

Agrupación de escuelas para intervenciones de política: análisis del caso chileno

**Ernesto Treviño
Francisca Donoso
Facultad de Educación, Universidad Diego Portales**

Introducción

El desarrollo reciente de la política educativa en Chile ha traído importantes cambios en la regulación aplicable al sistema escolar con miras a establecer mayores niveles de rendición de cuentas y apoyos para las escuelas. La Ley de Subvención Escolar Preferencial 20.248 y la Ley General de Educación 20.370 se aprobaron en el congreso en 2008 y 2009 respectivamente. La primera de ellas otorga un monto mayor de subvención por alumno prioritario (en condiciones de marginación) y otro monto en función del porcentaje de alumnos prioritarios en el colegio. La Ley General de Educación, por su parte, establece el marco global de operación del sistema escolar y la institucionalidad. En relación con la Ley General de Educación el poder legislativo tramita actualmente el Proyecto de Ley del Sistema Nacional de Aseguramiento de la Calidad de la Educación Parvularia, Básica y Media. A través de esta normativa se crearán la Agencia de Calidad de la Educación y la Superintendencia de Educación, como entidades encargadas de evaluar y fiscalizar el desempeño del sistema educativo. Una de las características de la propuesta del nuevo Sistema Nacional de Aseguramiento de la Calidad es que busca agrupar a los establecimientos escolares de acuerdo a criterios de desempeño con propósitos de rendición de cuentas y apoyo.

El presente trabajo analiza las alternativas de agrupación de escuelas básicas tomando en consideración distintos indicadores de resultados de aprendizaje que se construyen a partir del SIMCE y se hacen sugerencias para la aplicación de esta ley. A lo largo del mismo se ponen de manifiesto las dificultades técnicas y las consecuencias de política educativa de distintas formas de agrupación de establecimientos escolares.

Para ello se presenta, en primer lugar el contexto de la política educativa. En segundo lugar, una revisión de literatura. Después se explican los métodos y utilizados. Finalmente, se analizan los resultados, se hace una discusión y se presentan conclusiones y recomendaciones.

1. Contexto de política educativa

El año 2006 Chile vivió una serie de movilizaciones estudiantiles que pedían cambios profundos para mejorar la calidad y equidad de la educación. Este hito dio la partida a una serie de cambios en la regulación de la educación chilena. Primero a través de largas discusiones entre representantes de la sociedad civil en el Consejo Asesor

Presidencial para la Educación. Después en intensas negociaciones en el Congreso Nacional para crear un nuevo marco jurídico que rigiera al sistema educativo. En este marco se destacan las ya aprobadas Ley General de Educación y la Ley de Subvención Escolar Preferencial. Por otro lado, aún se encuentra en discusión el Proyecto de Ley del Sistema Nacional de Aseguramiento de la Calidad de la Educación Parvularia, Básica y Media, que es clave para el propósito de este artículo. En particular, es relevante destacar que en este proyecto de ley se propone que la Agencia de Calidad clasifique a “los establecimientos educacionales reconocidos por el Estado, de acuerdo a los resultados de aprendizaje de los alumnos, en función del grado de cumplimiento de los estándares de aprendizaje...”. Para ello, propone el texto, se considerarán “los resultados de aprendizaje de todas las áreas evaluadas censalmente en las mediciones nacionales, la distribución de los resultados de los alumnos en relación a los estándares de aprendizaje y las características de los alumnos del establecimiento educacional y, cuando sea posible, indicadores de valor agregado”. Además, agrega la propuesta que la metodología de clasificación se actualizará en un plazo entre cuatro y ocho años, y que la metodología deberá considerar que los cambios de los establecimientos en la clasificación no sean consecuencia de cambios en la composición del alumnado.

El proyecto de ley establece también que los establecimientos deben clasificarse en función del cumplimiento de los estándares de aprendizaje en cuatro categorías, a saber:

- a. Establecimientos Educacionales de Buen Desempeño;
- b. Establecimientos Educacionales de Desempeño Satisfactorio;
- c. Establecimientos Educacionales de Desempeño Regular, y
- d. Establecimientos Educacionales de Mal Desempeño con Necesidad de Acciones Intensivas de Mejoramiento.

De acuerdo al artículo 14 del proyecto de ley, la clasificación de escuelas se llevará a cabo anualmente “y considerará los resultados de los establecimientos educacionales en tres mediciones consecutivas válidas, en el caso que éstas sean anuales, y dos mediciones consecutivas válidas, en el caso que éstas se realicen cada dos años o más”. En el mismo artículo se dispone que en caso de que las escuelas tengan un número insuficiente de alumnos evaluados la Agencia deberá establecer una metodología pertinente a esa realidad considerando un mayor número de mediciones en el tiempo entre otros aspectos.

Por otro lado, la propuesta de ley indica que en caso de que los establecimientos ofrezcan educación básica y media, éstas deberán ser clasificadas por separado. Asimismo, se dispone que los establecimientos nuevos sean clasificados como satisfactorios provisoriamente hasta que cumplan con los requisitos para ser clasificados.

El texto propuesto en la ley indica que la agencia debe informar al Ministerio de educación y a los padres la clasificación de los establecimientos.

Uno de los puntos más interesantes de la ley consiste en el uso de la clasificación de escuelas para llevar a cabo intervenciones de política. Específicamente, el párrafo cuarto del texto propuesto habla de los efectos de la clasificación en los establecimientos financiados con aportes del Estado. En resumen, se propone que la agencia realice visitas evaluativas al menos cada dos años a los establecimientos con Mal Desempeño y con una frecuencia de entre dos y cuatro años a los de desempeño regular. Para el resto de los establecimientos queda a discreción de la agencia determinar el patrón de visitas, y los de buen desempeño sólo podrán evaluarse si el sostenedor lo solicita o si acepta la propuesta de visitas de la agencia.

De la propuesta de ley llaman la atención dos aspectos. En primer lugar, que los establecimientos más destacados no son sujetos de evaluación, cuando ellos pueden ser los que mayores lecciones arrojen para la mejora del sistema escolar en general. El segundo punto consiste en preguntarse qué se va a hacer con la educación parvularia y especial, ya que según la propuesta de ley éstas se evaluarán de acuerdo a un programa que debe diseñar y aprobar la agencia (artículo 21).

En el artículo 22 se expresa que, a partir de la evaluación, se deberá elaborar o revisar el plan de mejoramiento educativo (que debe ser de conocimiento público), para tomar acciones que solucionen las debilidades identificadas. Dichas acciones deberán ser explicitadas y contendrán los rubros de estrategias, actividades, metas y recursos.

La idea que reina detrás de esta propuesta, es que los establecimientos educacionales soliciten apoyo técnico para elaborar e implementar su plan al ministerio de educación o a una entidad del Registro Público de Entidades Pedagógicas y Técnicas de Apoyo del Ministerio de Educación, a su elección.

El párrafo 5° de la propuesta describe las medidas que se pretende tomar para los establecimientos de mal desempeño. Además de informar a los padres de su clasificación, los establecimientos de mal desempeño están obligados a recibir apoyo técnico pedagógico por un plazo máximo de cuatro años. Si en la siguiente rueda de clasificación los establecimientos no muestran una mejora significativa—el texto no indica qué se entiende por mejora significativa y se deja a la definición que generará la agencia—deberán recibir apoyo obligatoriamente por dos años más y la agencia debe informar de la situación a los padres y se les enviará información respecto de escuelas cercanas que tengan mejor desempeño, y se deberán otorgar facilidades de transporte para que los estudiantes accedan a esos colegios mejor ubicados en la clasificación.

La propuesta indica que si después de cuatro años de ser clasificado en la categoría de mal desempeño, el establecimiento se mantiene en esa condición, se deberá informar a la Superintendencia para que revoque el reconocimiento oficial al establecimiento.

2. Revisión de la literatura

En este apartado se intentará dar cuenta de las principales líneas investigativas desarrolladas entorno al sistema de accountability poniendo principal énfasis en la confiabilidad de las mediciones que este sistema requiere. Para ello se dará una breve descripción del concepto, señalando luego las discusiones generadas entorno a sus principales implicancias: la decisión informada de los padres y la política educativa. Finalmente, se presentarán los principales problemas que las mediciones propias de un sistema de rendición de cuentas presentan y los alcances para la correcta interpretación.

En términos generales el principio de “accountability” hace referencia al comportamiento por el cual los proveedores de educación “rinden cuentas” o son responsables de su actividad, siendo el objetivo final aumentar la calidad de la educación. Si bien en Chile aún no existe un adecuado sistema de accountability o rendición de cuentas, las reformas educacionales que hoy en día se vienen desarrollando consideran fortalecer este punto generando instituciones especializadas en cumplir esta labor.

Al respecto Hirschman (1970) plantea que los usuarios pueden realizar dos acciones cuando se encuentran insatisfechos con el producto que consumen: hacer uso de la voz o de la salida. En el caso del sistema educativos, si los padres están insatisfechos con la enseñanza que imparte el establecimiento pueden sacar a sus hijos de los establecimientos, o bien pueden exigir una mejora en la enseñanza de forma individual u organizándose colectivamente. En Chile no es posible llevar a cabo completamente este postulado, ya que por un lado existen fuertes barreras económicas y geográficas que impiden que los padres puedan cambiar libremente a sus hijos de establecimientos y, por otro lado, la relación entre padres y escuelas es una relación marcada por el verticalismo, con escasa participación en los directorios y sin una figura legal que represente sus intereses.

Existe una amplia línea de investigación que intenta cuantificar el efecto de la rendición de cuentas de los establecimientos sobre la decisión de los padres. En un extremo de la discusión se ubica aquellos que postulan que los padres no se interesan en el resultado grupal (escuela) sino el desempeño individual su hijo (Hallgarten, 2001), por otro lado, la rendición de cuentas solamente del rendimiento promedio pareciera ser menos trascendente en la decisión de los padres de los quintiles I a IV que la calidad de los profesores y en el quintil V la señalización social del grupo de pares que asiste al colegio parece ser lo que marca la elección de colegio (Thieme y Treviño, 2010.)

Por otro lado se encuentran la línea argumentativa que consideran que la rendición de cuentas sí afecta las decisiones de los padres. Sapelli y Torche (2002) muestran que los resultados SIMCE tienen un efecto positivo y significativo en la decisión de los padres. A su vez ellos manifiestan que ciertos subsidios entregados a las escuelas y focalizados a los niños de las familias más pobre terminan siendo focalizados a las escuelas públicas más que a los niños de escasos recursos. Este hecho actuaría como un impuesto al sistema de subvenciones e impediría una completa movilidad de los estudiantes, ya que muchos padres optan por no cambiar a sus hijos a establecimientos particular subvencionados.

Así también Gallego y Hernando (2008) encontraron que los atributos más valorados por los padres a la hora de escoger un establecimiento educativo son la calidad y la cercanía, pero también consideran factores como la composición socioeconómica de los establecimientos, el costo monetario y las características de los pares. A su vez, ellos señalan que a mayor ingreso y educación de los padres mayor es la valoración por la calidad.

El caso de Polonia sirve para ejemplificar como los sistemas de rendición de cuentas pueden adaptarse a las demandas que la opinión pública genera, siendo altamente valorados por ésta y por el círculo educativo en general. En este país se ha desarrollado una institución especializada en rankear establecimientos de educación secundaria y superior, debido al aumento de universidades privadas acontecido en los '90. La principal dificultad que esta institución ha tenido que sortear es cómo preparar un ranking objetivo, pues consideraban que los problemas de objetividad se presentaban al menos en dos partes de la elaboración de sus rankings: una de ellas es la elaboración de indicadores y la otra el peso asignado a dichos indicadores (Siwiski, 2002). Ante la imposibilidad de lograr la completa objetividad en la elaboración de los rankings la institución resolvió crear un cuerpo consultivo conformado por prestigiosos representantes de distintas esferas del país, quienes se encargaban de establecer y evaluar las metodologías y criterios que año a año se utilizan. Así también fue necesario construir un ancla que permitiera considerar los resultados de años anteriores. Por medio de esta procedimiento no sólo se hacía uso de datos de años anteriores, sino también era posible controlar las variaciones en el ranking producto de eventos externos (Siwiski, 2002). Refinando y estabilizando la metodología, ésta institución ha ganado un lugar en el corazón de la opinión pública polaca y guía las apreciaciones y decisiones referentes al ámbito educativo de toda la comunidad.

Otra línea investigativa ha intentado medir el impacto del sistema de rendición de cuentas en el rendimiento de los estudiantes y las escuelas para guiar la política educativa. En este sentido Hanushek (2004) señala que el sistema de rendición de cuentas si tiene impacto en el rendimiento y este impacto es mayor en grupos de estudiantes de bajo rendimientos, e incluso en el rendimiento de grupos de minoría étnicas de USA. El autor es enfático en señalar que los sistemas de rendición de cuenta generan impacto siempre que la información venga aparejada de un sistema de

incentivos y sanciones, ya que la consecuencia deseable del sistema de rendición de cuentas es dejar que los otros actores actúen como respuesta a los resultados observados. La principal dificultad que enfrentan estas investigaciones vuelve a ser cómo generar una medida confiable del impacto del sistema de rendición de cuentas.

En busca de mediciones confiables y certeras capaces de capturar en esencia los desafíos que enfrenta la política educativa se han ido desarrollando una serie de medidas, algunas de ellas han surgido como una forma de hacerse cargo de las preguntas que las medidas existentes no podían responder. En este sentido se puede hacer la siguiente distinción en las medidas de desempeño (Wilson and Piebalga, 2008):

- 1) Niveles (Levels): mide el porcentaje de alumnos (de un establecimiento) que alcanza cierto objetivo establecido.
- 2) Cambios (Changes): mide las mejoras que experimenta un curso de determinado establecimiento en distintos momentos.
- 3) Ganancias (Gains): mide el progreso de un grupo de estudiantes entre dos puntos del tiempo.

