

**UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL**

**EDUCACIÓN Y EL MUNDO RURAL:
"ESTUDIO DE VARIABLES QUE INCIDEN EN LA CALIDAD Y EN LA
ELECCIÓN DE ESTABLECIMIENTOS RURALES PARA EL
DESARROLLO DE UN SISTEMA DE LOCALIZACIÓN"**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL INDUSTRIAL

CONSUELO ISABEL ZÚÑIGA PACHECO

PROFESOR GUÍA
RICARDO SAN MARTÍN ZURITA

MIEMBROS DE LA COMISIÓN
JOSÉ MOSQUERA CÁDIZ
FRANCISCO TUBINO CORTÉS

SANTIAGO DE CHILE
MAYO DE 2009

RESUMEN DE LA MEMORIA PARA OPTAR AL
TÍTULO DE INGENIERA CIVIL INDUSTRIAL
POR: CONSUELO ZÚÑIGA PACHECO
FECHA: 03/07/2009
PROF. GUÍA: RICARDO SAN MARTÍN

EDUCACIÓN Y EL MUNDO RURAL:
"ESTUDIO DE VARIABLES QUE INCIDEN EN LA CALIDAD Y EN LA ELECCIÓN DE ESTABLECIMIENTOS RURALES PARA EL DESARROLLO DE UN SISTEMA DE LOCALIZACIÓN"

Calidad y equidad son el desafío que tiene la educación pública. Cómo alcanzar la calidad y qué es lo que busca la sociedad de la educación son cuestionamientos difíciles de responder. Éstos han sido, y seguirán siendo, temas que movilizan tanto a los agentes del Estado como a la ciudadanía en sí misma. Es en este ámbito que se desarrolla el presente trabajo de título, el cual tiene como principal objetivo determinar las variables que influyen en la calidad de la educación y en la elección del establecimiento por parte de los padres, en el ámbito de la educación rural, con el fin de encontrar componentes estructurales y de calificación de los establecimientos para integrarlas a un modelo de optimización de localización. Además, es en estas zonas donde los establecimientos presentan ineficiencias debido a la baja matrícula que tienen al estar emplazadas en lugares lejanos o de difícil acceso, por lo que medidas que contribuyan a mejorar esta situación son necesarias y de gran impacto para las personas a las que beneficia.

La parte inicial de este trabajo consiste en determinar los factores que han sido estudiados en Chile y el mundo sobre calidad de la educación y elección de establecimientos educacionales. Posteriormente, se evalúa la función de producción de la educación en base al rendimiento del estudiante y las características del establecimiento y su familia a través de los métodos de regresión lineal y lineal jerárquico. En conjunto con variables familiares se logra determinar una serie de características propias de los establecimientos que favorecen el rendimiento con lo que se logra establecer que el establecimiento sí importa, tanto estructural como culturalmente, sobre todo para los grupos socioeconómicos más desposeídos, a pesar de que los modelos obtenidos presentan un bajo ajuste, pero que están dentro del rango de las investigaciones actuales.

Luego, se aplica un modelo logarítmico multinomial para establecer las causas que determinan la toma de decisiones de los padres. Éstos al ser consultados por las razones que los llevaron a elegir el actual establecimiento de sus hijos mencionan principalmente la cercanía, mientras que el resultado en el SIMCE es mencionado por una fracción menor de ellos, en un 67,3% y un 9,8% respectivamente. Al aplicar el modelo se logra esclarecer que, independientemente del grupo socioeconómico, los padres eligen los establecimientos con mejor rendimiento, enfrentándose a un *trade-off* entre cercanía y calidad, siendo importante el segmento que viaja más en pos de una mejor educación, y que prefieren establecimientos con educación secundaria, cursos simples y de tamaño medio.

Mediante la presente investigación se determina que, de un total de 21 características propias de cada establecimiento, las variables matrícula, curso combinado, tamaño del curso, dotación docente, educación secundaria, tiempo de viaje y puntaje SIMCE se pueden incluir en el modelo de optimización de modo de representar el rendimiento y las preferencias de los padres, los cuales se recomienda incluir a través de restricciones que permitan simular distintos escenarios, dado el bajo ajuste del modelo obtenido para el rendimiento. Además, deja abiertas líneas futuras de investigación que permitan medir variables que no son cuantificables actualmente dada su subjetividad, como, por ejemplo, la disciplina del establecimiento, la participación de los padres en la educación de sus hijos, etc., lo que hará posible alcanzar modelos que presenten un mejor ajuste, así como también queda abierta la posibilidad de elaborar nuevas metodologías para el estudio de la calidad de la educación.

Donde hay educación no hay distinción de clases.

Confucio

AGRADECIMIENTOS

Hoy puedo mirar hacia atrás para recordar a todos quienes estuvieron en las distintas etapas de mi vida como estudiante, donde muchos no entendieron nunca lo que es una integral, ni la covarianza o un Garch, pero sí supieron darme todo el apoyo que necesité en cada uno de esos momentos en que uno cree que el mundo se viene abajo, sobre todo al escribir esta memoria.

Agradezco a mis padres, quienes contribuyeron a que no pasara por la vida sin tener una opinión, aunque fuera absolutamente distinta a la suya, por su paciencia, dedicación y por haber estado ahí cada vez que los necesité, a pesar de todo.

Hoy quisiera que estuviera presente la Nani y agradecerle por todo su amor y sus enseñanzas, pero sé que desde donde esté, me ha acompañado todos estos años y, en especial, cuando se trata de trabajar por superar las diferencias sociales y por crear un país más justo.

A mi hermana Ji y a Guille, quienes me dieron su amor incondicional y me acompañaron siempre, a toda hora y en todo lugar.

Agradezco enormemente a quienes participaron del proyecto SIGER, de quienes aprendí muchísimo, recibí sus consejos y respaldo. A los profes Andrés Weintraub, Vladimir Marianov, y en especial a los profes Pedro Donoso y Francisco Martínez.

A mi profesor guía, quien siempre demostró su confianza en mí y me aconsejó en los momentos en que las cosas perdían el rumbo.

A todos mis amigos, a quienes ya traía y mantengo conmigo, y a quienes tuve la suerte de conocer en esta Escuela y que hicieron que mi paso por ella fuera una de las mejores experiencias que he tenido, los que me apoyaron y me alentaron a seguir adelante a pesar de todos los obstáculos, brindándome muchísimo más que su apoyo, sorprendiéndome muchas veces con su presencia, preocupación y cariño. Si me sigo rodeando de gente como ustedes, así de seguro que voy a encontrar lo que ando buscando.

ÍNDICE

INTRODUCCIÓN.....	1
1 OBJETIVOS.....	3
1.1 OBJETIVO GENERAL.....	3
1.2 OBJETIVOS ESPECÍFICOS.....	3
2 ANTECEDENTES GENERALES.....	3
2.1 MOTIVACIÓN DEL ESTUDIO.....	3
2.2 ESTRUCTURA DEL ESTUDIO.....	5
3 MARCO CONCEPTUAL.....	5
3.1 ANÁLISIS DE LA SITUACIÓN ACTUAL DE LA EDUCACIÓN RURAL.....	5
3.1.1 Rendimiento Académico.....	5
3.1.2 Elección de los Padres.....	6
3.1.3 Caracterización de la Oferta y la Demanda Académica Rural.....	7
3.2 MARCO TEÓRICO.....	9
3.2.1 Metodología de Estimación de la Función de Producción de la Educación.....	9
3.2.2 Metodología de Estimación de los Factores de Elección.....	13
4 REVISIÓN BIBLIOGRÁFICA.....	16
4.1 ESTUDIOS SOBRE EL RENDIMIENTO EDUCACIONAL.....	16
4.1.1 Aspectos Importantes de Rendimiento.....	20
4.2 ESTUDIOS SOBRE FACTORES DE ELECCIÓN ACADÉMICA.....	24
4.2.1 Aspectos Importantes de Elección.....	27
5 ESTIMACIÓN DE VARIABLES QUE INFLUYEN EN EL RENDIMIENTO ACADÉMICO.....	27
5.1 VARIABLES Y DESCRIPCIÓN DE LA INFORMACIÓN UTILIZADA.....	28
5.2 REGRESIÓN LINEAL.....	31
5.2.1 Especificación de Modelos.....	31
5.2.2 Resultados Obtenidos.....	32
5.2.3 Segmentaciones estudiadas.....	40
5.3 MODELO LINEAL JERÁRQUICO.....	44
5.3.1 Especificación de Modelos.....	44
5.3.2 Resultados Obtenidos.....	46
5.3.3 Segmentaciones Estudiadas.....	52
5.4 RESUMEN DE RESULTADOS.....	57
6 ESTIMACIÓN DE VARIABLES QUE INFLUYEN EN LA ELECCIÓN DE ESTABLECIMIENTOS EDUCATIVOS.....	59
6.1 VARIABLES Y DESCRIPCIÓN DE LA INFORMACIÓN UTILIZADA.....	60

6.2	ESPECIFICACIÓN DEL MODELO UTILIZADO.....	62
6.3	RESULTADOS OBTENIDOS.....	62
6.3.1	<i>Estimación del modelo de elección</i>	64
6.4	SEGMENTACIONES ESTUDIADAS	67
6.4.1	<i>Segmentación por GSE del estudiante</i>	67
6.5	RESUMEN DE RESULTADOS.....	70
7	ANÁLISIS DE FACTIBILIDAD	71
7.1	VARIABLES DE RENDIMIENTO	71
7.2	VARIABLES DE ELECCIÓN	72
7.3	INCLUSIÓN DE VARIABLES EN EL MODELO DE OPTIMIZACIÓN.....	72
	CONCLUSIONES	79
	BIBLIOGRAFÍA	83
	ANEXOS	87
	ANEXO I: METODOLOGÍA DE ESTIMACIÓN DE VARIABLES	87
	ANEXO II: ESTADÍSTICOS DESCRIPTIVOS RENDIMIENTO	91
	ANEXO III: MATRICES DE CORRELACIONES Y ANÁLISIS FACTORIAL RENDIMIENTO	92
	ANEXO IV: : MATRICES DE CORRELACIONES Y ANÁLISIS FACTORIAL ELECCIÓN	96
	ANEXO V: ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO DE ELECCIÓN	100
	ANEXO VI: ESTADÍSTICOS DE MODELOS DE ELECCIÓN SEGÚN GSE	103

INTRODUCCIÓN

¿Es posible lograr una educación de calidad y con equidad para todos los niños de Chile? Sin duda, gran parte de la ciudadanía y quienes trabajan para mejorar la educación creen que sí es factible, pero probablemente nadie tiene la solución que logre consenso y, menos, que brinde la seguridad de encontrar una fórmula “ideal”. Probablemente, esta solución única no exista, pero lo que es claro, es que muchos de los esfuerzos que se han desarrollado y se están implementando sirven para contribuir al menos con un paso más en la construcción de un camino que lleve al país más cerca de una educación igualitaria para todos sus hijos.

Es por medio de la educación -del latín *educere*: guiar, conducir o de *educare*: formar o instruir- que se persigue que todos los niños y jóvenes tengan igualdad de oportunidades para surgir en el futuro. Sin embargo, lo que no hay que perder de vista es que a través de ella es cómo la sociedad mantiene y transmite sus conocimientos, valores y cultura, afectando mental, emocional, moral y socialmente a las personas. Es por ello que el cómo se entregue esa educación influye en lo que Chile y su pueblo puedan lograr en un futuro. Sin lugar a dudas, el brindar una educación de calidad y con equidad logrará que se desarrollen los talentos y aptitudes de cada uno de los niños de cualquier nación, con lo que los países llegarán, más allá del desarrollo económico, a construir una sociedad justa y con oportunidades para todos, ya sea en la ciudad o en las afueras de ésta.

La siguiente pregunta es casi tan difícil de consensuar como la anterior, ¿puede la ingeniería ser capaz de aportar en un tema social como la educación?

Mediante la presente investigación se pretende acercar las ciencias sociales a las ciencias exactas, a través de modelos matemáticos que buscan encontrar cuáles son los factores que afectan la calidad de la educación rural y qué es lo que buscan o necesitan los padres al elegir un establecimiento educacional emplazado en una zona rural.

El fin último que persigue este trabajo es apoyar el desarrollo de un modelo de optimización que busque acercarse al óptimo social de localización y tamaño de los establecimientos rurales, lo que contribuirá a superar las ineficiencias que se dan a nivel rural sobre la provisión de la educación.

Por lo tanto, el objetivo general de este estudio es determinar las variables que inciden en la calidad y elección de los establecimientos emplazados en zonas rurales para incluirlas en un modelo de optimización de localización, el que se alcanzará a través de los siguientes objetivos específicos:

- Establecer, a través de una revisión bibliográfica, los factores que han sido investigados en Chile y el mundo, sobre los que se cree tienen influencia en los resultados educacionales y en la elección del establecimiento por parte de los padres.
- Especificar qué variables son significativas a nivel de estudiantes y a nivel de establecimientos en la realidad rural chilena para la calidad de la educación y la elección de establecimientos.
- Definir las variables que son posibles de incorporar en el modelo de optimización.

La forma en que se aborda este trabajo es, en primera instancia, a través de la presentación de antecedentes que permitan reflejar la situación actual de la educación rural, tanto en la calidad como en las características de la oferta y la demanda educativa, para luego dar paso a la presentación del marco teórico que sostiene esta investigación mediante distintos métodos econométricos que permitan modelar las características de la calidad de la educación a través del rendimiento de los estudiantes de establecimientos rurales y los factores que determinan la elección por parte de los padres.

La siguiente parte del trabajo de título está dedicada a una extensa revisión bibliográfica sobre calidad de la educación y elección de establecimientos, con lo que se pretende encontrar las variables que se han investigado en Chile y el mundo que hayan mostrado una evidencia significativa de su influencia en cada uno de los temas a estudiar, con el fin de identificar variables a estudiar en la realidad rural chilena.

Luego se presenta la sección correspondiente al estudio de la calidad de la enseñanza, que se basa en el análisis del rendimiento académico a través de los resultados para 2006 del SIMCE, Sistema de Medición de la Calidad de la Educación, que es la herramienta nacional de medición de resultados de aprendizaje del Ministerio de Educación, Mineduc. Este análisis se realiza a través de dos modelos econométricos: regresión lineal multivariada y modelo lineal jerárquico.

La regresión lineal multivariada ha sido de amplia utilización en la investigación de la función de producción de la educación, pero cuenta con una serie de supuestos que hacen que muchas veces se contradigan con lo que sucede en la realidad.

El modelo lineal jerárquico permite realizar una exploración en distintos niveles -los que en este caso son estudiante y establecimiento- lo cual permite aislar el efecto directo que tiene el establecimiento en el rendimiento de sus estudiantes, especificando sus variables significativas y el porcentaje de varianza que el establecimiento es capaz de explicar. Dicho modelo también es capaz de explicar la interacción y el comportamiento que tienen las variables dentro de cada nivel y entre niveles, es decir, la interacción y comportamiento de las características del estudiante, establecimiento y de este último con las características del estudiante.

El siguiente apartado abarca el estudio desde los intereses que tiene la demanda por educación rural, donde se utiliza una regresión logarítmica multinomial, modelo econométrico que permite analizar las variables que toman en cuenta los consumidores -en este caso padres y apoderados- al momento de elegir frente a varias opciones. Las distintas alternativas que tienen los padres son establecimientos de diferentes categorías, las cuales son determinadas a través de un análisis de conglomerados con k-medias, el que permite segmentar la oferta académica en cinco tipos de establecimientos dadas sus características socioeconómicas, recursos y localización.

Posteriormente, se presenta un análisis de factibilidad para incluir dentro del modelo de optimización los factores que hayan sido encontrados como relevantes a lo largo de esta investigación en cuanto a las características de los establecimientos que influyen en el rendimiento y a las que son valoradas por los padres para elegir un establecimiento educacional, lo cual se lleva a cabo a través de la incorporación de ellas en la función objetivo del problema de optimización o como políticas que se pueden adoptar a través de restricciones del modelo.

Finalmente se presentan las conclusiones acerca de los resultados obtenidos mediante el presente trabajo de título.

1 OBJETIVOS

1.1 Objetivo General

Determinar las variables que inciden en la calidad y elección de los establecimientos emplazados en zonas rurales para incluirlas en un modelo de optimización de localización.

1.2 Objetivos Específicos

- Establecer, a través de una revisión bibliográfica, los factores que han sido investigados en Chile y el mundo, sobre los que se cree tienen influencia en los resultados educacionales y en la elección del establecimiento por parte de los padres.
- Especificar qué variables son significativas a nivel de estudiantes y a nivel de establecimientos en la realidad rural chilena para la calidad de la educación y la elección de establecimientos.
- Definir las variables que son posibles de incorporar en el modelo de optimización.

2 ANTECEDENTES GENERALES

2.1 Motivación del Estudio

Durante mucho tiempo la discusión política del país se ha centrado en la calidad de la educación, desarrollándose diversos movimientos, tanto estudiantiles como políticos, que buscan mejorar la calidad y la equidad de la educación para Chile.

En las políticas públicas impulsadas por los distintos gobiernos democráticos el objetivo principal ha sido el desarrollar proyectos que aporten al mejoramiento del sistema educativo y a la equidad de éste, con lo que el Ministerio de Educación, Mineduc, busca constantemente diversas formas de alcanzar este objetivo.

Con este afán es que surge la idea de diseñar e implementar un Sistema de Información Geográfica para la Educación Rural, SIGER, como un apoyo a futuras políticas elaboradas por el Mineduc, por parte del Instituto Milenio de Sistemas Complejos de Ingeniería y la Universidad de Chile, el cual está integrado por un sistema de información geográfico y un modelo de optimización de la localización de establecimientos educacionales rurales en Chile.

El desarrollo de este sistema busca generar una herramienta que permita evaluar la localización actual de los establecimientos en el área rural y analizar las posibles modificaciones de la oferta actual, mostrando cuáles establecimientos cerrar, modificar y/o donde sería necesario abrir nuevos establecimientos, ya que éstos presentan una serie de características que generan ineficiencias posibles de corregir, con lo cual se apunta a mejorar tanto la gestión como la calidad brindada a los alumnos que asisten a los establecimientos emplazados en las zonas rurales.

En la actualidad los establecimientos educacionales ubicados en zonas rurales son 4.426¹, los cuales representan el 37,6% del total de establecimientos país. Durante el año 2007, asistieron

¹ Fuente: Tabla Directorio 2007 y Matrículas 2007.

a establecimientos rurales 348.306 estudiantes, lo que representa al 9,32% del total de alumnos que pertenece al sistema educativo².

Dichas escuelas, en general, atienden a pocos alumnos y se caracterizan por la pérdida de ellos, muchas veces por la lejanía que presentan con respecto al hogar del estudiante; por una relación alumnos por profesor extremadamente baja, donde además es común que haya sólo un profesor a cargo, lo cual entrega menos opciones que un establecimiento con mayor número de docentes, pues la calidad de la educación impartida dependerá fuertemente de lo que pueda hacer este único docente.

Por otra parte, existen características adicionales de estos establecimientos rurales que los diferencian del sector urbano, específicamente, la existencia de los cursos combinados o multigrado, es decir, cursos de 2 hasta 6 grados simultáneamente.

Los problemas descritos afectan principalmente al sector municipal, ya que éste se hace cargo del 77,9% de los establecimientos educacionales rurales, los que atienden al 71,65% de la matrícula en esta área geográfica.

Para eliminar o disminuir las ineficiencias dadas en las zonas rurales se hace necesario realizar intervenciones como las que se obtendrán a través del SIGER, las que permitirán optimizar el sistema educativo. Sin embargo, las intervenciones deben surgir de un proceso de planificación de calidad, que permita considerar las características específicas del sector rural y los objetivos del Estado acerca del sistema educativo.

Los antecedentes previos muestran la necesidad de diseñar y aplicar un modelo de optimización geográfica que defina el nivel y tipo de intervenciones en cada localidad rural del país, respecto al tamaño, ubicación y acceso a los establecimientos. Sin embargo, no basta con determinar ubicaciones y rutas óptimas, ya que intervenciones como éstas, necesariamente tienen impacto en la forma de vida de las personas, siendo importante también el preguntarse qué factores valoran los habitantes de las zonas rurales, qué necesitan de la oferta educativa y cómo contribuir a mejorar la calidad de la enseñanza.

Por todo lo anterior, se hace necesario que dentro del desarrollo del modelo de optimización se tomen en cuenta las necesidades y preferencias de los habitantes del mundo rural, incorporando también las variables que influyen en la calidad de la educación rural y cuáles son los factores importantes para los padres en cuanto a la elección del establecimiento donde se educarán sus hijos con el fin de incluirlos en el modelo.

Mediante la presente investigación, se abordará el ámbito de la calidad de la educación, medida a través del rendimiento de los estudiantes en el SIMCE, Sistema de Medición de la Calidad de la Enseñanza, y los factores que son importantes, según los padres y apoderados, para la elección de un establecimiento educacional. Con esto se busca apoyar el desarrollo de un modelo que se acerque al óptimo social de localización y tamaño de los establecimientos rurales, lo que contribuirá a superar las ineficiencias que se dan a nivel rural sobre la provisión de la educación.

² Fuente: Tabla Matrículas 2007, Mineduc.

2.2 Estructura del Estudio

Este trabajo de título está compuesto por 3 grandes áreas con las que se busca llegar a una respuesta sobre qué variables deben ser incluidas en el modelo de optimización del SIGER.

La primera etapa consiste en investigar acerca de las variables que han sido estudiadas y encontradas como relevantes en Chile y el mundo en los temas de rendimiento académico y las preferencias de los padres y apoderados.

Se consultan publicaciones referentes a la educación en general, con especial énfasis en los estudios realizados en la población rural, desarrolladas a nivel nacional e internacional, respecto a ambos temas, de modo de visualizar el estado del arte en cuanto a lo que se ha estudiado y las variables en las que se ha encontrado evidencia empírica de su significancia como agentes que influyen en la calidad y en la elección de la educación.

La siguiente etapa consiste en el análisis de datos referentes a las variables encontradas en la literatura como significativas para el rendimiento y la elección con los datos provistos por Mineduc, como resultados del SIMCE e información de los establecimientos, entre otros. Dicho análisis se basa en la aplicación de dos métodos econométricos que permiten evaluar la relación de una variable dependiente con un conjunto de variables explicativas en el caso del rendimiento, mientras que la elección es analizada a través de un modelo que permite encontrar las variables que toman en cuenta los padres y apoderados cuando eligen un establecimiento frente a varias alternativas.

Finalmente, se evalúa la factibilidad de incluir las variables encontradas como significativas en ambos temas dentro del modelo de optimización que se desarrollará para la implementación del SIGER. Concretamente se busca establecer qué variables introducir en el modelo y proponer algún tipo de función a incorporar en el modelo de optimización para lograr establecimientos de mejor calidad.

3 MARCO CONCEPTUAL

Para tener una visión general de la situación actual de la educación en zonas rurales, se realizan análisis descriptivos del rendimiento académico, los factores declarados por los padres que fueron considerados para matricular a sus hijos en el establecimiento al cual asisten y cómo están conformadas y caracterizadas la oferta académica y la demanda educacional rural. Además, se presenta el marco teórico de los métodos econométricos a utilizar en el análisis de datos.

3.1 Análisis de la Situación Actual de la Educación Rural

3.1.1 Rendimiento Académico

Diversas investigaciones han establecido que las zonas rurales tienen un rendimiento académico menor que las zonas urbanas, siendo atribuible en un 85% esa diferencia al hecho de que los estudiantes rurales provienen de hogares con un menor nivel socioeconómico, lo que se traduce como un menor capital humano que poseen las familias en estas zonas.³

³ Ver Gallego, Rodríguez y Sauma, 2007.

A continuación se muestra cómo se distribuye el rendimiento académico en zonas urbanas y rurales en términos de estudiantes aprobados, reprobados y retirados, donde se evidencia una desventaja para el sector rural. Además, en el caso de los estudiantes reprobados, se analizan las razones que los hacen caer en esa categoría, donde el sector rural tiene una tasa mayor para los repitentes por inasistencia, lo que podría indicar que al encontrarse en zonas donde los establecimientos quedan en lugares lejanos o poco accesibles se da un aumento de estos estudiantes, pues las condiciones ambientales adversas hacen que no puedan asistir regularmente a clases.

Cuadro 1: Rendimiento en zonas urbanas y rurales

Rendimiento	Estudiantes Urbanos	Estudiantes Rurales
Aprobados	94,51%	93,15%
Reprobados	4,13%	5,55%
Retirados	1,36%	1,30%

Fuente: Tabla Rendimiento 2007

Cuadro 2: Reprobación en zonas urbanas y rurales

Causa de Reprobación	Estudiantes Urbanos	Estudiantes Rurales
Inasistencia	15,19%	27,97%
Rendimiento	84,81%	72,03%

Fuente: Tabla Actas 2007

Al analizar la situación urbano versus rural tomando los puntajes del SIMCE, se produce la misma situación que con las tasas de aprobación y reprobación. Es decir, la zona urbana mantiene su hegemonía frente al rendimiento académico de las zonas rurales.

En este caso se muestra la proporción de establecimientos que presentan puntajes muy bajos y muy altos, esto considerando que el SIMCE es una prueba donde los 250 puntos se asignan a quienes están en la media de las respuestas correctas, con lo que, quienes obtienen bajo 200 puntos tienen un rendimiento deficiente y quienes tienen sobre 300 son los establecimientos de calidad sobresaliente.

Cuadro 3: Establecimientos con bajo y alto rendimiento

Tipo de Establecimiento	Puntaje SIMCE	<= 200	300<=
Rural	Lenguaje	3,5%	1,4%
	Matemática	11,9%	1,5%
Urbano	Lenguaje	1,1%	4,6%
	Matemática	1,2%	4,9%

Fuente: Tabla Establecimientos SIMCE 2007

3.1.2 Elección de los Padres

En este ámbito, muchas veces se cree que los padres no tienen opción de elegir a qué establecimiento enviar a sus hijos en las zonas rurales por lo lejos que quedaría la siguiente opción. Sin embargo, al momento de preguntar a los padres cuáles fueron las 3 razones principales por las que eligieron el establecimiento actual de sus hijos, sólo un 29,4% de las veces se menciona como la única alternativa que tenían en la comuna, siendo la cercanía la razón con más menciones, seguida de la familiaridad que tienen con el establecimiento, la cual tiene relación con que algún familiar estudie o haya estudiado en dicho establecimiento.

En el siguiente cuadro se muestra la frecuencia con que fueron mencionadas las razones de los padres en cuanto a la elección del establecimiento.

Cuadro 4: Razones de los padres de zonas rurales para elegir el establecimiento actual de sus hijos

Razón	Mención de los padres
Cercanía	67,3%
Familiaridad con el establecimiento	37,2%
Costo	35,5%
Único en la comuna	29,4%
Entrega de valores	21,8%
Prestigio	16,3%
Otra	12,0%
Buenos resultados en el SIMCE	9,8%
Formación Profesional	7,2%

Fuente: Elaboración propia a partir de la Tabla Cuestionario Padres SIMCE 2006

En general, la principal razón de elección es la cercanía del establecimiento, al mismo tiempo, los padres no toman los resultados del SIMCE como una de las razones para elegir el establecimiento de sus hijos, lo que se podría esperar que fuese una razón importante al pensar que el SIMCE representa la calidad de los establecimientos, pero la calidad muchas veces se valora a través de otros atributos como en este caso lo hacen el prestigio del establecimiento y la entrega de valores.

3.1.3 Caracterización de la Oferta y la Demanda Académica Rural

3.1.3.1 Oferta Académica

En la actualidad existen 4.426⁴ establecimientos educacionales funcionando en zonas rurales, representando el 37,6% del total de establecimientos educacionales del país.

Los establecimientos rurales que cuentan con una jornada escolar completa son el 22,8%, mostrando un claro déficit de dotación e infraestructura, al menos, para poder cumplir con una de las metas de la reforma educacional que se lleva a cabo desde fines de los años noventa.

Una de las principales características de los establecimientos rurales es la escasa matrícula con que cuentan, lo que se muestra en la siguiente tabla que agrupa los establecimientos según tramos de matrícula.

Cuadro 5: Matrícula rural por tramos de estudiantes

Alumnos	Establecimientos
de 1 a 10	8,98%
11 a 30	29,04%
31 a 60	18,37%
61 a 100	14,01%
más de 100	29,61%

Fuente: Tabla Matrícula 2007

⁴ Fuente: Tabla Directorio 2007 y Matrículas 2007, Departamento de Estudios y Desarrollo, Mineduc.

