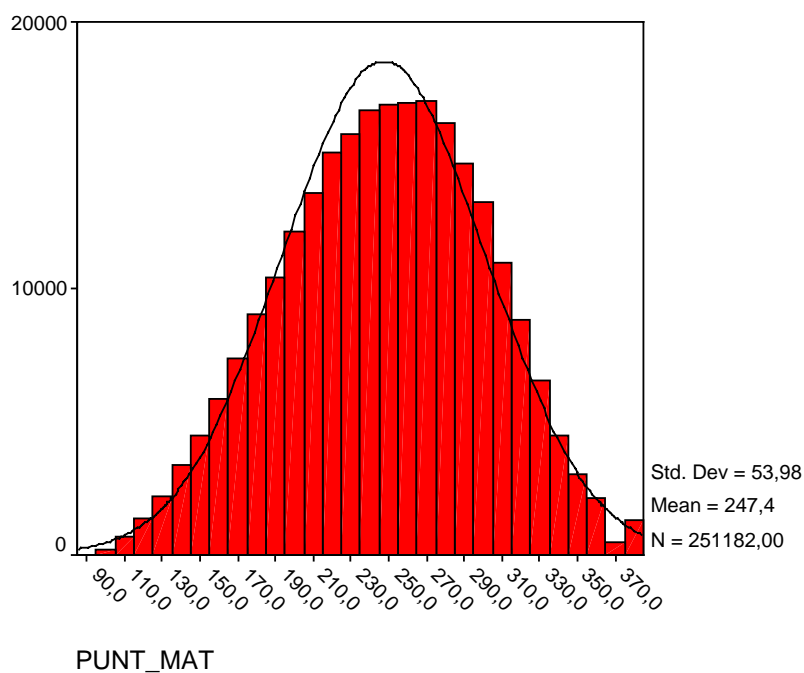
Población Total

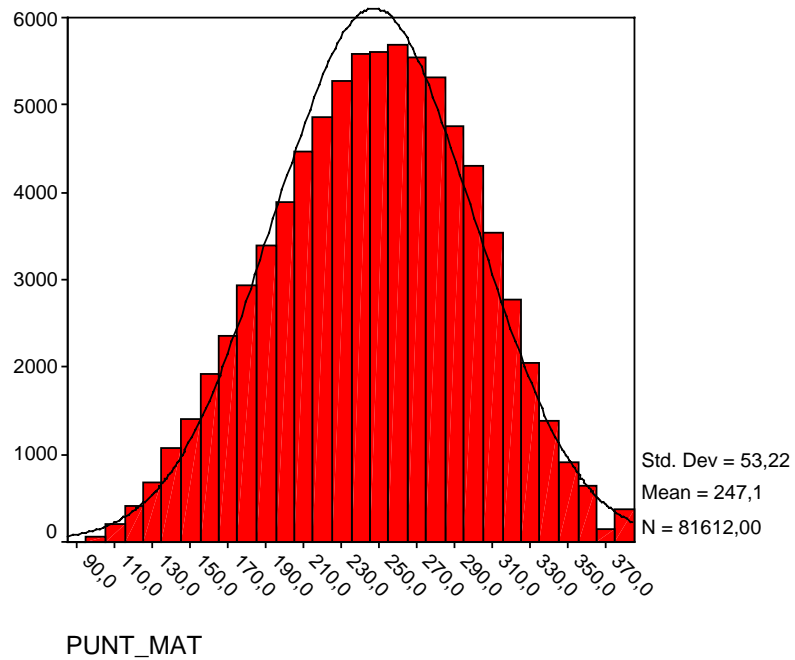
	Estadísticas Descriptivas						
	N	Mínimo	Máximo	Media	DS	Skewness	Kurtosis
Puntaje Matemáticas	251182	94.14	379.26	247.43	53.98	-0.112	-0.48
Preesc	220831	0.00	1.00	0.93	0.25		
Mujer	254389	0.00	1.00	0.49	0.50		
IP900	254520	0.00	1.00	0.18	0.38		
PP900	254520	0.00	1.00	0.06	0.25		
SP900	254520	0.00	1.00	0.14	0.34		
PRUR	254520	0.00	1.00	0.04	0.19		
JEC	254520	0.00	1.00	0.38	0.49		
CRIT	254520	0.00	1.00	0.01	0.11		
EXP	244028	0.00	49.00	17.54	7.01		
MAT	247146	2.00	4777.00	694.26	512.17		
MAT2	247146	4.00	22819729.00	744313.82	1715610.83		
CAP	242876	1.90	2673.75	33.75	73.49		
PSUB	254520	0.00	1.00	0.40	0.49		
PPAG	254520	0.00	1.00	0.07	0.26		
FPAD	254520	0.00	1.00	0.36	0.48		
RURI	254126	0.00	1.00	0.12	0.32		
ISE	117436	-2.30	4.98	0.01	1.00		
ISE2	117436	0.00	24.79	1.00	1.98		
CONF	207568	1.00	3.00	2.74	0.33		
COB	190924	1.54	4.00	3.31	0.45		