Las principales críticas a las medidas en niveles están asociadas a que estas medidas podrían generar incentivos perversos para la consecución del objetivo deseado, asignando menos recursos a aquellos alumnos con peores resultados que no contribuyen a acercar al establecimiento al objetivo deseado. Así también, esto podría generar una selección de alumnos al interior de los establecimientos o forzar a una autoselección de éstos (Wilson, 2004). Debido a la alta correlación entre el NSE (nivel socioeconómico) y el desempeño de los estudiantes, se considera que las medidas en niveles no logran transmitir información acerca de la eficacia de los establecimientos pues establecimientos con mayor NSE tendrán mayor facilidad, dada su composición, para alcanzar los objetivos (Wilson and Piebalga, 2008).

Al respecto Goldschmidt y Choi (2007) concuerdan con los riesgos señalados por Wilson (2004) y consideran que la tarea por alcanzar un porcentaje de alumnos en un nivel de logro dado termina por enfatizar en las características de los alumnos. Observando la experiencia de *No Child Left Behind* (NCLB) los autores señalan que dentro de las grandes falencias no sólo se encuentra el hecho que establecimientos con mayor cantidad de alumnos pertenecientes a subgrupos tienen mayor dificultades para alcanzar cierto porcentaje de rendimiento, sino también que este modelo considera que el rendimiento de los alumnos es función únicamente de un año de aprendizaje y no de los últimos años del proceso de aprendizaje.

Los autores consideran que con este modelo se pierda mucha información pues una vez alcanzado el porcentaje de alumnos con el desempeño deseado no importa cómo cambia el rendimiento mientras no se ubique bajo el porcentaje establecido, dejando muchas preguntas sin responder. Goldschmidt y Choi (2007) señalan que dentro de las

principales preguntas que no es posible contestar con modelos que utilizan medidas en niveles se encuentra ¿en qué medida los desempeños anteriores del alumno influyen en el desempeño actual?, ¿qué factores están influyendo en el logro de los estudiantes?, ¿qué tan exacto es el modelo en la identificación de escuelas que necesitan mejorar? y para gusto de los autores los modelos de crecimientos o ganancia parecen una opción mucho más prometedora en este sentido.

Respecto a los modelos en cambios Goldschmidt y Choi (2007) consideran que tiene ciertas ventajas sobre el modelo en niveles, estas ventajas se asocian a tres elementos. El primer elemento refiere a la escala de medición, ya que al ser una medida para el mismo grado y la misma escuela en el tiempo los temas referentes a la escala de medición pierden importancia. En segundo lugar, resulta ser una medida informativa fácil de reportar ya que permite saber si los progresos de las escuelas son suficientes para alcanzar el rendimiento objetivo de largo plazo. Y en tercer lugar, la medida entrega información sobre un estado actual y sobre un posible desenlace futuro. Lamentablemente esta medida no dice nada sobre cómo cambia el desempeño del estudiante en forma individual en el tiempo, lo que lleva nuevamente a centrar las expectativas por mayor información sobre las medida de ganancia.

De esta forma se reconoce que las medidas en ganancia son una medida más informativa acerca de la eficacia de los establecimientos, aislando el efecto de la escuela del progreso de los alumnos (Wilson and Piebalga, 2008) y permiten reducir los incentivos a comportamientos deshonesto. Goldschmidt y Choi (2007) concuerdan con este último punto, señalan que los modelos de crecimiento permiten mejorar los problemas de los subgrupos pues si el crecimiento de cada estudiante se utiliza para generar las estimaciones del crecimiento global de la escuela, los subgrupos no necesariamente jugarán en contra de los resultados de la escuela. Finalmente los autores señalan que las medidas de diseño longitudinal producen información diferente, con una correlación que varía entre 0.25 y 0.75, dependiendo de la prueba y el período de tiempo analizado y sugieren usar más de un indicador para monitorear el desempeño de los establecimientos.

Hanushek (2010) buscando una medida confiable para cuantificar la calidad de los profesores señala que las medidas de valor agregado tienen ciertas falencias para cumplir con esta labor. Por un lado, el desempeño de los estudiantes está muy relacionado con el logro de la escuela, con el ingreso de de la familia y con indicadores económicos agregados de la escuela y, por otro lado, se encuentran los problemas asociados al error de medición y al sesgo de especificación. Pero la principal reticencia del autor a generar pagos a los profesores de acuerdo a medidas de valor agregado se relacionan con la sospecha de una enseñanza acotada solamente a los tópicos evaluados en las pruebas de medición general.

Como un refinamiento de las medidas de ganancia se ha surgido el CVA (contextual value added). El CVA se utilizó por primera vez el año 2006 en las escuelas Británicas,

con esta medida se pretendía asilar el impacto real de la escuela sobre el progreso del alumno. En este sentido existe evidencia de la relación que existe entre las características del establecimiento y las mejoras que éstos experimentan. Levacic y Woods (2002) señalan que elementos como el rendimiento del establecimiento en el año base de la medición o la cantidad de niños en desventaja social y la variación en el presupuesto, tienen una incidencia significativa en el desempeño que alcanzan los establecimientos.

El cálculo del CVA incluye una estimación de los resultados según una serie de factores como lo son el sexo, la etnia, la lengua de origen, entre otros. La diferencia entre la estimación del modelo y el resultado que realmente obtiene el alumno se denominó CVA (Wilson and Piebalga, 2008). La principal dificultad de las medidas de valor agregado radica en la naturaleza de los datos necesarios para su elaboración, en Chile recién a mediados del 2010 será posible contar con este tipo de datos. Otra dificultad asociada a dicha medida es que dada la complejidad del constructo resulta más difícil de comprender e interpretar. De esta forma, esta medida podría contribuir menos a la decisión de los padres y más a la orientación de las políticas públicas en educación.

La discusión sobre la confiabilidad de las mediciones en materia educativa resuena con urgencia hoy día en nuestro país debido a la necesidad de establecer una clasificación general de los establecimientos educativos que permita identificar aquellos establecimientos que requieren de algún tipo de intervención, tal como lo señala el Proyecto de Ley para el Aseguramiento de la Educación Parvularia, Básica y Media.

Para lograr una clasificación de los establecimientos confiable es necesario considerar las características de los datos que se manejan en Chile. A la fecha sólo es factible trabajar con datos de cohorte transversal, lo que obliga a trabajar con medidas en niveles y cambios. La experiencia en el uso de datos SIMCE señala dos grandes dificultades a sortear para lograr identificar los establecimientos de buen y mal desempeño: la volatilidad de los datos y la gran influencia del NSE del alumno en la clasificación de los establecimientos (Mizala, Romaguera, Urquiola; 2006).

Se reconoce que la volatilidad de los resultados generalmente está asociada a dos fuentes, la primera de ellas son las fluctuaciones producto de la naturaleza de los datos - ya que el grupo de estudiantes que rinden la prueba cambia de un año a otro- y la segunda fuente corresponde a las fluctuaciones producto de ruidos en las mediciones generadas por factores temporales externos a la escuela (por ejemplo: ruidos en el ambiente en el momento de rendir la prueba que actúan como agentes distractores) (Kane, Staiger, Geppert; 2002). De esta forma los datos resultantes de las mediciones son en esencia volátiles, los efectos de dicha volatilidad en los datos dependerá de cómo varían las características de los alumnos que conforman los establecimientos entre un año y otro, y del error asociado a las mediciones.

Para mitigar el error asociado a las mediciones Hanushek (2010) propone trabajar un gran número de alumno y varios años. Solución que coincide con los

hallazgos de Chay, McEwan y Urquiola (2005) quienes demostraron que los efectos de la volatilidad en la confiabilidad de los resultados están estrechamente relacionados con el tamaño de los establecimientos educativos. De esta forma, las fluctuaciones de los puntajes en establecimientos educativos de menor tamaño es un hecho aún más preocupante. La experiencia del programa P-900 da cuenta de la alta probabilidad de que escuelas con estas características sean seleccionadas, el criterio de selección del programa enfatiza en las escuelas que se encuentran bajo cierto puntaje (el puntaje de corte se establece de manera independiente en cada región). De esta forma escuelas pequeñas que tuvieron bajos puntajes el año de evaluación experimentan altos puntajes en los años posteriores. Los autores son enfáticos en señalar que la causa del alza en el puntaje es la fuerte “reversión a la media” que experimentan los establecimientos con menos de 30 alumnos inscrito (Chay, McEwan, Urquiola; 2005). De esta forma, escuelas con pocos alumnos tienen mayor probabilidad de ser seleccionadas para recibir algún tipo de recompensa o sanción, ya que la fluctuación de sus resultados es más extrema que las fluctuaciones en escuelas de mayor tamaño.

Kane, Staiger y Geppert (2002) proponen como medidas para aminorar la volatilidad de los resultados trabajar con un promedio ponderado de los resultados de varios años. De esta forma escuelas de mayor tamaño, que sufren menos fluctuaciones producto de ruido en las mediciones, asignan mayor peso a los puntajes más reciente e, inversamente, escuelas pequeñas con fuertes fluctuaciones deben asignar mayor peso a los resultados de años anteriores. En este sentido, los autores señalan, que un promedio simple de los resultados de varios años también podría resultar un camino eficiente para aminorar la volatilidad de los resultados.

Como solución para el problema de la sobre representación de los establecimientos pequeños se ha propuesto rankear por separado establecimientos de distintos tamaños, esto podría contribuir a que establecimientos con mayor cantidad de alumnos inscritos puedan optar a participar de programas, premios o sanciones asociados a sus resultados (Kane, Staiger, Geppert; 2002).

Otra dificultad que surge a la hora de generar una clasificación confiable de los establecimientos en Chile es la alta correlación que existe entre NSE del estudiante y el puntaje que este obtiene. De hecho, “entre un 75% y un 85% de la variación del puntaje de la prueba se explica por escolaridad de los padres y el ingreso del hogar” (Mizala, Romaguera, Urquiola; 2007). De esta forma un ranking elaborado con los resultados SIMCE estaría ordenando las escuelas según el NSE de los alumnos que la componen.

Para superar este obstáculo se ha propuesto controlar por los factores determinantes en la clasificación, lo que equivale a correr una regresión de dos niveles- alumno y escuela-. Una opción es controlar por el NSE del alumno y con el residuo del nivel de alumno rankear los establecimientos, con esto es posible disminuir la influencia del NSE pero como contrapartida se obtiene un ranking sumamente volátil,

hecho atribuible a que un ranking así construido no considera el efecto de la escuela (Mizala, Romaguera, Urquiola; 2007). La otra opción es realizar el mismo ejercicio pero esta vez controlando por el NSE del establecimiento y rankeando con el residuo del nivel de escuela, con lo cual se obtiene un mejor ajuste que en la primera opción planteada pero aumenta la volatilidad de los datos, “los movimientos de la distribución se asemejan a una lotería. De los 8 años de los que se tienen datos, usando esta medida podría tenerse como resultado que más del 65% de las escuelas se mueve a un punto del primer quintil. Además, el 36% de las escuelas podría aparecer tanto en el peor como en el mejor grupo” (Mizala, Romaguera, Urquiola; 2007).

Los resultados anteriormente expuestos dan cuenta del tradeoff existente entre la elaboración de un ranking más representativo (menos determinado por el NSE del establecimiento) y la volatilidad de éste. De esta forma al intentar solucionar los problemas de representatividad de los rankings se está generando un problema mayor, el de la volatilidad de los datos.

En lo que sigue si intentará elaborar una clasificación de establecimientos que sortee de la mejor forma posible las dificultades aquí expuesta, para ello se hará uso de algunas de las alternativas anteriormente planteadas.

3. Métodos

En esta investigación se utilizan los datos de las pruebas del Sistema de Medición de la Calidad de la Educación (SIMCE) para las áreas de matemáticas y lenguaje de los años 2006, 2007 y 2008 y la información recolectada por medio del cuestionario de apoderados y la base de establecimientos. La investigación sólo considera la población de escuelas que tiene 30 o más cuestionarios de alumnos válidos durante los tres años. De esta forma 1946 escuelas conforman la población de estudio, de ellas 897 son Municipales, 927 Particular Subvencionadas y 122 Particular pagada, tal como se observa en la tabla 1

Tabla 1. Población de Alumnos y Establecimientos que Conforman el Análisis

Alumno / Dependencia // Año	2006	2007	2008
N° Alumnos	139,248	122,882	130,523
N° Establecimientos Municipal	897	897	897
N° Establecimientos Particular Subvencionado	927	927	927
N° Establecimientos Particular Pagado	122	122	122
Total Establecimientos	1,946	1,946	1,946

Los establecimientos se agrupan por medio de una técnica estadística multivariante conocida como análisis de cluster, esta herramienta permite agrupar de forma objetiva y replicable en el tiempo un grupo de elementos en torno a ciertas variables de agrupación. El análisis de *cluster no jerárquicos disjuntos* determina, según la naturaleza de los casos, el grupo de pertenencia de cada establecimiento en base a

métodos heurísticos que agrupan objetos de alta homogeneidad interna en grupos diferenciados por la alta heterogeneidad externa que existe entre ellos, excluyendo la posibilidad de que un grupo dependa de otro y que un caso pueda pertenecer a más de un grupo (Pérez, 2004).

El algoritmo seleccionado para generar el agrupamiento de los establecimientos es el Método de K-medias, este método busca maximizar la distancia entre los grupos disminuyendo la varianza al interior de cada grupo. Para ello el algoritmo sigue los siguientes pasos: (Affifi y Clark, 1997, p. 395)

1° Divide los datos en k grupos iniciales.

2° Calcula el centroide (media) de cada grupo.

3° Para cada caso calcula la distancia entre el caso y el centroide. Y asigna el caso al grupo cuya distancia con el centroide sea menor.

4° Luego se repite el mismo procedimiento para cada caso hasta que ningún caso pueda disminuir la varianza residual.