En la actualidad existen 2.623 establecimientos que tienen menos de 4 docentes y representan el 59,26% de la oferta educacional rural actual, los que en su mayoría atienden cursos multigrado y cuentan con un solo docente para todo el establecimiento. Por otra parte, los cursos multigrado corresponden al 69% del total.

Cuadro 6: Establecimientos rurales con baja dotación docente

Dotación Docente	
Unidocente	31,73%
Bidocente	13,57%
Tridocente	7,28%
Tetradocente	6,69%

Fuente: Tabla Dotación 2007

Mediante la tabla anterior se puede apreciar que la cantidad de establecimientos rurales que cuentan con un escaso número de profesores es bastante grande y afecta a gran parte de la matrícula de estos establecimientos lo que muestra que la calidad de la enseñanza en ellos se debe en gran parte a lo que puedan hacer esos pocos profesores.

3.1.3.2 Demanda Académica

La población rural está compuesta principalmente por los dos primeros grupos socioeconómicos, lo que se puede cuantificar a través de la información entregada por el Censo 2002 de la población rural.

Cuadro 7: Composición Socioeconómica de la población rural

Grupo Socioeconómico	Habitantes
Bajo	52,7%
Medio Bajo	35%
Medio	7,2%
Medio Alto	3,8%
Alto	1,3%

Fuente: Censo 2002

A través de la información recogida por el SIMCE 2006 se puede visualizar cuál es el estado de aprobación de los padres sobre diversos aspectos de la educación y el establecimiento en que están sus hijos, siendo la infraestructura la característica que tiene mayor tasa de desaprobación, seguida por la gestión de los establecimientos, la acogida a la opinión de los padres y el apoyo a los alumnos de bajo rendimiento, lo que nuevamente muestra que hay mucho por hacer en las zonas rurales para mejorar la calidad, la equidad y la oferta educativa.

Cuadro 8: Satisfacción de los padres

Característica	Muy insatisfecho	Insatisfecho	Satisfecho	Muy satisfecho
Calidad de la educación	4%	13%	57%	25%
Entrega de valores	3%	11%	57%	29%
Apoyo a alumnos de bajo rendimiento	5%	16%	50%	28%
Buena disposición del profesor	4%	10%	51%	35%
Información resultados SIMCE	4%	16%	58%	22%
Información en reunión de apoderados	3%	7%	53%	37%
Gestión del establecimiento	5%	19%	60%	16%

Característica	Muy insatisfecho	Insatisfecho	Satisfecho	Muy satisfecho
Acogida de la opinión de los padres	5%	19%	55%	21%
Infraestructura	9%	20%	47%	23%
Relación del profesor con el alumno	4%	6%	44%	46%
Desempeño del profesor	4%	12%	51%	33%

Fuente: Tabla Cuestionario Padres SIMCE 2006

3.2 Marco Teórico

La investigación sobre la calidad de la educación ha sido desarrollada a lo largo de los años en base a mediciones de cada país y mediciones internacionales del rendimiento de sus estudiantes a través de la regresión lineal principalmente. Este tipo de modelo tiene ventajas sobre su aplicación e interpretación de resultados, pero examina todas las variables dentro de un mismo nivel, sin darle mayor o menor jerarquía a unas u otras, es por ello que también se ha utilizado en este tipo de mediciones el modelo anidado o lineal jerárquico, el que permite agrupar a los individuos en grupos, con lo que se logra aislar el efecto unitario del efecto grupal, lo cual permite examinar y determinar la importancia que tiene el establecimiento en el rendimiento de los estudiantes.

Con respecto a la elección se utiliza un modelo que represente la probabilidad de elegir cierto establecimiento sobre otros, con lo que se busca determinar qué factores son los más gravitantes al momento de tomar una decisión. Para estudiar dicho fenómeno se han construido modelos que se basan en la utilidad que genera tomar cierta opción sobre las otras, y asignan una probabilidad de elección a cada alternativa. En este caso el modelo a utilizar es un modelo logarítmico multinomial o Logit, el que permite obtener estimaciones de probabilidad de un suceso, además de identificar los factores que determinan dichas probabilidades, así como la influencia o peso relativo que tienen éstos sobre las mismas.

3.2.1 Metodología de Estimación de la Función de Producción de la Educación

3.2.1.1 Regresión Lineal con Mínimos Cuadrados Ordinarios

Este método se basa en la idea de expresar un vector \mathbf{y} , el que en este caso está compuesto por el puntaje obtenido en el SIMCE, a través de una combinación lineal de las columnas de la matriz \mathbf{X} , donde cada columna representa los valores observados de cada variable estudiada, dando paso a un modelo matricial de la forma:

$$\mathbf{y} = \mathbf{Xb} + \varepsilon$$

donde ε es la diferencia entre \mathbf{y} y \mathbf{Xb} .

El método de los mínimos cuadrados ordinarios consiste en hacer mínima la suma de los cuadrados residuales, es decir, se trata de hallar los estimadores que hagan que esta suma sea lo más pequeña posible:

$$\begin{aligned}\sum_{i=1}^n \hat{\varepsilon}_i^2 &= \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \\ &= \sum_{i=1}^n (Y_i - \hat{\beta}X_i)^2\end{aligned}$$

con lo que los estimadores de β quedan definidos como:

$$\hat{\beta} = (X'X)^{-1} X'y$$

Dichos estimadores permiten cuantificar la proporción en que las variables significativas afectan a la variable dependiente, la que en este caso es el puntaje obtenido por el estudiante en la prueba SIMCE.

Este tipo de modelo cuenta con una serie de supuestos que son prácticamente imposibles de cumplir en la realidad, por lo que muchas veces sus resultados no presentan un buen ajuste.

Una vez obtenida la recta que se ajusta a la nube de puntos que representa cada observación es necesario obtener un indicador que muestre qué tanto se ajusta esta recta a las observaciones. En este caso, se utiliza el coeficiente de determinación como medida de bondad de ajuste, el cual está definido como,

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Este coeficiente de determinación mide el porcentaje de variabilidad de la variable y explicada por el modelo tomando valores entre 0 y 1. Cuando todas las observaciones se encuentran sobre la recta se tiene un ajuste perfecto y este coeficiente toma el valor 1.

Además, el test F permite rechazar la hipótesis nula de que el R^2 poblacional es cero, con lo que la significancia global del modelo especificado se comprueba al obtener un F significativo.

3.2.1.2 Modelo Lineal Jerárquico

Este tipo de modelo considera una estructura de anidamiento de las unidades muestrales dentro de grupos y permite la utilización de variables explicativas tanto al nivel de las unidades como a nivel de grupos. Las unidades utilizadas en esta investigación son los estudiantes y los grupos son los diversos establecimientos, lo que permite estimar el efecto específico que tiene el establecimiento sobre el rendimiento individual.

Cada uno de los niveles de la estructura jerárquica es representado formalmente con su propio submodelo. Cada submodelo expresa las relaciones entre variables dentro de un determinado nivel y el conjunto de submodelos especifica de qué modo, variables de un nivel influyen en las relaciones que ocurren a otro nivel, quedando de este modo formalizada esta estructura anidada de los datos.

El mayor aporte de este modelo es que cuenta con una serie de características que favorecen este estudio para encontrar resultados más robustos:

- Considera las diferencias de contexto; en el estudio del rendimiento académico se necesita considerar simultáneamente las variables individuales y las variables contextuales, como, por ejemplo, los recursos del establecimiento.
- Permite analizar simultáneamente contextos y heterogeneidad individual. Las diferencias entre contextos son tan importantes como las posibles diferencias entre los sujetos.
- Permite formular y probar hipótesis sobre los efectos formulados entre niveles, realizar la partición de componentes de varianza y covarianza entre niveles y una adecuada estimación de los términos de error del modelo.

La ecuación del primer nivel es de la forma:

$$Y = \beta X + r$$

Uno de los supuestos del modelo es que los β varían de grupo en grupo, de modo que pueden ser tratados como aleatorios en un segundo nivel del modelo, cuya ecuación es,

$$\beta = \gamma Z + \mu$$

El modelo más sencillo es el incondicional o modelo vacío, el cual no contiene predictores en ningún nivel y sus parámetros son útiles para comprender más a fondo el problema estudiado. La ecuación del modelo incondicional se formula del siguiente modo,

Nivel 1

$$Y = \beta_0 + r$$

donde β_0 es la intersección con el origen y r los errores atribuibles al nivel 1, los que se suponen siguen una distribución normal $N(0, \sigma^2)$.

Nivel 2

$$\beta_0 = \gamma_{00} + \mu_0$$

El parámetro γ_{00} es la media poblacional y μ_0 es el efecto aleatorio del segundo nivel o la desviación de la media de grupo respecto a la media poblacional y se asume que tiene una media de 0 y una varianza de τ_0 .

La función de este modelo es obtener una estimación puntual de la media poblacional γ_{00} y entregar información sobre la variabilidad del resultado en cada uno de los niveles. El coeficiente de correlación intraclases, CCI, es el cociente entre la variabilidad entre grupos y la variabilidad total, quien muestra la variabilidad que se puede atribuir a diferencias entre grupos. Este coeficiente va de 0 a 1, y mientras más cercano a 1 se puede afirmar que el grupo al que pertenece el individuo es muy importante para lograr diferencias.

En este tipo de modelo, se pueden utilizar las variables sin centrar, centradas en la media del grupo o centradas en la media poblacional, también llamada gran media. Dependiendo del nivel en que se esté interesado, se utilizan distintos métodos de centrado. En el caso del primer nivel, investigaciones han determinado que el centrado en la media del grupo permite obtener los resultados más robustos cuando se está interesado en las variables del primer nivel.⁵

El siguiente paso entonces, es encontrar las variables significativas del primer nivel, por lo que el modelo que se construye es de la forma:

$$\begin{aligned}
 Y &= \beta_0 + \beta_1(\text{var individuo1}) + \beta_2(\text{var individuo2}) + \dots + \beta_k(\text{var individuok}) + r \\
 \beta_0 &= \gamma_{00} + \mu_0 \\
 \beta_1 &= \gamma_{10} + \mu_1 \\
 &\vdots \\
 \beta_k &= \gamma_{k0} + \mu_k
 \end{aligned}$$

con esto, es el intercepto β_0 y las pendientes $\beta_1, \beta_2, \dots, \beta_k$ son a su vez las variables dependientes del segundo nivel.

Una vez encontradas aquellas variables, se procede a especificar el siguiente nivel, donde la interacción del grupo con la variable dependiente se refleja a través del la intersección con el origen, pues al tener las variables predictoras del individuo en cero, es este elemento quien da la diferencia entre los individuos y este elemento se desprende de la interacción de las variables del segundo nivel con él. Esta vez, el centrado en la media del grupo no tiene sentido, por lo que las variables se centran en la gran media para realizar una interpretación más adecuada que al no centrarlas.

Como el objetivo de esta investigación es identificar las variables del establecimiento, es el intercepto quien dará la información de la influencia del establecimiento en el estudiante, lo cual se logra con la primera ecuación del segundo nivel. Al incluir variables en otra de las ecuaciones, se estará modelando la influencia de aquellas variables del nivel 2 sobre la variable del nivel 1 ligada a esa ecuación.

$$\begin{aligned}
 Y &= \beta_0 + \beta_1(\text{var individuo1}) + \beta_2(\text{var individuo2}) + \dots + \beta_k(\text{var individuok}) + r \\
 \beta_0 &= \gamma_{00} + \gamma_{01}(\text{var establecimiento1}) + \dots + \gamma_{0k}(\text{var establecimientok}) + \mu_0 \\
 \beta_1 &= \gamma_{10} + \gamma_{11}(\text{var establecimiento1}) + \dots + \gamma_{1k}(\text{var establecimientok}) + \mu_1 \\
 &\vdots \\
 \beta_k &= \gamma_{k0} + \gamma_{k1}(\text{var establecimiento1}) + \dots + \gamma_{kk}(\text{var establecimientok}) + \mu_k
 \end{aligned}$$

Como se ve en el sistema de ecuaciones anterior, β_0 es quien refleja la influencia del establecimiento en el rendimiento, mientras que β_k muestra la interacción del establecimiento frente a la variable k del estudiante, es decir, su pendiente. Por lo tanto, es el coeficiente β_0 y sus variables explicativas lo que se busca determinar a través de este estudio.

⁵ Ver Schumacker y Bembry 1995.

3.2.2 Metodología de Estimación de los Factores de Elección

El ámbito que abarca esta parte del estudio es un problema de respuesta categórica, ya que se quiere encontrar la relación entre variables que hacen que se elija cierto establecimiento o no, es decir, la variable dependiente es discreta por lo que los modelos clásicos de regresión no resultan adecuados. Específicamente, la elección de establecimientos por parte de los padres es una elección binaria, donde se busca encontrar la relación entre las características de ellos y el establecimiento que hacen que sea elegido o no, es decir, se buscan las variables que afectan la función de utilidad que tratan de maximizar los padres al momento de elegir un establecimiento. Esto parte del supuesto de que la utilidad derivada de una elección, U_{ij} , es función de las variables explicativas de dicha decisión, que son tanto las características del individuo como las de cada alternativa de elección, de manera que la utilidad toma la forma

$$U_{ij} = \alpha_j + X_{ij}\beta + \varepsilon_{ij}$$

donde los ε_{ij} recogen las desviaciones que los individuos tienen respecto a lo que sería el comportamiento del individuo medio y que se debe a factores aleatorios. El individuo i , entonces, optará por la opción j si su utilidad es superior a la de las otras opciones y viceversa, de manera que,

$$Y_i = \begin{cases} 1 & \text{si } U_{ij} > U_{ik} \\ 0 & \text{si } U_{ik} < U_{ij} \end{cases} \quad \forall j \neq k$$

y el modelo queda definido por,

$$\begin{aligned} \text{Prob}(Y_i = 1) &= \text{Prob}(U_{ij} > U_{ik}) \\ &= \text{Prob}(U_{ij} - U_{ik} > 0) \\ &= F(X_i\beta) \end{aligned}$$

Como lo que se busca es la probabilidad de realizar cierta elección, se debe utilizar un modelo de este tipo, es decir, un modelo Logit o Probit, donde Probit tiene como distribución de probabilidad una Normal Estándar, mientras que Logit tiene una distribución Logística Estándar.

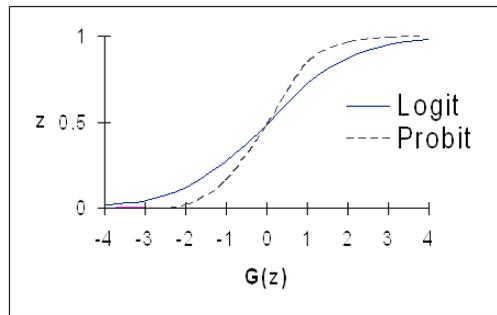
La diferencia principal del modelo Logit con respecto al modelo Probit, es que la distribución logística tiene colas más planas y altas, lo que hace que ambos modelos entreguen resultados similares en el centro, pero no en los extremos, donde Logit entrega una probabilidad mayor, por lo que al tener una muestra con pocas respuestas afirmativas o negativas ($Y=1$ ó $Y=0$) este modelo funciona mejor.

Los coeficientes obtenidos de Logit y Probit son también similares, pero no son directamente comparables entre sí. Sin embargo, la principal razón de la ventaja del Logit es que es más fácil de interpretar. Además, la elección de un modelo Probit es mucho más apropiada cuando existe una fuerte creencia de que la distribución subyacente es normal o que el verdadero acontecimiento estudiado no es un resultado binario sino que una proporción, como por ejemplo, proporción de la población en diferentes niveles de deuda.⁶

⁶ http://en.wikipedia.org/wiki/Predictive_analytics#Logit_vs._Probit

A continuación se muestra un gráfico con ambas distribuciones para un mejor entendimiento del lector:

Figura 1: Distribución Logit versus Probit



Fuente: <http://www.hsrmethods.org/Glossary/Terms/P/Probit.aspx>

El modelo Logit posee una serie de ecuaciones que proporcionan un conjunto de probabilidades para las a J+1 alternativas de elección y que se pueden tomar con X_i características individuales.

Por definición un modelo lineal de probabilidad toma la siguiente forma:

$$P(y = 1 | X) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

donde y es una variable aleatoria que toma valores $\{0, 1, 2, \dots, J\}$ y los coeficientes β reflejan la variación en la probabilidad de un cambio unitario en X .

El problema que esto presenta es que la probabilidad no está acotada entre 0 y 1. La solución es que se tenga una función $F(\cdot)$ que tome valores entre (0,1).

$$\text{Prob}(y = 1 | X) = F(X\beta), \text{ con } 0 < F(z) < 1$$

La función del modelo Logit es una distribución logística, la que tiene la siguiente forma funcional:

$$F(z) = \frac{\exp(z)}{[1 + \exp(z)]}$$

la cual entrega una probabilidad de $P(y = j | x_i) = \frac{\exp(\beta'_j x_i)}{1 + \sum_{k=1}^J \exp(x\beta_k)}$ con $j=1,2,\dots,J$

$$P(y = 0 | x_i) = \frac{1}{1 + \sum_{k=1}^J \exp(x_i \beta_k)}$$

Así la probabilidad de no ir al establecimiento j está dado por $1 - P(y = j | x)$, con que la razón de probabilidades de elegir un establecimiento es:

$$\frac{P_{ij}}{1 - P_{ij}} = \frac{1 + \exp(x_i \beta_j)}{1 + \exp(-x_i \beta_k)} = \exp(x_i \beta_k)$$

Con esta formulación, los J logaritmos de los cuocientes de probabilidades que se obtienen son $\ln \left[\frac{P_{ij}}{P_{i0}} \right] = \beta'_j x_i$ lo que al ser normalizado queda $\ln \left[\frac{P_{ij}}{P_{ik}} \right] = x_i (\beta_j - \beta_k)$. De este modo,

el cuociente de probabilidades $\frac{P_j}{P_k}$ no depende del resto de las alternativas, lo que se basa en el supuesto de independencia de los errores, con lo que se obtiene la solución del modelo.

Uno de los estadísticos que buscan aprobar o rechazar la hipótesis de este modelo es el test χ^2 para contrastar la razón de verosimilitud, quien busca rechazar la hipótesis nula que postula que todos los coeficientes $\beta_i = 0, \forall i$. Dicha razón es la diferencia entre el modelo sin variables explicativas y el modelo final, aplicando la función -2Log verosimilitud, la que se acepta o rechaza con este estadístico. En otras palabras el estadístico χ^2 es la diferencia en las -2Log verosimilitudes entre el modelo final y el modelo reducido.

Además, el ajuste del modelo se verifica a través de dos pruebas χ^2 de bondad de ajuste, las que buscan comparar los valores pronosticados con los valores observados, los que se postulan son similares y su diferencia es nula, por lo que lo que busca el estadístico es no rechazar la hipótesis nula, es decir que los test no sean significativos, demostrando que las desviaciones son pequeñas y el modelo se ajusta a los datos. Estas 2 pruebas de bondad de ajuste son la de desviación y la de Pearson.

Ya que este modelo no es lineal, el coeficiente de determinación R^2 no sirve para explicar la proporción de varianza explicada por el modelo, por lo que se han ideado indicadores alternativos llamados pseudo R^2 , los que están entre 0 y 1, siendo el valor más cercano a 1 el de un buen modelo.

Para definir qué parámetros son significativos se utiliza el contraste de razón de verosimilitud, es decir, busca cuáles variables independientes son influyentes en la decisión tomada por el consumidor, en este caso los padres del estudiante. Este contraste nuevamente se basa en la diferencia del modelo a través de -2Log verosimilitud y el estadístico χ^2 , donde se compara cómo cambia el modelo con y sin la variable que se está contrastando.

Al tratarse de una decisión con más de 2 alternativas, el cálculo de un vector de coeficientes β no es directo como en la regresión lineal, pues se tienen tantos vectores como alternativas -1, pues se establece una opción como la categoría de referencia. Esto implica que cada coeficiente β de una alternativa representa el efecto o probabilidad de estar en esa alternativa o la de referencia.

4 REVISIÓN BIBLIOGRÁFICA

Esta sección busca reflejar el estado del arte de los temas a investigar en Chile y el mundo.

En primera instancia se muestran las variables de calidad encontradas en la literatura y una descripción de los fenómenos más importantes encontrados para luego mostrar lo que se ha investigado en el tema de elección de establecimientos educacionales, con sus variables relevantes y resultados de algunas investigaciones que son relevantes.

4.1 Estudios sobre el Rendimiento Educativo

Los estudios sobre el rendimiento han ido evolucionando en sus enfoques y planteamientos, lo que comienza en los años sesenta con el estudio de Coleman de 1966. El enfoque utilizado por él es la estimación de la función de producción educacional con el rendimiento académico como resultado y las características socioeconómicas y del establecimiento como variables de entrada. Gracias a esto, los primeros modelos tienen una orientación Proceso-Producto, con los que se identificaron factores contextuales y escolares como determinantes en la eficacia.

En un principio, prácticamente sólo se desarrollan estudios sobre la eficacia escolar para países desarrollados, hasta que en 1978 se realiza el primer estudio que incluye varios países de Latinoamérica.⁷

Luego, aparecen nuevos enfoques de investigación, basados en sistemas relacionales de eficacia institucional, donde destaca el trabajo desarrollado por Murphy, Hallinger y Mesa (1985), donde se toma una clasificación entre factores de contexto, entrada, proceso y producto y se identifican 3 subsistemas relacionados: tecnología del centro, entorno del centro y resultados de los alumnos.

Finalmente, aparecen los sistemas de indicadores de contexto, proceso y producto; con lo que se dejan atrás los sistemas relacionales de características, lo que permite incorporar nuevos factores y dimensiones de eficacia, reuniendo tanto elementos teóricos que admiten analizar interrelaciones entre variables de distintos niveles, como un marco general de referencia que facilita incorporar y determinar la posición cualquier variable a estudiar.

Las principales metodologías que se utilizan en la literatura son regresiones lineales con mínimos cuadrados ordinarios, análisis factorial, modelos lineales jerárquicos y análisis descriptivos. De la revisión realizada, se presentan los resultados obtenidos de la investigación sobre las posibles variables que pudieran afectar el rendimiento escolar. Luego se presentan algunos aspectos destacados surgidos de esta revisión.

⁷ “The determinants of School Achievement: A Review of the Research”, Simmons, J y Alexander, L.

Cuadro 9: Características del establecimiento

Variable	Significado	Estudios que la encuentran significativa	Relación
Escala	A través de mayor escala hay más profesores, los que tienen mayores posibilidades de especializarse y además los recursos pueden ser utilizados por más alumnos	5	Positiva
Competencia	La oferta permite la competencia.	1	Positiva
Matrícula	Total de alumnos del establecimiento	9	Positiva
Razón alumno-profesor	Cantidad de Alumnos por profesor	13	Negativa
Establecimiento multigrado	Escuela con cursos que combinan 2 o más grados	3	Incierta, 2 de los 3 estudios la encuentran positiva
Gasto por estudiante	Cantidad de dinero invertida en el centro por estudiante	2	Positivo, pero hay una serie de estudios que no lo encuentran significativo
Actividades extra programáticas	Actividades extracurriculares de libre elección	3	Positiva
Establecimiento bilingüe	Establecimientos que tienen un idioma nativo como lengua de enseñanza	1	Positiva, tiene gran efecto en la población indígena
Aumento de la jornada escolar	Establecimientos con más horas de clases	7	Positiva
Disciplina	Normas y regulaciones internas	3	Positiva
Programas de Alimentación	Alimentación brindada en el establecimiento	2	Positiva

Fuente: Elaboración propia

Cuadro 10: Materiales educativos

Variable	Significado	Estudios que la encuentran significativa	Relación
Infraestructura	Incluye calidad de la construcción, mobiliario y servicios básicos	7	Positiva
Recursos Pedagógicos	Acceso a materiales educativos como libros y computadores	13	Positiva, sobretodo en recursos TIC ⁸

Fuente: Elaboración propia

Cuadro 11: Características del Profesor

Variable	Significado	Estudios que la encuentran significativa	Relación
Sueldo	Remuneración del docente	5	Positiva
Calidad	Calificación del profesor	8	Positiva
Experiencia	Años de docencia	10	Positiva
Antigüedad	Años trabajando en el mismo establecimiento	6	Positiva, aunque muchos estudios no lo encuentran relevante
Género	Género del profesor	3	Positiva para mujeres
Cursos de actualización	Formación permanente de los docentes	6	Positiva

Fuente: Elaboración propia

Cuadro 12: Características del Director

Variable	Significado	Estudios que la encuentran significativa	Relación
Antigüedad	Años en el cargo	3	Incierta, 2 de los 3 estudios la encuentran positiva
Género	Género del director	1	Positiva para Mujeres
Formación	Estudios cursados, sobre todo en educación	2	Positiva

Fuente: Elaboración propia

⁸ TIC: Tecnologías de la Información y la Comunicación.

Cuadro 13: Características del estudiante y su familia

Variable	Significado	Estudios que la encuentran significativa	Relación
GSE ⁹	Características socioeconómicas, dentro de las que están: ingresos, educación y tipo de trabajo de los padres	15	Positiva
Lengua materna distinta	La madre tiene una lengua distinta a la de enseñanza, principalmente indígena	6	Negativa
Trabajo infantil	El estudiante trabaja	2	Negativa
Género	Influencia del género en el aprendizaje	5	Positiva en lenguaje para las mujeres y en matemáticas para los hombres
Tamaño familia y hogares completos	Número de integrantes de la familia y con ambos padres presentes	3	Positiva, aunque se vuelve negativa al pasar cierto número de integrantes
Tasa de Asistencia	Asistencia a clases	2	Positiva

Fuente: Elaboración propia

Cuadro 14: Motivación e inserción del estudiante en el sistema educativo

Variable	Significado	Estudios que la encuentran significativa	Relación
Expectativas del profesor, alumno y padres	Se refiere a lo que esperan del rendimiento del alumno	7	Positiva
Apoyo y participación de los padres	Compromiso e involucramiento de los padres en la educación y desarrollo del alumno	8	Positiva

Fuente: Elaboración propia

⁹ GSE: Grupo Socioeconómico

4.1.1 Aspectos Importantes de Rendimiento

En el estudio “Determinantes de la Calidad: ¿Qué falta por mejorar?”¹⁰ se establece que la brecha entre establecimientos municipales y subvencionados, con respecto a los particulares ha disminuido entre 1989 y 1997.

Dentro de los factores responsables de la disminución de la brecha se mencionan el aumento de las remuneraciones docentes; traspaso masivo de alumnos desde el sector municipal al particular subvencionado (20% de la matrícula total), que habría reducido el nivel socioeconómico promedio de sus padres; las políticas de mejoramiento educativo implementadas por el Estado; la opinión pública ha comenzado a interesarse en conocer las cifras del SIMCE con lo que se ha educado la demanda para seleccionar buenos establecimientos; los alumnos de 4° Básico de los años ‘90 tienen padres que, en una gran mayoría, lograron una mayor educación familiar que les permite apoyar el aprendizaje de los hijos en mejor forma que en el pasado; el acceso al computador del 75% de los alumnos de los recintos municipales, que antes sólo tenían los establecimientos privados¹¹.

Además se estudia el impacto de las características del aula, donde se sostiene que la diversidad de estudiantes con distintas capacidades hace que sea difícil seguir el ritmo del alumno promedio a los menos aventajados, siendo el material didáctico y cursos con pocos alumnos los que ayudan a mejorar el aprendizaje de los estudiantes con peor rendimiento.

Los establecimientos con pocos alumnos tienden a tener salas multigrado, lo que produce una gran heterogeneidad en el grupo curso, los que, además, suelen tener escasos docentes e incluso pueden ser unidocentes, lo que implica que un solo profesor atiende a todos los alumnos que asisten al establecimiento, debiendo enseñar a alumnos de varios grados distintos en la misma sala. Aunque en Chile hay cursos multigrado con excelente rendimiento,¹² en promedio tienen un resultado bajo.

Al existir una alta heterogeneidad dentro del grupo curso, se postula que se obtienen menores resultados, pues al ser un curso más homogéneo, los estudiantes aprenden al mismo ritmo, sin dejar rezagados a los menos aventajados.

En el estudio “Eficiencia escolar y diferencias socioeconómicas: a propósito de los resultados de las pruebas de medición de la calidad de la educación en Chile” se analizan para el sistema educacional chileno, los resultados obtenidos en el SIMCE en matemática y lenguaje por los Octavos Básicos en 2000. Lo más relevante de este estudio es que el nivel socioeconómico es el factor más importante en el rendimiento, respecto a la dependencia administrativa del establecimiento, por lo que sería incorrecto afirmar que los menores resultados de los establecimientos municipales se debe a que son públicos, sino que “las escuelas municipales tienen inferiores resultados porque atienden a los hijos de los más pobres del país”.