Datos regresión

	Estadísticas Descriptivas						
	N	Mínimo	Máximo	Media	DS	Skewness	Kurtosis
Puntaje Matemáticas	81612	94.14	379.26	247.13	53.22	-0.116	-0.463
Preesc	81612	0.00	1.00	0.93	0.25		
Mujer	81612	0.00	1.00	0.49	0.50		
IP900	81612	0.00	1.00	0.19	0.39		
PP900	81612	0.00	1.00	0.07	0.25		
SP900	81612	0.00	1.00	0.15	0.36		
PRUR	81612	0.00	1.00	0.03	0.18		
JEC	81612	0.00	1.00	0.38	0.48		
CRIT	81612	0.00	1.00	0.01	0.11		
EXP	81612	0.83	45.00	17.87	6.88		
MAT	81612	2.00	4777.00	711.44	496.04		
MAT2	81612	4.00	22819729.00	752198.97	1587360.25		
CAP	81612	0.00	1.00	0.38	0.49		
PSUB	81612	0.00	1.00	0.05	0.21		
PPAG	81612	0.00	1.00	0.32	0.47		
FPAD	81612	0.00	1.00	0.11	0.32		
RURI	81612	-2.30	4.98	-0.01	0.98		
ISE	81612	0.00	24.79	0.96	1.92		
ISE2	81612	1.40	3.00	2.75	0.32		
CONF	81612	1.54	4.00	3.32	0.45		

A continuación se presentan, además, las distribuciones de puntajes para los casos totales y los datos utilizados en la regresión.





A continuación se presentan los resultados de los tests. Se concluye que las distribuciones no difieren entre ellas.

	Puntaje Matemáticas
Mann-Whitney U	10215857152,000
Wilcoxon W	13546157056,000
Z	-1,421
Asymp. Sig. (2-tailed)	,155
Kolmogorov-Smirnov Z	1,526
Asymp. Sig. (2-tailed)	,019

ANEXO 2: Estadísticas Descriptivas MCO

Estadísticas Descriptivas Modelos MCO y Cuantilas para la Prueba de Lenguaje					
Variable	N	Mínimo	Máximo	Promedio	DS
Puntaje	84821	94,14	379,26	246,67	53,18
PREESC	85641	0,00	1,00	0,93	0,25
MUJER	85641	0,00	1,00	0,49	0,50
IP900	85641	0,00	1,00	0,19	0,39
PP900	85641	0,00	1,00	0,07	0,25
SP900	85641	0,00	1,00	0,15	0,36
PRUR	85641	0,00	1,00	0,04	0,19
JEC	85641	0,00	1,00	0,38	0,49
CRIT	85641	0,00	1,00	0,01	0,11
EXP	85641	0,00	45,00	17,86	6,89
MAT	85641	2,00	4777,00	706,08	497,67
MAT ²	85641	4,00	22819729,00	746220,33	1605574,21
CAP	85641	1,90	2673,75	32,92	67,79
PSUB	85641	0,00	1,00	0,38	0,49
PPAG	85641	0,00	1,00	0,04	0,21
FPAD	85641	0,00	1,00	0,31	0,46
RUR	85641	0,00	1,00	0,12	0,32
ISE	85641	-2,30	4,98	-0,02	0,97
ISE ²	85641	0,00	24,79	0,94	1,85
CONF	85641	1,00	3,00	2,62	0,49
COB	85641	1,55	4,00	3,32	0,46