En este procedimiento se hace uso de la distancia euclídea, ya que minimizar la varianza residual es equivalente a conseguir que sea mínima la suma de distancias al cuadrado desde los casos a la media de los grupos (Pérez, 2004). Para el caso de dos objetos medidos sobre dos variables se tiene que el cálculo de la distancia euclídea es:

$$\text{Ecuación 1: } \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

El algoritmo de K-medias tiene como limitación que el orden de los establecimientos puede alterar el cálculo de los grupos. En esta investigación se define como punto de partida la primera observación de un orden ascendente por el RBD del establecimiento. De igual forma se repitió el cálculo con elecciones aleatorias de puntos de arranques y se verificó que las alteraciones eran menores para los fines de esta investigación, dado que se dejó iterar el algoritmo hasta conseguir la convergencia, definiendo como máximo número de iteraciones 1000.

Mediante esta técnica se simularán de 4 a 9 grupos de establecimientos agrupados en torno a una sola variable para cada ranking y se realizará un análisis en torno a la cantidad de grupos óptima para clasificar los establecimientos. Debido a que la elección de la cantidad de grupos óptima debe realizarse con una adecuada combinación de criterios matemáticos y de interpretabilidad, el análisis no sólo contemplará los principales descriptivos de los grupos creado, sino también un indicador de adecuación denominado Pseudo F (Pérez, 2004). El índice Pseudo F fue propuesto por Calinski y Harabas en 1974 como medida para identificar el número óptimo de grupos, éste índice es una medida de la proporción entre la similitud entre grupos y la similitud al interior del grupo (Jain y Koronios, 2008). La ecuación 2 detalla la composición del Pseudo F:

$$\text{Ecuación 2: } \frac{\frac{B(k)}{(k-1)}}{\frac{W(k)}{(n-k)}}$$

Donde $B(k)$ denota la suma de cuadrados entre grupos y $W(k)$ denota la Suma de cuadrados al interior de los grupos, de modo que un óptimo número de grupos es aquel que maximiza el valor del índice (Jain y Koronios, 2008).

Con los datos y metodología anteriormente señaladas se compararán tres formas de clasificar el desempeño de los establecimientos:

- A. Agrupación por Puntaje SIMCE
- B. Agrupación por Nivel de Logro Inicial
- C. Agrupación por Eficacia

La principal diferencia entre estas tres formas de clasificar radica es la información en torna a la que se clasifican las escuelas. Mientras la agrupación por puntaje SIMCE utiliza el puntaje SIMCE de los establecimientos como variable de agrupación, la agrupación por Nivel de Logro Inicial utiliza el porcentaje de alumnos del establecimiento educativo que se ubica en el nivel de logro inicial y la agrupación por eficacia utiliza el error del nivel de escuela de una regresión lineal jerárquica en dos niveles, nivel estudiante y nivel escuela.

La variable de agrupación de las tres formas de clasificar son el promedio de las variables para los años 2006, 2007 y 2008, al promediar los datos se busca disminuir la volatilidad de la clasificación, pues con ellos se elimina el ruido asociado al error de las mediciones. La especificación de las variables de agrupación se detalla en la tabla 1.

Tabla 1. Variables de agrupación utilizadas para la agrupación de establecimientos.

Variable Agrupación	Definición	Media y desviación estándar
Puntaje SIMCE Lenguaje	Media de los años 2006, 2007 y 2008 del puntaje promedio de los estudiantes de 4° grado por escuela en la prueba SIMCE Lenguaje.	$\mu = 259.054$ $\sigma = 23.827$
Puntaje SIMCE Matemática	Media de los años 2006, 2007 y 2008 del puntaje promedio de los estudiantes de 4° grado por escuela en la prueba SIMCE matemática.	$\mu = 250.607$ $\sigma = 27.417$
Error Lenguaje	Error de estimación promedio de los años 2006, 2007 y 2008 de la regresión lineal jerárquica a nivel escuela que predice el puntaje de los alumnos de 4°	$\mu = 0$

	básico en la prueba SIMCE Lenguaje para cada año.	$\sigma = 10.573$
Error Matemática	Error de estimación promedio de los años 2006, 2007 y 2008 de la regresión lineal jerárquica a nivel escuela que predice el puntaje de los alumnos de 4° básico en la prueba SIMCE Matemática para cada año.	$\mu = 0$ $\sigma = 12.822$
Nivel de Logro Inicial Lenguaje	Promedio de los años 2006, 2007 y 2008 de la proporción de alumnos de 4° básico por escuela que se ubica en el nivel de logro inicial según los resultados en la prueba SIMCE Lenguaje para cada año.	$\mu = 0.366$ $\sigma = 0.175$
Nivel de Logro Inicial Matemática	Promedio de los años 2006, 2007 y 2008 de la proporción de alumnos de 4° básico por escuela que se ubica en el nivel de logro inicial según los resultados en la prueba SIMCE Matemática para cada año.	$\mu = 0.378$ $\sigma = 0.198$

Para la elaboración de la Agrupación por Eficacia fue necesario generar un modelo jerárquico lineal en dos niveles¹, estudiante y escuela. La metodología empleada para su realización se resume en las siguientes ecuaciones:

$$\text{Ecuación 1: } Y_{ij} = \beta_{0j} + \beta_{nj}ESTUDIANTE_{ij} + r_{ij}$$

$$\text{Ecuación 2: } \beta_{0j} = \gamma_{0j} + \gamma_{nj}ESCUELA_j + u_{0j}$$

$$\text{Ecuación 3: } \beta_{nj} = \gamma_{n0} + u_{nj}$$

Las ecuaciones 1, 2 y 3 indican la forma general de los modelos jerárquicos lineales utilizados. La ecuación 1 corresponde al nivel de estudiante, donde el aprendizaje Y_{ij} depende de un vector de variables relativas al estudiante, cuyos coeficientes son β_{nj} , donde n corresponde a cada uno de los coeficientes asociados a las variables de nivel estudiante dentro de cada escuela j . La ecuación incluye un error estocástico de predicción para el nivel de estudiantes representado por r_{ij} . Las ecuaciones 2 y 3 son una formulación general para el modelo de nivel de escuela. En la ecuación 2 se observa un vector de variables escolares que predicen el rendimiento de los estudiantes sobre el intercepto, considerando un error estocástico representado por u_{0j} . La ecuación 3, por su parte, establece el comportamiento de las variables del nivel estudiante al estar agrupadas en escuelas e incluye su correspondiente error estocástico a nivel escuela representado por u_{nj} . Es necesario recordar que la n representa a cada uno de los

¹ La estimación del modelo se realiza con el programa HLM

coeficientes asociados a variables del nivel de estudiante. Finalmente, y una vez que el modelo ha sido estimado, se calcula el error asociado al nivel de escuela para cada prueba, es decir u_{0j} , y se guarda como variable de agrupación.

4. Resultados

A continuación se expondrán los principales resultados de las distintas formas de agrupación de escuelas desarrolladas, según: Puntaje SIMCE, Nivel de Logro Inicial, Error de Estimación. Se expondrá en detalle el análisis acerca de la pertinencia de trabajar con 4, 7 u 8 grupos.

Las escuelas agrupadas según el Puntaje SIMCE, tanto de Lenguaje como de Matemática, se ordenan de tal forma que el primer grupo concentra las escuelas con mejores rendimientos y el último grupo las escuelas de peores rendimientos.

La tabla 1 muestra los principales resultados de clasificar en 4, 7 y 8 grupos según la media del puntaje que obtienen los establecimientos en la prueba de Lenguajes de los años 2006, 2007 y 2008 para 4to básico. En términos generales es posible señalar que los establecimientos se concentran mayormente en los grupos del centro independiente la cantidad de grupos, las medias aumentan y las varianzas disminuyen a medida que aumenta la cantidad de grupos.

El gráfico ubicado en la esquina izquierda superior muestra cómo cambia el porcentaje de establecimientos que componen los grupos al aumentar la cantidad de grupos en la clasificación. Como patrón común a los 3 formas de clasificar se observa que los grupos del centro concentran el mayor porcentaje de establecimientos y los grupos ubicados a los extremos concentran el menor porcentaje, además se observa que al aumentar la cantidad de grupos en la clasificación los grupos con peor desempeño son los que concentran el menor porcentaje de establecimientos.

Para el caso de la clasificación en 4 grupos se observa que un tercio de los establecimientos se ubican en el grupo 3, éste grupo es el que concentra el mayor porcentaje de establecimientos y en términos del Proyecto de Ley de Aseguramiento de la Calidad equivale al grupo de establecimientos con desempeño regular. El grupo 1 o de buen desempeño concentra el menor porcentaje de establecimientos con sólo un 16.3% del total.

Al aumentar los grupos de la clasificación se observa que todos los grupos disminuyen el porcentaje de establecimientos que concentran y los grupos de peores rendimientos son los grupos que menor porcentaje de establecimientos concentran. Para el caso de la clasificación en 7 grupos el mayor porcentaje de establecimientos se encuentra en el grupo 4 y en una clasificación en 8 grupos el mayor porcentaje de establecimientos se ubica en el grupo 5. El aumento en un grupo al pasar de una clasificación de 7 a otra en 8 grupos produce una pequeña disminución del porcentaje de establecimientos en los

grupos del 1 al 6 y un notorio aumento en el porcentaje de establecimientos que concentra el grupo 7.

En la esquina inferior izquierda de la tabla 1 se observa el comportamiento de las medias de los grupos para las distintas clasificaciones consideradas en el análisis. En términos generales es posible señalar que al aumentar la cantidad de grupos en la clasificación las medias de los grupos aumentan y la distancia entre la media de un grupo y el grupo que le sigue disminuye. De esta forma las medias de una clasificación en 4 grupos poseen un rango que se encuentra entre los 297.4 y los 229 puntos, mientras el mismo rango para una clasificación en 7 grupos es de 304.6 a 217.5 puntos y para una clasificación en 8 grupos es de 305.1 a 215.4 puntos. La diferencia de medias para una clasificación en 4 grupos es en promedio de 22.8 puntos, si la clasificación se realiza con 7 grupos el promedio disminuye a 14.51 y con 8 grupos el promedio de estas diferencias es de 12.81 puntos.

En la esquina inferior derecha se observa el gráfico de las medias para las distintas clasificaciones. Como patrón común a las tres clasificaciones se observa que las varianzas siguen una distribución similar, siendo mayor las medias de los grupos extremos y menor en los grupos del centro, además del fuerte descenso que experimentan las varianzas al pasar de 4 a 7 u 8 grupos. A su vez llama la atención que el grupo de peor desempeño en una clasificación de 7 u 8 grupos tenga una varianza mayor que el grupo de mejor desempeño, ya que en una clasificación en 4 grupos se observa el escenario inverso.

Las varianzas de una clasificación en 4 grupos son bastante elevadas especialmente en los grupos extremos, la mayor varianza la presenta el grupo 4 y la menor varianza el grupo 3. En una clasificación en 7 u 8 grupos las varianzas disminuyen sustantivamente, los grupos que presentan la menor varianzas para cada clasificación son el grupo 4 y 5 respectivamente, mientras que los grupos que presentan la mayor varianza son el grupo 7 y 8 respectivamente. Para la clasificación en 7 u 8 grupos la varianza de los grupos 1 y 4 son bastante parecidas, visualmente se observa que sus líneas prácticamente se traslapan.

Finalmente, en la esquina superior derecha de la tabla 1 se encuentra el gráfico del índice Pseudo F, en este gráfico es posible observar un claro aumento en el índice a medida que aumenta la cantidad de grupos. El aumento del índice puede interpretarse como señal de la gran diversidad de escuelas que se está intentando clasificar, las cuales requieren de una clasificación en la mayor cantidad de grupos posibles para que dicha clasificación logre representarlas. A su vez es posible señalar que el índice asociado a clasificaciones entre 9 y 50 grupos sigue aumentando, observándose por ejemplo que entre 45 y 50 grupos comienza a estabilizarse en un valor aproximado a 25.000.

La tabla 2 repite el ejercicio de clasificación realizado anteriormente, pero con la media del Puntaje SIMCE Matemáticas 4to básico por establecimientos para los años 2006, 2007 y 2008.

En términos generales es posible señalar que las agrupaciones generadas por medio de los puntajes SIMCE Lenguaje y Matemáticas son bastante similares, la principal diferencia entre ellos radica en que la varianza al interior de los grupos elaborados con el puntaje de matemática es mayor que la de los grupos elaborados con el puntaje de lenguaje. Otro hecho que salta a la vista al realizar un paralelo entre ambas pruebas es que las agrupaciones por puntaje lenguaje generan grupos con medias más elevadas para todas las agrupaciones. Esto puede sugerir que, empíricamente, la clasificación de resultados de matemática necesite más categorías que en lenguaje, aunque esto no necesariamente es práctico para una política central.

En la esquina superior izquierda de la tabla 2 se presenta el gráfico con el porcentaje de establecimientos por grupo. El porcentaje de establecimientos por grupo presenta una estructura y una distribución similar a la prueba de Lenguaje, de esta forma los grupos extremos poseen un menor porcentaje de establecimientos que los grupos del centro y al aumentar el número de grupos en la clasificación el grupo con mejor rendimiento supera en porcentaje de establecimientos al grupo de peor desempeño. Los grupos con mayor y menor porcentaje de establecimientos para las tres clasificaciones son los mismo que en las agrupaciones hechas a partir del puntaje en la prueba SIMCE lenguaje.

En la esquina inferior izquierda de la tabla 2 se ubica el gráfico con las medias de las distintas clasificaciones. Si bien las medias son un poco más bajas que las obtenidas en la prueba de Lenguaje, éstas presentan un comportamiento muy similar. Las medias aumentan a medida que a clasificación se realiza con más grupos y la diferencia entre la media de un grupo y otro disminuye. En este sentido se observa que el promedio la diferencia de medias en las tres clasificaciones es superior a las diferencias promedio de la prueba de lenguaje. De esta forma, para una clasificación en 4 grupos la diferencia promedio de medias es de 25.7, mientras que en una clasificación en 7 grupos es de 16.7 y en una clasificación en 8 grupos es de 15.