Dentro de las características estructurales de la educación rural, se produjo el estudio “Escuelas rurales y financiamiento”¹³, donde se destacan hechos relacionados con la realidad del funcionamiento de este tipo de establecimientos, en que la menor matrícula hace que exista una menor posibilidad de cubrir los costos fijos y, además, se evidencia que el subsidio de ruralidad

¹⁰ Ver Schiefelbein y Schiefelbein, 2000.

¹¹ Ver Potashnik, 1996; Álvarez et al, 1998

¹² Ver Arancibia et al, 1998

¹³ Ver Mizala y Melis 2007

no alcanza a cubrir estos costos donde existe una baja densidad poblacional y las escuelas no tienen el adecuado sistema de administración financiera, por lo que el consolidar escuelas resultaría también beneficioso en cuanto a la sustentabilidad de los establecimientos.

Además, se hace un diagnóstico de la distribución de los establecimientos rurales y su financiamiento, donde se establece que el 77,5% de los establecimientos rurales municipales se ubica en las regiones IV, VII, VIII, IX y X, dentro de las cuales las regiones VIII y X tienen la mayor participación en el total; los establecimientos pequeños (de 1 a 10 y de 10 a 17 alumnos) se concentran en IV, VII, VIII, IX, X y la mitad de los establecimientos tienen menos de 31 alumnos, concentrando el 8,9% de la matrícula municipal rural; los ingresos de los establecimientos municipales rurales menos las remuneraciones docentes arrojaron a 980 establecimientos con saldo negativo, lo que representa el 25,8% matrícula municipal rural y 27,9% de los establecimientos; y la mayor proporción de los establecimientos con saldo negativo (47,2%) está en los establecimientos de 100 a 250 alumnos, los que representan el 63,2% de la matrícula de los establecimientos con déficit.

Otro de los estudios revisados, “Estudio sobre la educación para la población rural en Chile”¹⁴, muestra la realidad que se vive en las zonas rurales y posee buenas definiciones que permiten entender mejor la dinámica de la educación rural. Dentro de ellas detalla el contexto de las escuelas multigrado.

(...) “Se trata de establecimientos situados en contextos de alta pobreza, economías tradicionales, con poblaciones indígenas, con baja densidad poblacional. Se han estructurado, para fines de incentivos y reglamentación, en escuelas de mediano o difícil acceso, dependiendo de su aproximación o lejanía a centros urbanos de convergencia poblacional y de comunicaciones. Esto ha hecho que los niveles de control directo y asesoría técnica respecto del desarrollo educacional de estas escuelas haya sido precario. Son aproximadamente 3.300 establecimientos, con 5.068 profesores y 96.500 alumnos distribuidos a lo largo del territorio nacional (al año 2003).

Son escuelas incompletas, generalmente llegan hasta 6° año básico, organizados en cursos combinados y cuentan con uno, dos o tres profesores.

En general la formación inicial docente no considera conceptual ni metodológicamente esta situación pedagógica, que los docentes deben enfrentar desde la práctica, desde su formación empírica, combinada con participación en eventos de perfeccionamiento que, en la última década, ha impulsado el Ministerio de Educación a través del Programa de Educación Básica Rural.”

Además, se muestra que los internados rurales tienen mejores resultados en el SIMCE que las escuelas rurales, mientras que en las zonas urbanas, los internados muestran un rendimiento menor al de escuelas sin internado, por lo que resultaría ventajoso para los niños rurales asistir a este tipo de régimen pues les brinda una mayor y mejor oportunidad de aprendizaje.

Un estudio que muestra evidencia empírica de la importancia de los recursos del establecimiento es “Factores escolares asociados a los aprendizajes en la educación primaria mexicana”¹⁵, donde la principal conclusión es que la dotación de recursos materiales y humanos

¹⁴ Ver Williamson, G. 2003

¹⁵ Ver Blanco, E. 2008

de las escuelas es un factor importante para explicar las diferencias en los aprendizajes. Los factores que más importan en este sentido son una mejor infraestructura, mayor cantidad de recursos de aprendizaje, docentes con mayor experiencia y que permanezcan más tiempo en el establecimiento.

En “Informe de capital humano en Chile”¹⁶ se determina que, con respecto a los estudiantes y su entorno, en Chile el peso de las variables de origen socio-económico, de recursos del hogar y de distribución del capital social, se encuentra condicionado por las profundas desigualdades de la sociedad.

Al contrario de otros estudios, éste establece que “la escuela sí importa”, especialmente para los alumnos de bajo nivel socioeconómico:

(...) “la mayor ganancia neta en términos de aprendizaje se produce entre los alumnos de más bajo nivel socioeconómico que estudian en una escuela efectiva a cargo de un profesor efectivo. Cuando se reúne esa doble condición, la educación tiene un impacto decisivo en el aprendizaje de los alumnos, especialmente de aquellos con bajo nivel de logro inicial.

Lo contrario ocurre cuando el establecimiento y el profesor son ineficaces. En este caso los alumnos de más bajo nivel socioeconómico son los que más pierden; en efecto, terminan peor de lo que habían comenzado.”,

con lo que el preocuparse por una mejor calidad de los establecimientos rurales, que suelen ser de bajos recursos, tendría un impacto importante en sus estudiantes.

Finalmente, el estudio que más información ha aportado hasta el momento es la publicación “Investigación Iberoamericana sobre eficacia escolar”¹⁷ que entrega una visión global de los factores asociados con el logro cognitivo, los cuales agrupa en 3 conjuntos:

- **Factores del alumnado**

- El nivel socio-económico y el nivel cultural de las familias son los factores que más ayudan a explicar el logro académico de los alumnos. Además, características personales del alumno –como el género, ser inmigrante o indígena o tener una lengua materna diferente de enseñanza– cumplen también un papel determinante como elementos explicativos del logro académico.
- El trabajo infantil hace que el rendimiento, comportamiento y satisfacción con el establecimiento se vean gravemente afectados.
- La asistencia regular del alumno a clases también incide de manera positiva.
- Los hábitos culturales, especialmente los hábitos lectores y, en menor medida, el deporte, realización de actividades extraescolares son factores asociados con el logro académico.

¹⁶ Ver Brunner, J. y Elacqua, G. 2003

¹⁷ Ver Murillo, F. et al, 2007

- La relación del alumno con los padres, el apoyo de ellos y de sus compañeros resultan también ser factores asociados con el logro.
- La satisfacción del alumno con sus compañeros y la ausencia de problemas en el establecimiento están asociados con el buen rendimiento.
- Por último, las expectativas del profesor, de sus padres y del mismo alumno sobre el nivel máximo de estudio que podrá alcanzar están fuertemente asociadas con el rendimiento.

- **Factores del docente y el aula**

- La experiencia docente, antigüedad en el establecimiento y el género son factores asociados con el logro. Los profesores más experimentados y las mujeres docentes tienen mejores resultados con sus alumnos.
- El número de horas de formación permanente del docente en los últimos años es un factor que explica un alto porcentaje de varianza del rendimiento cognitivo del alumno.
- Los docentes que están más satisfechos con el establecimiento y con sus condiciones laborales tienen alumnos con mejor logro.
- Los profesores más represivos afectan negativamente el resultado académico.
- En el caso de alumnos por sala, los que asisten a salas muy numerosas o con muy pocos compañeros tienen resultados más bajos, o visto de otro modo, una razón alumno-profesor extrema, produciría menores resultados. En este sentido, la UNICEF ha manifestado que clases con más de 25 alumnos tienen 1,5 veces más de posibilidades de tener menores resultados y de reprobado el año.¹⁸ Sin embargo, se ha encontrado evidencia empírica sobre la conveniencia de cursos pequeños para los estudiantes menos aventajados,¹⁹ por lo que en el caso de establecimientos con bajo rendimiento, sería indicado comenzar con cursos pequeños para luego ir incrementando el número de alumnos.
- La metodología desarrollada en las clases es uno de los factores que determinan en mayor medida los resultados académicos de los alumnos.

- **Factores escolares**

- El nivel socio-cultural del establecimiento explica un porcentaje de varianza significativo del rendimiento.
- Los directivos mujeres y aquellos que llevan más años en el cargo generan dinámicas que favorecen el logro académico.
- Una mayor participación de padres, madres y alumnos hace que ellos obtengan mejores resultados.

¹⁸ Ver UNICEF 2000

¹⁹ Ver Krueger, A. 1997

- Instalaciones y recursos, sobre todo la limpieza y calidad de las instalaciones, y la adecuación de los recursos TIC también aumentan el rendimiento escolar.
- Los alumnos de docentes que están satisfechos con su remuneración, obtienen mejores resultados.

El estudio desarrollado por Mizala y Romaguera en 2000, “Determinación de factores explicativos de los resultados escolares en educación media en Chile” encuentra evidencia empírica de un tamaño óptimo de establecimiento, pues la matrícula afecta positivamente los puntajes del SIMCE, pero no de forma lineal, si no que cóncava, con lo que pasado cierto número de alumnos el establecimiento tiene menores resultados.

Así también, los beneficios detectados de un establecimiento con menos alumnos son: una mejor conducta y relaciones entre compañeros y alumno profesor, más participación en actividades extra programáticas, menor ausentismo y tasas de abandono, mayor sentido de pertenencia y participación de los padres.²⁰

El último punto, acerca del tamaño de los establecimientos, en términos de número de alumnos, es fundamental para el desarrollo del proyecto SIGER, pues existe un *trade-off* importante entre tamaño del establecimiento, costos operacionales y distancia recorrida por el estudiante (medida en unidades de tiempo), con lo que encontrando el punto de equilibrio entre tamaño y distancia se podría obtener un tamaño ideal para cada establecimiento.

Finalmente, cabe destacar que la aplicación de métodos econométricos en las investigaciones del ámbito social, que involucran variables cuantificables y no cuantificables, es tomado como válido a pesar de que no presentan un ajuste demasiado alto²¹, dado que se trata de modelos que son difíciles de calibrar al incluir características personales. En general, este tipo de estudios presentan un ajuste del orden del 20%, pero se han identificado como una buena herramienta y forma de aproximación para determinar los factores que influyen en el rendimiento.

4.2 Estudios sobre Factores de Elección Académica

El contexto educacional chileno tiene una oferta compuesta por 3 tipos de establecimientos: particulares pagados, particulares subvencionados y municipales. Esto se debe a que, a través de la creación de los establecimientos particulares subvencionados se buscó fomentar la competencia entre los establecimientos por retener a sus estudiantes, ya que se abrían más posibilidades de elección para los padres y estudiantes, con lo que se mejoraría el servicio brindado por los establecimientos, es decir, se mejoraría la calidad de la educación.

Si lo que se busca es saber lo que motiva la elección por parte de las familias, se debe tener en cuenta que requieren información para tomar una decisión, el punto a considerar es qué tipo de información es relevante para los padres al momento de elegir un establecimiento.

Con el fin de ser un medio de información sobre la calidad fue creado el SIMCE, pero luego es natural que surjan interrogantes como: ¿es la calidad un factor importante en la elección?, ¿los padres consideran el puntaje del SIMCE como una medida de calidad?, ¿qué

²⁰ Ver Maradona y Calderón 2004.

²¹ Por ejemplo, ver Gallego et al 2007, Mizala y Romaguera 2000, Mizala y Melis 2007

factores son los que en realidad toman en cuenta para tomar su decisión? Y lo más importante, en consideración al ámbito de este estudio, ¿qué pasa cuando la oferta es reducida, como lo es en el ámbito rural?, ¿cuáles son los factores que afectan la decisión en el mundo rural?

Lo que se ha observado es que mientras más alto el nivel socioeconómico, es más importante la *performance* del establecimiento, mientras que en los niveles socioeconómicos más desaventajados, la calidad es percibida, aparte del rendimiento, un reflejo de la disciplina del establecimiento y la composición socioeconómica de éste.

Las principales metodologías que se utilizan en la literatura están basadas tanto en técnicas cualitativas de investigación de mercados, como entrevistas dirigidas con los padres y *focus group*, como en técnicas econométricas que buscan un modelo de elección de los establecimientos a través de modelos de ecuaciones simultáneas, funciones de utilidad de los padres con respecto al establecimiento, modelos Logit, Probit y Tobit.

A continuación se presenta un resumen que cuantifica las variables encontradas en la literatura y la relación que tienen con la elección hecha por los padres con respecto a la unidad educativa donde envían a sus hijos.

Cuadro 15: Características de la Oferta

Variable	Significado	Estudios que la encuentran significativa	Relación
GSE	Los padres buscarían establecimientos con alumnos que compartan sus características socioeconómicas	3	Positiva dentro de los distintos niveles
Accesibilidad	Cercanía y disponibilidad de medios de transportes o camino para llegar al establecimiento	5	Positiva
Disponibilidad de escuelas y liceos	Existencia de establecimientos que permitan continuar los estudios en un nivel superior	2	Positiva

Fuente: Elaboración propia

Cuadro 16: Características del estudiante y su familia

Variable	Significado	Estudios que la encuentran significativa	Relación
GSE	Características socioeconómicas, como ingresos y ocupación de los padres	8	Positiva
Hogar Monoparental	Generalmente son hogares de madres solteras	3	Positiva
Educación padres	Nivel más alto de educación alcanzado por ambos padres	4	Positiva
Tamaño familiar	Número de integrantes de la familia.	4	Negativa

Variable	Significado	Estudios que la encuentran significativa	Relación
Familiaridad con el establecimiento	Conocimiento del establecimiento a través de algún pariente o conocido que estudia o trabaja ahí	1	Positiva

Fuente: Elaboración propia

Cuadro 17: Características del establecimiento

Variable	Significado	Estudios que la encuentran significativa	Relación
Calidad Enseñanza	Resultados obtenidos en el SIMCE o prueba similar de cada país	8	Positiva
Infraestructura	Incluye infraestructura de calidad, servicios básicos y materiales de aprendizaje	3	Positiva
Disciplina	Normas y regulaciones internas	3	Positiva
Aumento jornada escolar	Establecimientos con mayor número de horas de clases	1	Positiva
Orientación valórica	Orientación religiosa o de otra índole	3	Positiva para los padres que valoran esta característica
Establecimiento bilingüe	Establecimientos que tienen un idioma nativo como lengua de enseñanza	1	Positiva, desde la perspectiva de hijos de indígenas
Estudiantes de minorías étnicas	Presencia de estudiantes indígenas o de otras razas	1	Negativa para los padres que no pertenecen a esas minorías
Costo	Cantidad de dinero para el pago de la mensualidad	3	Negativo
Género Profesor	Influencia del género en la elección del establecimiento, Mujer =1, Hombre =0	1	Positiva

Fuente: Elaboración propia

4.2.1 Aspectos Importantes de Elección

Dentro de las características del mundo rural, existe una tasa de reprobación mucho más alta que en la educación en establecimientos urbanos. Este fenómeno se produce por la escasa oferta de establecimientos en las zonas rurales, por lo que los estudiantes repiten “administrativamente” para seguir recibiendo la subvención, pues al no haber establecimientos donde el estudiante pueda seguir avanzando en los niveles educacionales, es decir, si no hay un liceo o una escuela completa, los padres siguen mandando a sus hijos al mismo establecimiento con la idea de que “algo van a aprender mientras van al colegio”.

Además en el estudio de Mizala y Melis 2007 se demuestra que hay sobre dotación docente y, en algunos casos, sobre oferta, por lo que en conjunto a la problemática anterior, es sumamente importante para el mundo rural el disponer de una oferta real de educación, es decir, con escuelas y liceos donde sus hijos puedan completar sus estudios.

Según el estudio de Gallego 2007 en cuanto a la característica más importante que se da en la ruralidad para la elección del establecimiento, está, como es de esperar, la cercanía y accesibilidad, que en contraste con las localidades urbanas tiene un 60% de importancia versus un 40%.

En estudios realizados en Chile²², se ha determinado que es muy importante el nivel socioeconómico de los estudiantes de un establecimiento para la elección de los padres. Esto se basa en que los padres buscan un establecimiento donde los estudiantes compartan el mismo nivel de ingresos y, sobre todo, el mismo nivel cultural y conductual.

La permanencia de los estudiantes en estos establecimientos se ve afectada por la migración intra rural, la que hace que los estudiantes no completen sus estudios por la movilidad de sus padres, con lo que, lógicamente, la demanda por educación es menor que la esperada, sobre todo en zonas donde se dan muchos trabajos temporales.

Como se menciona anteriormente, la calidad no tan sólo es vista como los resultados de los establecimientos, si no que incluye percepciones de disciplina, como la expulsión de los malos alumnos por ejemplo, la capacidad del colegio para retener y atraer a buenos profesores y otra serie de variables subyacentes.

Finalmente, hay que notar que la demanda por educación depende de las expectativas de los padres y de la función utilidad que ellos tengan, la que está compuesta por los beneficios que reporta versus los costos que involucra el lograrla.

5 ESTIMACIÓN DE VARIABLES QUE INFLUYEN EN EL RENDIMIENTO ACADÉMICO

A través de dos métodos econométricos se busca encontrar cuál es la influencia de diversas variables estructurales en el rendimiento académico de los estudiantes de zonas rurales. Dichas variables son las que se puedan describir a través de las bases de datos proporcionadas por Mineduc que contengan o puedan reflejar las características señaladas como relevantes en el capítulo anterior.

²² Ver Elacqua et. al 2006 y Córdoba, C. 2006

Además, se estudian las variables propias de los estudiantes y sus familia, pues el omitirlas induce a errores de estimación, obteniendo una regresión con menor ajuste y, lo más importante, que no refleja la realidad, pues los estudiantes están insertos en un medio que condiciona su rendimiento como se ha descrito en la literatura.

La variable dependiente es el puntaje obtenido por el estudiante en la prueba SIMCE de matemáticas 2006. Se toma esta variable como representante del rendimiento ya que se ha demostrado que el puntaje en matemáticas es el mejor predictor del capital humano²³, nivel socioeconómico que alcanzan los estudiantes cuando llegan a adultos. Dada esa información se opta por estudiar sólo este resultado, omitiendo del estudio el puntaje obtenido en la prueba de lenguaje.

5.1 Variables y Descripción de la Información Utilizada

Dentro del conjunto de variables encontradas en la literatura como relevantes, se busca el año en que hubiere mayor cantidad de ellas en las bases de datos disponibles desde 2002 a 2007. Esta búsqueda arroja como resultado que el año con mayor cantidad de información es el año 2006, por lo que es ese el año en que se basa este trabajo de título.

Las únicas variables que no son obtenidas directamente de las bases de datos son la tasa alumno profesor, obtenida del cruce de las bases dotación y matrícula; el grupo socioeconómico de la familia, elaborado a través de la técnica de k-medias, los que representan los 5 grupos socioeconómicos existentes en población. Esta técnica es la misma que utiliza el Mineduc para calcular el grupo socioeconómico del curso en la prueba SIMCE; y, por último, a modo de asignar a los establecimientos un indicador de accesibilidad, se calcula el índice de exceso de viaje, que refleja el porcentaje de alumnos que viaja más allá de cierto límite de tiempo, dependiendo del medio de transporte utilizado. La información está disponible para 2004 a través de la pregunta en qué medio viaja y cuánto tiempo demora su hijo en llegar al establecimiento hecha a los padres de 8° Básico. Si bien estos alumnos son mayores que los de 4° Básico, su tiempo de viaje muestra una cota superior, pues al ser mayores, estos alumnos pueden viajar un tiempo superior, con lo que se puede obtener en el extremo, la peor situación de accesibilidad. La determinación de dicho estimador y de los grupos socioeconómicos se encuentra disponible en el Anexo I.

A continuación se presentan las variables utilizadas, sus significados, fuentes y valores para luego hacer un análisis descriptivo de la base a utilizar.

Cuadro 18: Variables utilizadas en el estudio

Variable	Significado	Fuente	Valores
Disponibilidad de Recursos	Disponibilidad declarada por el profesor en cuestionario SIMCE sobre textos escolares, textos de lectura, libros de consulta, diarios y revistas, Internet, recursos multimedia, recursos audiovisuales, mapas, biblioteca, juegos didácticos, etc.	Cuestionario Profesor SIMCE 2006	Entre 0 y 1
Índice de Vulnerabilidad	% de estudiantes en riesgo de deserción.	InfoEscuela	Entre 0 y 1

²³ Ver Hanushek y Kimko (2000) y Barro (2001)

Variable	Significado	Fuente	Valores
Municipal	Establecimiento de dependencia municipal, 1; particular subvencionado o pagado, 0.	InfoEscuela	1 ó 0
Selección	Si el establecimiento hace examen de conocimientos para recibir a los estudiantes, 1; si no, 0.	Cuestionario Padres SIMCE 2006	1 ó 0
GSE Curso	Grupo socioeconómico al que pertenece en promedio el curso.	Resultados SIMCE 2006	Entre 1 y 5
Exceso de Viaje	% de estudiantes que traspasan el umbral de viaje según medio de transporte.	Elaboración Propia	Entre 0 y 1
Cantidad de grados	Grados en una misma aula.	RECH	Entre 1 y 6
Tamaño curso	Estudiantes por curso.	RECH	>1
Alumnos indígenas curso	Fracción de estudiantes del curso descendientes de algún pueblo originario.	RECH	Entre 0 y 1
Ed. Secundaria	Si el establecimiento ofrece educación media o técnico profesional 1; si no 0.	InfoEscuela	1 ó 0
Matrícula	Cantidad de estudiantes del establecimiento.	Matrículas	>1
% Retirados	% de estudiantes que abandonaron sus estudios en el establecimiento.	Rendimiento	Entre 0 y 1
Centro de padres	Si el establecimiento tiene centro de padres 1; si no 0.	InfoEscuela	1 ó 0
Horas jornada escolar	Cantidad de horas de clases declarada por el establecimiento.	RECH	Entre 4 y 9,5
Jornada escolar completa	Si la jornada dura 7:30 horas o más, 1; si dura menos, 0.	RECH	1 ó 0
Pago mensual	Cantidad de dinero que paga mensualmente por la colegiatura.	Cuestionario Padres SIMCE 2006	≥ 0
Género Profesor	Si es mujer, 1; si es hombre, 0.	Cuestionario Profesor SIMCE 2006	1 ó 0
Antigüedad Profesor	Cantidad de años que el profesor lleva en el establecimiento.	Cuestionario Profesor SIMCE 2006	≥ 0
Expectativas Profesor	Nivel educacional al que espera que lleguen sus estudiantes.	Cuestionario Profesor SIMCE 2006	1: No termina la Ed. básica; 2: Completa la Ed. Básica, 3: No terminará la Ed. Media; 4: Terminará la Ed. Media en un Técnico Profesional; 5: Terminará la Ed. Media en un Científico Humanista; 6: Completará carrera técnica; 7: Completará universidad; 8: Completará un postgrado.

Variable	Significado	Fuente	Valores
Uso de Recursos	Frecuencia de uso de los recursos disponibles.	Cuestionario Profesor SIMCE 2006	Entre 0 y 1
Tasa Alumno-Profesor	Razón entre estudiantes del establecimiento y docentes de aula.	Elaboración Propia	Entre 0 y 1
Rendimiento	Puntaje obtenido por el estudiante en SIMCE matemáticas.	Resultados SIMCE 2006	>70
Género Alumno	Si es mujer, 1; si es hombre, 0.	Cuestionario Padres SIMCE 2006	1 ó 0
Educación padre	Años de estudios.	Cuestionario Padres SIMCE 2006	Entre 0 y 21
Educación madre	Años de estudios.	Cuestionario Padres SIMCE 2006	Entre 0 y 21
Ingresos	Ingreso mensual promedio de la familia.	Cuestionario Padres SIMCE 2006	Entre 100.000 y 2.000.000
Tamaño Familiar	Número de integrantes de la familia.	Cuestionario Padres SIMCE 2006	≥ 2
Hogar Incompleto	Si el estudiante no tiene padre o madre, 1; si no, 0.	Cuestionario Padres SIMCE 2006	1 ó 0
GSE Alumno	Grupo socioeconómico al que pertenece el estudiante.	Elaboración Propia	1 a 5
Padres indígenas	Si alguno de los padres pertenece a algún pueblo originario.	Cuestionario Padres SIMCE 2006	1 ó 0
Expectativas padres	Nivel educacional al que espera que llegue su hijo.	Cuestionario Padres SIMCE 2006	Los mismos del profesor
Satisfacción Calidad	Satisfacción de los padres con respecto a la calidad del establecimiento.	Cuestionario Padres SIMCE 2006	1: muy insatisfecho; 2: insatisfecho; 3: satisfecho; 4: muy satisfecho

Fuente: Elaboración propia.

Las características de la base a utilizar²⁴ muestran que el promedio de los establecimientos en el SIMCE de Matemáticas es de 234 puntos, mientras que a nivel de estudiantes el promedio sube a 239, pero con una mayor dispersión.

En general los establecimientos tienen recursos insuficientes, tienen una gran proporción de estudiantes en riesgo de deserción escolar, son principalmente de dependencia municipal, sin

²⁴ Ver Anexo II

educación secundaria, con cerca de un 3% de retirados y su nivel socioeconómico está entre el estrato bajo y medio bajo.

A nivel de curso, tienen un promedio de 26 alumnos y son combinados, tienen más niños que niñas, sus profesores son mujeres principalmente, tienen una jornada de menos de 7 horas.

En cuanto a las familias, la educación del padre es un poco menor que la de la madre, quien tiene 9,2 años de estudios frente a los 8,5 años de él. El ingreso familiar promedia unos \$234.000, las familias tienen 5 integrantes y el 5% de los hogares no tienen padre o madre, mientras que el 13% tiene ascendencia de un pueblo originario.

5.2 Regresión Lineal

Con este método se prueban cuatro versiones de un modelo de regresión lineal para el rendimiento, donde se prueban diversas combinaciones de variables para llegar a un modelo más parsimonioso, es decir, se busca un modelo con menos variables que presente un ajuste similar a los otros modelos.

Además, se tiene otro modelo donde se incluyen los términos cuadráticos que logren un coeficiente de determinación mayor que la relación lineal, es decir, se incluyen las variables presentan un mejor ajuste al ser expresadas como un polinomio cuadrático.

Finalmente, se utiliza k-Medias para realizar diversas segmentaciones y estudiar el perfil de los estudiantes de cada categoría como forma de ahondar en los distintos perfiles de cada segmentación.

5.2.1 Especificación de Modelos

- **Modelo I: Modelo Base**

Este modelo es especificado con las variables linealmente relacionadas con el rendimiento. En este caso se tiene un modelo con todas las variables disponibles,

- **Modelo II: Modelo Base Reducido**

Esta versión contiene las variables significativas en el modelo general y variables seleccionadas según su correlación.

- **Modelo III: Modelo Factorial**

Se trata de un modelo donde las variables son elegidas con análisis factorial, conservando las que aportan mayor variabilidad al modelo.

- **Modelo IV: Modelo Factorial Reducido**

Es un modelo con las variables anteriores reducidas según la correlación entre ellas.²⁵

²⁵ El análisis factorial y dichas correlaciones se encuentran disponibles en el anexo III

- **Modelo con Relaciones no Lineales**

De existir relaciones no lineales entre las variables explicativas y el rendimiento, el ajuste de un modelo que omita esas relaciones no lineales es menor que el modelo que sí las contempla. Debido a esto es que se estudian este tipo de relaciones que pudiesen existir para incluirlas en un nuevo modelo, es decir, su ajuste con respecto al rendimiento es mejor al incluir la variable y la variable al cuadrado o la variable como función logarítmica de la misma.

5.2.2 Resultados Obtenidos

Con los modelos estructurados como se señala en el apartado anterior, se ejecuta la aplicación del Software SPSS para el cálculo de cada uno de ellos.

A continuación se presentan los modelos con las variables utilizadas en cada uno de ellos, donde las que presentan sus coeficientes β significativas al 5% se encuentran destacadas y la desviación estándar se muestran en paréntesis.