Estadísticas Descriptivas Modelos MCO y Cuantilas para la Prueba de Matemáticas					
Variable	N	Promedio	DS	Promedio	DS
Puntaje	81612	94,14	379,26	247,13	53,22
PREESC	81612	0,00	1,00	0,93	0,25
MUJER	81612	0,00	1,00	0,49	0,50
IP900	81612	0,00	1,00	0,19	0,39
PP900	81612	0,00	1,00	0,07	0,25
SP900	81612	0,00	1,00	0,15	0,36
PRUR	81612	0,00	1,00	0,03	0,18
JEC	81612	0,00	1,00	0,38	0,48
CRIT	81612	0,00	1,00	0,01	0,11
EXP	81612	0,83	45,00	17,87	6,88
MAT	81612	2,00	4777,00	711,44	496,04
MAT ²	81612	4,00	22819729,00	752198,97	1587360,25
CAP	81612	1,90	2673,75	33,19	69,84
PSUB	81612	0,00	1,00	0,38	0,49
PPAG	81612	0,00	1,00	0,05	0,21
FPAD	81612	0,00	1,00	0,32	0,47
RUR	81612	0,00	1,00	0,11	0,32
ISE	81612	-2,30	4,98	-0,01	0,98
ISE ²	81612	0,00	24,79	0,96	1,92
CONF	81612	1,40	3,00	2,75	0,32
COB	81612	1,54	4,00	3,32	0,45

ANEXO 3: Estadísticas de Multicolinealidad

A continuación se presentan los valores VIF (Variance Inflation Factors) y 1/VIF (tolerancia). Tal como se desprende de los cuadros los valores VIF y 1/VIF, no muestran indicios de multicolinealidad.

VIF LENGUAJE			VIF MATEMATICAS		
Variable	VIF	1/VIF	Variable	VIF	1/VIF
MAT	5.330	0.188	MAT	5.320	0.188
MAT2	4.670	0.214	MAT2	4.700	0.213
PSUB	3.010	0.332	PSUB	3.030	0.330
FPAD	2.750	0.363	FPAD	2.780	0.360
PPAG	2.550	0.392	PPAG	2.670	0.375
ISE	1.850	0.540	ISE	1.910	0.525
ISE2	1.800	0.557	ISE2	1.890	0.530
RUR	1.700	0.587	EXP	1.690	0.590
EXP	1.690	0.591	RUR	1.680	0.594
PRUR	1.540	0.651	PRUR	1.520	0.657
IP900	1.410	0.712	IP900	1.400	0.714
SP900	1.280	0.782	COB	1.300	0.770
PP900	1.220	0.823	CONF	1.280	0.784
Preesc	1.180	0.847	SP900	1.270	0.788
COB	1.160	0.859	PP900	1.210	0.825
CONF	1.150	0.873	Preesc	1.170	0.851
JEC	1.130	0.882	JEC	1.130	0.883
CRIT	1.030	0.968	CRIT	1.030	0.967
CAP	1.020	0.978	CAP	1.020	0.978
Mujer	1.000	0.999	Mujer	1.000	0.999
Promedio VIF	1.920		Promedio VIF	1.950	

ANEXO 4: Estadísticas Descriptivas HLM

Estadísticas Descriptivas Modelo HLM para la Prueba de Lenguaje					
Variable	N	Promedio	DS	Mínimo	Máximo
NIVEL 1					
Puntaje	88846	251,70	52,60	100,27	376,35
PREESC	88846	0,94	0,24	0,00	1,00
MUJER	88846	0,49	0,50	0,00	1,00
ISE	88846	0,00	0,96	-2,30	4,98
ISE ²	88846	0,92	1,87	0,00	24,79
NIVEL 2					
IP900	4143	0,18	0,38	0,00	1,00
PP900	4143	0,08	0,27	0,00	1,00
SP900	4143	0,18	0,38	0,00	1,00
PRUR	4143	0,16	0,36	0,00	1,00
JEC	4143	0,47	0,50	0,00	1,00
CRIT	4143	0,01	0,11	0,00	1,00
EXP	4143	17,53	7,14	0,83	45,00
MAT	4143	437,01	382,05	2,00	4777,00
CAP	4143	28,15	58,88	1,90	2673,75
PSUB	4143	0,34	0,47	0,00	1,00
PPAG	4143	0,07	0,25	0,00	1,00
FPAD	4143	0,28	0,45	0,00	1,00
RUR	4143	0,30	0,46	0,00	1,00
CONF	4143	2,59	0,46	1,00	3,00
COB	4143	3,28	0,44	1,72	4,00
ISEP	4143	-0,14	0,86	-2,30	3,96
ISEP ²	4143	0,76	1,50	0,00	15,68