En la esquina inferior derecha de la tabla 2 se encuentra el gráfico de las varianzas para las tres clasificaciones realizadas. Si bien las varianzas de los grupos que se conforman a partir de los resultados en la prueba de matemática son mayores que las varianzas de los grupos elaborados a partir de los resultados en lenguaje, se observa que las varianzas poseen una estructura muy similar, siendo mayores en los grupos extremos que en los grupos del centro. La menor y mayor varianza de una clasificación en 4 grupos coincide con lenguaje, ya que corresponde al grupo 3 y 1 respectivamente. Esta misma coincidencia ocurre con la clasificación en 8 grupos, donde el grupo de mayor varianza es el 1 y el de menor varianza es el 5. La clasificación en 7 grupos tiene una pequeña diferencia en términos de varianza con la prueba de lenguaje, si bien el grupo con mayor varianza sigue siendo el grupo 1, el grupo con menor varianza es el grupo 5 y no el grupo 4.

El análisis del índice Pseudo F ubicado en la esquina superior derecha de la tabla 2, muestra un continuo aumento a medida que aumenta la cantidad de grupos en la

clasificación, tomado valores desde 6,337 a 11,318 al aumentar paulatinamente los grupos de 4 a 9. De igual forma, se observa que la pendiente es menos inclinada la pasar de 6 a 7 grupos, lo podría indicar que una mejor clasificación se consigue con más de 8 grupos. Si el análisis del estadígrafo se extiende a 50 grupos se observa que el índice presenta pequeños descensos en su valor al pasar de 45 a 48, volviendo a aumentar hacia los 50 grupos.

A continuación se presentarán los resultados de agrupar los establecimientos según el Error de Estimación de un modelo lineal jerárquico que explica el resultados de la prueba SIMCE lenguaje y matemática. Los establecimientos así agrupados se ordenan de tal forma que el primer grupo concentra las escuelas con mejores rendimientos, pues su rendimiento se encuentra muy por sobre el rendimiento estimado, y el último grupo concentra las escuelas de peor rendimiento, pues su rendimiento se encuentra muy por debajo del rendimiento estimado.

La tabla 3 muestra los principales resultados de clasificar en 4, 7 y 8 grupos según la media del error de estimación de los establecimientos para los resultados en la prueba de Lenguajes de los años 2006, 2007 y 2008 para 4to básico. En términos generales se observa que el mayor porcentaje de establecimientos se concentra en los grupos del centro para las 3 clasificaciones, al aumentar la cantidad de grupos en la clasificación las varianzas disminuyen pero son muy similares en las clasificaciones hechas con 7 y 8 grupos y las medias de los grupos aumentan, volviéndose incluso positivas en algunos grupos del centro.

El gráfico ubicado en la esquina izquierda superior muestra cómo cambia el porcentaje de establecimientos que componen los grupos al aumentar la cantidad de grupos en la clasificación. Tal como ocurría con las clasificaciones hechas a partir del puntaje de los establecimientos se observa que los grupos del centro concentran el mayor porcentaje de establecimientos y los grupos ubicados a los extremos concentran el menor porcentaje, este hecho es un patrón común a las 3 formas de clasificar aquí empleadas.

En una clasificación en 4 grupos se observa que el mayor porcentaje de establecimientos se concentra en el grupo 3 con más de un tercio del total de establecimientos y el menor porcentaje de establecimientos lo concentra el grupo 1.

Al aumentar los grupos de la clasificación a 7 el mayor porcentaje de establecimientos lo concentra el grupo 4 con un quinto de los establecimientos y el menor porcentaje de establecimientos lo presenta el grupo 1 con menos del 4%. Al pasar de una clasificación en 7 a otra en 8 grupos, del grupo 1 al 4 se observa una reducción en el porcentaje de establecimientos y en los grupos 6 y 7 se observa un alza, la cual es más notoria en el grupo 7. En este escenario es el grupo 5 el que concentra un mayor porcentaje de establecimientos con menos de un quinto del total de establecimientos y el grupo 1 el menor porcentaje con menos del 3%.

En la esquina inferior izquierda de la tabla 3 se observa el gráfico de las medias de los grupos que se obtienen de una clasificación en 4, 7 y 8 grupos según el error de estimación de la prueba de lenguaje. Al aumentar la cantidad de grupos se observa que la media de los grupos aumentan y llegan a ser positivas en algunos grupos en que al clasificar en 4 grupos eran negativa, como es el caso de los grupo 3 y 4.

El rango de las medias de una clasificación en 4 grupos va desde los 17 a los -14.5 puntos, mientras que en una clasificación de 7 grupos va desde los 23.7 a los -11.8 puntos y en una clasificación en 8 grupos el mismo rango se encuentra entre los 25.8 y los -19.4 puntos. El promedio de las diferencias de las medias entre un grupo y el que le sigue es de 10.5 para el caso de una clasificación en 4 grupos. Si la clasificación se realiza con 7 grupos este mismo promedio es de 7.2, mientras que en una clasificación en 8 grupos el promedio corresponde a 6.5.

En la esquina inferior derecha de la tabla 3 se encuentra el gráfico de las varianzas de los grupos. Nuevamente se observa que las varianzas de los grupos extremos son mucho más elevadas que las varianzas de los grupos del centro para los tres tipos de clasificación aquí empleadas.

La clasificación en 4 grupos presenta la mayor varianza en el grupo 1 y la menor en el grupo 2. Al aumentar los grupos a 7 u 8 se observa que la varianza del grupo 1 sigue siendo la mayor en cualquiera de los dos casos, mientras que la menor varianza la presenta el grupo 4 para una clasificación en 7 grupo y el grupo 4 y 5 para una clasificación en 8 grupos. Llama la atención que la varianza disminuyan al aumentar la cantidad de grupos en la clasificación para casi todos los grupos, salvo para el grupo 1. Además el comportamiento de las varianzas de una clasificación en 8 y 7 grupos es bastante similar, la principales diferencias se encuentran en los grupos extremos, como lo son el grupo 1 y 7.

Finalmente, en la esquina superior derecha de la tabla 3 se encuentra el índice Pseudo F asociado a las distintas clasificaciones. El índice muestra un continuo aumento entre los 4 y 9 grupos. Si bien la pendiente entre los últimos grupos presenta una menor inclinación, al calcular el índice para más grupos se observa que entre los 40 y 45 grupos el índice presenta un descenso, teniendo un pick en los 30 grupos.

La tabla 4 muestra los principales resultados tras repetir el ejercicio anterior con el error de estimación asociado a la prueba de matemáticas. En términos generales se observan resultados similares a la anterior agrupación en términos de porcentaje de establecimientos por grupo y media, pero en términos de varianza existen ciertas diferencias relacionadas con los cambios que experimenta al aumenta la cantidad de grupos en la clasificación.

En la esquina superior izquierda de la tabla 4 se observa la composición de los grupos para las distintas clasificaciones. El porcentaje de establecimientos en los distintos grupos sigue un comportamiento muy similar a la anterior agrupación ya que los grupos

que concentran el mayor porcentaje de establecimientos son los del centro, mientras que los grupos de los extremos concentran el menor porcentaje. El grupo que concentra el menor porcentaje de establecimientos es el grupo 1 en las tres clasificaciones, mientras que el grupo que concentra el mayor porcentaje de establecimientos en las clasificaciones en 4, 7 y 8 grupos son el grupo 3, 4 y 5 respectivamente. Cabe señalar que en la agrupación hecha con el error de estimación de lenguaje se tenían los mismos resultados con pequeñas diferencias en los porcentajes.

En la esquina inferior izquierda de la tabla 4 se ubica el gráfico con las medias de los grupos para las tres clasificaciones empleadas. Se observa que las medias en matemática son más altas que en lenguaje para el caso de los grupos con medias positivas y más bajas que en lenguaje para el caso de grupos con medias negativas, de esta forma el rango de las medias es mayor para las 3 clasificaciones analizadas. La media de la clasificación en 4 grupos tiene un rango que se encuentra entre los 23.2 y los -16.8 puntos, el mismo rango para una clasificación en 7 grupos se encuentra entre los 28.9 y los -22.5 puntos, y para el caso de la clasificación en 8 grupos se tiene un rango de medias entre los 31.3 y los -22.5 puntos. El promedio de las diferencias de medias entre un grupo y el siguiente para una misma clasificación es menor al de la prueba de lenguaje para el caso de las clasificaciones con 4 y 7 grupos y mayor para la clasificación en 8 grupos.

En la esquina derecha inferior se observa el gráfico con las varianzas. El comportamiento de la varianza al interior de los grupos para las 3 clasificaciones empleadas es distinto al comportamiento de las varianzas en la agrupación por error lenguaje ya que al aumentar los grupos la varianza de todos los grupos disminuye. A pesar de esta diferencia la estructura de las varianzas posee un comportamiento similar, ya que son más elevadas en los casos extremos y menores en los casos del centro. En el caso de la clasificación en 4 grupos la menor varianza la presenta el grupo 3, para la clasificación en 7 grupos la menor varianza la presenta el grupo 5 y para la clasificación en 8 grupos la menor varianza la presenta el grupo el grupo 4 y 5, a pesar de esto las varianzas de los grupos intermedios para la clasificación de 7 y 8 grupos son similares. En todas las clasificaciones es el grupo 1 el que presenta la mayor varianza.

En la esquina superior derecha se observa el índice Pseudo F, éste índice, muestran un claro y constante aumento de su valor a medida que aumenta el número de grupos de la clasificación. El estadígrafo toma valores desde 4,970 a 7,640. Al calcular el valor del índice para una clasificación con más grupos se observa que hacia los 50 grupos el índice continúa aumentando su valor, aunque sufre un pequeño descenso al pasar de los 45 a los 48 grupos tras lo cual sigue aumentando.

A continuación se presentarán los resultados de agrupar los establecimientos con el promedio de la proporción de alumnos por establecimiento en nivel de logro inicial según los resultados de la prueba SIMCE de matemáticas y lenguaje de los años 2006, 2007 y 2008. Los establecimientos así agrupados se ordenan de tal forma que el primer

grupo concentra los establecimientos con peores rendimientos, pues poseen un mayor porcentaje de alumnos en nivel de logro inicial, y el último grupo concentra los establecimientos con mejor rendimiento, pues poseen un menor porcentaje de alumnos en dicho nivel de logro.

La tabla 5 muestra los principales resultados de clasificar en 4, 7 y 8 grupos con el promedio del porcentaje de alumnos en nivel de logro inicial según los resultados de la prueba SIMCE lenguaje de los años 2006, 2007 y 2008. En general se observa que el porcentaje de establecimientos se distribuye de forma homogénea al interior de los grupos, las medias disminuye al aumentar los grupos y estas disminuciones son más significativas en los grupos de mejor rendimiento, las varianzas parecen muy similares en los grupos del centro y se traslapan en más de un grupo.

El gráfico ubicado en la esquina izquierda superior de la tabla 5 muestra las variaciones que experimenta el porcentaje de establecimientos que componen los grupos al aumentar la cantidad de grupos en la clasificación. Se observa que el porcentaje de establecimientos en los grupos intermedios es mayor que en los grupos extremos para las 3 clasificaciones y que la distribución de establecimientos en los grupos para las tres clasificaciones es bastante homogénea. En la clasificación en 4 grupos el mayor porcentaje de establecimientos se concentra en el grupo 3, concentrando cerca de un tercio de los establecimientos, y el menor porcentaje de establecimientos se encuentra en el grupo 1, con un poco más de un quinto de los establecimientos. En la clasificación en 7 y 8 grupos 5 y 4 respectivamente, mientras que el menor porcentaje de establecimientos se encuentra en el grupo 1 para las dos clasificaciones.

El gráfico ubicado en la esquina inferior izquierda de la tabla 5 muestra las medias de los grupos para las distintas clasificaciones. En términos generales se observa que las medias aumentan a medida que aumenta el número de grupos en la clasificación y que las medias de las clasificaciones en 7 y 8 grupos tiene pequeñas diferencias hasta el grupo 5, del grupo 6 en adelante la diferencia entre la media de los grupo se vuelve mayor. En la clasificación en 4 grupos el rango de las medias se encuentra entre 0.61 y 0.13, si la clasificación se hace en 7 grupos el rango se ubica entre 0.70 y 0.10 y para una clasificación en 8 grupos el rango se encuentra entre los 0.71 y 0.09.

La diferencia entre la media de un grupo y el siguiente disminuye al aumentar el número de grupos, de esta forma una clasificación con 4 grupos tiene como promedio de la diferencia de medias 0.16, mientras una clasificación en 7 grupo tiene como promedio de dichas diferencias 0.1 y para una clasificación en 8 grupos el promedio es de 0.09.

El gráfico de la esquina inferior derecha muestra las varianzas obtenidas en los grupos de las tres clasificaciones. Si bien las varianzas disminuyen al aumentar la cantidad de grupos, este hecho no se cumple en todos los grupos de las clasificaciones en 7 y 8 grupos.

La varianza de la clasificación en 4 grupos es la más elevada para todos los grupos, el grupo con mayor varianza es el grupo 1 y el grupo con menor varianza es en grupo 2. En una clasificación en 7 grupos la mayor varianza la presenta el grupo 1 y la menor el grupo 4. Si la clasificación se realiza en 8 grupos sigue siendo el grupo 1 el que presenta mayor varianza, mientras que el grupo con menor varianza es el grupo 6. Las varianzas de los grupo 1 y 3 para las clasificaciones en 7 y 8 grupos son las mismas, tal como se señaló más arriba.

En la esquina superior derecha se ubica el gráfico del índice Pseudo F, el cual muestra un continuo aumento a medida que aumenta la cantidad de grupos en la clasificación, tomado valores desde 6,797 a 12,451 al aumentar paulatinamente los grupos de 4 a 9. Si el análisis del índice se extiende a 50 grupos se observa que éste presenta pequeños descensos en su valor al pasar de 45 a 48 grupos, volviendo a aumentar hacia los 50 grupos.

La tabla 6 muestra los principales resultados de clasificar en 4, 7 y 8 grupos con el promedio del porcentaje de alumnos en nivel de logro inicial según los resultados de la prueba SIMCE matemática de los años 2006, 2007 y 2008.