Cuadro 19: Resultados Modelos Lineales

VARIABLES	Modelo			
	I	II	III	IV
(Constante)	185,31 (8,38)	172,95 (7,62)	148,27 (5,52)	147,01 (5,32)
Disponibilidad Recursos	6,47 (2,97)	3,34 (2,69)		
Índice de Vulnerabilidad	-0,17 (0,07)	-0,17 (0,07)	-0,12 (0,04)	-0,13 (0,03)
GSE Curso	-0,21 (1,30)	-0,25 (1,28)		
Exceso de Viaje	37,19 (6,61)	35,61 (6,59)		
Cantidad de grados	-3,06 (0,977)	-3,10 (0,96)	-4,81 (1,01)	-4,67 (0,95)
Tamaño Curso	-0,17 (0,07)	-0,12 (0,07)	-0,22 (0,08)	-0,21 (0,08)
% Alumnos indígenas curso	-8,43 (2,12)	-8,38 (2,12)	-6,9 (1,98)	-6,57 (1,97)
Género Profesor	-0,34 (1,33)	-0,2 (1,32)	-0,01 (0,92)	
Antigüedad Profesor	0,18 (0,05)	0,18 (0,05)	0,19 (0,05)	0,19 (0,05)
Expectativas Profesor	2,01 (0,43)	1,78 (0,43)	2,39 (0,32)	2,35 (0,32)
% Alumnos retirados	-37,81 (14,82)	-35,40 (14,79)		
Centro de Padres	1,97 (1,99)	2,08 (1,95)		
Horas Jornada Escolar	0,0 (0)	0,0 (0)		
Matrícula	-0,01 (0,004)	-0,01 (0,003)	-0,01 (0,002)	-0,01 (0,002)

VARIABLES	Modelo			
	I	II	III	IV
Selección de alumnos	8,72 (2,24)	8,34 (2,15)	12,29 (3,10)	13,1 (2,93)
Municipal	0,29 (1,72)			
Pago Mensual	-0,06 (0,05)		-1,01 (1,14)	
Uso Recursos	-13,34 (7,53)		-5,10 (3,27)	
Biblioteca	-2,68 (1,38)			
Educación Secundaria	-2,99 (1,91)		-1,64 (1,47)	
Razón Alumno/Profesor			-0,16 (0,08)	-0,20 (0,08)
Alumnos por computador	-0,14 (0,09)		-0,02 (0,03)	
Género Alumno			-2,82 (0,75)	-2,88 (0,75)
Educación padre	0,98 (0,18)	0,66 (0,151)	0,84 (0,13)	0,88 (0,12)
Educación madre	1,38 (0,18)	1,54 (0,174)	1,49 (0,13)	1,53 (0,13)
Ingresos	0,01 (0,003)	0,01 (0,002)	7,17 (3,95)	
Expectativas Padres	7,12 (0,35)	7,36 (0,347)	6,92 (0,26)	6,97 (0,26)
Satisfacción Calidad	5,6 (0,71)	5,49 (0,7)	16,12 (1,35)	16,28 (1,33)
Tamaño Familiar	-1,9 (0,31)		-1,97 (0,24)	-1,94 (0,24)
Hogar Incompleto	10,39 (2,55)		10,09 (2,05)	10,33 (2,05)
Padres indígenas	-0,82 (1,71)		-2,77 (1,24)	-2,87 (1,24)
N	10814	10814	10814	10814
R ²	0,150	0,145	0,149	0,149
Test F	65,70	91,21	187,27	127,70

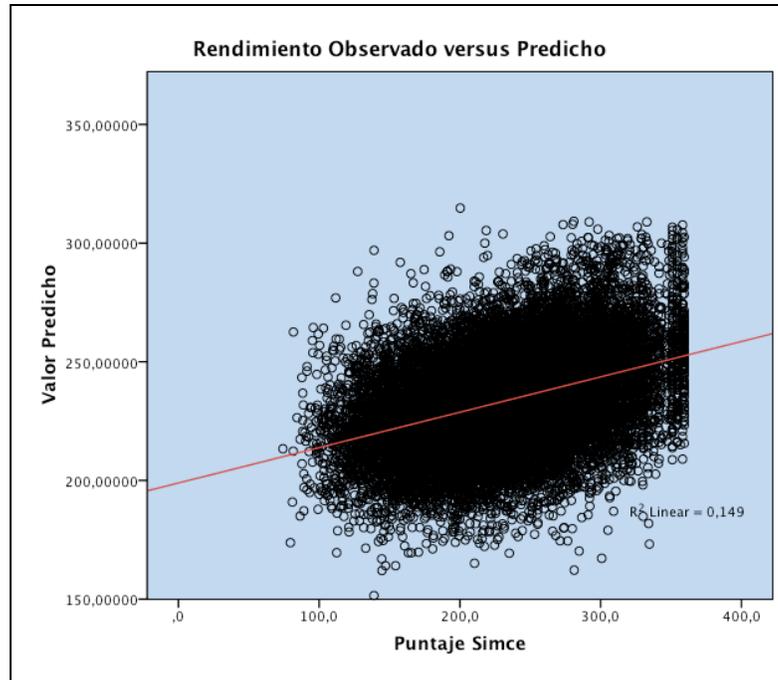
Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

En primer lugar, se tienen modelos con un ajuste bastante pequeño, sin embargo, todos tienen coeficientes de determinación significativos como lo muestra el test F. Al centrarse en un modelo parsimonioso, se tiene que el mejor de ellos es el IV, presentando un R² muy similar al modelo base con menos variables y un test F mucho mayor.

Si bien el ajuste de estos modelos es bajo, en la literatura se comprueba que en este tipo de investigaciones los resultados son similares, lo que plantea la inquietud sobre qué tan válido es aplicar técnicas matemáticas para representar fenómenos sociales. Al observar el comportamiento

de los modelos frente a los datos reales se puede comprobar que la nube de puntos a ajustar tiene gran dispersión, con lo que se evidencia lo errático de los datos. En el siguiente gráfico se muestra el comportamiento de rendimiento observado frente al pronosticado.

Figura 2: Rendimiento Observado versus Predicho



Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

Como se observa en el gráfico, los datos que se ajustan son sumamente pocos, esto es debido a que se trata de un estudio multivariado, y, por sobre todo, a que en el modelo se omiten variables importantes, pero no cuantificables, como lo son el talento innato de cada estudiante, la disciplina del establecimiento, el tipo de enseñanza, el proyecto educativo del establecimiento, exigencia de los padres, etc. Bajo estas circunstancias es que el modelo trata de ajustar datos que tan sólo tienen una parte de la información necesaria para poder predecir cómo será el rendimiento de cada estudiante.

Si esta investigación tratase acerca de un tema comercial, de gestión o de cualquier otra índole netamente ingenieril, claramente el modelo encontrado no tendría validez, pero al tratarse de un tema social, que involucra a personas e instituciones con distintas características tangibles e intangibles, hace sentido que el modelo logrado sólo pueda replicar una fracción de la realidad. Por lo tanto, es esa fracción la que trata de ser analizada y entendida a través de estos modelos a pesar de la falta de ajuste en que incurrir.

Al analizar las variables se tiene que existe un conjunto de ellas que son significativas en todos los modelos, además de tener signos y magnitudes concordantes en todos los modelos, con lo que se pueden tomar como válidas en cuanto a su influencia respecto al rendimiento. En base a aquellas variables se realiza un análisis de quiénes y cómo se relacionan con el rendimiento.

El resultado más importante de estos modelos es que en todos ellos la cantidad de grados combinados en una misma aula presenta un comportamiento negativo, siendo categórico este resultado al contrario de la ambigüedad de opiniones encontradas en la literatura. En el modelo

no lineal este punto es analizado en profundidad al tratar el comportamiento curvilíneo de esta relación.

El índice de vulnerabilidad resulta negativo, lo que es coherente, pues representa el porcentaje de alumnos en riesgo de deserción, lo cual está muy ligado con el nivel socioeconómico, por lo que no es de extrañar que el GSE del curso no resulte significativo ya que tienen una correlación alta entre ellos²⁶ y presenta menos variabilidad al ser una variable categórica en vez de una continua como lo es el índice de vulnerabilidad.

La proporción de alumnos indígenas en un curso, provoca un menor rendimiento entre sus integrantes. Esto es provocado por el rendimiento inferior de los niños indígenas respecto a sus compañeros, ya que ellos presentan desventajas propias de este sector de la población por el hecho de pertenecer a una minoría étnica, como por ejemplo, hablar un idioma distinto al que se utiliza en clases, tener padres con un capital humano que se considera menor que el de personas que no pertenecen a estas minorías, etc., lo que se comprueba directamente al ver que el tener padres indígenas resulta significativo y negativo en dos de los tres modelos donde fue incluida esta variable. El hecho de que estos estudiantes tengan un rendimiento menor afecta a todos los integrantes del curso, pues se retrasa el rendimiento de todos al tener que esperar a los menos aventajados. Esto es conocido como el efecto *peer*, el que implica que el rendimiento de un estudiante también depende de el rendimiento de sus compañeros.

El que un establecimiento tenga procesos de selección de sus estudiantes a través de sus capacidades cognitivas, hace que sus estudiantes tengan un mayor rendimiento, pero también tiene contribuye con el efecto *peer* explicado anteriormente, ya que sus estudiantes se benefician de tener compañeros con capacidades altas y similares a las de ellos en favor de su rendimiento. Se podría pensar que la selección sólo ocurre en establecimientos particulares subvencionados o pagados, pero a nivel municipal esto también sucede cuando los establecimientos tienen una capacidad menor a su demanda.

Con respecto a las características del profesor, son sus expectativas y la antigüedad que tiene en el establecimiento las que favorecen un mejor rendimiento. Al referirse a las expectativas es el profesor quien da su opinión del nivel educacional que alcanzarán sus estudiantes, lo que refleja de algún modo las capacidades de los niños y las oportunidades de educarse que tienen a futuro. En este caso, muchos profesores saben que el proseguir estudios de enseñanza media muchas veces depende de los recursos económicos que tengan sus padres para enviarlos a un poblado más grande o a alguna ciudad cercana. Finalmente, esto podría implicar que el profesor puede perder la motivación por enseñar a estudiantes “sin futuro” lo que merma, aún más, las bajas posibilidades que tienen los estudiantes con menos oportunidades.

En el caso de la antigüedad del profesor en el establecimiento se dan fenómenos en dos áreas, la primera tiene que ver con la experiencia del maestro en la enseñanza, lo que hace que esté mejor preparado para dictar sus clases y ayudar a los estudiantes, y la segunda es que el profesor al permanecer más tiempo en la misma escuela se involucraría más con la comunidad, con lo que su trabajo tendría un mayor compromiso.

Las variables familiares que más influyen son sin duda la educación de los padres, siendo la más importante la educación de la madre, pues es habitual que sea ella quien más se preocupe de ayudar a los niños en sus deberes escolares y quien tenga más contacto con el establecimiento

²⁶ Ver Anexo III

y la educación que sus hijos reciben. Además, el hecho de que los padres tengan una mayor expectativa del nivel que alcanzarán sus hijos refleja las capacidades de los estudiantes, la calidad del establecimiento, las oportunidades que tiene el estudiante en cuanto a llegar a niveles más altos, entre otras situaciones, por lo que mayores expectativas tendrían un efecto positivo en el rendimiento del estudiante, pues aparte de contar con mayores capacidades y oportunidades, contarían con un mayor apoyo familiar. Así mismo, como la satisfacción con la calidad del establecimiento tiene una relación positiva con el rendimiento, si bien tiene diversas implicancias, hace pensar que una de ellas es que el establecimiento sí importa.

A través de los distintos modelos hay una serie de variables que presentan significancia sólo algunas veces o sólo están presentes en algunos de ellos, por lo que sacar conclusiones a través de ellas podría resultar equívoco, sobre todo al fijarse en los coeficientes β . Sin embargo, se podría afirmar que hay indicios de que la cantidad de estudiantes en un curso tiene una relación negativa con respecto al rendimiento, al igual que la razón alumno profesor, variables que no deben ser confundidas, pues los establecimientos al tener grados más altos, requieren de profesores especializados por área, por lo que la razón profesor por curso no es uno a uno. En cuanto a la matrícula es difícil establecer algún tipo de relación, pues a pesar de resultar negativa en todos los modelos, su coeficiente es muy pequeño y esta variable tiene alta correlación con otras presentes en los modelos tales como la selección de estudiantes.

En el caso del exceso de viaje se observa significativo en los dos modelos que lo utilizan. Puede sorprender la magnitud de sus coeficientes, pero hay que recordar que el rango de esta variable es entre 0 y 1. Finalmente, que no se puede concluir directamente que a mayor viaje mayor rendimiento, pues esta variable puede ser un efecto más que una causa del rendimiento.

Cabe destacar el comportamiento de la variable hogar incompleto, pues presenta un efecto positivo y de bastante envergadura. Lo que se puede inferir a través de esta situación es que estos estudiantes tendrían más atención de ambos padres al estar separados o de parte del sostenedor del hogar, padre o madre, en cuanto a su educación. La lectura de este resultado es errada si se piensa que la causa de un mejor rendimiento es la separación de los padres, pues en el fondo lo que reflejaría esta variable es el involucramiento de los padres con respecto a la educación de sus hijos.

A continuación se presentan los resultados obtenidos para el modelo no lineal. Para una mejor comprensión del lector se presentan los coeficientes de las variables simples y cuadráticas en la misma fila.

Cuadro 20: Resultados modelo no lineal

VARIABLES	Modelo no Lineal	
	Variable lineal	Variable ²
(Constante)	218,53 (9,74)	
Disponibilidad de Recursos	9,99 (10,86)	-3,6 (9,543)
Índice de Vulnerabilidad	-0,51 (0,16)	0,003 (0,001)
Cantidad de grados	-5,63 (1,52)	0,73 (0,22)

VARIABLES	Modelo no Lineal	
	Variable lineal	Variable ²
Tamaño Curso	-6,05 (9,51)	-5,39 (8,05)
% Alumnos indígenas curso	-8,77 (2,07)	
Género Profesor	0,66 (0,94)	
Antigüedad Profesor	43,19 (5,99)	-49,45 (9,09)
Expectativas Profesor	4,65 (1,89)	-0,14 (0,2)
Log(Matricula)	-0,67 (0,12)	
Selección de alumnos	12,43 (3,24)	
Municipal	-0,92 (1,26)	
Pago Mensual	-38,15 (9,51)	35,52 (10,9)
Uso Recursos	47,0 (19,22)	-50,31 (17,86)
Educación Secundaria	-3,03 (1,41)	
Razón Alumno/Profesor	-0,67 (0,33)	0,01 (0,01)
Género Alumno	-1,89 (0,77)	
Educación padre	0,96 (0,36)	0,02 (0,02)
Educación madre	1,44 (0,43)	0,05 (0,02)
Ingresos	4,47 (0,77)	-3,36 (1,69)
Expectativas Padres	7,53 (1,31)	-0,09 (0,09)
Satisfacción Calidad	20,3 (10,8)	
Tamaño Familiar	-0,58 (1,14)	-0,13 (0,09)
Hogar Incompleto	12,58 (2,289)	
Padres indígenas	-1,89 (1,27)	
N	10814	
R ²	0,155	
Test F	66,75	

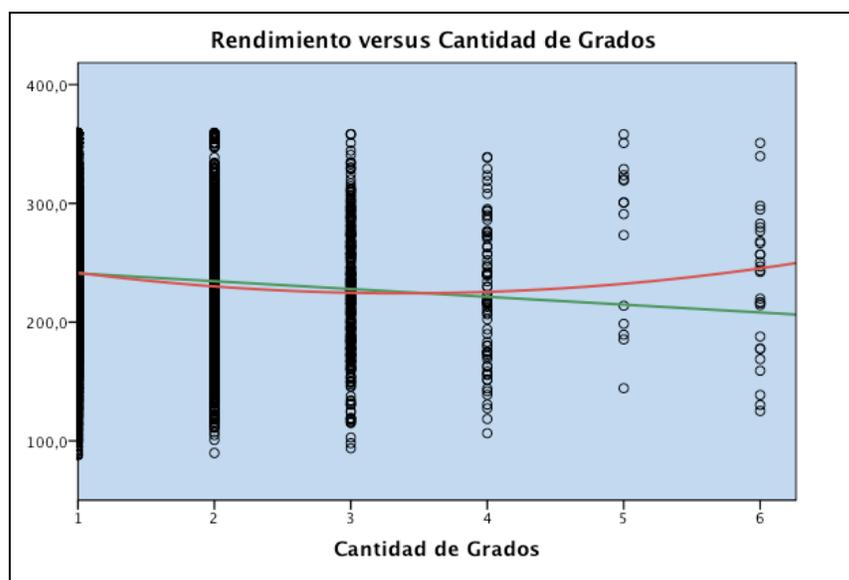
Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

Este modelo también tiene un escaso ajuste, pero en términos del R^2 es el mejor. En él se ve claramente la relación que tiene el índice de vulnerabilidad con respecto al rendimiento, pues si bien linealmente es negativo, aquí se ve que su función cuadrática es significativa y presenta una forma convexa, indicando que al estar en los extremos se tienen mejores resultados. Esto es evidente en el caso de tener el índice en cero o en un valor pequeño, pero ¿qué pasa en el otro extremo? Si bien no es posible afirmar que los resultados de establecimientos en este tramo tengan iguales resultados que los del extremo inferior, sí se puede entender que en comparación con quienes están en el rango medio ellos obtienen mejores resultados.

El cómo sucede esto es explicado por la cantidad de programas sociales a los que se adscriben los establecimientos en estas condiciones, recibiendo mayores beneficios para educar a los hijos más necesitados del país, quienes a su vez, al estar en riesgo reciben apoyo de programas y ayudas familiares impulsadas por el gobierno. Dadas esas circunstancias es que en relación a la parte media de la curva, los estudiantes de estos establecimientos obtendrían un mejor rendimiento.

La cantidad de grados combinados tiene un comportamiento similar, siendo los no combinados y los con más grados combinados quienes tendrían mejor rendimiento que los que se ubican en el tramo medio. Esto se debe a que los cursos con mayor cantidad de grados combinados están conformados por grados con pocos alumnos, donde prácticamente los alumnos tienen educación personalizada dadas las condiciones del aula donde el profesor no puede hacer su clase en la forma tradicional, teniendo que enseñar y supervisar a cada nivel por separado, con lo que recibirían un apoyo e involucramiento del profesor mucho mayor que en una clase tradicional. La relación lineal y cuadrática de los datos se verifica al observar el gráfico de dispersión con las relaciones lineales y cuadráticas de la cantidad de grados y el rendimiento, el cual una vez más muestra que el ajustar los datos a una función específica es bastante difícil por lo erráticos que éstos resultan. Sin embargo, al ser significativas ambas componentes de la relación cuadrática, se puede establecer que los grados combinados presentan un comportamiento convexo en relación al rendimiento de sus estudiantes.

Figura 3: Rendimiento según cantidad de grados

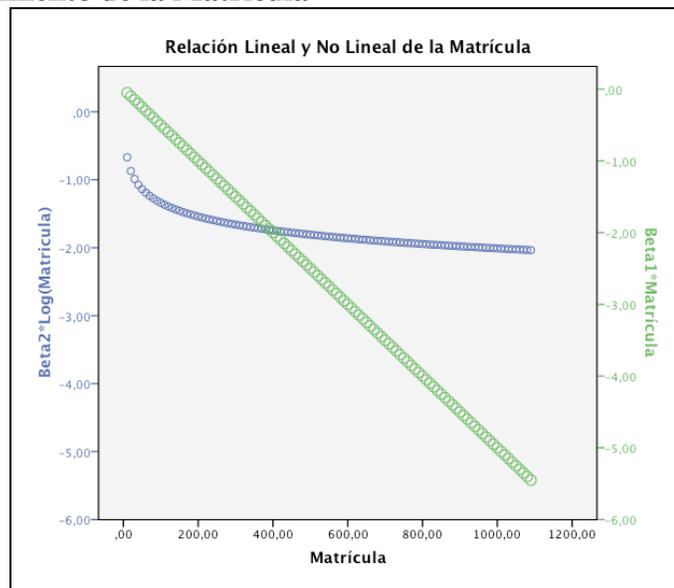


Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

Con este modelo se logra aclarar lo que sucede con el uso de recursos, pues al incluir su componente cuadrática se vuelven ambas variables significativas, dejando fuera a la disponibilidad de recursos. El comportamiento del uso de recursos es una función cóncava, lo que significa que hay un máximo en que el uso de ellos aumenta el rendimiento y por mucho que aumentase el uso de recursos, llega un momento en que no es posible aumentar el rendimiento por medio de esta variable. Además, cuando se está más arriba en la curva, el aumentar el uso tiene un efecto menor en el aumento del rendimiento que cuando se está al inicio de la curva, pero ¿en qué se traduce esto? En que el aumento del uso, y lógicamente, la disponibilidad, en los estudiantes que tienen menor acceso a ellos tiene una repercusión superior, por lo que al dotar de recursos a las escuelas más pobres o de peor rendimiento influye positivamente en el rendimiento de sus estudiantes.

Finalmente, a través de este modelo se logra encontrar una nueva forma de relacionar la matrícula con el rendimiento, ya que ésta se relaciona significativamente al aplicar la función Logarítmica, con lo que se observa que la disminución en el rendimiento de los estudiantes al aumentar la matrícula no es tan drástica como lo es en el modelo lineal. Esto queda demostrado al graficar ambas funciones de la matrícula con sus coeficientes β respectivos.

Figura 4: Comportamiento de la Matrícula



Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

Gracias al gráfico anterior se constata que la relación lineal de la matrícula es mucho más drástica que la logarítmica. Además, una disminución lineal del rendimiento resulta poco intuitivo, por lo que la relación que se logra establecer en este modelo no lineal es más sensata, pues el cambio marginal de integrar a un nuevo estudiante al establecimiento no debiera ser el mismo para todos los tamaños del establecimiento, y el modelo completo presenta un mejor ajuste que los modelos lineales estudiados anteriormente, con lo que se puede establecer que este modelo genera un aporte frente a lo encontrado en los modelos lineales.

5.2.3 Segmentaciones estudiadas

5.2.3.1 Segmentación por grupo socioeconómico

Para analizar detenidamente lo que sucede según el nivel socioeconómico se presenta la siguiente segmentación, donde se utiliza el método de pasos sucesivos para introducir las variables, el que permite tener un modelo sólo con variables significativas. De esta forma se elimina el exceso de información para claridad del lector. Cabe recordar que a medida que aumenta el nivel socioeconómico, se asciende en la numeración de los GSE.

Cuadro 21: Modelo por GSE

Variables	GSE 1	GSE 2	GSE 3	GSE 4	GSE 5
(Constante)	173,82 (6,051)	128,86 (13,22)	159,08 (26,34)	139,56 (46,54)	102,14 (25,68)
Disponibilidad Recursos	1,95 (2,9)			54,18 (22,32)	
Índice de Vulnerabilidad	-0,18 (0,05)				
Exceso de Viaje	35,24 (6,91)				
Cantidad de grados	-3,05 (1,03)				
Tamaño curso	-0,16 (0,08)				
Alumnos indígenas curso	-0,34 (0,11)	-1,02 (0,49)			
Antigüedad	0,16 (0,05)	0,48 (0,18)			
Expectativas profesor	1,68 (0,45)	3,59 (1,35)			
% Retirados	-26,65 (15,53)				-31,2 (19,32)
Matrícula	-0,01 (0,003)				
Educación padre	0,59 (0,16)				
Educación madre	1,43 (0,19)	1,51 (0,57)	3,74 (1,09)		
Ingresos	0,03 (0,01)			0,07 (0,03)	
Selección	7,59 (2,54)		22,0 (6,94)		
Expectativas Padres	6,99 (0,37)	10,10 (1,54)	8,8 (4,28)	18,22 (3,16)	27,23 (3,45)
Satisfacción Calidad	5,51 (0,75)	7,87 (2,50)			
N	9557	819	171	92	175
R ²	0,103	0,13	0,219	0,405	0,333
Test F	68,68	20,3	15,6	19,94	42,96

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

El principal resultado obtenido a través de esta segmentación es que a medida que se sube de GSE el ajuste es mayor, lo cual se debe a que la variabilidad de los datos es mucho menor en los niveles más ricos, es decir, los individuos suelen ser más homogéneos en la variable explicada y en la explicativa, por lo que se ajustan mucho mejor al modelo de su segmento.

Además, al ir de los más pobres a los más ricos van ausentándose las variables que tienen que ver con los establecimientos, lo que podría reflejar que el establecimiento importa mucho más para los estudiantes más pobres, pues es en los segmentos más desposeídos donde sus características se hacen presentes en el modelo. Esto muestra que al mejorar las condiciones de los establecimientos que atienden a los más pobres, su rendimiento se verá directamente afectado positivamente. Además, esto se da porque a medida que se aumenta de GSE los establecimientos suelen ser más homogéneos en cuanto a sus características, por lo que dentro de cada grupo ellas no serían las responsables de los cambios en el rendimiento.

5.2.3.2 Segmentación por rendimiento

Esta segmentación busca establecer qué factores hacen que los estudiantes pasen de un nivel a otro de rendimiento. De igual forma que con la segmentación anterior, se utiliza el método de pasos sucesivos para obtener un modelo sólo con las variables significativas. A continuación se presenta la caracterización de la segmentación para luego dar paso al análisis de los modelos obtenidos

Cuadro 22: Segmentos según rendimiento

	Rendimiento			
	Bajo	Medio bajo	Medio alto	Alto
Rendimiento	155	211	259	314
N estudiantes	1.746	3.258	3.594	2.216

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

Cuadro 23: Modelo por segmento de rendimiento

Variables	1	2	3	4
(Constante)	137,9 (4,42)	253,36 (7,35)	200,07 (1,52)	297,34 (2,83)
Disponibilidad de Recursos			5,09 (1,43)	
Disponibilidad ²				5,45 (2,35)
Exceso de viaje		11,09 (3,46)		
Cantidad de grados		-0,98 (1,48)		
Tamaño Curso	0,67 (0,37)			
Tamaño curso ²	-0,01 (0,01)			
% Retirados ²		-179,89 (67,74)		
Antigüedad	0,10 (0,05)			

Variables	1	2	3	4
Antigüedad ²				0,003 (0,001)
Género Alumno			-1,11 (0,53)	-2,93 (0,91)
Educación padre	0,53 (0,12)			0,27 (0,12)
Educación madre		0,25 (0,08)	0,25 (0,09)	
Tamaño familiar		-0,42 (0,16)		
Hogar incompleto	5,23 (2,55)			
Expectativas Padres	1,10 (0,35)	0,62 (0,18)	0,70 (0,18)	1,45 (0,33)
Satisfacción Calidad		1,02 (0,36)	0,96 (0,37)	1,68 (0,67)
N	1.746	3.258	3.594	2.216
R ²	0,025	0,020	0,021	0,031
Test F	6,89	9,93	12,97	10,91

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

Con esta segmentación no es mucho lo que se logra aclarar sobre el rendimiento, pues el modelo para cada segmento presenta un ajuste excesivamente bajo pues se segmenta sobre la variable explicada, con lo que la varianza de ella es poca dentro de un segmento con características heterogéneas. A pesar de aquello, se puede rescatar que dentro del grupo de menor rendimiento tendrían un mejor desempeño al tener cursos de no más de 24 estudiantes, pues el máximo de la función tamaño del curso se obtiene al derivar la expresión dada por los coeficientes β :

$$\text{máx } y(x) = \frac{d(0,67x - 0,014x^2)}{dx} = 0$$

$$0,67 = 0,028x \Rightarrow x = 23,929$$

dado este resultado, no sería conveniente conformar cursos con más de 24 estudiantes para que los estudiantes con más dificultades puedan potenciar sus capacidades.

5.2.3.3 Segmentación por jornada escolar

Si bien en los diversos modelos la extensión de la jornada escolar no ha tenido resultados significativos, es importante ahondar más en ella, pensando que uno de los objetivos de la reforma educacional es contar con jornada escolar completa, JEC, en todo los establecimientos del país.

Cuadro 24: Modelos según jornada escolar

Variables	Con JEC	Sin JEC
(Constante)	196,34 (9,83)	208,95 (7,06)
Disponibilidad ²	34,93 (9,19)	
Índice de vulnerabilidad		-0,12 (0,05)
Exceso de viaje	36,33 (14,35)	33,16 (7,54)
Cantidad de grados	-19,3 (6,57)	-2,15 (1,1)
Cantidad grados ²	2,95 (1,28)	
Tamaño curso		-1,29 (0,34)
Tamaño curso ²		0,02 (0,01)
Alumnos indígenas curso		-0,27 (0,12)
Género Profesor	6,87 (3,44)	
Antigüedad		0,23 (0,05)
Expectativas Profesor		2,90 (0,47)
% Retirados ²	-650,16 (234,69)	
Uso de recursos ²	-65,5 (21,63)	
Género Alumno		-4,06 (1,11)
Expectativas Padres	10,26 (0,87)	6,69 (0,39)
GSE alumno	6,17 (1,22)	9,4 (0,85)
Tamaño familiar		-2,10 (0,34)
Hogar incompleto		11,62 (2,61)
Satisfacción Calidad	3,71 (1,74)	6,33 (0,79)
N	1874	8940
R ²	0,224	0,13
Test F	44,29	89,32

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

A través de esta segmentación se logra un modelo con el mejor ajuste de todos en el caso de los estudiantes que cuentan con JEC. Esto se puede deber a que los estudiantes que cuentan con

JEC son menos, presentando datos con menor variabilidad, lo que hace posible alcanzar un mayor ajuste.