Estadísticas Descriptivas Modelo HLM para la Prueba de Matemáticas					
Variable	N	Promedio	DS	Mínimo	Máximo
NIVEL 1					
Puntaje	88846	247,58	52,96	94,14	379,26
PREESC	88846	0,94	0,24	0,00	1,00
MUJER	88846	0,49	0,50	0,00	1,00
ISE	88846	0,00	0,96	-2,30	4,98
ISE ²	88846	0,92	1,87	0,00	24,79
NIVEL 2					
IP900	4143	0,18	0,38	0,00	1,00
PP900	4143	0,08	0,27	0,00	1,00
SP900	4143	0,18	0,38	0,00	1,00
PRUR	4143	0,16	0,36	0,00	1,00
JEC	4143	0,47	0,50	0,00	1,00
CRIT	4143	0,01	0,11	0,00	1,00
EXP	4143	17,53	7,14	0,83	45,00
MAT	4143	437,01	382,05	2,00	4777,00
CAP	4143	28,15	58,88	1,90	2673,75
PSUB	4143	0,34	0,47	0,00	1,00
PPAG	4143	0,07	0,25	0,00	1,00
FPAD	4143	0,28	0,45	0,00	1,00
RUR	4143	0,30	0,46	0,00	1,00
CONF	4143	2,71	0,32	1,60	3,00
COB	4143	3,25	0,44	1,54	4,00
ISEP	4143	-0,14	0,86	-2,30	3,96
ISEP ²	4143	0,76	1,50	0,00	15,68

ANEXO 5: Estadísticas Descriptivas Estimaciones para Diferencias 1999-2002

**Estadísticas Descriptivas
Estimaciones para la Diferencias 1999-2002
Lenguaje**

	N	Mínimo	Máximo	Promedio	DS
DIFLEN	4132	-118,36	92,11	1,02	20,90
IP900	4132	0,00	1,00	0,20	0,40
PP900	4132	0,00	1,00	0,09	0,28
SP900	4132	0,00	1,00	0,19	0,39
PRUR	4132	0,00	1,00	0,10	0,30
JEC	4132	0,00	1,00	0,45	0,50
CRIT	4132	0,00	1,00	0,01	0,12
EXP	4132	1,00	36,89	17,69	7,00
MAT	4132	18,00	4777,00	457,82	364,14
MAT ²	4132	324,00	22819729,00	342163,52	740653,38
CAP	4132	2,85	2673,75	28,28	58,34
PSUB	4132	0,00	1,00	0,34	0,47
PPAG	4132	0,00	1,00	0,08	0,27
FPAD	4132	0,00	1,00	0,29	0,45
RUR	4132	0,00	1,00	0,26	0,44
CONF	4132	1,00	3,00	2,60	0,46
COB	4132	1,55	4,00	3,29	0,43
ISEP	4132	-2,07	3,96	-0,09	0,87
ISEP ²	4132	0,00	15,68	0,77	1,58

**Estadísticas Descriptivas
Estimaciones para la Diferencias 1999-2002
Matemáticas**

	N	Mínimo	Máximo	Promedio	DS
DIFMAT	3944	-147,60	104,23	-4,15	22,01
IP900	3944	0,00	1,00	0,20	0,40
PP900	3944	0,00	1,00	0,09	0,28
SP900	3944	0,00	1,00	0,18	0,39
PRUR	3944	0,00	1,00	0,10	0,30
JEC	3944	0,00	1,00	0,45	0,50
CRIT	3944	0,00	1,00	0,01	0,12
EXP	3944	1,00	36,89	17,72	6,99
MAT	3944	18,00	4777,00	468,30	368,29
MAT ²	3944	324,00	22819729,00	354914,96	756131,55
CAP	3944	3,44	2673,75	28,53	59,59
PSUB	3944	0,00	1,00	0,34	0,47
PPAG	3944	0,00	1,00	0,08	0,27
FPAD	3944	0,00	1,00	0,29	0,45
RUR	3944	0,00	1,00	0,25	0,43
CONF	3944	1,40	3,00	2,72	0,31
COB	3944	1,68	4,00	3,27	0,43
ISEP	3944	-2,07	3,96	-0,09	0,88
ISEP ²	3944	0,00	15,68	0,77	1,63