Si se comparan los resultados que se obtienen según los puntajes de matemáticas y lenguaje se observan varianzas y distribuciones de establecimientos bastante similares, mientras que las medias de los grupos son más elevadas para los grupos conformados con los resultados de matemáticas. Este último hallazgo indica que existen más establecimientos con elevados porcentajes de alumnos en nivel de logro inicial en matemáticas que en lenguaje.

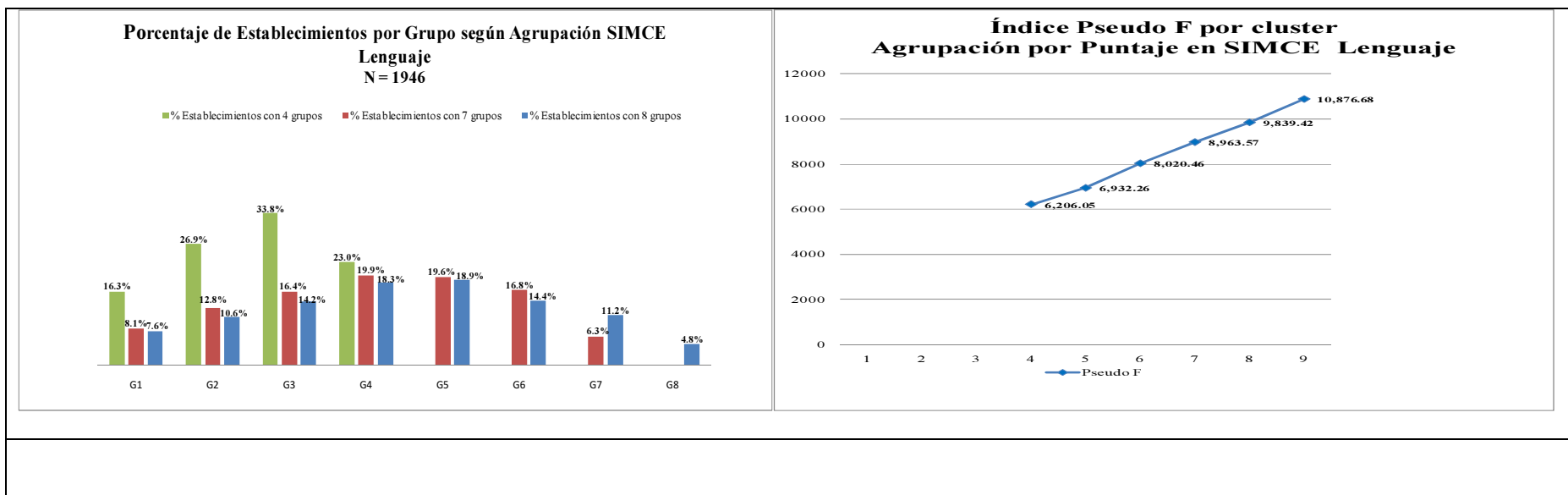
En la esquina superior izquierda de la tabla 6 se observa la distribución de establecimientos por grupo. La forma de distribución de los establecimientos al interior de los grupos posee una estructura bastante similar a lenguaje, ya que la distribución es bastante homogénea siendo levemente mayor el porcentaje de establecimientos que concentran los grupos del centro. Los grupos que concentran el mayor porcentaje de establecimientos son los grupos 2, 3 y 5 para las clasificaciones en 4, 7 y 8 grupos respectivamente, mientras que el grupo con menor porcentaje para cualquiera de las 3 clasificaciones es el grupo 1.

En la esquina inferior izquierda de la tabla 6 se encuentra el gráfico con las medias de los grupos. Si bien la media de los grupos son más elevadas que en el caso anterior se observa que las medias siguen el mismo patrón. De esta forma aumentan a medida que aumenta la cantidad de grupos en la clasificación, pero el aumento es menor en los primeros 3 grupos para el paso de una clasificación en 7 a otra de 8 grupos. En promedio la diferencia entre la media de un grupo y el grupo que le sigue en una misma clasificación es mayor en la prueba de matemáticas que en lenguaje para clasificaciones en 4 y 7 grupos e igual que en lenguaje para clasificaciones en 8 grupos.

En la esquina inferior derecha de la tabla 6 se observa el gráfico de las varianzas para las distintas clasificaciones empleadas. Las varianzas poseen una estructura bastante similar al escenario generado a partir del error de estimación asociado a la prueba de lenguaje. La clasificación en 4 grupos presenta una varianza mayor para todos los grupos, la menor varianza de la clasificación en 4 grupos la presenta el grupo 2 y la mayor el grupo 1. En una clasificación en 7 u 8 grupos la mayor varianza la presenta el grupo 1, mientras que la menor varianza la presenta 3 en las dos clasificaciones. La clasificación en 7 y 8 grupos presenta la misma varianza en el grupo 1 y muy similares a lo largo de los grupos, aunque estas diferencias comienzan a ser más elevadas a partir del grupo 5.

En la esquina superior derecha de la tabla 6 se observa el gráfico del índice Pseudo F, el cual toma valores desde 7,749 a 13,856 entre el grupo 4 y 9 respectivamente, presentando un claro aumento a medida que aumenta la cantidad de grupos. Las pendientes entre un grupo y otro son similares, presentando una pequeña disminución al pasar de una clasificación en 6 a otra en 7 grupos. Al calcular el índice asociado a una clasificación 10, 20, 30, 40, 45, 48 y hasta 50 grupos se observa un claro aumento en el valor del estadígrafo, tomando el valor de 56,862 para el caso de 50 grupos. Este hecho señalaría la necesidad de generar una clasificación de establecimientos en un mayor número de grupos.

Tabla 1. Resultados de agrupar según el promedio del Puntaje SIMCE Lenguaje 4to básico de los establecimientos en los años 2006, 2007 y 2008.



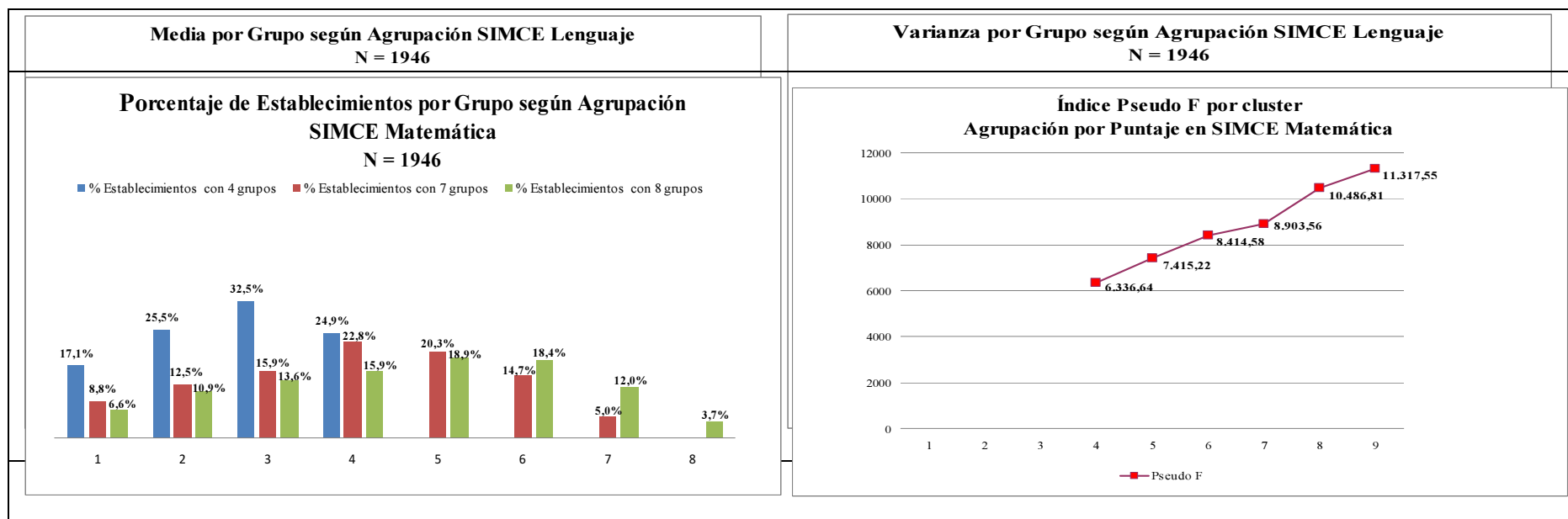
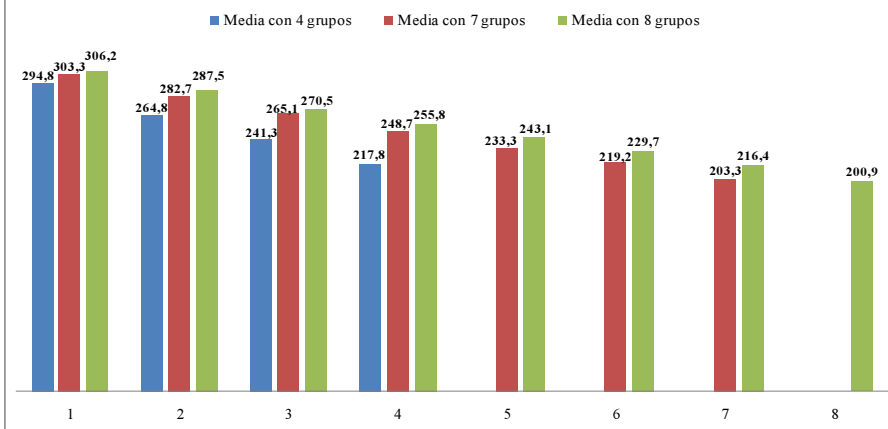


Tabla 2. Resultados de agrupar según el promedio del Puntaje SIMCE Matemática 4to básico de los establecimientos en los años 2006, 2007 y 2008.

Media por Grupo según Agrupación SIMCE Matemática
N = 1946



Varianza por Grupo según Agrupación SIMCE Matemática
N = 1946

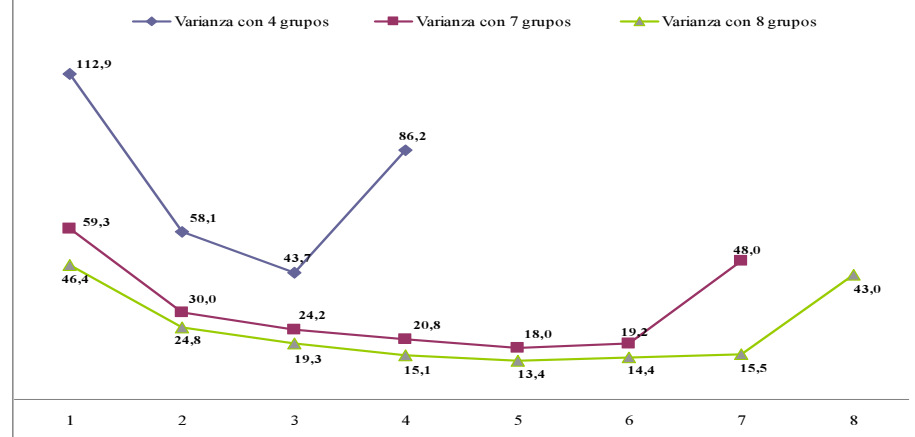
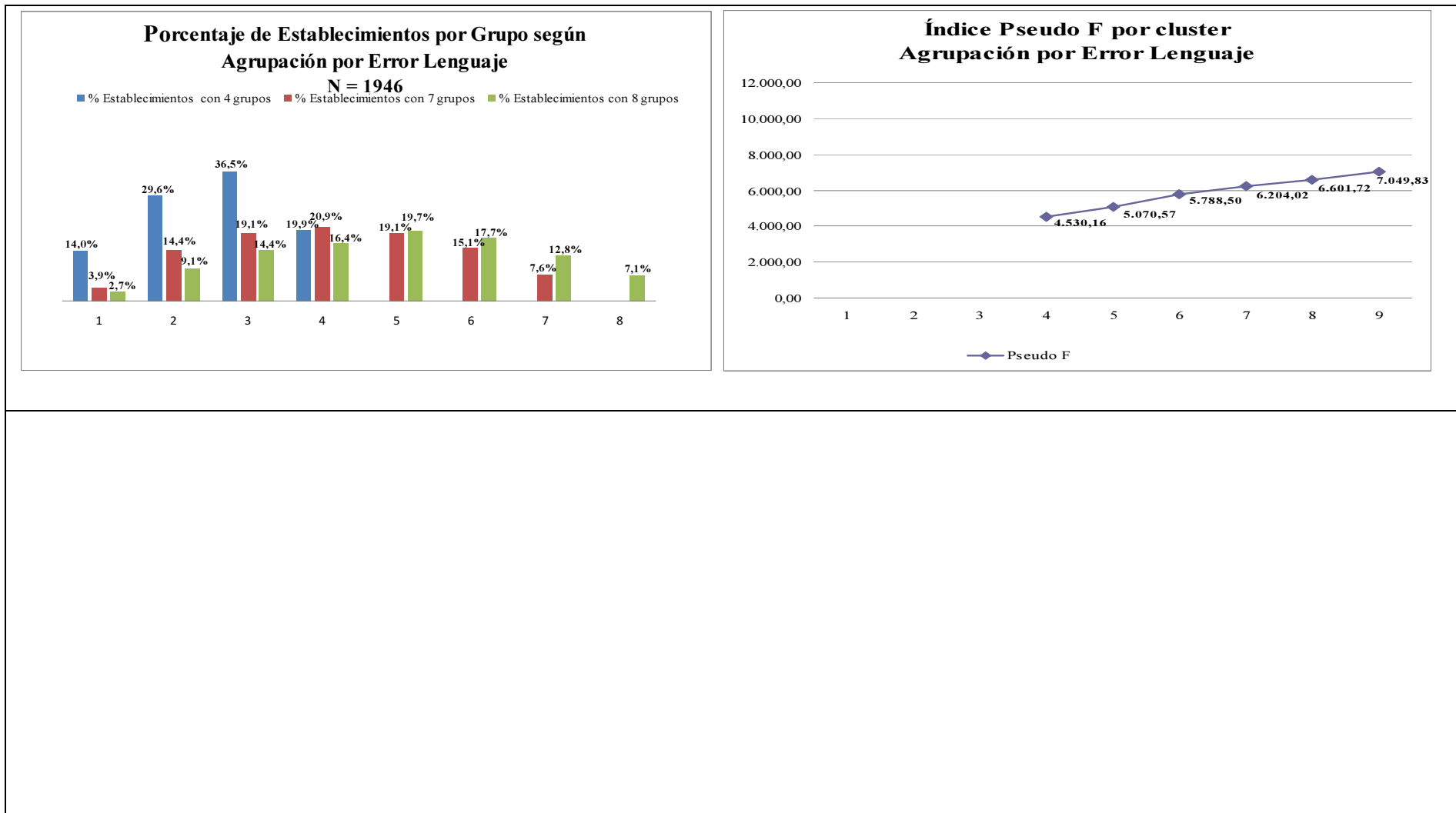
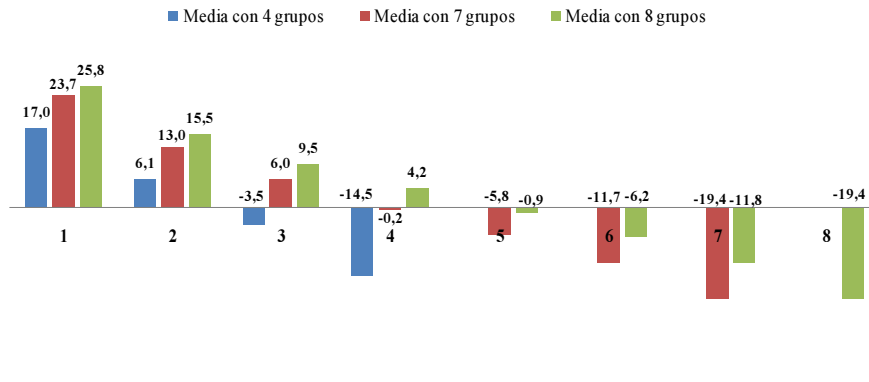