En cuanto a las diferencias que presentan ambos grupos cabe notar que los estudiantes con JEC se ven influenciados por más características del establecimiento que los estudiantes sin JEC, lo que abre espacio para un nuevo ámbito de investigación, pues al parecer, los estudiantes que pasan más tiempo en sus establecimientos logran superar algunas diferencias dadas por sus características socioeconómicas, lo que, de ser posible, sería una gran señal para que dichas diferencias se pudiesen superar dando una educación de en un ambiente de equidad para los niños más desposeídos.

5.3 Modelo Lineal Jerárquico

Para la aplicación de este modelo, se aplica el mismo conjunto de variables utilizadas en la regresión lineal y se utiliza el Software HLM, el que está especialmente diseñado para modelos jerárquicos.

Los modelos que son utilizados en esta fase son una combinación del modelo base y factorial, más el modelo no lineal utilizando variables estructurales, siendo además estudiadas las segmentaciones socioeconómicas y según tipo de jornada.

El procedimiento a seguir es el descrito en el marco teórico para determinar la variabilidad atribuible a los establecimientos, variables significativas del primer nivel y, finalmente, el efecto establecimiento.

A continuación se presenta la especificación matemática de los modelos y el resultado obtenido de ellos.

5.3.1 Especificación de Modelos

- **Modelo incondicional**

El primer modelo a probar es el incondicional, donde no hay variables explicativas en ninguno de los dos niveles, por lo que tiene la siguiente formulación:

$$Y = \beta_0 + r$$

$$\beta_0 = \gamma_{00} + \mu_0$$

- **Modelo con variables familiares**

Este modelo se construye para encontrar las variables significantes del estudiante y su familia, así el modelo queda especificado como:

$$\begin{aligned}
Y &= \beta_0 + \beta_1(\text{GéneroEstudiante}) + \beta_2(\text{EducaciónPadre}) + \beta_3(\text{EducaciónMadre}) \\
&+ \beta_4(\text{Ingresos}) + \beta_5(\text{TamañoFamiliar}) + \beta_6(\text{HogarIncompleto}) + \beta_7(\text{PadresIndígenas}) \\
&+ \beta_8(\text{ExpectativasPadres}) + \beta_9(\text{SatisfacciónCalidad}) + r \\
\beta_0 &= \gamma_{00} + \mu_0 \\
\beta_1 &= \gamma_{10} + \mu_1 \\
&\vdots \\
\beta_9 &= \gamma_{90} + \mu_9
\end{aligned}$$

- **Modelo Base**

Para la construcción de este modelo se mantienen las variables que sean significativas en el nivel 1 y se agregan en el nivel 2 las variables a estudiar con respecto al establecimiento, por lo cual el modelo queda construido como:

$$\begin{aligned}
Y &= \beta_0 + \beta_1(\text{Var1}) + \dots + \beta_k + r \\
\beta_0 &= \gamma_{00} + \gamma_{01}(\text{HorasJornadaEscolar}) + \gamma_{02}(\text{DisponibilidadRecursos}) + \gamma_{03}(\text{IVE}) \\
&+ \gamma_{04}(\text{Municipal}) + \gamma_{05}(\text{ExcesoViaje}) + \gamma_{06}(\text{CantidadGrados}) + \gamma_{07}(\text{TamañoCurso}) \\
&+ \gamma_{08}(\text{IndígenasCurso}) + \gamma_{09}(\text{GéneroProfesor}) + \gamma_{010}(\text{AntigüedadProfesor}) \\
&+ \gamma_{011}(\text{ExpectativasProfesor}) + \gamma_{012}(\text{UsoRecursos}) + \gamma_{013}(\%Retirados) \\
&+ \gamma_{014}(\text{EdSecundaria}) + \gamma_{015}(\text{Matrícula}) + \gamma_{016}(\text{Selección}) + \gamma_{017}(\text{PagoMensual}) + \mu_0 \\
&\vdots \\
\beta_k &= \beta_{k0} + \mu_k
\end{aligned}$$

- **Modelo No Lineal**

En este modelo se trabaja de igual forma que en el modelo base, pero además se agregan variables estructurales al cuadrado y en logaritmo, con lo que el modelo queda conformado por las siguientes ecuaciones:

$$\begin{aligned}
Y &= \beta_0 + \beta_1(\text{Var1}) + \dots + \beta_k + r \\
\beta_0 &= \gamma_{00} + \gamma_{01}(\text{DisponibilidadRecursos}) + \gamma_{02}(\text{IVE}) + \gamma_{03}(\text{Municipal}) \\
&+ \gamma_{04}(\text{ExcesoViaje}) + \gamma_{05}(\text{CantidadGrados}) + \gamma_{06}(\text{TamañoCurso}) + \gamma_{07}(\text{IndígenasCurso}) \\
&+ \gamma_{08}(\text{GéneroProfesor}) + \gamma_{09}(\text{AntigüedadProfesor}) + \gamma_{010}(\text{ExpectativasProfesor}) \\
&+ \gamma_{011}(\text{UsoRecursos}) + \gamma_{012}(\%Retirados) + \gamma_{013}(\text{EdSecundaria}) + \gamma_{014}(\text{Matrícula}) \\
&+ \gamma_{015}(\text{Selección}) + \gamma_{016}(\text{PagoMensual}) + \gamma_{017}(\text{DisponibilidadRecursos})^2 + \gamma_{018}(\text{CantidadGrados})^2 \\
&+ \gamma_{019}(\text{TamañoCurso})^2 + \gamma_{020}(\text{UsoRecursos})^2 + \gamma_{021}\text{Log}(\text{Matrícula}) + \mu_0 \\
&\vdots \\
\beta_k &= \beta_{k0} + \mu_k
\end{aligned}$$

5.3.2 Resultados Obtenidos

Los resultados que se presentan en esta sección son el coeficiente estimado y su desviación estándar entre paréntesis, destacando los coeficientes que resultan significativos al 5%.

- **Modelo incondicional**

En esta sección se muestran los resultados obtenidos para los efectos fijos y aleatorios del modelo vacío, junto al coeficiente de correlación intraclase obtenido.

Cuadro 25: Resultados modelo incondicional

Efectos fijos	Coeficiente
Modelo para rendimiento promedio del establecimiento (β_0)	
Intercepto 2 (γ_{00})	238 (0,98)
Efectos aleatorios	Componente de varianza
Var. Entre establecimientos (τ_0)	542,71 (23)
Var. Entre estudiantes (σ^2)	2.461,62 (50)

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa HLM.

Los resultados anteriores muestran que la varianza entre establecimientos es mucho menor que la varianza entre los estudiantes como es intuitivo, pues es entre los individuos donde existe un mayor rango de puntajes. Al añadir variables en cada nivel en los modelos sucesivos se reduce la varianza entre establecimientos y entre estudiantes, lo que se comprueba más adelante.

El objetivo de este modelo es encontrar la proporción de varianza explicada directamente por el establecimiento la que se calcula a través del coeficiente de correlación intraclase. Dadas estas condiciones, con el set de variables y data utilizado, el aporte del establecimiento a la varianza es de 18,1% lo que significa que el establecimiento influiría en esa proporción en el puntaje del estudiante. Lo importante es entonces, saber cómo hacer para que los establecimientos de peor rendimiento pasen a entregar una educación de calidad, lo que se analiza del punto de vista de las características que posean y sean factibles de mejorar para que sus estudiantes alcancen mejores rendimientos

- **Modelo con variables familiares**

Este modelo sólo utiliza las variables explicativas del nivel 1 para evaluar cuáles de ellas son influyentes y deben ser incluidas en el modelo que incluye variables explicativas en el nivel 2.

Cuadro 26: Resultados del modelo multinivel con variables familiares

Efectos fijos	Coeficiente
Modelo para rendimiento promedio del establecimiento (β_0)	
Intercepto (γ_{00})	237,46 (0,98)
Modelo para pendiente género alumno (β_1)	
Intercepto (γ_{10})	-3,78 (0,97)

Efectos fijos	Coefficiente
Modelo para pendiente ed. Padre (β_2)	
Intercepto (γ_{20})	0,99 (0,17)
Modelo para pendiente ed. Madre (β_3)	
Intercepto (γ_{30})	1,45 (0,17)
Modelo para pendiente ingresos (β_4)	
intercepto (γ_{40})	0,01 (0,003)
Modelo para pendiente tamaño familiar (β_5)	
intercepto (γ_{50})	-1,36 (0,30)
modelo para pendiente hogar incompleto (β_6)	
intercepto (γ_{60})	10,75 (2,53)
Modelo para pendiente padres indígenas (β_7)	
intercepto (γ_{70})	0,06 (1,76)
Modelo para pendiente expectativas padres (β_8)	
intercepto (γ_{80})	6,9 (0,35)
Modelo para pendiente satisfacción (β_9)	
intercepto (γ_{90})	2,17 (0,73)
Efectos aleatorios	Componente de Varianza
Var. Entre establecimientos (τ_0)	567,53 (23,82)
Var. Entre estudiantes (σ^2)	2245,71 (47,39)

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa HLM.

Con los resultados anteriores, la única variable que no es significativa a nivel estudiante es la variable padres indígenas, por lo que es eliminada en los modelos siguientes. Con esto se puede afirmar que los estudiantes de pueblos originarios no presentan resultados disímiles con respecto al resto de los estudiantes.

Al incluir estas variables, la varianza entre estudiantes muestra una disminución del 9% con respecto al modelo anterior, con lo que su CCI es de 20,17% lo que representa el ajuste del modelo.

Finalmente, las variables familiares que se mantendrán en los modelos siguientes son género del alumno, educación del padre, educación de la madre, ingresos, tamaño familiar, hogar incompleto, expectativas y satisfacción.

- **Modelo base**

Con este modelo se busca establecer qué variables a nivel establecimiento son importantes para que los estudiantes tengan un mejor rendimiento. A continuación se muestran los resultados obtenidos del modelo que incluye variables explicativas en ambos niveles.

Cuadro 27: Resultados del modelo base

Efectos fijos	Coefficiente
Modelo para rendimiento medio establecimiento (β_0)	
Intercepto (γ_{00})	236,43 (0,92)
Horas Jornada escolar (γ_{01})	1,24 (2,18)
Disponibilidad de recursos (γ_{02})	12,37 (5,63)
Índice de vulnerabilidad (γ_{03})	-0,42 (0,08)
Municipal (γ_{04})	3,84 (2,86)
Exceso de viaje (γ_{05})	36,33 (10,43)
Cantidad de grados (γ_{06})	-3,79 (1,75)
Tamaño curso (γ_{07})	-0,08 (0,15)
Indígenas curso (γ_{08})	-0,26 (0,17)
Género profesor (γ_{09})	3,07 (2,25)
Antigüedad profesor (γ_{10})	0,30 (0,08)
Expectativas profesor (γ_{011})	3,44 (0,70)
Uso de recursos (γ_{012})	-22,68 (13,69)
% Retirados (γ_{013})	-37,04 (23,75)
Ed. Secundaria (γ_{014})	0,45 (3,86)
Matrícula (γ_{015})	-0,01 (0,01)
Selección (γ_{016})	9,97 (3,97)
Pago mensual (γ_{017})	0,16 (0,06)

Efectos fijos	Coefficiente
Modelo para pendiente género alumno (β_1)	
Intercepto (γ_{10})	-3,78 (1,01)
Modelo para pendiente ed. Padre (β_2)	
Intercepto (γ_{20})	0,99 (0,19)
Modelo para pendiente ed. Madre (β_3)	
Intercepto (γ_{30})	1,45 (0,17)
Modelo para pendiente ingresos (β_4)	
Intercepto (γ_{40})	0,01 (0,01)
Modelo para pendiente tamaño familiar (β_5)	
Intercepto (γ_{50})	-1,35 (0,32)
Modelo para pendiente hogar incompleto (β_6)	
Intercepto (γ_{60})	10,75 (2,62)
Modelo para pendiente expectativas padres (β_7)	
Intercepto (γ_{70})	6,9 (0,36)
Modelo para pendiente satisfacción (β_8)	
Intercepto (γ_{80})	2,17 (0,77)
Efectos aleatorios	Componente de varianza
Var. Entre establecimientos (τ_0)	408,85 (20,22)
Var. Entre estudiantes (σ^2)	2245,71 (47,39)

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa HLM.

En este caso, el modelo presenta un bajo ajuste - su CCI es de 15,4% - al igual que en la regresión lineal, lo cual es esperable, pues el modelo adolece de las mismas falencias al omitir variables que son intangibles, las que son sumamente importantes al momento de predecir el rendimiento de los estudiantes. Sin embargo, al utilizar este modelo lo que se busca es una mejor forma de estimar el efecto que tiene el establecimiento en el rendimiento de los estudiantes.

Al evaluar las características del establecimiento que afectan el rendimiento este modelo arroja como significativas las variables disponibilidad de recursos, índice de vulnerabilidad, cantidad de grados combinados, antigüedad y expectativas del profesor y selección de estudiantes al igual que en la regresión lineal, así como también resulta significativo el pago mensual. Además la varianza entre establecimientos disminuye en un 28% a través de este modelo.

Finalmente, se podría haber esperado una mayor cantidad de variables significativas o con un mayor poder explicativo a través de sus coeficientes, pero se debe recordar la cantidad de relaciones existentes entre las variables, como por ejemplo que la disponibilidad de recursos está

altamente ligada a la matrícula del establecimiento. Dicho esto es que hay muchas variables que no resultan significativas, pero están siendo representadas intrínsecamente a través de las variables correlacionadas con ellas y que resultan significativas.

- **Modelo no lineal**

El siguiente modelo agrega la versión cuadrática de las variables estructurales disponibilidad de recursos, cantidad de grados combinados, tamaño curso, uso de recursos y el logaritmo de la matrícula para probar su efecto en el rendimiento.

Cuadro 28: Resultados del modelo no lineal

Efectos fijos	Coefficiente
Modelo para rendimiento medio establecimiento (β_0)	
Intercepto (γ_{00})	236,42 (0,92)
Disponibilidad de recursos (γ_{01})	14,42 (17,87)
Índice de vulnerabilidad (γ_{02})	-0,41 (0,08)
Municipal (γ_{03})	3,38 (2,88)
Exceso de viaje (γ_{04})	36,27 (10,37)
Cantidad de grados (γ_{05})	-6,56 (4,79)
Tamaño curso (γ_{06})	-0,63 (0,62)
Indígenas curso (γ_{07})	-0,19 (0,17)
Género profesor (γ_{08})	2,93 (2,27)
Antigüedad profesor (γ_{09})	0,29 (0,08)
Expectativas profesor (γ_{010})	3,34 (0,70)
Uso de recursos (γ_{011})	75,84 (86,61)
% Retirados (γ_{012})	-34,75 (23,62)
Ed. Secundaria (γ_{013})	0,94 (3,93)
Log(Matricula) (γ_{014})	-0,04 (0,05)
Selección (γ_{015})	7,93 (4,07)

Efectos fijos	Coefficiente
Pago mensual (γ_{016})	0,17 (0,06)
Disponibilidad de recursos ² (γ_{017})	-14,4 (20,9)
Cantidad de grados ² (γ_{018})	1,45 (1,0)
Tamaño curso ² (γ_{019})	0,01 (0,01)
Uso ² (γ_{020})	-123,68 (107,2)
Modelo para pendiente género alumno(β_1)	
Intercepto (γ_{10})	-3,78 (1,01)
Modelo para pendiente ed. padre(β_2)	
Intercepto (γ_{20})	0,99 (0,19)
Modelo para pendiente ed. Madre(β_3)	
Intercepto (γ_{30})	1,45 (0,17)
Modelo para pendiente ingresos(β_4)	
Intercepto (γ_{40})	0,01 (0,0)
Modelo para pendiente tamaño familiar(β_5)	
Intercepto (γ_{50})	-1,35 (0,32)
modelo para pendiente hogar incompleto(β_6)	
Intercepto (γ_{60})	10,75 (2,62)
Modelo para pendiente expectativas padres(β_7)	
Intercepto (γ_{70})	6,9 (0,36)
Modelo para pendiente satisfacción(β_8)	
Intercepto (γ_{80})	2,12 (0,77)
Efectos aleatorios	Componente de varianza
Var. Entre establecimientos (τ_0)	407 (20,18)
Var. Entre estudiantes (σ^2)	2245,71 (47,39)

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa HLM.

En este modelo se tienen menos variables significativas que en el modelo no lineal utilizado en con el método de regresión lineal, las variables del estudiante mantienen su nivel y el ajuste del modelo también se mantiene cercano al encontrado con la regresión lineal, dado por el CCI en

este caso, el que es de un 15,34% lo que es un poco menor que el modelo lineal. El resultado más destacable que se obtiene con este modelo es con respecto a la influencia de los grados combinados, pues su coeficiente en este caso es el doble que en la regresión lineal, lo que hace pensar que el efecto negativo que tiene esta característica es mucho más importante que lo reflejado con la regresión lineal, pues aquí se logra aislar el efecto colegio por lo que se cuantificaría de mejor forma la influencia de las características del establecimiento.

En definitiva, el mayor resultado obtenido a través del uso de un modelo anidado es que el efecto del establecimiento sí es importante y va más allá de las características socioeconómicas de los estudiantes, es decir, su influencia no queda limitada al efecto *peer*.

Gracias a este análisis es que definitivamente se puede afirmar que los factores estructurales disponibilidad de recursos y cursos simples tienen un efecto positivo en el rendimiento. Incluso, al mejorar la calidad estructural y la oferta educacional es que los profesores pueden sentirse más motivados a hacer un buen trabajo, pues con la mejora sus estudiantes si tendrían un futuro educacional por delante.

Todos estos resultados se basan en modelos que nuevamente presentan un ajuste pequeño, pero logran dar una referencia de cómo y quiénes influyen en el rendimiento de los niños de establecimientos rurales.

5.3.3 Segmentaciones Estudiadas

Siguiendo con la dinámica del apartado de la regresión lineal, se prueban segmentaciones de los grupos socioeconómicos y tipo de jornada. La segmentación por rendimiento a través de este método no es posible de estudiar, pues por la poca variabilidad de la variable dependiente, el software no es capaz de realizar iteraciones para realizar el método de estimación de máxima verosimilitud.

5.3.3.1 Segmentación por grupo socioeconómico

Para segmentar por nivel socioeconómico, nuevamente se toma el GSE del estudiante, con lo que el GSE 1 representa al estrato social más bajo, subiendo sucesivamente hasta llegar al GSE 5 que es el estrato donde se encuentran los más ricos. A continuación se presentan los resultados obtenidos con esta segmentación.

Cuadro 29: Resultados de la segmentación socioeconómica.

Efectos fijos	GSE 1	GSE 2	GSE 3	GSE 4	GSE 5
Modelo para rendimiento medio establecimiento (β_0)					
Intercepto (γ_{00})	234,37 (0,9)	253,86 (2,2)	274,80 (3,9)	263,51 (7,9)	258,06 (6,4)
Horas Jornada escolar (γ_{01})	2,20 (2,2)	2,55 (5,0)	-31,93 (10,6)	-0,51 (26,2)	-1,79 (19,6)
Disponibilidad de recursos (γ_{02})	24,69 (18,3)	-8,11 (45,3)	1,24 (75,5)	59,55 (22,2)	214,54 (165,6)
Índice de vulnerabilidad(γ_{03})	-0,78 (0,5)	-0,23 (0,2)	0,16 (0,4)	0,77 (1,2)	0,41 (0,9)
Municipal (γ_{04})	4,7 (3,0)	3,0 (7,9)	9,23 (12,8)	-67,02 (56,7)	-72,06 (47,9)
Exceso de viaje (γ_{05})	34,11 (10,8)	56,46 (30,4)	90,35 (53,7)	-26,10 (203,3)	-89,32 (163,0)

Efectos fijos	GSE 1	GSE 2	GSE 3	GSE 4	GSE 5
Cantidad de grados (γ_{06})	-10,97 (5,3)	-18,24 (19,2)	-31,76 (24,6)	124,40 (96,6)	32,53 (79,1)
Tamaño curso (γ_{07})	-0,18 (0,7)	-0,86 (1,2)	1,30 (2,4)	-13,28 (6,6)	-6,50 (5,3)
Indígenas curso (γ_{08})	-0,23 (0,2)	-0,44 (0,5)	1,10 (1,0)	-6,48 (4,3)	-4,42 (2,9)
Género profesor (γ_{09})	3,62 (2,4)	-3,97 (6,0)	0,13 (10,1)	7,12 (25,5)	4,79 (19,0)
Antigüedad profesor (γ_{10})	0,25 (0,1)	0,48 (0,2)	0,20 (0,5)	-2,14 (1,5)	-1,04 (1,1)
Expectativas profesor (γ_{011})	3,10 (0,7)	3,05 (1,8)	8,84 (4,2)	-3,07 (12,7)	-1,6 (10,7)
Uso de recursos (γ_{012})	73,37 (80,7)	403,46 (231,6)	-874,61 (393,3)	1789,57 (1863,5)	643,94 (1407,2)
% estudiantes retirados (γ_{013})	-25,85 (23,7)	-100,77 (75,2)	-169,62 (157,8)	159,82 (443,1)	-22,14 (331,9)
Educación Secundaria (γ_{014})	0,52 (4,1)	-5,73 (8,8)	-20,12 (14,6)	-25,48 (34,0)	-31,42 (28,9)
Matrícula (γ_{015})	-0,03 (0,02)	-0,03 (0,05)	0,04 (0,08)	0,22 (0,2)	0,16 (0,2)
Selección de estudiantes (γ_{016})	7,46 (4,5)	15,64 (8,3)	14,47 (13,5)	5,68 (28,2)	-13,52 (12,2)
Pago mensual (γ_{017})	0,14 (0,1)	0,21 (0,2)	0,10 (0,2)	0,03 (0,5)	0,15 (0,4)
Disponibilidad de recursos ² (γ_{018})	-17,12 (21,3)	31,53 (49,6)	21,27 (75,5)	-39,63 (28,9)	-200,15 (189,1)
Cantidad de grados ² (γ_{019})	1,31 (1,0)	1,60 (4,0)	4,87 (4,1)	-23,89 (16,9)	-6,09 (13,6)
Tamaño curso ² (γ_{020})	0,003 (0,01)	0,01 (0,02)	-0,03 (0,04)	-0,19 (0,1)	0,09 (0,1)
Uso de recursos ² (γ_{021})	-117,88 (97,9)	-557,35 (285,1)	963,15 (479,5)	-2157,08 (2376,8)	-695,3 (1762,5)
Log(Matricula) (γ_{022})	0,0 (0,0)	0,0 (0,0)	0,0 (0,0)	0,0 (0,0)	0 (0,0)
Modelo para pendiente género alumno(β_1)					
Intercepto (γ_{10})	-3,94 (1,08)	-3,6 (5,27)	-0,29 (12,24)	17,01 (14,47)	2 (5,54)
Modelo para pendiente ed. padre(β_2)					
Intercepto (γ_{20})	0,92 (0,21)	1,85 (0,96)	-4,03 (2,55)	-17,19 (64,91)	0,19 (1,9)
Modelo para pendiente ed. Madre(β_3)					
Intercepto (γ_{30})	1,37 (0,18)	1,06 (0,76)	5,16 (2,36)	0,05 (0,05)	1,76 (1,8)
Modelo para pendiente ingresos(β_4)					
Intercepto (γ_{40})	0,03 (0,01)	-0,003 (0,03)	0,05 (0,06)	0,98 (4,38)	0,001 (0,01)
Modelo para pendiente tamaño familiar(β_5)					
Intercepto (γ_{50})	-1,55 (0,34)	2,29 (1,29)	-2,79 (5,55)	-17,19 (64,91)	0,31 (1,98)

Efectos fijos	GSE 1	GSE 2	GSE 3	GSE 4	GSE 5
modelo para pendiente hogar incompleto(β_6)					
Intercepto (γ_{60})	10,53 (2,78)	31,10 (21,75)	-27,91 (69,35)	1,15 (13,40)	-23,45 (57,62)
Modelo para pendiente expectativas padres(β_7)					
Intercepto (γ_{70})	6,62 (0,38)	8,16 (2,09)	12,19 (10,76)	1,15 (13,40)	26,87 (4,97)
Modelo para pendiente satisfacción(β_8)					
Intercepto (γ_{80})	2,33 (0,83)	-0,69 (3,19)	8,11 (13,44)	4,95 (10,95)	2,64 (4,75)
Efectos aleatorios	Componente de varianza				
	GSE 1	GSE 2	GSE 3	GSE 4	GSE 5
Var. Entre establecimientos (τ_0)	410 (20,26)	455 (21,32)	28 (5,29)	1.484,5 (38,53)	1.078,34 (32,84)
Var. Entre estudiantes (σ^2)	2.270,75 (47,65)	2.200,54 (46,91)	1.878,74 (43,34)	1.749,26 (41,82)	1.586,86 (39,84)

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa HLM.

A través de estos resultados se puede ver que hay una disminución fuerte de la cantidad de variables que influyen en cada GSE, además de presentar algunos resultados de magnitudes muy grandes como los coeficientes de uso de recursos. En general, los modelos presentan buenos ajustes, excepto el GSE 3 que tiene un ajuste de 1,5% y no es significativo al igual que en la regresión lineal, mientras que el GSE 1 y 2 tienen un ajuste cercano al modelo con todos los datos, con un CCI de 15,29% y 17,13% respectivamente. Para los GSE más ricos, el modelo nuevamente se ajusta mucho mejor con un CCI de 45,91% para el 4 y de un 40,46% para el GSE más rico.

Gracias a esta segmentación se puede establecer que a medida que se sube de GSE, las variables significativas van disminuyendo, con lo que se puede decir que para el estrato más rico el establecimiento no tiene una gran influencia, pero no se debe confundir esto con que el estudiante tendrá el mismo rendimiento en cualquier establecimiento, sino que indica que la variabilidad entre los establecimientos más ricos es menor y es en relación a esos establecimientos que sus estudiantes no se ven afectados por sus características.

En los GSE del 1 al 4 se ven ciertas influencias que tiene el establecimiento, siendo para el GSE 4, es decir, el GSE medio alto, la disponibilidad de recursos y el tamaño del curso quienes hacen la diferencia en el rendimiento.

Nuevamente es el GSE más pobre el que más influencias recibe por parte del establecimiento, a través del efecto *peer* dado por el IVE y la selección, la cantidad de grados combinados, la antigüedad y expectativas del profesor, así como también el GSE 3 no presenta un modelo con un buen ajuste y el GSE 2 también se ve bastante influido por el establecimiento, pero a través de menos variables que el más pobre.

Con este resultado se comprueba que el estrato más pobre es quien más influido se ve por parte del establecimiento, con lo que se puede tener la certeza de que el establecimiento sí importa y, lo que es aún más alentador, el establecimiento es fundamental para el buen rendimiento de los estudiantes más pobres del país.

5.3.3.2 Segmentación según jornada escolar

A continuación se presentan los resultados obtenidos al aplicar el modelo en establecimiento con y sin jornada escolar completa.