Tabla 3. Resultados de agrupar según el promedio del Error de Estimación del Puntaje SIMCE Lenguaje 4to básico de los establecimientos en los años 2006, 2007 y 2008.



Media por Grupo según Agrupación por Error Lenguaje
N = 1946



Varianza por Grupo según Agrupación por Error Lenguaje
N = 1946

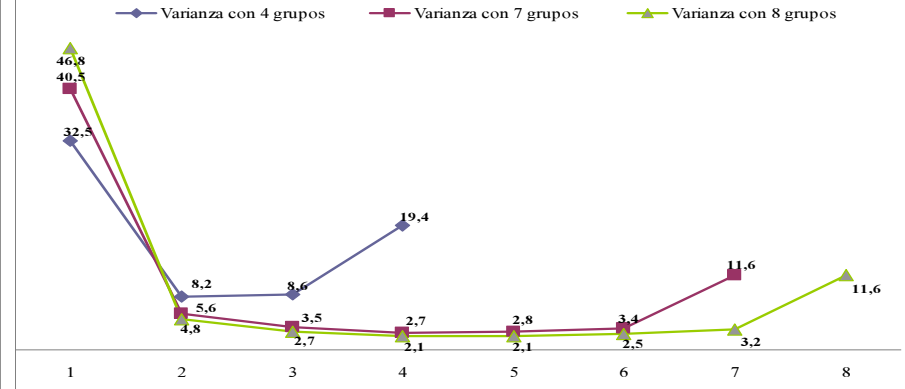


Tabla 4. Resultados de agrupar según el promedio del Error de Estimación del Puntaje SIMCE Matemática 4to básico de los establecimientos en los años 2006, 2007 y 2008.

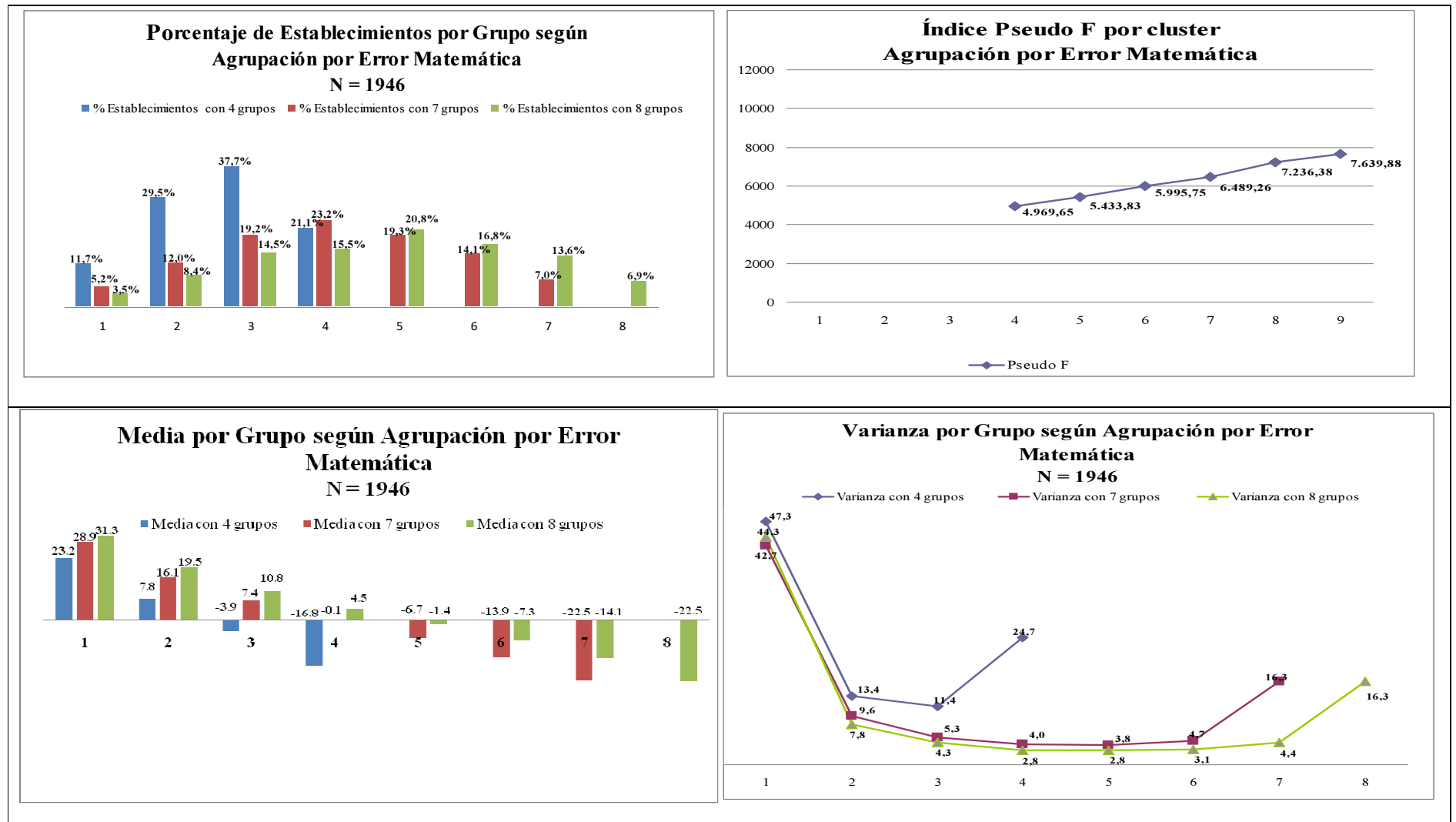


Tabla 5. Resultados de agrupar por el promedio del Porcentaje de alumnos en Nivel de Logro Inicial según el Puntaje SIMCE Lenguaje 4to básico de los establecimientos en los años 2006, 2007 y 2008.

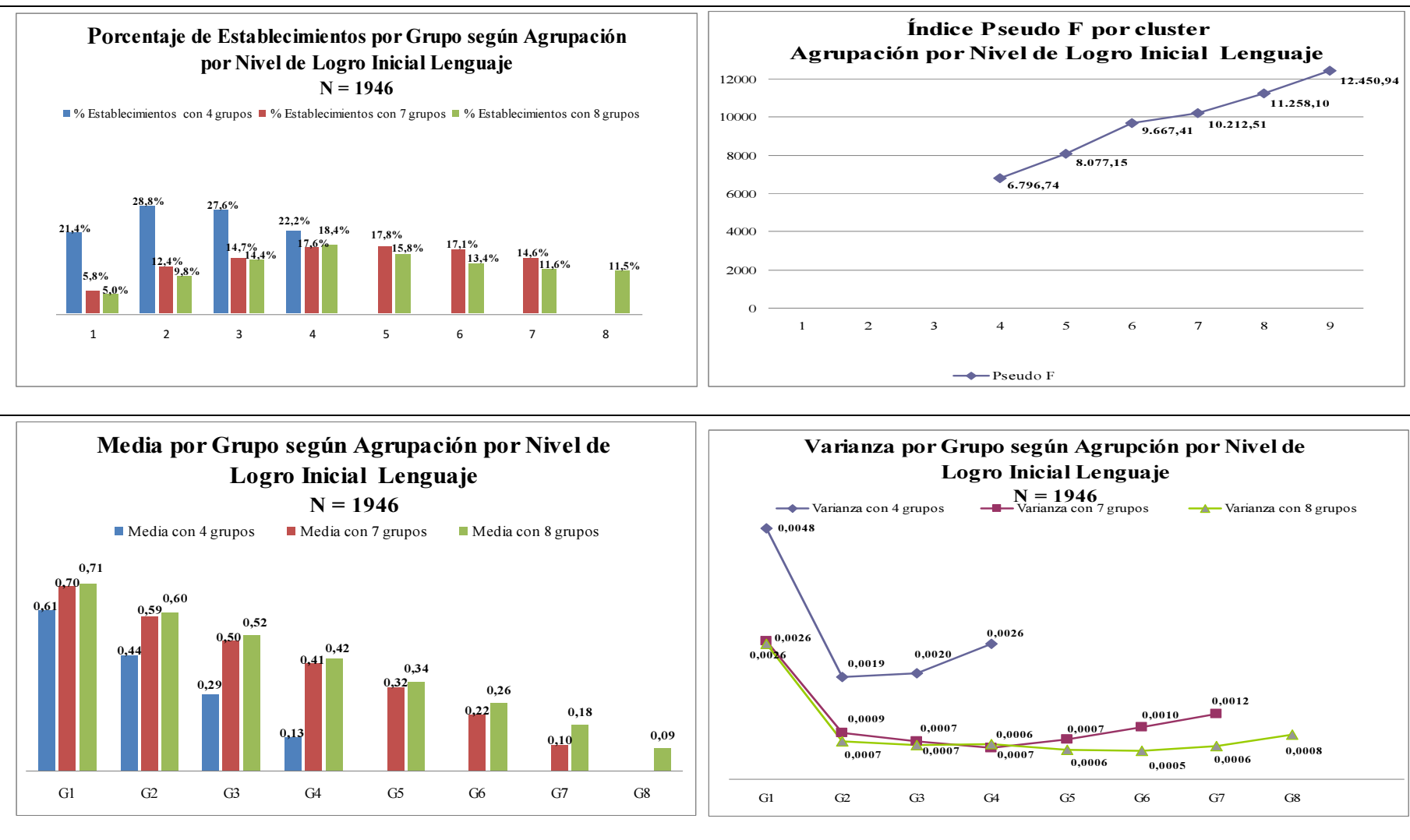
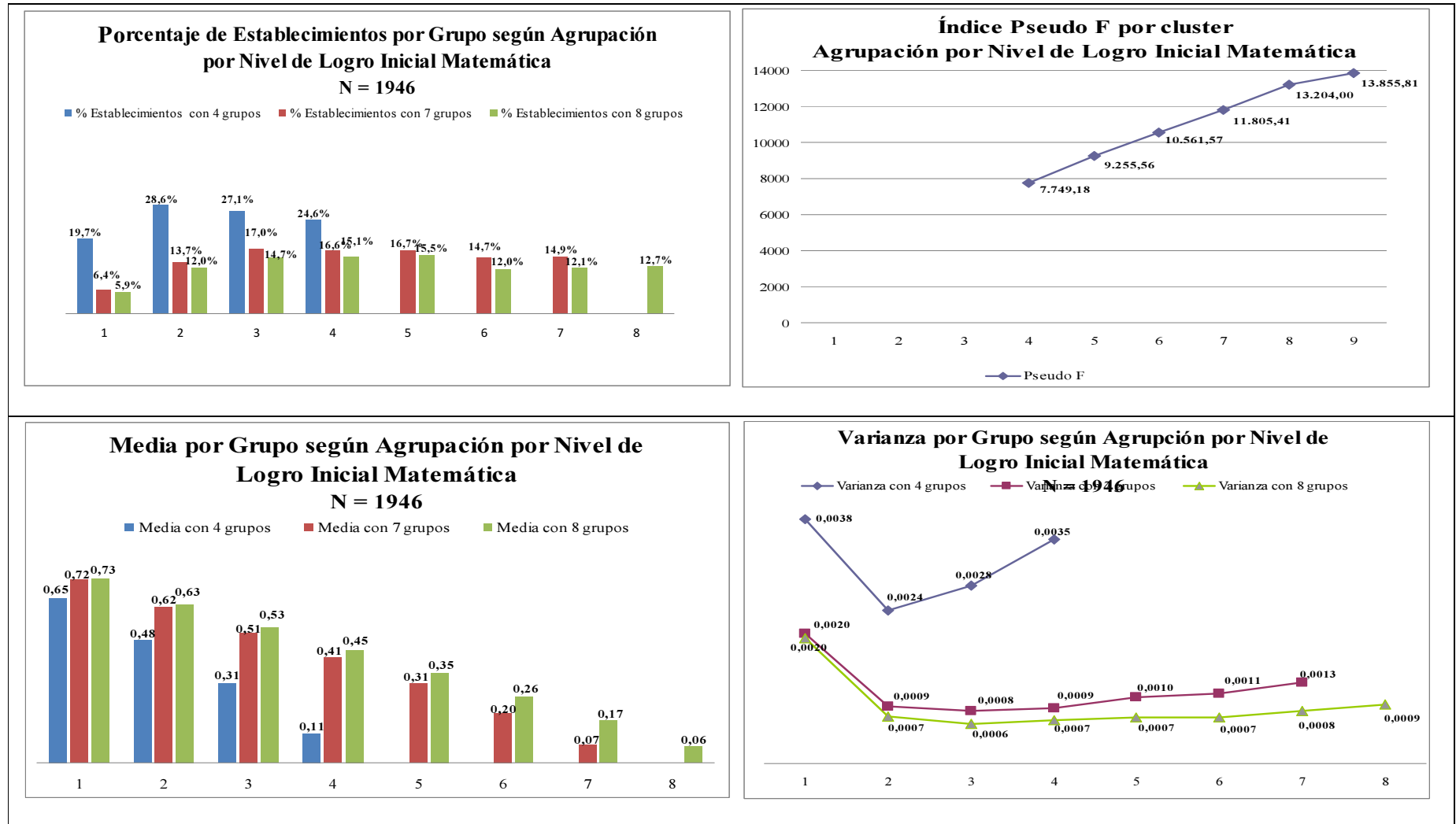


Tabla 6. Resultados de agrupar por el promedio del Porcentaje de alumnos en Nivel de Logro Inicial según el Puntaje SIMCE Matemática 4to básico de los establecimientos en los años 2006, 2007 y 2008.



5. *Discusión*

En este capítulo se expondrán la discusión de los principales resultados de la investigación. Se comenzará comentando las diferencias en las agrupaciones generadas por el uso de datos de distintas pruebas y posteriormente se compararán las distintas formas de clasificar.

Las principales similitudes y diferencias de agrupar según los datos de distintas pruebas se exponen en los siguientes tres puntos. En ellos se evidencia que independiente de la prueba que se utilice para generar la clasificación los resultados son bastante similares. Debido a que los resultados en matemática presentan mayor variabilidad las agrupaciones generadas con dicho puntaje presentan mayor varianza y debido a que los puntajes de matemática son más bajos las medias también difieren.

1) Agrupación por Puntaje SIMCE

En las clasificaciones hechas a partir de la media del puntaje SIMCE de los establecimientos para las pruebas de matemáticas y lenguaje 4to básico de los años 2006, 2007 y 2008 se observa un comportamiento muy similar en términos del porcentaje de establecimientos que componen los grupos, medias de los grupos y varianza. Los grupos elaborados a partir del puntaje en matemática presentan medias más bajas y mayor varianza que los contruidos a partir del puntaje lenguaje.

En ambas pruebas, la gran disminución de las varianzas junto al cambio más moderado en las medias señalan la necesidad de pasar de una clasificación en 4 grupos a una en 7 u 8 grupos. Por otro lado el índice Pseudo F muestra un progresivo aumento en su valor al aumentar los grupos de la clasificación. Todas estas señalan reflejan que la minimización de la varianza al interior de los grupos es mayor que la reducción de la brecha entre la media de un grupo y otro al aumentar los grupos en la clasificación, lo que se deduce del cálculo del pseudo-F. Este hecho se relaciona directamente con la gran diversidad de establecimientos educativos que existen en el país, por lo que una clasificación adecuada a esta naturaleza requiere de una mayor cantidad de grupos.

Por parte parece correcto asumir que dada la variabilidad de los resultados de la prueba de matemática, una clasificación de acuerdo a éstos resultados requiere de una mayor cantidad de grupo que que una clasificación hecha a partir de los resultados de la prueba de lenguaje.

En término prácticos trabajar con 7 u 8 grupos en vez de trabar con 4, significa que un establecimiento A que obtiene 240.48 puntos promedio de los tres años en la prueba SIMCE lenguaje se ubica en el último grupo en una clasificación en 4 categorías, al igual que un establecimiento B que obtiene 212.45 puntos. Pero si la misma clasificación se replica con 7 categorías el establecimiento A se ubicará en el grupo 5 mientras el

establecimiento B lo hará en el grupo 7. Ahora bien, con 8 categorías tanto el establecimiento A como el B se desplazan en un grupo, ubicándose en el grupo 6 y 8 respectivamente.

Con los resultados de la prueba SIMCE matemáticas se observa que al trabajar con 7 u 8 grupos en vez de 4, un establecimiento A que se encuentra en el 4 grupo en una clasificación de 4 con 227.95 puntos, se ubicaría en el grupo 5 en una clasificación de 7 y en el grupo 7 de una clasificación en 8 grupos. Mientras que un establecimiento B que tiene 199.42 puntos y se ubica en el grupo 4 de una clasificación de 4 se ubicará en el grupo 8 en una clasificación de 8 grupos.

2) Agrupación por Eficacia

En las clasificaciones hechas a partir de la media del error de estimación del puntaje SIMCE de los establecimientos para las pruebas de matemáticas y lenguaje 4to básico de los años 2006, 2007 y 2008 se observa un comportamiento muy similar en términos del porcentaje de establecimientos que componen los grupos pero las medias de los grupos y las varianzas experimentan pequeñas diferencias si se comparan los resultados entre ambas pruebas.

Los grupos elaborados con el error de estimación asociado a la prueba de matemática presentan medias más extremas, es decir las medias positivas son mayores y las negativas menores y las varianzas de todos los grupos, incluido el grupo 1, son menores al aumentar la cantidad de grupos en la clasificación. Además se observa que las varianzas de los grupos del centro de las clasificaciones en 7 y 8 grupos son muy próximas, esto quiere decir que la disminución de la varianza al aumentar en un grupo la clasificación es leve en estos grupos pero marcada en los grupos extremos. Esto puede deberse a que el ajuste del modelo en matemática es algo menor que en lenguaje.

El índice Pseudo F para el error de estimación asociado a ambas pruebas aumenta al aumentar la cantidad de grupos en la clasificación. Este hecho junto con la disminución de la varianza, nuevamente señalaría la necesidad de aumentar los grupos de las clasificaciones para dar cuenta de la verdadera diversidad de escuelas existentes en términos de eficacia.

A modo de ejemplo es posible señalar que un establecimiento A que obtiene 274.8 puntos como promedio de la prueba SIMCE lenguaje entre los años 2006, 2007 y 2008 se ubica en el grupo 2 de 4 grupos en la clasificación agrupando según puntaje SIMCE lenguaje, pero si la clasificación en 4 grupos se realizara agrupando por error lenguaje el establecimiento A se ubicará en el grupo 4 de 4.

Así también en una clasificación elaborada según el error de estimación asociado a la prueba SIMCE matemática, un establecimiento A que tiene un error de estimación de -18.43 se ubica en el grupo 4 de una clasificación en 4 grupos, al igual que un establecimiento B tiene un error de estimación de -29.3. Si la clasificación se realiza con 7 grupos el establecimiento A se ubicará en el grupo 7 al igual que el establecimiento B, pero si la clasificación se realiza en 8 grupos el establecimiento A permanecerá en el grupo 7 mientras el establecimiento B se ubicará en el grupo 8.

Ahora bien, si se compara la agrupación elaborada con el puntaje de matemática con la agrupación realizada con el error de estimación de matemática se observa que un establecimiento C que obtuvo 265.38 puntos como promedio de la prueba matemáticas en los 3 años analizados se ubica en el grupo 2 de una clasificación en 4 grupos, mientras el mismo establecimiento C en la agrupación realizada mediante el error se ubica en el grupo 4. Realizar la misma comparación con 8 grupos muestra que el establecimiento C agrupando según el puntaje se ubica en el grupo 4 y en la agrupación según error el mismo establecimiento se ubica en el grupo 6.

3) Agrupación por Nivel de Logro Inicial

En las clasificaciones hechas a partir de la media de la proporción de alumnos por establecimientos en el nivel de logro inicial según los resultados SIMCE para las pruebas de matemáticas y lenguaje 4to básico de los años 2006, 2007 y 2008 se observa que en ambas pruebas existe un comportamiento muy similar y homogéneo en el porcentaje de establecimientos que componen los grupos, pero las medias de los grupos conformados según el nivel de logro de la prueba de matemáticas son mayores que las medias conformadas según el nivel de logro de la prueba de lenguaje.

Por otra parte se observa un comportamiento de las varianzas distinto entre pruebas, ya que según los resultados en la prueba de lenguaje las varianzas del grupo 4 son mayores al aumentar la cantidad de grupos de 7 a 8 y son similares en el grupo 1 y 3, mientras que las varianzas de los grupos elaborados según los resultados de la prueba de matemática son similares son en el grupo 1 de las clasificaciones en 7 y 8 grupos.

El índice Pseudo F de ambas pruebas aumenta al aumentar la cantidad de grupos, por lo que nuevamente el índice da cuenta de la necesidad de aumentar la cantidad de grupos en la clasificación.

En términos prácticos es posible señalar que la clasificación en 4 grupos según el porcentaje promedio de alumnos en Nivel de Logro Inicial para SIMCE Lenguaje señala que un establecimiento A que concentra 56% de alumnos en dicho nivel de logro se ubica

en el grupo 1. Ahora bien, ese mismo establecimiento en una clasificación en 7 grupos se ubica en el grupo 2 al igual que en una clasificación en 8 grupos. Por otro lado, un establecimiento B que concentra el 71% de estudiantes en el nivel de logro inicial y se ubica en el grupo 1 de una clasificación en 4 grupos permanece en el grupo 1 aunque la clasificación se realice con 7 u 8 grupos.

Con los resultados de matemáticas, usar una clasificación con una mayor cantidad de grupos significa que un establecimiento A con 64% de alumnos en el nivel de logro inicial se ubica en el grupo 1 en una clasificación de 4 grupos, pero si la clasificación se realiza en 7 grupos el mismo establecimiento A se ubicará en el grupo 2 al igual que si la clasificación se realiza con 8 grupos. Sin embargo, un establecimiento B con 73% de alumnos en el nivel de logro inicial permanece en el grupo 1 independiente de si la clasificación se realiza con 4, 7 u 8 grupos.

Al comparar las distintas formas de agrupar se observa que todos los grupos presentan estructuras similares, con diferencias en el comportamiento de las medias y varianza al aumentar la cantidad de grupos. El patrón común a todas las clasificaciones es que los grupos con mayor varianza son los grupos que concentran el menor porcentaje de establecimientos, lo que indica la presencia de outliers en los grupos extremos de todas las clasificaciones.

De igual forma es válido para las agrupaciones por eficacia y nivel de logro que al aumentar la cantidad de grupos de 4 a 7 u 8 el grupo que menor porcentaje de establecimientos concentra es el grupo 1. En cambio para las agrupaciones por puntaje SIMCE el grupo que concentra el menor porcentaje de establecimientos al aumentar la clasificación a 7 u 8 grupos es el último grupo de la clasificación. Así también, los grupos con mayor porcentaje de establecimientos coinciden en las agrupaciones por puntaje SIMCE y eficacia pero no en las agrupaciones por nivel de logro, en esta última agrupación se observa que la distribución de los establecimientos al interior de los grupos son mucho más homogéneas que en el resto de las agrupaciones.

Los resultados de la prueba de matemáticas respecto a la de lenguaje genera diferencias en todas las agrupaciones, ya que las medias resultas más elevadas o extremas y en general se observa mayor variabilidad. Nuevamente, eso sugiere que empíricamente se requieren más grupos en matemática que en lenguaje.

En términos de varianza se observan estructuras similares a distintas escalas -dependiendo de la naturaleza de la variable de agrupación- salvo en el caso de las varianzas de las clasificaciones agrupadas según nivel de logro inicial. En este último caso se observa que las varianzas de algunos grupos de la clasificación en 7 y 8 grupos se traslapan.

Los índices Pseudo F de todas las agrupaciones señalan que las clasificaciones con mayor número de grupos son mejores. Este hecho se relaciona con la gran disminución que experimentan las varianzas de los grupos al pasar de 4 a 7 u 8 grupos. De esta forma una adecuada clasificación de las escuelas de Chile requieren de un mayor número de grupos para capturar y dar cuenta de la diversidad de necesidades propia de cada realidad.

Conclusiones y recomendaciones de política

El estudio arroja dos conclusiones centrales. La primera de ellas, que en la clasificación de escuelas por resultados se presentan una serie de alternativas. Claramente, los puntajes brutos de rendimiento en una prueba reflejan una mezcla de la composición socioeconómica del alumnado, los recursos de la escuela y, en última instancia, la calidad de la enseñanza de las escuelas. Por lo tanto, la decisión de clasificar escuelas por resultados brutos en las pruebas sería una forma de premiar simbólicamente la segregación social. La segunda conclusión es que, de acuerdo a los resultados de los análisis, tenemos una inmensa variabilidad en el comportamiento de nuestras escuelas, lo que hace difícil agruparlas en grupos homogéneos. Incluso tomando en cuenta que este estudio restringió la muestra a colegios con más de treinta estudiantes evaluados en los tres años bajo análisis. En total, se consideraron solamente 1946 colegios. Al intentar incluir establecimientos más pequeños, sin duda la clasificación se hará más compleja.

Finalmente, es preciso realizar estudios más profundos respecto de la validez y confiabilidad de las clasificaciones, lo que debiera considerarse como parte de la implementación de la ley de aseguramiento de la calidad, una vez que esta se apruebe.

ANEXO

La tabla 2 muestra las variables utilizadas para la elaboración del modelo en cada nivel. El modelo se calcula para las pruebas SIMCE de matemáticas, lenguaje y para el puntaje promedio de ambas pruebas entre los años 2006, 2007 y 2008.

Tabla 2. Variables utilizadas en el modelo jerárquico lineal en dos niveles.