Cuadro 30: Resultados de la segmentación por jornada escolar

Efectos fijos	Con JEC	Sin JEC
Modelo para rendimiento medio establecimiento (β_0)		
Intercepto (γ_{00})	236,19 (1,05)	237,38 (1,84)
Disponibilidad de recursos (γ_{02})	18,08 (18,56)	11,08 (37,40)
Índice de vulnerabilidad (γ_{03})	-0,37 (0,09)	-0,39 (0,17)
Municipal (γ_{04})	4,29 (3,22)	-5,29 (7,24)
Exceso de viaje (γ_{05})	37,11 (11,01)	7,2 (27,28)
Cantidad de grados (γ_{06})	-10,08 (5,68)	-30,83 (28,12)
Tamaño curso (γ_{07})	-0,19 (0,71)	-1,59 (1,24)
Indígenas curso (γ_{08})	-0,2 (0,18)	-0,08 (0,46)
Género profesor (γ_{09})	5,08 (2,59)	-2,63 (4,95)
Antigüedad profesor (γ_{10})	0,30 (0,1)	0,13 (0,17)
Expectativas profesor (γ_{011})	2,8 (0,79)	4,77 (1,5)
Uso de recursos (γ_{012})	98,87 (94,45)	91,51 (204,51)
% estudiantes retirados (γ_{013})	-56,08 (24,79)	89,43 (52,35)
Educación Secundaria (γ_{014})	0,56 (4,49)	-1,88 (9,29)
Matrícula (γ_{015})	-0,01 (0,03)	-0,03 (0,04)
Selección de estudiantes (γ_{016})	4,62 (5,24)	15,32 (6,82)
Pago mensual (γ_{017})	0,19 (0,06)	0,17 (0,13)
Disponibilidad de recursos ² (γ_{018})	1,61 (20,15)	-29,12 (44,99)
Cantidad de grados ² (γ_{019})	1,35 (1,03)	5,82 (8,02)
Tamaño curso ² (γ_{020})	0,001 (0,01)	0,03 (0,02)
Uso de recursos ² (γ_{021})	-152,12 (114,93)	-170,06 (267,34)
Matrícula ² (γ_{022})	0 (0,0)	0,0 (0,0)

Efectos fijos	Con JEC	Sin JEC
Modelo para pendiente género alumno(β_1)		
Intercepto (γ_{10})	-2,93 (1,17)	-5,97 (1,92)
Modelo para pendiente ed. padre(β_2)		
Intercepto (γ_{20})	1,13 (0,22)	0,57 (0,39)
Modelo para pendiente ed. Madre(β_3)		
Intercepto (γ_{30})	1,44 (0,20)	1,5 (0,33)
Modelo para pendiente ingresos(β_4)		
Intercepto (γ_{40})	0,003 (0,003)	0,01 (0,01)
Modelo para pendiente tamaño familiar(β_5)		
Intercepto (γ_{50})	-1,75 (0,37)	-0,32 (0,597)
Modelo para pendiente hogar incompleto(β_6)		
Intercepto (γ_{60})	10,25 (3,01)	12,48 (5,36)
Modelo para pendiente expectativas padres(β_7)		
Intercepto (γ_{70})	6,77 (0,41)	7,2 (0,73)
Modelo para pendiente satisfacción(β_8)		
Intercepto (γ_{80})	2,01 (0,89)	2,68 (1,54)
Efectos aleatorios		Componente de varianza
Var. Entre establecimientos (τ_0)	404,55 (20,11)	388,34 (19,71)
Var. Entre estudiantes (σ^2)	2.267,56 (47,61)	2.187,74 (46,77)

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa HLM.

Esta segmentación permite evaluar el efecto del establecimiento al tener una jornada normal o jornada completa. Ambos modelos obtenidos muestran un ajuste similar, con un CCI de 15,12% para los estudiantes con JEC y de un 15,07% para los sin JEC.

Para ambos segmentos se tiene una influencia similar del índice de vulnerabilidad, lo que indica que no hay gran diferencia a través del efecto *peer* dado por el nivel socioeconómico del establecimiento, además, para ambos segmentos tiene importancia la expectativa que tiene el profesor, sin embargo, se debe notar que para los sin JEC, ésta tiene casi el doble de la influencia que tiene para los con JEC, con lo que se puede vislumbrar que son los estudiantes de los establecimientos sin JEC son los que tienen menos oportunidades, pues al tenerlas, sus profesores tienen mayores expectativas de su desempeño a futuro.

Además, a través del modelo jerárquico se comprueba que los estudiantes que tienen JEC se ven influidos por más características del establecimiento, lo que podría indicar que los estudiantes de estos recintos logran superar diferencias familiares a través de una mayor estadía en los establecimientos. En cambio, para los estudiantes sin JEC se tiene que los mejores

resultados se dan donde se seleccionan los estudiantes, por lo que el efecto mayor que tiene el establecimiento es a través del efecto *peer* y no, en mayor medida, a las características de éste.

5.4 Resumen de Resultados

A través de los modelos y segmentaciones revisados, se han logrado resultados que no tienen un ajuste demasiado bueno para los métodos utilizados. En este sentido queda abierta la posibilidad de buscar métodos alternativos para encontrar un modelo que presente un mejor ajuste a los datos o una nueva metodología de predicción de los datos. Sin embargo, los resultados obtenidos están en el rango de resultados obtenidos para investigaciones sobre el rendimiento. Es en base a la evidencia bibliográfica que los resultados obtenidos se toman como válidos para hacer interpretaciones de cuales variables tienen importancia en el rendimiento de los estudiantes de las zonas rurales.

El principal resultado es que el establecimiento sí importa, lo que ha sido cuestionado desde el principio de las investigaciones sobre el rendimiento académico. Lo más importante es que hay mucho por hacer para entregarles una educación con calidad y equidad a los estudiantes del mundo rural en beneficio de su rendimiento. Es más, a través de las segmentaciones se comprueba que los estudiantes de los grupos socioeconómicos inferiores son quienes más ven afectados sus resultados por las características del establecimiento. Además, la gran mayoría de los estudiantes de bajo rendimiento son de esos grupos socioeconómicos, con lo que al mejorar la calidad de los establecimientos rurales a través de sus características estructurales, su rendimiento debiera tener un aumento importante.

Las segmentaciones estudiadas son según GSE, según rendimiento y según tipo de jornada, las que arrojan las variables que son influyentes para cada segmento, con lo que además de los modelos aplicados a la muestra completa, se puede corroborar la importancia de las variables estudiadas.

Lo que se expone en seguida es el resumen de las variables que tienen significancia en los modelos y segmentaciones utilizadas para entender en profundidad la relación entre las variables significativas y los distintos segmentos.

Con respecto a las características del establecimiento:

- La cantidad de grados en una misma aula es, sin duda, el atributo que resulta significativo en la mayoría de los modelos por lo que es de suma importancia el mejorar este aspecto en la educación rural, pues es sólo en este contexto donde se da este tipo de cursos y presenta una influencia negativa en los estudiantes.
- Los recursos disponibles y su uso también tienen una amplia presencia en los modelos, pues al menos uno de estos dos atributos aparece como significativo en cada modelo, con lo que se muestra que al tener establecimientos mejores dotados los estudiantes tienen un rendimiento mucho mejor.
- El tamaño del curso a pesar de no aparecer significativo en muchos modelos, tiene una influencia importante sobre todo en el primer GSE y en el segmento de menor rendimiento. El determinar un tamaño ideal no es simple dado que tiene un efecto directo en los costos de construir y operacionales de un establecimiento.

- El viaje que efectúan los estudiantes para llegar a sus establecimientos estaría dado por dos razones. La primera es, como resulta intuitivo, por la geografía y accesibilidad de las zonas rurales, donde los establecimientos efectivamente quedan distantes o en lugares de difícil acceso y la otra razón es que hay quienes buscan una mejor educación sacrificando el tiempo de viaje. El cuánto influye el viaje en el rendimiento es difícil de estimar entonces, pues hay estudiantes que se mueven en busca de una mejor calidad, teniendo resultados mucho mejores. Además, hay otros incentivos que hacen atractivos un establecimiento para los padres del primer GSE por ejemplo, pues en un establecimiento con JEC y vulnerable los estudiantes reciben alimentación gratuita durante todo el año, por lo que también este es un factor que puede influir en un mayor viaje.
- Las características socioeconómicas del establecimiento así como también la selección muestran la influencia en el rendimiento del efecto *peer*.

En el caso del profesor, las características que tienen influencia en el rendimiento de los estudiantes rurales están ligadas a:

- La antigüedad que éste tiene en el establecimiento, la que estaría vinculada con el efecto de compromiso con la comunidad y sus estudiantes, la experiencia del maestro y a la capacidad del establecimiento para retener a los buenos profesores.
- Las expectativas que se tengan de hasta qué nivel educacional llegarán sus estudiantes, lo cual puede estar influido por las capacidades de éstos así como también por sus oportunidades de seguir educándose.
- El caso del género no es concluyente, pues sólo en algunos modelos resulta significativo siendo las mujeres quienes tendrían una influencia positiva en el rendimiento de los estudiantes.

Cuando de la familia se trata, en general son todas sus características las que influyen en el desempeño de los niños, siendo sin duda las más importantes y destacables:

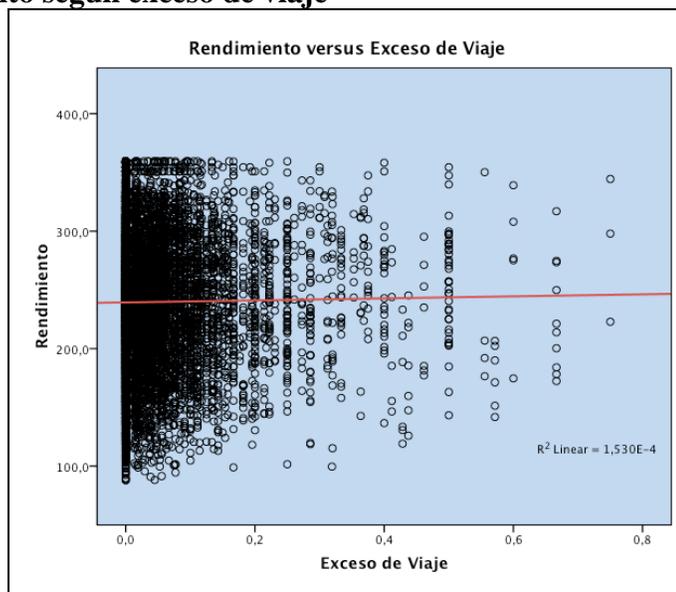
- El nivel educacional alcanzado por la madre.
- Las expectativas que se tengan del estudiante.
- La satisfacción con la calidad del establecimiento, lo que muestra que si los padres saben que sus hijos tienen una buena educación, éstos esperarían más de ellos y los estudiantes recibirían más apoyo y atención en sus estudios.
- Los hogares incompletos muestran tener un efecto positivo. Esto significa que, *ceteris paribus*, un estudiante de este tipo de hogar tiene un mejor rendimiento que el de un hogar con papá y mamá. Esto se ha investigado en algunos estudios y se ha establecido que quienes no tienen a uno de sus progenitores recibirían más apoyo y atención para que reciban una buena educación y obtengan buenos resultados.

Para dar término a este capítulo se hace necesario estudiar con mayor profundidad acerca de la relación entre el rendimiento y la localización y accesibilidad de los establecimientos, medidos a través del índice de exceso de viaje.

Como se puede observar en los diversos modelos obtenidos, el exceso de viaje muestra un efecto positivo sobre el rendimiento académico. Este comportamiento podría llevar a conclusiones equívocas, pues a simple vista se podría pensar que un mayor viaje favorece el rendimiento, lo cual no es correcto, pues a través de este estudio se ha determinado que esta forma de medir la distancia se trata de una causa más que un efecto, con lo que este comportamiento se trata más de una opción de los padres, pues preferirían enviar a sus hijos a estudiar más lejos en pos de conseguir una mejor educación, que de un efecto directo sobre el rendimiento.

La correlación entre el rendimiento y el exceso de viaje es positiva, la cual toma un valor de 0,018, con lo que se puede ver claramente que no tienen una relación marcada entre ellas. Esto mismo se puede observar en el gráfico de dispersión de ambas variables.

Figura 5: Rendimiento según exceso de viaje



Fuente: Elaboración propia, resultados arrojados por el programa SPSS.

El comportamiento del rendimiento frente al exceso de viaje no muestra un patrón definido que pudiera hacer pensar que hay algún tipo de influencia del viaje y el rendimiento. El punto aquí va más allá de eso, pues el viaje se ve más como una causa que un efecto, donde hay estudiantes que efectivamente viajan mucho por la falta de oferta y mala accesibilidad de los establecimientos y otros que viajan en busca de una educación de mejor calidad.

Finalmente, a través de los dos métodos utilizados se ha logrado definir qué variables influyen y de qué forma en el rendimiento académico de los estudiantes rurales a pesar de tener modelos con ajustes bajos.

6 ESTIMACIÓN DE VARIABLES QUE INFLUYEN EN LA ELECCIÓN DE ESTABLECIMIENTOS EDUCATIVOS

En este capítulo se busca determinar los factores que influyen en la decisión de los padres cuando eligen el establecimiento educacional para sus hijos con el fin de incluirlos en el modelo de optimización del SIGER.

Si bien es cierto que resulte intuitivo que la elección de establecimientos educacionales sea en función de su calidad, es este término es quien presenta una concepción ambigua en las percepciones de los padres. Dentro de lo encontrado en la literatura, muchos asimilan calidad a la entrega de valores o incluso a si el establecimiento tiene políticas de expulsión de los estudiantes con mal comportamiento, factores que poco tienen que ver con la calidad en cuanto a la *performance* educativa del establecimiento que es medida a través del SIMCE. En este sentido resulta fundamental el encontrar qué variables resultan ser importantes para los padres para la educación de sus hijos.

Las variables exploradas en este análisis son tanto familiares como propias del establecimiento, incluyendo, por cierto, las variables estructurales. Sin duda que el evaluar las opciones que tienen los padres resulta sumamente complicado sin saber a cuánta distancia viven de los establecimientos que están dentro de sus opciones reales de oferta educativa. Se podría haber optado por construir distintos modelos para las comunas con mayor cantidad de establecimientos, pero esto presenta varios problemas, que van desde la estimación de los modelos, pues resultan ser múltiples las alternativas y pocos los datos por comuna para que el software sea capaz de calcularlos, hasta problemas de sesgo por no incluir comunas donde la oferta es menor y no se recoge la información aportada por ellas.

La solución para obtener las múltiples alternativas de establecimientos pasa por trabajar con comunas que tengan al menos un par de establecimientos donde los padres pueden elegir enviar a sus hijos, lo que puede parecer poco, pero la comuna no está aislada, por lo que los padres pueden tener opciones en establecimientos urbanos o de una comuna vecina que esté cercano, con lo que sus opciones de elegir se amplían. Una vez conformada la data para trabajar se realiza una segmentación con k-medias para categorizar los establecimientos. Así se obtienen categorías para las cuales se calcula un modelo de elección que permite encontrar las variables relevantes para la decisión de los padres.

6.1 Variables y Descripción de la Información Utilizada

Las variables utilizadas en esta sección son tomadas nuevamente de lo recogido en la revisión bibliográfica, basándose en el año 2006 por contar con la mayor cantidad de información sobre estas variable de donde se estudian 865 establecimientos que agrupan a 10.728 estudiantes.

Las características que han sido dejadas de lado por la falta de datos, subjetividad de las preferencias o complejidad de su estimación, son disciplina del establecimiento, orientación valórica y establecimiento bilingüe.

Las variables utilizadas son obtenidas de forma directa de las bases de datos provistas para el estudio por Mineduc y en el caso del GSE y del exceso de viaje se calculan con la metodología mostrada en el Anexo I.

A continuación se presentan las variables utilizadas, sus significados, fuentes y valores para luego hacer un análisis descriptivo de la base a utilizar.

Cuadro 31: Resumen de variables

Variable	Significado	Fuente	Valores
SIMCE establecimiento	Puntaje obtenido por el establecimiento en prueba SIMCE Matemáticas.	Resultados SIMCE 2006	>120
Disponibilidad de Recursos	Disponibilidad declarada por el profesor en cuestionario SIMCE sobre textos escolares, textos de lectura, libros de consulta, diarios y revistas, Internet, recursos multimedia, recursos audiovisuales, biblioteca, juegos didácticos, etc.	Cuestionario Profesor SIMCE 2006	Entre 0 y 1
Índice de Vulnerabilidad	% de estudiantes en riesgo de deserción.	InfoEscuela	Entre 0 y 1
Municipal	Establecimiento de dependencia municipal, 1; particular subvencionado o pagado, 0.	InfoEscuela	1 ó 0
Jornada escolar completa	Si la jornada dura 7:30 horas o más, 1; si dura menos, 0.	RECH	1 ó 0
GSE Curso	Grupo socioeconómico al que pertenece en promedio el curso.	Resultados SIMCE 2006	Entre 1 y 5
Exceso de Viaje	% de estudiantes que traspasan el umbral de viaje según medio de transporte.	Elaboración Propia	Entre 0 y 1
Cantidad de grados	Grados en una misma aula.	RECH	Entre 1 y 6
Tamaño curso	Estudiantes por curso.	RECH	>1
% Indígenas curso	Fracción de estudiantes del curso descendientes de algún pueblo originario.	RECH	Entre 0 y 1
Ed. Secundaria	Si el establecimiento ofrece educación media o técnico profesional 1; si no 0.	InfoEscuela	1 ó 0
Matrícula	Cantidad de estudiantes del establecimiento.	Matrículas	>1
GSE Alumno	Grupo socioeconómico al que pertenece el estudiante.	Elaboración Propia	1 a 5
Hogar Incompleto	Si el estudiante no tiene padre o madre, 1; si no, 0.	Cuestionario Padres SIMCE 2006	1 ó 0
Cercanía	Si los padres manifiestan esta razón como importante al escoger el establecimiento, 1; si no, 0	Cuestionario Padres SIMCE 2006	1 ó 0
Familiaridad	Si los padres manifiestan esta razón como importante al escoger el establecimiento, 1; si no, 0	Cuestionario Padres SIMCE 2006	1 ó 0
Prestigio	Si los padres manifiestan esta razón como importante al escoger el establecimiento, 1; si no, 0	Cuestionario Padres SIMCE 2006	1 ó 0
Resultados SIMCE	Si los padres manifiestan esta razón como importante al escoger el establecimiento, 1; si no, 0	Cuestionario Padres SIMCE 2006	1 ó 0
Entrega de Valores	Si los padres manifiestan esta razón como importante al escoger el establecimiento, 1; si no, 0	Cuestionario Padres SIMCE 2006	1 ó 0
Único en la comuna	Si los padres manifiestan esta razón como importante al escoger el establecimiento, 1; si no, 0	Cuestionario Padres SIMCE 2006	1 ó 0

Fuente: Elaboración propia.

6.2 Especificación del Modelo Utilizado

Al igual que en el estudio del rendimiento académico, se busca un modelo que tenga un buen ajuste y sea parsimonioso, con lo que a través de un análisis de componentes principales de las variables se busca establecer las que entrarán definitivamente en el modelo porque aportan una mayor proporción de variabilidad. Dicho análisis se encuentra disponible en el Anexo III.

La forma de estimar el modelo global es a través de la incorporación de variables que se consideran relevantes de estudiar, las que son SIMCE establecimiento, cantidad de grados, cercanía, educación secundaria y exceso de viaje, y la entrada sucesiva de variables significativas a través de la prueba de la razón de verosimilitudes. La estimación de parámetros del modelo se realiza con un algoritmo iterativo de máxima verosimilitud.

De esta forma, el modelo tiene la forma siguiente para las J+1 opciones

$$\text{Logit}(p_i) = \left(\frac{p_i}{1-p_i} \right) = \beta_0 + \beta_1 \text{Caracter}'stical_i + \dots + \beta_k \text{Caracter}'sticak_i$$

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \text{Caracter}'stical_i + \dots + \beta_k \text{Caracter}'sticak_i)}} \quad \forall i, i = 1, \dots, J$$

6.3 Resultados Obtenidos

Para aplicar este método, se debe contar con categorías de elección que reflejen los gustos de los consumidores, en este caso, los padres o apoderados. En este caso se debe categorizar los establecimientos para evaluar qué atributos de ellos son tomados en cuenta por los padres al momento de tomar la decisión.

Dichas categorías han sido clasificadas con la técnica de segmentación de k-medias, con lo que se definen 5 tipos de establecimientos catalogados según sus características, después de haber realizado un estudio exploratorio con conglomerados jerárquicos, que sugiere un rango de 3 a 5 segmentos.

A continuación se presenta el análisis realizado para luego caracterizar cada tipo de establecimiento.

Cuadro 32: Centro de los conglomerados finales

	Conglomerado				
	1	2	3	4	5
JEC	0,81	0,74	0,53	0,38	0,67
SIMCE establecimiento	232,23	237,69	234,14	233,31	247,07
Disponibilidad de Recursos	0,29	0,32	0,39	0,38	0,46
Índice de Vulnerabilidad	63,69	55,52	47,47	36,88	36,57
Municipal	0,88	0,86	0,78	0,37	0,48
GSE Establecimiento	1,52	1,88	2,2	2,73	2,74

	Conglomerado				
	1	2	3	4	5
Exceso de Viaje	0,06	0,03	0,03	0,01	0,02
Cantidad de grados	1,41	1,01	1	1	1
Tamaño Curso	20	29,13	32,73	37,74	38,12
Indígenas Curso	2,9	1,91	2,05	1,53	1,58
Ed. Secundaria	0,02	0,15	0,29	1	0,59
Matrícula	132,22	281,47	541,47	1116,02	856,92
Cantidad de docentes	9,56	15,6	25,64	44,25	36,46
Selección	0,02	0,08	0,15	0,4	0,29
Mensualidad	3,16	7,32	8,94	9,42	6,01
N total estudiantes	5274	3267	1209	351	627
N total establecimientos	540	253	51	7	14

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

El análisis de los segmentos privilegia aspectos como la calidad, disponibilidad de recursos y composición socioeconómica, por lo que el orden de las categorías queda definido desde el conglomerado más precario hasta el mejor dotado.

Como se puede apreciar, existen grandes diferencias entre los establecimientos. En el caso del conglomerado 1 están los establecimientos más vulnerables y necesitados, pues todas sus características ocupan el primer lugar, yendo de peor a mejor, por lo que es el segmento de categoría Baja.

El conglomerado 2 tiene establecimientos con mejores resultados en el SIMCE con respecto a los conglomerados 3 y 4, pero al observar variables como disponibilidad de recursos, matrícula, GSE del establecimiento, su índice de vulnerabilidad, cantidad de docentes y que al igual que el segmento 1, tiene cursos combinados, se nota que son establecimientos sólo un poco mejores que los del segmento anterior, por lo que son de categoría Medio-Baja.

El conglomerado que agrupa a los establecimientos que están en una situación media es el segmento 3, el que puede parecer similar al 4, pero con su índice de vulnerabilidad, tipo de dependencia (municipal), exceso de viaje, matrícula y cantidad de docentes revelan que es inferior a nivel socioeconómico y a nivel de escala. En consecuencia, el segmento 4 agrupa a los establecimientos que se encuentran en un rango Medio-Alto al clasificarlos de peor a mejor, mientras que el segmento 3 queda en el centro, con una clasificación de Medio.

Finalmente, los establecimientos del segmento 5 alcanzan mejores puntajes y son superiores a nivel socioeconómico y de recursos, siendo eso sí, establecimientos más cercanos, más pequeños en matrícula, principalmente municipales y económicos -de ser pagados- que los

del segmento inmediatamente anterior. Este segmento es llamado Alto y es tomado como la categoría de referencia para estimar el modelo de elección.

6.3.1 Estimación del modelo de elección

Este tipo de modelo se ve afectado por dos tipos de variables, los factores, que en general son variables categóricas y las covariables, que son variables continuas, generalmente. En este caso, los factores son educación secundaria, cercanía, prestigio, resultados SIMCE, entrega valórica, JEC, hogar incompleto, familiaridad y único en la comuna. Cabe recordar que la variable único en la comuna es algo mencionado por los padres, pero no es efectivo en la muestra, pues ella sólo toma en cuenta las comunas donde hay opción de elegir el establecimiento.

Este modelo parte con la inclusión de las variables SIMCE establecimiento, cantidad de grados, cercanía, Educación secundaria, exceso de viaje, para luego incluir las variables que sean significativas para el modelo, con lo que este modelo se va desarrollando paso a paso, donde se evalúa qué variable incluir y si hay que eliminar alguna. En este caso sólo fueron incluidas las variables índice de vulnerabilidad, tamaño curso, JEC, GSE alumno, % alumnos indígenas curso, entrega valórica, resultados SIMCE, prestigio y único en la comuna.

La significancia de las variables del modelo se da para todas las variables, excepto para la variable cercanía, lo que muestra que muchos padres prefieren la calidad u otras características por sobre la cercanía.

Cuadro 33: Contrastes de la razón de verosimilitud

Efecto	Criterio de ajuste del modelo	Contrastes de la razón de verosimilitud		
	-2 Log verosimilitud del modelo reducido	Chi-cuadrado	g.l.	Sig.
Intersección	12537,493	0	0	.
SIMCE establecimiento	13043,817	506,324	4	0,00
Cantidad de grados	14321,816	1784,323	4	0,00
Cercanía	12541,572	4,079	4	0,40
Ed. Secundaria	14440,445	1902,952	4	0,00
Exceso de Viaje	12651,406	113,913	4	0,00
Prestigio	12557,961	20,468	4	0,00
Resultados SIMCE	12566,501	29,008	4	0,00
Entrega Valórica	12565,459	27,965	4	0,00
JEC	12797,086	259,592	4	0,00

Efecto	Criterio de ajuste del modelo	Contrastes de la razón de verosimilitud		
	-2 Log verosimilitud del modelo reducido	Chi-cuadrado	g.l.	Sig.
Único en la comuna	12557,874	20,381	4	0,00
Índice de Vulnerabilidad	13643,425	1105,932	4	0,00
GSE Alumno	12786,823	249,33	4	0,00
Tamaño Curso	16882,724	4345,23	4	0,00
% Indígenas Curso	12580,505	43,012	4	0,00
Modelo	Criterio de ajuste del modelo	Contrastes de la razón de verosimilitud		
	-2 Log verosimilitud	Chi-cuadrado	g.l.	Sig.
Sólo la intersección	25505,707	12968,21	56	0,000
Final	12537,493	4		

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

El modelo obtenido tiene un ajuste significativo, pues al considerar el test χ^2 se puede afirmar que los coeficientes β calibrados para el modelo son distintos de cero. Además, presenta un buen ajuste a través de la mayoría de los estadísticos, siendo la bondad de ajuste no significativa, lo que corrobora que los datos observados y los valores predichos no son significativamente diferentes. El ajuste del pseudo R^2 de Cox y Snell y el de Nagelkerke son bastante altos y cercanos a 1, con lo que muestran un buen comportamiento del modelo encontrado, mientras que el pseudo R^2 de McFadden presenta cerca de un 50% de ajuste. Con todo esto se puede afirmar que el modelo encontrado es válido ya que presenta un buen comportamiento a través de casi todos los estadísticos de prueba.

Cuadro 34: Bondad de ajuste

Test	Chi-cuadrado	g.l.	Sig.
Pearson	21336,406	26940	1,000
Desviación	12537,493	26940	1,000

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

Cuadro 35: Pseudo R-cuadrado

Pseudo R-cuadrado	
Cox y Snell	0,721
Nagelkerke	0,785
McFadden	0,508

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

La interpretación de los coeficientes β de este método no es directa como en la regresión lineal, ya que se trata de una función logarítmica que da una probabilidad y además hay una categoría de referencia con la que se compara el coeficiente, lo cual significa que el cambio provocado por el coeficiente es sobre la probabilidad de elegir la alternativa j frente a la categoría de referencia. En el Anexo IV se muestran los coeficientes obtenidos y sus estadísticos, donde es posible comprobar que muchos coeficientes por sí solos no son significativos. Esto es resultado de la alta cantidad de variables a estimar y los datos en cada categoría para poder estimar el coeficiente de forma significativa resultan ser insuficientes.²⁷

Las variables que resultan significativas por categoría son las que hacen que un apoderado pase de cierta opción a la de referencia o viceversa.

En el caso de la categoría Baja las variables que son capaces de cambiar la elección frente a la categoría Alta son, el puntaje SIMCE del establecimiento, la cantidad de grados, la educación secundaria, el prestigio, los resultados SIMCE, la entrega de valores, la JEC, el índice de vulnerabilidad, el GSE del alumno, el tamaño del curso y el % de indígenas en el curso, sin embargo, esta última variable no es demasiado importante, pues su intervalo de confianza contiene al 1, lo que significa que su aporte podría ser $\text{Ln}(\exp(\beta))=\text{ln}(1)=0$.

Los padres al momento de elegir entre los establecimientos Medio-Bajo y Alto, están influenciados por las variables puntaje SIMCE del establecimiento, educación secundaria, prestigio, entrega de valores, índice de vulnerabilidad, GSE del alumno, tamaño del curso y % de indígenas.

En el caso de la probabilidad de elegir un establecimiento Medio o uno Alto, lo que entra en juego es la educación secundaria, el exceso de viaje, el prestigio, la entrega de valores, la JEC, el índice de vulnerabilidad, el GSE del alumno y el tamaño del curso.

Los factores que inclinan la balanza sobre un establecimiento Medio-Alto frente a uno Alto, o viceversa, son el puntaje SIMCE del establecimiento, el exceso de viaje, el resultado SIMCE, la JEC, si es considerado como el único en la comuna, el índice de vulnerabilidad, el GSE del alumno, el tamaño del curso y % de indígenas.

En resumen, son las características tamaño del curso, educación secundaria, SIMCE establecimiento y cursos combinados las variables que son posibles de incluir en el modelo de optimización para representar las preferencias de los padres.