Variable Nivel Estudiante	Definición	Métrica y Centrado
Nivel socioeconómico del estudiante	Índice compuesto por la educación de los padres y el ingreso del hogar.	Índice con media cero y desviación estándar 1. Centrado en la media del grupo.
Gastos en educación	Índice compuesto por la cantidad de libros en el hogar, mensualidad.	Índice con media cero y desviación estándar 1. Centrado en la media del grupo.
Género	Identificación del género del estudiante.	Variable dicotómica que toma valor 1 para los hombres y 0 para las mujeres. Se utiliza en los modelos con su métrica original.
Baja expectativas de educación del estudiante	Identificación de las expectativas que los padres tienen sobre la educación que alcanzará su hijo.	Variable dicotómica que toma valor 1 si el alumno no termina la educación media o termina la educación media en TP y 0 en otro caso. Se utiliza en los modelos con su métrica original.
Mediana expectativa de educación del estudiante	Identificación de las expectativas que los padres tienen sobre la educación que alcanzará su hijo.	Variable dicotómica que toma valor 1 si el alumno completa la educación media humanista o una carrera en IP y 0 en otro caso. Se utiliza en los modelos con su métrica original.
Alta expectativa de educación del	Identificación de las expectativas que los padres tienen sobre la educación	Variable dicotómica que toma valor 1 si el alumno completa una carrera universitaria o

estudiante	que alcanzará su hijo.	un postgrado en la universidad y 0 en otro caso. Se utiliza en los modelos con su métrica original.
Variabes Nivel Escuela	Definición	Métrica y Centrado
Nivel socioeconómico del establecimiento	Promedio por establecimiento del índice compuesto por la educación de los padres y el ingreso del hogar.	Índice con media cero y desviación estándar 1. Centrado en la gran media.
Escuela subvencionada	Identificación de escuelas subvencionadas.	Variable dicotómica que toma valor 1 para las escuelas subvencionadas y 0 para otro tipo de establecimiento. Se utiliza en los modelos con su métrica original.
Escuela particular pagada	Identificación de escuelas particulares pagadas.	Variable dicotómica que toma valor 1 para las escuelas particular pagadas y 0 para otro tipo de establecimiento. Se utiliza en los modelos con su métrica original.
Selectividad	Índice compuesto por los requisitos de selectividad de los establecimientos.	Índice con media cero y desviación estándar 1. Centrado en la gran media.
Título de profesor	Índice compuesto por las distintas formas de obtener el título de profesor.	Índice con media cero y desviación estándar 1. Centrado en la gran media.
Especialización del profesor	Identificación de aquellos profesores que tienen especialización.	Variable dicotómica que toma valor 1 para aquellos establecimientos que cuentan con profesores con diplomado o magíster 0 en otro caso. Se utiliza en los modelos con su métrica original.

La elaboración del modelo se detalla a continuación, comenzando por un análisis de gradiente y posteriormente se realiza un análisis de efectos composicionales para finalizar con el modelo seleccionado.

Del análisis de gradientes expuesto en la tabla 3 se observa que el puntaje que obtienen los alumnos está estrechamente relacionado con su NSE, el cual explica en gran medida la

diferencia de resultados entre los establecimientos educativos. De esta forma, aunque un pequeño porcentaje de la variación del puntaje en el nivel de estudiante se explica por el NSE del alumno, dicha variable explica entre el 44 y el 53% de la variación del puntaje entre escuelas.

Tabla 3. Gradientes Socioeconómicos

Coefficientes y varianzas	MAT 06	MAT 07	MAT 08	LENG 06	LENG 07	LENG 08	PROM 06	PROM 07	PROM 08
Constante	252.65	250.27	250.63	257.07	258.16	263.50	254.86	254.21	257.05
NSE alumno	16.34	14.45	14.56	15.11	13.59	13.17	15.57	13.85	13.7
% varianza explicada nivel 1 (estudiantes)	3.60%	2.66%	2.81%	2.96%	2.31%	2.03%	3.86%	2.89%	2.82%
% varianza explicada nivel 2 (escuela)	52.33%	44.65%	45.59%	53.19%	47.92%	46.56%	53.56%	46.59%	46.66%

Variables significativas al 0.05%

La tabla 4 contiene el modelo de efectos composicionales en el cual se evidencia el fuerte impacto del NSE del establecimiento en el puntaje que obtienen los alumnos, ya que al incluir en el modelo el NSE del establecimiento no sólo aumenta el porcentaje de varianza explicada en el nivel de escuela, sino también es posible constatar que la magnitud del coeficiente asociado a esta variable es mucho mayor que el asociado al NSE del estudiante. Si bien, tanto el NSE del estudiante como de la escuela son significativos para explicar el puntaje, la incidencia del NSE del establecimiento es notoriamente superior. Para ejemplificar esto es posible señalar que un estudiante cuyo NSE está una desviación estándar por debajo del NSE promedio obtiene entre 12 y 15 puntos menos en la prueba de matemáticas y entre 11 y 14 puntos menos en la prueba de lenguaje que estudiantes con NSE igual al promedio de Chile. Así también la inclusión del NSE del establecimiento produce un aumento en el porcentaje explicado de la varianza de los puntajes entre establecimientos, la cual es fuertemente superior al porcentaje explicado de varianza a nivel de alumno. El NSE del establecimiento explica entre un 61 y un 64% de la variación del puntaje de la prueba de matemáticas entre escuelas, y entre un 60 y un 64% de dicha variación para el puntaje de la prueba de lenguaje.

Tabla 4. Efectos Composicionales del NSE sobre el Puntaje

Coefficientes y varianzas	MAT 06	MAT 07	MAT 08	LENG 06	LENG 07	LENG 08	PROM 06	PROM 07	PROM 08
Constante	252.22	249.82	249.88	256.65	257.72	262.81	254.43	253.77	256.35
NSE alumno	14.82	12.78	12.98	13.52	11.78	11.40	14.17	12.28	12.19
NSE escuela	28.54	30.17	30.14	25.3	26.17	24.93	26.91	28.16	27.54

% varianza explicada nivel 1 (estudiantes)	3.63%	2.70%	2.86%	3.00%	2.36%	2.08%	3.89%	2.93%	2.87%
% varianza explicada nivel 2 (escuela)	64.20%	61.53%	62.38%	63.66%	62.52%	60.04%	65.29%	63.05%	62.59%

Variables significativas al 0.05%

Finalmente se presenta la tabla 5 que contiene en detalle la especificación del modelo para cada prueba y año, así como para el puntaje promediado de ambas pruebas. El modelo elegido considera variables teóricamente relevantes para explicar el puntaje a nivel estudiante y escuela y, a su vez dichas variables son las mismas independiente de la prueba y el año a modelar. De esta forma se presenta un modelo estable y replicable en el tiempo.

En todos los modelos queda en evidencia el alto porcentaje de varianza explicada a nivel de escuela, entre un 66 y un 69% en la prueba de matemáticas y entre 67 y un 70% en la prueba de lenguaje. A nivel de estudiante el porcentaje de varianza explicada sólo alcanza un 7% en la prueba de matemáticas y un 6% en la prueba de lenguaje.

En el nivel de escuela las variables NSE del establecimiento y selectividad son significativas al 5% para todas las pruebas y años, dejando en evidencia el alto coeficiente asociado al NSE del establecimiento. Si bien el coeficiente asociado a la especialización del profesor es pequeño, es significativo al 5% para todos los años de la prueba de matemáticas y en lenguaje lo es al 5 y al 10%. La variable título del profesor es significativa al 5% sólo en matemáticas 2006 y al 10% en el puntaje promediado de matemáticas y lenguaje el mismo año. A pesar de ser una variable relevante en la teoría, esta variable parece no ser significativa en todas las pruebas y año por que tan sólo entre un 0,6 y un 1,5% de los profesores² –en los tres años considerados- no tienen título.

La variable dependencia administrativa presente a través de las dummies particular pagado y particular subvencionado resulta significativa sólo en este último caso para los años 2007 y 2008. Es muy probable que esta situación se deba a que el resto de las variables introducidas en el modelo han logrado capturar las diferencias que existen entre estos tipo de establecimiento.

² Ver anexo.

Tabla 5. Especificación del Modelo

Coefficientes del modelo multinivel	PROM06	PROM07	PROM08	LENG 06	LENG 07	LENG 08	MAT 06	MAT 07	MAT 08
<i>Nivel escuela</i>									
Especialización del profesor	1.89	1.57	1.69	1.61	1.18	1.37	2.16	1.95	1.99
Título del profesor	5.65	2.67	4.78	3.70	1.74	2.34	7.60	3.60	4.90
Selectividad	5.19	4.52	4.78	5.03	4.26	4.24	5.36	4.78	5.33
Particular Pagado	-2.27	-1.68	0.91	-1.80	-1.12	0.08	-2.77	-2.28	1.72
Particula Subvencionado	0.83	3.24	4.96	1.00	3.11	4.68	0.65	3.35	5.24
NSE	17.00	18.48	18.00	15.55	16.77	16.08	18.46	20.21	19.91
<i>Nivel estudiante</i>									
Alta expectativas de educación del estudiante	31.07	31.69	27.21	27.98	29.01	25.36	34.21	34.42	29.12
Mediana expectativas de educación del estudiante	14.42	14.82	13.40	11.65	12.60	11.78	17.21	17.07	15.05
Baja expectativas de educación del estudiante	10.59	10.50	8.20	7.89	8.48	6.39	13.30	12.54	10.01
Hombre	-0.18	-0.55	-2.42	-6.52	-6.42	-10.93	6.14	5.31	6.09
Gastos en educación	0.63	0.84	3.57	1.02	1.16	3.44	0.24	0.51	3.69
NSE	9.21	7.63	5.35	8.73	7.33	4.83	9.67	7.93	5.86
Intercepto	225.33	226.28	230.66	235.50	236.57	244.74	215.12	215.96	216.64
Significativa al 5%									
Significativa al 10%									

Parte Aleatoria

Varianza por Niveles

Nivel 2	186.37	225.15	228.94	172.95	195.22	204.10	225.90	281.62	283.51
Nivel 1	1817.69	1841.42	1775.16	2167.96	2120.76	2141.99	2137.43	2169.78	2021.42

Varianza Explicada Modelo

Nivel 2	70.32	68.37	68.05	69.65	68.51	66.69	68.52	66.35	67.07
Nivel 1	7.08	6.31	5.72	5.99	5.57	5.59	6.88	6.02	5.69

Las variables presentes en el nivel de estudiante son en su gran mayoría significativas. De esta forma el NSE del estudiante, los gastos en educación en que incurren los apoderados, el sexo de los estudiantes y las expectativas de los padres contribuyen de buena forma a explicar el puntaje alcanzado por el alumno. Respecto a esta última variable mencionada, existe un fuerte discusión acerca de la supuesta endogeneidad de las expectativas de los

padres sobre el nivel de estudios que alcanzarán sus hijos, en esta investigación se supondrá que las calificaciones que obtienen estudiantes de cuarto básico no son determinantes en las expectativas de los padres pues en ese curso aún no se ha cursado la mitad de la enseñanza que es obligatoria por ley.

Finalmente, y una vez que el modelo ha sido estimado, se calcula el error asociado al nivel de escuela para cada prueba y se guarda como variable de agrupación.

BIBLIOGRAFÍA

- Affifi, A., y Clark, V. (1997). Computer-aided multivariate analysis. Third Ed. Chapman & Hall/CRC, New York.
- Chay, K., McEwan, P., Urquiola, M. (2005) *The central Role of Noise in Evaluating Interventions That Use Test Scores to Rank Schools*. American Economic Review, 95 (4), 1237-1258.
- Gallego, F. y Hernando, A. (2009) *School Choice in Chile: Looking at the Demand Side*. Documento de trabajo 356, Instituto de Economía. Pontificia Universidad Católica de Chile
- Goldschmidt, P. y Choi, K. (2007) *The Practical Benefits of Growth Models for Accountability and the Limitations Under NCLB*. Página de la Universidad de UCLA <http://www.cse.ucla.edu/products/policy.html>
- Hallgarten, J. (2001). School league tables: have they outlived their usefulness? New Economy, 8 (4), 189-196.
- Hanushek, E. y Rivkin, S. (2010) *Generalizations about Using Value-Added Measures of Teacher*. American Economic Review (saldrá en Mayo en esta revista pero yo lo saqué de la página web)
- Hanushek, E. (2004) *United States Lessons about School Accountability*. (lo saqué de la página web)
- Hirschman, A. (1970) *Exit, Voice and Loyalty: Responses to Decline in Firms, Organizations and States*. Cambridge, MA: Harvard University Press.
- Jain, R. y Koronios, A. (2008) *Innovation in the cluster validating techniques*. Springer+Business Media
- Kane, T., Staiger, D. y Geppert, J. (2002) *Failing to account for natural fluctuations in test scores could undermine the very idea of holding schools accountable for their efforts – or lack thereof*. www.educationnext.org
- Levacic, R. y Woods P. (2002) *Raising School Performance in the League Tables (Part 1): disentangling the effects of social disadvantage*. British Educational Research Journal, Vol. 28, No. 2
- Mizala, A., Romaguera, P. y Urquiola, M. (2006) *Socioeconomic status or noise? Tradeoffs in the generation of school quality information*. Journal of Development Economics 84 (2007) 61-75.

- Pérez López, C. (2004) *Técnicas de Análisis Multivariante de Datos* PEARSON EDUCACIÓN, S.A: Madrid,2004 (ISBN:84-205-4104-04 MATERIA: Estadística Matemática 519.2)
- Sapelli, C. y Torche, A. (2002) *Determinantes de la Selección de Tipo de Colegio*. Cuadernos de Economía Año 39, N° 117, pp.175-202. Pontificia Universidad Católica de Chile.
- Siwiski, W. (2002) *Perspektywy --Ten Years of Rankings*. Higher Education in Europe,27:4,399 — 406
- Wilson, D. (2004) *Which Ranking? The Impact of a Value-Added Measure of Secondary School Performance*. Public Money & Management January.
- Wilson, D. y Piebalga, A. (2008) *Performance Measures, Ranking and Parental Choice: An Analysis of the English School League Tables*. International Public Management Journal,11:3,344 — 366.