En el caso del tamaño del curso, los padres prefieren el tamaño de curso de su establecimiento actual con respecto al establecimiento de referencia, quien presenta cursos más grandes, por lo que se estima que los cursos más pequeños son preferidos por los padres.

La presencia de educación secundaria presenta un comportamiento donde los coeficientes β van disminuyendo a medida que se acerca a la categoría de referencia, por lo que se prefiere la presencia de educación secundaria en los establecimientos, pues la categoría de referencia posee, en su mayoría, establecimientos con este tipo de educación.

²⁷ Ver Guisande, C. Et al, 2006

Con respecto al SIMCE del establecimiento resulta negativo para el caso del segmento medio alto con respecto al alto, es decir, se prefiere el tipo de establecimiento Medio-Alto. Esto sin duda es reflejo de que esta categoría tiene un buen desempeño en el SIMCE y, además, los padres que eligen este tipo de establecimientos realizan su elección en base a otras características que van más allá de los resultados obtenidos. Para el caso del segmento medio β es cero y no significativo, mientras que los segmentos Medio-Bajo y Bajo preferirían un mejor desempeño, ya que sus coeficientes son positivo, con lo que la probabilidad de escoger la categoría de referencia es mayor que para sus categorías de origen.

En último término con respecto a las variables estructurales, se tiene que los cursos combinados son importantes para el segmento bajo, pues es significativo y positivo, con lo que se preferirían los establecimientos con cursos simples, como lo son los de referencia. Es más, al ver el comportamiento a través de todas las categorías se ve que el segmento Bajo y Medio-Bajo prefieren cursos simples y no combinados como los de sus establecimientos de origen, en cambio los segmentos Medio y Medio-Alto preferirían sus establecimientos de origen, que en general son de cursos simples. Es por esta misma razón que sus coeficientes no resultan significativos ya que presentan similares condiciones al segmento de referencia.

Con respecto a las demás variables, a medida que se acerca hacia la categoría de referencia, se prefieren establecimientos con un índice de exceso de viaje medio, se hace más importante el prestigio, el GSE del alumno y el IVE del establecimiento. En el caso de los resultados SIMCE, entrega de valores, JEC, único en la comuna y el % de indígenas, las categorías más bajas prefieren la categoría de referencia respecto a las de origen y a medida que se van acercando a los establecimientos de referencia, se prefieren las características del tipo de establecimiento de origen.

6.4 Segmentaciones Estudiadas

6.4.1 Segmentación por GSE del estudiante

Como es sabido, el comportamiento varía entre cada GSE, por lo que resulta necesario indagar sobre qué sucede dentro de cada uno de ellos al momento de elegir un establecimiento. Es por ello que se genera un modelo para cada GSE de modo de entender de mejor forma lo que sucede en la realidad rural. Como primera aproximación de lo que prefieren los padres de estudiantes rurales se muestra el porcentaje de mención que tienen las variables cercanía, prestigio del establecimiento, resultados SIMCE, entrega valórica, único en la comuna y familiaridad con el establecimiento, como una de las 3 razones que lo llevaron a inscribir al estudiante en su establecimiento actual.

Cuadro 36: Preferencias declaradas según GSE

	GSE 1	GSE 2	GSE 3	GSE 4	GSE 5
Cercanía	70,6%	63,6%	56,3%	45,1%	42,8%
Prestigio	18,1%	24%	34,1%	36,3%	43,4%
Resultados SIMCE	9,8%	10,4%	9,6%	8,8%	5,8%
Entrega Valórica	21,3%	26,9%	34,1%	48,4%	67,6%
Único en la comuna	29,1%	18,4%	17,4%	10,9%	9,8%
Familiaridad	42,3%	38,6%	23,4%	35,2%	31,8%

Fuente: elaboración propia en base a Cuestionario Padres SIMCE 2006

Según las preferencias declaradas, la cercanía descende en importancia a medida que se sube en la escala social, por lo que los más ricos la sacrificarían en virtud de otras características, como el prestigio y la entrega valórica, variables que aumentan su mención a medida que se sube de GSE. En el caso de resultados SIMCE tiene una baja mención en todos los estratos, sin mostrar ninguna tendencia a través de ellos. En el caso de la familiaridad y que sea considerado el único en la comuna se observa que en los grupos más pobres tienen una mayor participación, por lo que para los más pobres no se ve una oferta efectiva de establecimientos y, además, para ellos es importante tener una referencia del establecimiento donde envían a sus hijos.

A continuación se presentan las variables estudiadas para cada nivel socioeconómico; cabe recordar que la construcción de los modelos se hace en base a variables que se incluyen siempre - SIMCE establecimiento, cantidad de grados, cercanía, educación secundaria y exceso de viaje- y otras que se incluyen sólo si son significativas para el modelo, por lo que no es necesario que estén presentes en todos los modelos. El detalle de los estadísticos de ajuste del modelo se encuentra disponible en el Anexo V.

Cuadro 37: Contrastes de la razón de verosimilitud

Efecto	Significancia de contrastes de la razón de verosimilitud				
	GSE 1	GSE 2	GSE 3	GSE 4	GSE 5
Intersección
SIMCE Establecimiento	0	0	0	0,005	0
Cantidad de grados	0	0	0	0,002	0
Cercanía	0,395	0,782	0,652	0,539	0,043
Ed. Secundaria	0	0	0	0	0
Exceso de Viaje	0	0	0,944	0,028	0,984
Prestigio	0	---	0	---	---
Resultados SIMCE	0	0,044	---	---	---
Entrega Valórica	0	0,049	---	---	---
JEC	0	0	0	---	0,001
Único en la comuna	0	---	0	---	---
Índice de Vulnerabilidad	0	0	---	---	---
GSE Alumno	0	---	---	---	---
Tamaño Curso	0	0	0	0	---
% Indígenas Curso	0	0,002	---	---	---

Ajuste	Significancia de contrastes de la razón de verosimilitud				
	GSE 1	GSE 2	GSE 3	GSE 4	GSE 5
Modelo Final	0	0	0	0	0

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

• GSE Bajo

Para las personas más desposeídas del mundo rural se produce que los factores gravitantes al momento de escoger el establecimiento donde enviarán a estudiar a sus hijos son puntaje SIMCE del establecimiento, cantidad de grados combinados, educación Secundaria, exceso de viaje, prestigio, resultados SIMCE, entrega de valores, JEC, único en la comuna, familiaridad con el establecimiento, IVE, % indígenas curso y tamaño curso.

En este segmento resultan ser significantes los mismos factores del modelo por pasos de toda la muestra, lo que lo valida aún más y verifica que la calidad sí es un atributo importante para el ámbito rural. Además el modelo presenta buen ajuste general y buena bondad de ajuste, a pesar de tener pseudos R^2 bajos, pero no menores a 0,5 y su precisión en la predicción es de un 75%, por lo tanto, el modelo tiene un buen comportamiento para este GSE.

• GSE Medio-Bajo

Al tratarse del ajuste que tiene el modelo de este estrato, presenta un buen ajuste general y bondad de ajuste, tiene sus pseudos R^2 por sobre el 50%, lo que indica que, en general, tiene un buen comportamiento.

En la elección de los padres de este segmento se tienen similares variables que en el modelo definido para toda la muestra. La diferencia radica en que para estos padres el prestigio, que sea único en la comuna y GSE del alumno no son significativos, lo que implica que en este segmento existiría una opción real de elegir el establecimiento por el lado de la oferta, es decir, ellos tendrían más opciones de donde elegir, dado su nivel socioeconómico.

• GSE Medio

El modelo para este segmento tiene un comportamiento aún mejor que el de los segmentos anteriores, mostrando pseudos R^2 mucho más altos, además de buen ajuste general y con tests de bondad de ajuste que prueban que la predicción hecha por el modelo no es significativamente distinto a las preferencias observadas.

Además en este segmento se reducen considerablemente las variables influyentes, siendo ellas el puntaje SIMCE del establecimiento, cantidad de grados combinados, educación Secundaria, prestigio, resultados SIMCE, entrega de valores, JEC, único en la comuna y el tamaño del curso.

• GSE Medio-Alto

El modelo encontrado para este GSE muestra que sólo hay 5 factores que influyen en la decisión de los padres cuando de elegir el tipo de establecimiento al que irán sus hijos se trata. Los factores importantes son el puntaje SIMCE del establecimiento, la cantidad de grados combinados, si tiene educación Secundaria, el exceso de viaje y el tamaño del curso.

El modelo en sí tiene un buen comportamiento, siendo significativo su ajuste y la diferencia entre el modelo final y el modelo vacío no es significativa al revisar su bondad de ajuste. Los pseudo R^2 son más bajos que en el modelo anterior, pero siguen presentando valores razonables sobre 0,5.

- **GSE Alto**

En este caso, el modelo tiene un buen ajuste y pseudos R^2 , sin embargo, uno de los test de bondad de ajuste dice que la diferencia entre las observaciones y las predicciones hechas por el modelo si son significativas.

Las variables que hacen la diferencia en este GSE son el puntaje SIMCE del establecimiento, la cantidad de grados, la cercanía, siendo el único modelo que la encuentra significativa, la educación Secundaria y la JEC.

En el caso de la cercanía no se debe confundir con que los padres prefieran los establecimientos cerca, si no que este es un factor que influye en su decisión, de hecho al ver las preferencias declaradas, en este GSE son más quienes no eligen el establecimiento por cercanía, es decir, se sacrificaría la cercanía en virtud de otras características por una parte importante de los padres de este segmento. Además el exceso de viaje no es significativo, pero hay que recordar que existe una proporción importante de este segmento que viaja más para encontrar mejor calidad, lo que se reafirma más aún, que los padres prefieren un viaje más extenso para lograr una mejor educación.

6.5 Resumen de Resultados

A través del análisis efectuado, se ha logrado establecer que las variables estructurales que afectan la decisión de los padres:

- Puntaje SIMCE del establecimiento.
- Presencia de educación Secundaria.
- Tamaño del curso.
- Cantidad de grados combinados.

Mediante de estas variables se logra establecer que idealmente lo que prefieren los padres son cursos simples y medianos, establecimientos con educación secundaria y con buen rendimiento en el SIMCE.

Con respecto a las otras variables, para los padres de los establecimientos de tipo Bajo y Medio-Bajo, ellos eligen los establecimientos en base a los resultados SIMCE, la entrega de valores, que sea percibido el único en la comuna, que tenga JEC y un porcentaje de indígenas. Esto último se debe a que es en esos establecimientos donde se concentra la mayor cantidad de estudiantes de pueblos originarios, y como se comenta en la revisión bibliográfica, los padres tienden a buscar establecimiento homogéneos en cuanto a sus variables socioeconómicas, en busca de un sentido de pertenencia e identificación con el entorno del establecimiento.

Para los padres de los establecimientos de tipo Medio y Medio-Alto, ellos prefieren un índice de exceso de viaje moderado, tiene mucha importancia el GSE del estudiante y el IVE del

establecimiento, esta preferencia es por la búsqueda de características socioeconómicas similares, por lo que prefieren establecimientos con GSE de los alumnos más alto y un IVE. Además, el prestigio del establecimiento es una variable fundamental dentro de estos segmentos, por lo que la calidad de los establecimientos está también asociada a este concepto declarado por los padres.

Al analizar lo que sucede según GSE se tiene que a medida que se asciende en la escala social van disminuyendo los factores que son relevantes al momento de elegir un establecimiento. Sin embargo, las variables estructurales son importantes transversalmente, salvo el tamaño del curso en el caso de los más ricos. En cuanto a las preferencias declaradas por los padres, las que realmente tienen efecto sobre la elección son el resultado SIMCE y el prestigio para los 3 GSE más pobres, en cambio para los 2 GSE superiores es importante lo que respecta a la ubicación de los establecimientos, a través de el índice de exceso de viaje en el caso del GSE medio alto y de la cercanía para el GSE alto.

7 ANÁLISIS DE FACTIBILIDAD

El presente capítulo busca encontrar la mejor forma de incluir las variables de calidad y elección en el modelo de optimización que utilizará el SIGER. El objetivo es qué incluir y cómo en el modelo de optimización de localización de modo de complementar el desarrollo de éste.

Para la incorporación de ambos temas investigados se pueden definir políticas a través de restricciones del modelo de optimización o definir una parte de la función objetivo que permita optimizar sobre el rendimiento y las preferencias de los padres.

Este capítulo se compone de un análisis de las variables a incluir ligadas al rendimiento y a la elección para luego dar paso a cómo incluir aquellas variables en el modelo de optimización. Lógicamente, las variables a incluir son las que tengan que ver con infraestructura y localización, ya que no es posible modelar, por ejemplo, las características familiares o socioeconómicas del establecimiento, ya que ellas se dan a través de la asignación de los estudiantes, lo que es justamente el objetivo del SIGER.

7.1 Variables de Rendimiento

Para evaluar las características que influyen en el rendimiento de los estudiantes se utilizaron 2 métodos y varios modelos, por lo que se debe escoger uno que presente los mejores resultados y refleje de la mejor forma posible las variables significativas que se puedan incluir en el modelo de optimización.

En el caso de la regresión lineal quienes presentan un mejor comportamiento, respecto a su R^2 y el test de Fisher, son los modelos I, III y IV, donde el primero tiene un coeficiente R^2 de 0,15 y los otros dos de 0,149. En cuanto al test de Fisher el modelo III es quien presenta un mejor comportamiento, seguido por el modelo IV y I. En base a estos estadísticos es que el modelo III y IV son los modelos que presentan mejores resultados globales. Para definir cuál de los dos presenta mejores características es necesario recordar que el modelo IV tiene un subconjunto de las variables del modelo III, incluyendo sólo las variables que resultan significativas. De este modo, se obtiene un modelo más parsimonioso con el resultado del modelo IV, por lo que es el elegido a través de este método para ser incluido en el modelo de optimización.

Entre las múltiples variables que tiene el modelo IV, las variables que son posibles de incluir en el modelo de optimización son: cantidad de grados, cuyo coeficiente β es de -4,67; tamaño curso con un coeficiente β de -0,21; razón Alumno-Profesor con un β igual a -0,20; y Matrícula, que tiene un coeficiente de -0,005.

A través de la regresión lineal también se prueba un modelo no lineal que resulta con mejor ajuste que los modelos lineales, el que tiene un coeficiente R^2 de 0,155 y con un test de Fisher mayor que el del modelo I. En este caso, las variables que son posibles de insertar en el modelo de optimización son: cantidad de grados, con un β de -5,63 para la variable lineal y de 0,73 para la variable al cuadrado; razón Alumno-Profesor con un coeficiente β igual a -0,67; y Matrícula, que tiene un coeficiente β de -0,67 para la variable al aplicarle la función logarítmica.

El segundo método utilizado para la estimación de las variables que influyen en el rendimiento entrega dos modelos que tienen un ajuste de 15,4% y de 15,34% para el modelo lineal y no lineal respectivamente. En el caso del modelo lineal, sólo presenta dos variables que pueden ser incluidas en el modelo, las que son disponibilidad de recursos y cantidad de grados, mientras que el modelo no lineal agrega una tercera variable que es la matrícula en su versión logarítmica. Al tener más variables significativas y muy poca diferencia con el ajuste del modelo lineal, es que el modelo no lineal presenta ventajas para ser incluido en el modelo de optimización. Los coeficientes que acompañan a las variables son 14,42 para la disponibilidad de recursos, -6,56 para los grados combinados y de -0,04 para el logaritmo de la matrícula.

Entonces, los dos modelos susceptibles a ser incluidos en el modelo de optimización son los modelos no lineales estimados con cada método econométrico. Si bien sólo se utiliza una parte del modelo generado en cada caso, en la sección de inclusión de variables se explica el cómo corregir este sesgo al incorporar el modelo de rendimiento truncado en el modelo de optimización.

7.2 Variables de Elección

A través de la aplicación de un modelo logarítmico multinomial, se ha determinado que las variables que importan para los padres son Puntaje SIMCE, donde más es mejor; Educación secundaria, donde la presencia de ella en el establecimiento es una característica favorable; Tamaño del curso, característica que es valorada al tener cursos de tamaño medio; y la cantidad de grados, donde los padres se inclinan por elegir establecimientos que cuenten con cursos simples.

El modelo obtenido para la elección de los padres presenta un buen ajuste y un buen nivel de sus estadísticos por lo que sus resultados son sólidos y serán incluidos en el modelo directamente. El incluir los coeficientes obtenidos con el modelo no es posible ya que no tienen una interpretación directa de sus magnitudes, por lo que sólo se toman las variables que pueden ser incluidas en el modelo de optimización sin un coeficiente. Las otras variables encontradas significativas por el modelo no son posibles de incluir en el modelo de optimización, por lo que también serán dejadas de lado.

7.3 Inclusión de Variables en el Modelo de Optimización

El modelo de optimización está basado en la minimización de costos de los establecimientos. Dentro de aquellos costos están el costo operativo del establecimiento, el costo de abrir un nuevo establecimiento, el costo de ampliar un establecimiento, el costo (o ganancias)

de cerrar un determinado establecimiento educacional y el costo monetario y social que tiene la distancia recorrida por los estudiantes.

Para incluir las variables encontradas a través de esta investigación, ellas se deben incluir en el modelo a través de políticas si es que no es posible representarlas a través de un costo. En este sentido se tienen una serie de variables que sólo serán incluidas a través de restricciones que reflejen ciertas políticas que pueden ser adoptadas por la autoridad al momento de evaluar las distintas soluciones que pueda entregar el modelo según la relajación o aumento de ciertas restricciones.

La inclusión de las variables ligadas con la función objetivo son las primeras en ser mostradas respecto del cómo incluirlas en el modelo de optimización. Para lograr esto, se muestra una versión generalizada del modelo de optimización para luego comenzar a construir las secciones correspondientes a la calidad, reflejada a través del rendimiento que se toma como el costo por unidad de rendimiento, mientras que la elección queda representada a través de políticas que se representan por restricciones y por el rendimiento directamente en la función objetivo.

La función objetivo del problema de optimización queda definida en forma general como un *trade-off* entre los costos y una función que refleje el rendimiento de cada establecimiento, la que está acompañada de un parámetro $\lambda \geq 0$ que permita modelar distintos escenarios al ir cambiando el parámetro que indica la importancia que se le da al rendimiento en la optimización.

$$F.O. = \text{Min} \sum_j^J (\text{costos} - \lambda \text{rendimiento})_j$$

El cómo incluir sólo un segmento de los modelos obtenidos para el rendimiento produce un problema relacionado con el escaso ajuste que tiene el modelo completo, ya que al tomar sólo una fracción de él, su ajuste empeora mucho más. En este sentido es que una solución válida es introducir el puntaje SIMCE actual del establecimiento para corregir en parte el sesgo del modelo en sí y el sesgo de truncarlo.

La forma de incluir el rendimiento podría hacerse imputando directamente el puntaje SIMCE del establecimiento y agregando la función obtenida para el rendimiento a través de los coeficientes β y el valor que tomen las variables significativas, pero si se incluye el SIMCE actual, éste es dado ciertas características del establecimiento, por lo que se considera que la inclusión del puntaje y las variables debiera ser desde las características actuales, por lo que al cambiar la magnitud de las variables es se tendrá un aumento o disminución del rendimiento según sea el caso. Para ello se incorpora el rendimiento a través de la función que se define a continuación.

Sean I estudiantes, J grados, K cursos, L establecimientos y M profesores. Sean las siguientes variables:

A_{ijkl} 1 si el alumno i pertenece al grado j del curso k del establecimiento l , 0 de lo contrario.

$G_{c_{jkl}}$ 1 si el grado j pertenece al curso k del establecimiento l , 0 de lo contrario.

$Prof_{ml}$ 1 si el profesor m pertenece al establecimiento l , 0 de lo contrario.

$Recursos_l$ asignación de recursos para el establecimiento l .

Además, se definen las condiciones iniciales de cada establecimiento como:

$Matr_l^0$ matrícula actual del establecimiento l .

Gc_{kl}^0 los grados combinados que tiene actualmente el curso k del establecimiento l .

$Dotación_l^0$ dotación docente actual del establecimiento l .

$Recursos_l^0$ disponibilidad de recursos actual para el establecimiento l .

Con lo anterior, se define la función del rendimiento para el establecimiento l en base a los resultados de la regresión lineal como:

$$Rendimiento_l = SIMCE_l + \beta_1 \Delta Gc_l + \beta_2 (\Delta Gc_l)^2 + \beta_3 \text{Log}(\Delta Matr_l) + \beta_4 \Delta (Al/Prof)_l$$

y según lo obtenido con el modelo lineal jerárquico, el modelo del rendimiento para el establecimiento l queda definido como:

$$Rendimiento_l = SIMCE_l + \beta_1 \Delta Rec_l + \beta_2 \Delta Gc_l + \beta_3 \text{Log}(\Delta Matr_l)$$

Donde

$$\Delta Gc_l = \sum_k \left(\sum_j Gc_{jkl} - Gc_{kl}^0 \right),$$

$$\Delta Matr_l = \sum_{k=1}^K \sum_{j=1}^J \sum_{i=1}^I Al_{ijkl} - Matr_l^0$$

$$\Delta (Al/Prof)_l = \frac{\sum_{k=1}^K \sum_{j=1}^J \sum_{i=1}^I Al_{ijkl}}{\sum_{m=1}^M Prof_{ml}} - \frac{Matr_l^0}{Dotación_l^0} \text{ y}$$

$$\Delta Rec_l = (Recursos_l - Recursos_l^0)$$

El problema principal de esta forma de incluir la calidad, es que ésta está modelada como el rendimiento individual de un estudiante, por lo que en estricto rigor, el rendimiento del establecimiento l queda definido por alguna de las funciones presentadas anteriormente por el número total de estudiantes, lo cual implica que el problema se vuelve no lineal, por lo que no es posible llegar a una solución en la optimización.

Para solucionar este problema, se podría incluir otra solución desde el enfoque del establecimiento como un todo, por lo que se opta por construir una nueva regresión para el establecimiento como unidad de observación.

Debido a la metodología que utiliza el modelo lineal jerárquico, los establecimientos deberían ser agrupados a través de una unidad de agrupación mayor, lo que es una alternativa válida, pero al no encontrar una unidad de agrupación innata o que sea interesante de estudiar, se opta por la modelación del rendimiento a través de la regresión lineal solamente.

Para construir un nuevo modelo por establecimiento, se incluyen nuevas variables dentro de las que están el tiempo promedio de viaje por establecimiento y su versión al cuadrado, con lo que se busca un mejor indicador de la relación viaje-rendimiento en vez de utilizar el índice de exceso de viaje, se utiliza una variable *Dummy* que indica si el curso es combinado o no, la razón de profesores con asignación por excelencia y total de profesores del establecimiento, una variable categórica del tamaño del establecimiento, donde los establecimientos se dividen en 5 clusters, con lo que el segmento 1 agrupa a los establecimientos con menos de 202 estudiantes, el 2 a los que tienen entre 202 y 424, el 3 a los que tienen 425 y 770, el 4 a los que tienen entre 771 y 1170 y el 5 a los que tienen más de 1170 estudiantes, la dotación docente del establecimiento y una variable categórica para el tamaño del curso donde los cursos muy pequeños tienen hasta 10 estudiantes, los pequeños de 11 a 20, los medianos de 21 a 30, los grandes de 31 a 40 y los muy grandes más de 40 estudiantes. A continuación se presentan los resultados alcanzados con esta nueva regresión.

Cuadro 38: Regresión lineal del rendimiento promedio establecimiento

Variables	Coefficientes
Constante	238,92 (7,13)
Tiempo de Viaje Promedio	0,97 (0,27)
Curso Combinado	-5,29 (2,03)
Indígenas por Curso	-0,5 (0,12)
Género Profesor	4,89 (1,83)
IVE	-0,61 (0,06)
Antigüedad Profesor	0,24 (0,07)
(Tiempo de Viaje Promedio) ²	-0,01 (0,005)
% Profesores de Excelencia	29,20 (14,81)
Tamaño Establecimiento	-5,36 (1,85)
Alumnos Retirados	-0,12 (0,05)
Dotación Docente	0,49 (0,19)
Expectativas Profesor	3,41 (0,63)
Tamaño curso	-1,05 (1,00)

Variables	Coefficientes
N	1150
R ²	0,241
Test F	27,81

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

A través de este modelo se obtiene un resultado con un R² mayor que los obtenidos anteriormente en los modelos de rendimiento por estudiante y con variables que permiten aseverar que un establecimiento con una mayor dotación y, lo más importante, con profesores de calidad obtiene mucho mejores resultados. Un rendimiento mejora al tener cursos simples, y que éstos sean pequeños, al igual que la matrícula. En el caso del tiempo de viaje, se tiene que al incluir esta variable se logra identificar una relación no lineal con el, la cual es una relación cuadrática cóncava, que logra su máximo en 48,5 minutos de viaje. A través de este fenómeno se sigue observando que en muchos casos hay quienes buscan un establecimiento con mejor calidad sacrificando la cercanía, pero llega un momento en que el rendimiento se ve mermado por viajes excesivos, con lo que en el caso de los estudiantes que tienen un bajo capital social y no tienen más alternativa que asistir a un establecimiento lejano su rendimiento queda aún más limitado por tener que viajar mucho.

Con la regresión anterior es posible incorporar al modelo las variables estructurales curso combinado, tamaño del establecimiento y tamaño curso, así como también es posible incorporar el tiempo promedio de viaje y la dotación del establecimiento, con lo que la función del rendimiento queda como sigue:

$$\begin{aligned} \text{Rendimiento}_l = & \text{SIMCE}_l - 5,29\text{Combinado}_l - 5,36\Delta\text{TamañoEstablecimiento}_l \\ & - 1,05\Delta\text{TamañoCurso}_l + 0,97\Delta\text{TiempoViaje}_l - 0,01\Delta\text{TiempoViaje}_l^2 + 0,49\Delta\text{Dotación}_l \end{aligned}$$

$$\text{donde } \text{Combinado}_l = \sum_{k=1}^K \text{CC}_k, \text{ con } \text{CC}_k = \begin{cases} 1 & \text{si el curso } k \text{ tiene más de un grado} \\ 0 & \text{si no} \end{cases}$$

$$\Delta\text{TamañoEstablecimiento} = \text{TipoEstablecimiento}_l - \text{TipoEstablecimiento}_l^0$$

$$\Delta\text{TamañoCurso}_l = \sum_{k=1}^K (\text{TipoCurso}_{kl} - \text{TipoCurso}_{kl}^0)$$

$$\Delta\text{TiempoViaje}_l = \text{TiempoViajeProm}_l - \text{TiempoViajeProm}_l^0$$

$$\Delta\text{TiempoViaje}_l^2 = \text{TiempoViajeProm}_l^2 - (\text{TiempoViajeProm}_l^0)^2$$

$$\Delta\text{Dotación}_l = \text{CantProf}_l - \text{CantProf}_l^0$$

con

$\text{TipoEstablecimiento}_l^0$ el segmento de matrícula actual al que pertenece el establecimiento,

TipoCurso_{kl}^0 el segmento de tamaño de curso actual de cada curso k ,

$\text{TiempoViajeProm}_l^0$ el viaje promedio actual del establecimiento l ,

$CantProf_l^0$ la cantidad actual de profesores del establecimiento l .

Con lo anterior la relación del modelo con la calidad queda completa y no presenta problemas para la optimización, con lo que se da paso a la inclusión de las variables de elección. En el caso de la presencia educación secundaria, ésta se da naturalmente a través de las economías de escala que se dan al tener un establecimiento más grande y al haber demanda para ella, al igual que el tamaño de curso así como también la dotación y el tiempo de viaje son objetos de optimización, pues ellos están relacionados con un costo que se incluye en la función objetivo. Sin embargo, las características que no están representadas por costos se pueden obtener a través de políticas que permitan simular diversos escenarios para que la autoridad tome una decisión. En este caso se puede establecer un tamaño máximo de curso y una penalización por sobrepasar ese máximo y una penalización por no tener educación secundaria. En el caso de los cursos combinados se puede establecer una matrícula máxima con la que se pueden abrir este tipo de cursos, y también un tamaño máximo de estudiantes en este tipo de cursos. Las restricciones enunciadas anteriormente quedan como sigue. Sean las siguientes variables:

$$EdS_l = \begin{cases} 1 & \text{si el establecimiento } l \text{ tiene educación secundaria} \\ 0 & \text{si no} \end{cases}$$

$$CostoTC_{kl} = \begin{cases} 1 & \text{si el curso } k \text{ del establecimiento } l \text{ tiene más de } TCMáx \\ 0 & \text{si no} \end{cases}$$

$$CostoCC_{kl} = \begin{cases} 1 & \text{si el curso } k \text{ del establecimiento } l \text{ es combinado} \\ 0 & \text{si no} \end{cases}$$

donde $TCMáx$ es el tamaño máximo de un curso. Las políticas de compensaciones y penalizaciones se incorporan a través de parámetros y las variables en la función de costos:

$$Costos_l = CostosEstablecimiento_l - \alpha CostoTC_{kl} - \beta CostoCC_{kl} + \delta EdS_l, \quad \alpha, \beta, \delta \geq 0$$

además, se puede restringir la posibilidad de que un establecimiento tenga cursos combinados si sobrepasa una cierta matrícula, de modo que no sólo sea a través de la variable $CostoCC$ que se penalice este tipo de cursos. Esto se obtiene a través de la siguiente restricción que establece una matrícula máxima que permite la existencia de cursos combinados:

$$CC_{kl} = 0 \quad \text{si } \sum_{i=1}^I Al_{il} \leq MatrMáx$$

Mediante las variables y parámetros anteriores se logra incorporar las variables de rendimiento y elección para los establecimientos existentes. En el caso de los nuevos establecimientos también se aplican las mismas políticas. La diferencia radica en la función del rendimiento, por lo que el rendimiento asignado a esos establecimientos es en base a la media de los resultados SIMCE obtenidos por los establecimientos que cumplan con el mismo perfil al que pertenezca el nuevo establecimiento, de modo de asignarle un puntaje SIMCE y haciendo cero la diferencia entre la situación “nueva” y “actual” del establecimiento nuevo.

Con el planteamiento anterior se recogen e incorporan todas las variables estructurales que se han encontrado con la presente investigación quedando así una serie de parámetros que permiten simular diferentes escenarios que privilegien de distintas características para tomar una decisión. Esta es una forma de incorporar lo obtenido, sin embargo, dado el ajuste del modelo de rendimiento el incorporar una función de él en la optimización puede llevar a estimaciones erradas, pues no se sabe a ciencia cierta cuál es la elasticidad de una unidad de calidad.

Si bien se han tratado de corregir al máximo los errores al incluir las diferencias de las variables con la situación actual y el puntaje actual de cada establecimiento, no se puede especificar la magnitud de la corrección que logran estas imputaciones. Sin embargo, a través de la investigación se han establecido resultados consistentes que han permitido identificar factores que inciden en el rendimiento, por lo que si no se quiere incorporar la función obtenida, al saber que estos elementos tienen efectivamente una influencia sobre el rendimiento se pueden incorporar a través de políticas que permitan modelar escenarios variando los parámetros o restricciones impuestas a las variables que influyen en el rendimiento de los estudiantes. Con esta estrategia, se logra disponer de la información de costos asociados a diferentes escenarios en los cuales se controlan algunos atributos incidentes en el rendimiento de los alumnos. De este modo, se puede estimar no sólo el valor absoluto de estos costos, sino que también el cambio en ellos ante diferentes escenarios de valores de los atributos mencionados.

Finalmente, es esta última forma de incluir los resultados obtenidos mediante el presente trabajo de título la que se recomienda para complementar el modelo de optimización desarrollado para el SIGER ya que logra recoger los atributos que influyen en el rendimiento y en la elección de los padres y consigue generar distintos escenarios que asocian un costo a la variación de éstos sin incluir los parámetros encontrados por un modelo que sólo logra ajustar un pequeño porcentaje de los datos y, más aún, está truncado al sólo incluir las componentes estructurales de los establecimientos.

CONCLUSIONES

El objetivo de esta investigación es definir e incorporar las variables significativas de calidad de la educación y elección de establecimientos en el mundo rural en un modelo de optimización de localización. Dicho objetivo ha sido alcanzado a través de una revisión bibliográfica que permitió definir las variables que se han estudiado en el mundo sobre ambos temas de modo de tener una noción de cuáles variables incluir en los modelos econométricos aplicados durante el estudio, gracias a los cuales se ha logrado establecer qué variables son importantes en la calidad, estimada a través del rendimiento de los estudiantes, y las características que son valoradas por los padres para tomar la decisión acerca del establecimiento educacional de sus hijos para luego mostrar distintas alternativas acerca del cómo incluirlas en el modelo de optimización.

Variables de Calidad

Los factores que han sido estudiados en diversas investigaciones se pueden clasificar en características del establecimiento, materiales educativos, características del profesor y del director, características familiares del estudiantes y motivación e inserción del estudiante en el sistema educativo.

A través de la revisión bibliográfica se constata que la investigación de este tema ha sido ampliamente abordada, pero no hay resultados que concilien la opinión de los distintos investigadores.

La dificultad de esta área de investigación está relacionada con que existen una serie de variables que no son fáciles de medir porque no existen las herramientas adecuadas o son basadas en apreciaciones subjetivas. Específicamente, el talento de los estudiantes, la disciplina del establecimiento, la conducta del estudiante, la metodología de enseñanza y la calidad del cuerpo docente son factores que, se cree, tienen una influencia importante en el rendimiento, pero su inclusión en modelos matemáticos ha sido difícil y cuestionada dadas las características subjetivas de ellas. Ésta es la principal razón por la que las investigaciones econométricas sobre educación presentan un ajuste sumamente bajo, donde los mejores resultados se acercan al 20%. Sin embargo, el nivel de ajuste de los modelos de calidad calibrados en este trabajo logran un nivel cercano al de las investigaciones que han sido consultadas en la revisión bibliográfica.

El escaso nivel predictivo que tienen los modelos calibrados ha dado paso a la inquietud sobre cuál o cuáles son esas variables que no se han incorporado y aportan parte importante del ajuste al modelo. A pesar de las falencias que tienen los modelos obtenidos al omitir variables importantes y al obtener resultados, en cuanto al ajuste y variables, similares a los de las investigaciones nacionales e internacionales, se obtienen factores que se presentan en gran parte de los modelos con coeficientes que concuerdan en valor y signo, con lo cual se puede aseverar que ellos sí tienen una influencia en el rendimiento de los estudiantes.

Con el trabajo desarrollado se puede afirmar que el establecimiento es importante para los estudiantes, el cual según el modelo lineal jerárquico sería responsable en un 18% del rendimiento escolar rural aproximadamente. Esto resulta sumamente importante, pues desde el principio de las investigaciones acerca del rendimiento se ha cuestionado si el establecimiento hace la diferencia o no. Además, al segmentar por GSE se comprueba que quienes más afectados

se ven por las características del establecimiento son los más pobres, quienes a su vez presentan los rendimientos más bajos. Es gracias a estos resultados que se puede seguir pensando, creando y trabajando para que Chile le entregue una educación de calidad y con equidad a sus niños.

Los factores que se encuentran como determinantes del rendimiento a nivel del establecimiento son: la existencia de cursos multigrado, lo que afecta de forma negativa a los estudiantes por lo que es de suma importancia mejorar este aspecto; los recursos educativos con que cuenta el establecimiento hacen que sus estudiantes tengan un mejor rendimiento si es que están mejores dotados; el tamaño del curso, que al ser mayor tendría un efecto negativo en el rendimiento al igual que el tamaño de la matrícula; el índice de vulnerabilidad escolar, el cual tiene, como resulta intuitivo, una relación negativa con el rendimiento; y finalmente, la dotación docente y la razón de docentes de excelencia que posea el establecimiento, los que son muy importantes para obtener mejores resultados, por lo que mejorar la calidad de los profesores es una tarea sumamente necesaria para lograr superar los niveles actuales de rendimiento.

Las características del profesor que afectan al rendimiento de los estudiantes son la antigüedad que tengan en el establecimiento, la que estaría ligada a la experiencia, compromiso con la comunidad en la que se desenvuelven y a la capacidad de los establecimientos para retener a los buenos profesores; las expectativas que el profesor tenga de sus estudiantes, lo que puede hacer que su trabajo sea de mejor calidad ya que sabe que sus alumnos tienen un futuro por delante y el género, el cual muestra una ventaja de las mujeres por sobre los hombres cuando de conseguir mejores resultados académicos de los estudiantes se trata.

En el caso de las características propias del estudiante, las que resultan más influyentes en el rendimiento son la educación alcanzada por la madre, las expectativas que tengan del estudiante sus padres y la satisfacción con la calidad del establecimiento. Si el estudiante pertenece a un hogar mono parental, su rendimiento es mejor en comparación a un estudiante que sólo difiere en esta característica, lo que estaría dado por la mayor atención que ponen las familias en la educación de los niños de hogares incompletos.

Cómo afecta el exceso de viaje al rendimiento individual no logra ser aclarado en esta investigación, pues se detecta que en algunos casos es una causa y en otros un efecto, ya que hay estudiantes que tendrían un bajo rendimiento por asistir a establecimientos lejanos y otros estudiantes viajan mucho para asistir a establecimientos de buena calidad, lo que les permitiría obtener mejores resultados.

Finalmente y a pesar de haber obtenido modelos con un bajo ajuste, se logra establecer qué variables son las que ayudarían a mejorar la calidad de la educación de los estudiantes de zonas rurales, encontrando un conjunto de ellas que son factibles de incluir en el modelo de optimización de localización.

Variables de Elección

Los factores que afectan la elección por parte de la demanda, se pueden agrupar en características de la oferta, familiares y del establecimiento.

Es importante notar que las características de la oferta condicionan el comportamiento de la demanda, así como también la calidad, pues al no existir oferta suficiente para que los estudiantes continúen sus estudios en la educación secundaria, padres y profesores ven afectadas sus

expectativas, haciendo incluso que el estudiante abandone sus estudios antes de completar la educación básica.

Las percepciones y preferencias de los consumidores no son un tema fácil de abordar, pues suelen ser subjetivas y muchas veces difieren del comportamiento real del consumidor. Esto ha llevado que se desarrollen más bien métodos de investigación cualitativa, con lo que la opinión de los investigadores afecta bastante a estos estudios, sin embargo la aplicación de modelos matemáticos ha ido en aumento en este tipo de investigaciones.

En el caso de las preferencias declaradas por los padres, independientemente del GSE, ellos responden que la calidad es lo que más les importa, pero es lo que cada uno entiende sobre calidad lo que realmente los hace elegir. Estudios han demostrado que los más ricos entienden calidad como los resultados del establecimiento, ya sea en SIMCE o PSU en el caso de Chile, mientras que para los más pobres la calidad tiene que ver con la disciplina y con el nivel socioeconómico del establecimiento.

En cuanto a la elección de los establecimientos, los padres manifiestan que la cercanía es una de las razones más importantes al momento de elegir el establecimiento, pero se ha detectado que existe un segmento que viaja mucho más para recibir una educación de mayor calidad, con lo que se demuestra que las preferencias reveladas muchas veces no coinciden con el comportamiento final del consumidor, en este caso sobre qué variable hacen la elección los padres. Además, son muy pocos los que manifiestan el puntaje SIMCE como un factor de elección, pero a través de la aplicación del modelo Logit se logra establecer que los padres prefieren establecimientos con un mejor rendimiento. Además, el tamaño de éste también tiene una influencia importante.

Las razones que hacen la diferencia en la preferencia de una de las opciones disponibles de establecimientos son además, la cantidad de grados combinados, la presencia de educación secundaria y el tamaño del curso, con lo que se puede concluir que, en general, lo que prefieren los padres de zonas rurales son establecimientos con cursos simples, de tamaño medio, con educación secundaria y buen rendimiento en el SIMCE.

Se analiza también el efecto que tienen las variables dentro de los grupos socioeconómicos, constatándose que, en general, influyen las mismas características a nivel transversal, por lo que se puede concluir que la calidad del establecimiento es una razón clara de elección en todos los estratos.

Además, el hecho de acercar la oferta de establecimientos a los centros donde se emplaza la demanda a través del SIGER, está acorde con las preferencias declaradas de los padres, con lo que al tomar en consideración el costo de transporte para la cercanía y el rendimiento del establecimiento, es coherente con los intereses de las personas del mundo rural.

Inclusión de Variables en el Modelo de Optimización

Con el desarrollo de este capítulo se detecta que el incluir el modelo de rendimiento en la función objetivo obtenido en el capítulo de calidad implica que se convierta en un problema no lineal, pues se debe incluir a su vez el rendimiento de cada estudiante. Como una forma de enfrentar este obstáculo para la resolución del problema de optimización se realiza una nueva regresión a nivel de establecimiento con lo que el modelo obtenido puede ser imputado directamente en el modelo de optimización para cada establecimiento.

Para construir esta nueva regresión, se modifican ciertas variables, como si el curso es combinado o no, en vez de la cantidad de grados, el tiempo promedio de viaje del establecimiento, ya que el índice no ha logrado hasta el momento hacer una predicción clara, una variable categórica para el tamaño del establecimiento y otra para el curso, la dotación docente y la razón de profesores de excelencia frente al total de profesores. Con estas modificaciones se logra definir un modelo con un ajuste mayor para los establecimientos y que no presenta problemas de no linealidad en la función objetivo.

Las variables que son tomadas en cuenta por el modelo en el caso de la calidad de la educación son el tiempo de viaje promedio, quien presenta una relación cóncava con el rendimiento, por lo que pasado cierto límite éste influye negativamente; cursos combinados, que tienen un efecto negativo sobre el rendimiento; el tamaño del establecimiento y del curso, quienes influyen negativamente al aumentar su magnitud; y la dotación docente, que tiene un efecto positivo al ser mayor.

Los factores que tratan de privilegiarse para incorporar las preferencias sobre la elección de establecimientos, son los resultados del SIMCE, donde más es mejor, la ausencia de cursos combinados, presencia de educación secundaria en los establecimientos y cursos de tamaño medio.

Después de analizar distintas estrategias para incorporar las variables antes mencionadas, se recomienda que esto se lleve a cabo a través de restricciones y parámetros que permitan simular distintos escenarios para tomar una decisión al hacer variar a éstos. Un ejemplo de ello es incluir un tamaño máximo de cursos como restricción y variar ese máximo para obtener distintas soluciones y que sea la autoridad competente quien tome la decisión frente a un conjunto de soluciones.

La forma de abordar la inclusión de variables aborda todos los factores encontrados como relevantes en el estudio sin tener que incorporar parte del modelo de rendimiento encontrado, el cual al incluir todas las variables tiene un bajo ajuste, por lo que al truncarlo, su ajuste sería aún peor, con lo que se incorporan todos los elementos importantes, sin introducir coeficientes asociados a ellos que estén errados y puedan llevar a una resolución equívoca del modelo de optimización.

Finalmente, a través de esta investigación se logra definir las variables que son importantes para acercarse al óptimo social de localización de establecimientos rurales. Sin embargo, los modelos obtenidos, en particular el de rendimiento, no cuentan con una serie de variables que no han sido posibles de medir dada la subjetividad de la que son sujeto, por lo cual queda abierta una gran área de investigación que permita determinar las variables que faltan en el modelo y cómo medirlas, donde se pueden probar e idear nuevas metodologías de medición de calidad y de las variables omitidas para desarrollar modelos más completos y con mejor ajuste. Sin lugar a dudas, el encontrar estas metodologías permitirá encontrar qué es lo que falta para implementar una educación de calidad y con equidad para todos los niños del país, en especial para los que tienen más carencias, de modo de reducir la brecha de pobreza y construir un país más justo y humano.

BIBLIOGRAFÍA

ABDI, H. 2003. Least squares. Encyclopedia of Social Sciences Research Methods. Thousand Oaks (CA)

ABERG-BENGSTSSON, L. 2004. Do small rural schools differ? A comparative two-level model of reading achievement among Swedish 9-years-old. Scandinavian Journal of Educational Research 48 (1): 19-33

ÁLVAREZ, M.I “et al”. 1998. Computers in schools: a qualitative study of Chile and Costa Rica”, Education and Technology Series, The World Bank.

ARANCIBIA, V., SCHMIDT, P., y SANCHO, A. 1998. Proyecto efectividad escolar. Libertad y Desarrollo, Santiago.

BLANCO, E. 2008. Factores escolares asociados a los aprendizajes en la educación primaria mexicana: un análisis multinivel. Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación 6(1):58-84

BRUNNER, J. y ELACQUA, G. 2003. Informe de capital humano en Chile. Santiago, Universidad Adolfo Ibáñez, Escuela de Negocios.

CASTRO, M. 1999. Proyecto docente. Universidad de Murcia.

CIDE. 2003. VI encuesta nacional "actores del sistema educativo". Universidad Alberto Hurtado.

CLAVEL, C. y SCHIEFELBEIN, E. 1979. Factores que inciden en la demanda por educación. Programa ECIEL.

CÓRDOBA, C. 2006. Elección de escuela en sectores pobres de Chile. Simposio: Políticas educacionales y gestión de sistemas escolares. Universidad Complutense de Madrid

DE LA ORDEN, A. “et. al”. 2001 Modelos de investigación del bajo rendimiento, Revista Complutense de Educación 12 (1): 159-178

DOSTIEY, B. Y JAYARAMAN, R. 2006. Determinants of school enrollment in Indian villages. Journal Economic Development and Cultural Change 54 (2): 405-421

DONOSO, S y HAWES, G. 2002. Eficiencia escolar y diferencias socioeconómicas: a propósito de los resultados de las pruebas de medición de la calidad de la educación en Chile.

EDWARDS, J. 2003. Causas de los bajos rendimientos en la educación rural. Tesis de Ingeniero Comercial, mención Economía. Santiago, Universidad de Chile, Facultad de Ciencias Económicas y Administrativas, Escuela de Economía y Administración.

ELACQUA, G., SCHNEIDER, M. y BUCKLEY, J. 2006. School choice in Chile: is it class or the classroom? Journal of Policy Analysis and Management 25 (3): 577 -601.

ESCRIBANO M. y FERNÁNDEZ, G. 2003. El análisis de la robustez y la ayuda a la decisión multicriterio discreta. Anales de Economía Aplicada, San Pablo, Universidad San Pablo,

Facultad de CC. Económicas y Empresariales, Departamento de Métodos Cuantitativos para la Economía.

GALLEGO, F., RODRÍGUEZ, C. y SAUMA, E. 2007. Provisión de educación en zonas rurales de Chile: incentivos, costos y calidad. Estudio de la Pontificia Universidad Católica de Chile.

GALLEGO, F y HERNANDO, A. 2007. School choice in Chile: looking at the demand side. Estudio de la Pontificia Universidad Católica de Chile.

GREENE, W.1998. Análisis Econométrico, 3ª ed. Madrid, Prentice Hall.

GUIMBERT, S. “et al”. 2008. Back to school in Afghanistan: Determinants of school enrollment. International Journal of Educational Development 28: 419- 434

GUJARATI, D. 1997. Econometría. 3ª ed. Santa Fe de Bogotá, Mc Graw Hill.

GUISANDE, C. Et al, 2006. Tratamiento de datos. España, Díaz de Santos.

HERNÁNDEZ, M “et. al”. 2003. Anales de Psicología. Servicio de Publicaciones de la Universidad de Murcia, 19,(1) :159-171.

KRUEGER, A. 1997. Experimental Estimates of Education Production Functions. NBER Working paper N° 6051.

KUZIEMKO, I. 2006. Using shocks to school enrollment to estimate the effect of school size on student achievement. Economics of Education Review 25: 63-75.

LEON, G. 2006. Metodología para implementar políticas de superación de pobreza en el ámbito de gestión municipal en educación, Tesis de Ingeniero Civil Industrial. Santiago, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas, Departamento de Ingeniería Industrial.

LIN, T. 2008. Local taxes and parental choice of education: the case of Pennsylvania. Economics Bulletin, 9(24): 1- 7.

MARADONA, G y CALDERÓN, I. 2004. Una aplicación del enfoque de la función de producción en educación .Revista de Economía y Estadística Vol. XLII. Instituto de Economía y Finanzas. Facultad de Ciencias Económicas, Universidad Nacional de Córdoba - Argentina

MAYSTON, D. 2000. The demand for education and the production of local public goods. Discussion Papers N° 00/50, Department of Economics, University of York.

MELLA, O. 2006, Factores que afectan los resultados de la escuela pública chilena. REICE - Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación 4 (1): 29- 37.

MIZALA, A y MELIS, F. 2007. Escuelas rurales y su situación financiera. Departamento de Ingeniería Industrial Universidad de Chile.

MIZALA, A. y ROMAGUERA, P.1998b. Desempeño escolar y elección de colegios: La experiencia chilena, Documentos de Trabajo N° 85, Centro de Economía Aplicada, Departamento de Ingeniería Industrial, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile.

MIZALA, A. y ROMAGUERA, P. 2000. Determinación de factores explicativos de los resultados escolares en educación media en Chile. Serie Economía No. 36, Centro de Economía Aplicada, Departamento de Ingeniería Industrial, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile.

MIZALA, A., ROMAGUERA, P. y REINAGA, T. 1999. Factores que Inciden en el Rendimiento Escolar en Bolivia. Documentos de Trabajo N° 61, Centro de Economía Aplicada, Departamento de Ingeniería Industrial, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile.

MUÑOZ-REPISO, M., CERDÁN, J. y MURILLO, F. 1995. Calidad de la educación y eficacia de la escuela. Estudio sobre la gestión de los recursos educativos. Madrid: CIDE.

MURILLO, F. 1999. Los modelos jerárquicos lineales aplicados a la investigación sobre eficacia escolar. *Revista de Investigación Educativa*, 17(2): 453-460.

MURILLO, F. "et al". 2007. Investigación iberoamericana sobre eficacia escolar. Bogotá, Convenio Andrés Bello.

OJEDA, M. 2000. Biplots para los residuos del segundo nivel de un modelo lineal jerárquico con dos niveles. *Revista Investigación Operacional*, Universidad Veracruzana, 21(3): 224-235.

OSTOIC, C. 2003. Nuevas Metodologías para el análisis de los resultados educacionales: una aplicación de modelos lineales jerárquicos a los datos de Chile. Tesis de Magíster en Economía Aplicada. Santiago, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas, Departamento de Ingeniería Industrial.

PAVEZ, M. 2003. Municipios Efectivos en Educación. Tesis de Magíster en Gestión y Políticas Públicas. Santiago, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas, Departamento de Ingeniería Industrial.

PETERS, P. y HALL, G. 2004. Evaluation of education quality and neighborhood well-being: a case of study of Independencia, Peru. *International Journal of Educational development* 24: 85-102

POTASSHNIK, M. 1996. Chile's learning network. Education Technology Series, 1(2), The World Bank.

SCHAFFT, K. 2006. Poverty, residential mobility, and student transiency within a rural New York school district. *Rural Sociology* 71 (2): 212- 231.

STAIR, A. "Et al". 2006 Demand for public education: Evidence from a rural school district. *Economics of Education Review* 25: 521 – 531.

SCHIEFELBEIN, E. y SCHIEFELBEIN, P. 2000. Determinantes de la calidad: ¿qué falta mejorar?. *Revista Perspectivas*, Departamento de Ingeniería Industrial, Universidad de Chile. 4(1):37-64.

SCHUMACKER, R. Y BEMBRY, K. 1995. Centering effects in HLM level-1 predictor variables. American Educational Research Association annual meeting. San Francisco, USA.

ULUBASOGLU, M. y CARDAK, B. 2007. Internationals comparisons of rural-urban educational attainment: Data and Determinants. *European Economic Review* 51: 1828 -1857

UNICEF 2000. Defining Quality in Education. Meeting of the International Working Group on Education Florence, Italy June 2000.

VEDDER, R. 2004. Effective, efficient, fair: Paying for public education in Texas. Ohio University and Joshua Hall Buckeye Institute.

VELEZ, E.; SCHIEFELBEIN, E. y VALENZUELA, J. 1994. Factores que afectan el rendimiento académico en la educación primaria (Revisión de la Literatura de América Latina y el Caribe). Revista Latinoamericana de Innovaciones Educativas pp: 1-16.

WHITE, F. y TWEETEN, L. 1973. Optimal School district size emphasizing rural areas, American Journal of Agricultural Economics, 55 (1): 45-53.

WILLIAMSON, G.2000. Estudio sobre la educación para la población rural en Chile. En: Educación para la población rural en Brasil, Chile, Colombia, Honduras, México, Paraguay y Perú. Paris, FAO y UNESCO. pp. 97 a 169.

ANEXOS

ANEXO I: Metodología de estimación de variables

I. Índice de Exceso de Viaje

Este análisis se basa en el concepto de que el tiempo de viaje y medio de transporte son indicadores válidos de la accesibilidad de los distintos establecimientos dando una noción de la facilidad o dificultad de acceder a éstos por parte de sus estudiantes.

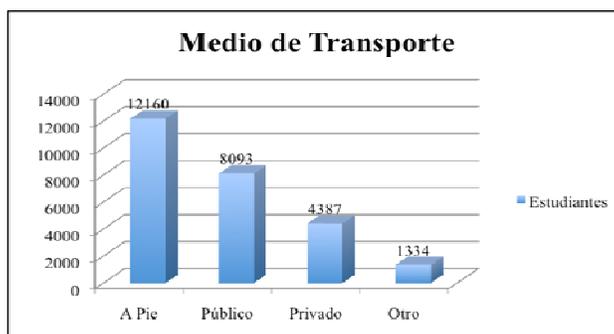
El medio de transporte utilizado por la mayor parte de los estudiantes es la caminata hacia sus establecimientos, seguido por el uso del transporte público. El uso de la caminata u otro, ya sea bicicleta, caballo, etc. superan el 50% de las preferencias, lo que se puede deber principalmente a que estos alumnos son menores, lo que hace suponer que ellos tienen menos libertad para utilizar el transporte público para trasladarse, pues la mayoría lo hace acompañado de un adulto, lo que encarece esta opción, dejándola en el segundo lugar de las preferencias.

Cuadro 1: Estudiantes según Medio de Transporte.

Medio de	Estudiantes	%
A Pie	12160	46,82%
Público	8093	31,16%
Privado	4387	16,89%
Otro	1334	5,14%

Fuente: Elaboración propia en Base Cuestionario Padres SIMCE 2004

Figura 1: Estudiantes según Medio de Transporte



Fuente: Elaboración propia en Base Cuestionario Padres SIMCE 2004

Los diversos grupos socioeconómicos muestran comportamientos diferentes en cuanto al medio de transporte escogido, siendo el transporte privado escogido por los dos grupos socioeconómicos más ricos, mientras que la caminata es el medio de transporte que utilizan los 3 primeros grupos socioeconómicos. Además, al analizar los tiempos medios de transporte de los distintos grupos socioeconómicos, llama la atención la alta cantidad de estudiantes que viajan en “otro” y la cantidad de tiempo que esto les toma, pues, en los 3 primeros grupos socioeconómicos superan el umbral de exceso de viaje.

Cuadro 2: Medios de Transporte según Grupo Socioeconómico

Medio de Transporte	A	B	C	D	E
A Pie	47,20%	46,40%	46,40%	23,10%	15,40%
Público	31,90%	28,30%	28,30%	5,80%	12,80%
Privado	15,50%	21,40%	21,40%	71,20%	70,90%
Otro	5,40%	3,90%	3,90%	0,00%	0,90%
TOTAL	86,90%	11,10%	11,10%	0,20%	0,50%

Fuente: Elaboración propia en Base Cuestionario Padres SIMCE 2004

Cuadro 3: Tiempo de Viaje según Medio de Transporte

Medio de Transporte	Tiempo [Minutos]	A	B	C	D	E
A Pie	14	15	14	12	8	9
Público	29	30	32	53	38	28
Privado	24	35	24	31	12	18
Otro	26	23	25	31	0	0

Fuente: Elaboración propia en Base Cuestionario Padres SIMCE 2004

En cuanto al transporte público y privado se tiene que el promedio general y de todos los grupos socioeconómicos no supera el umbral, pero en el caso del transporte público el promedio general y en casi todos los grupos socioeconómicos es bastante alto, lo que no debiera suceder pues estos alumnos son menores y no debieran demorarse tanto en el traslado a sus establecimientos.

Para finalizar, se analizan los casos donde se excede el tiempo de viaje fijado como umbral para poder evaluar más a fondo la situación de la Educación Básica.

Cuadro 4: Cantidad de estudiantes que sobrepasan el umbral de tiempo de viaje

Medio de Transporte	Total Estudiantes	Viajes sobre umbral	Índice de Exceso de Viaje
A Pie	12160	622	5%
Público	8093	449	6%
Privado	4387	254	6%
Otro	1334	102	8%
TOTAL	25974	1427	5%

Fuente: Elaboración propia en Base Cuestionario Padres SIMCE 2004

Cuadro 5: Ranking Regional según Exceso de Viaje

Región	Total Estudiantes	Viajes sobre Umbral	Índice de Exceso de Viaje
X	3905	318	8,10%
VIII	2115	206	9,70%
IX	4842	97	2,00%
XIII	1626	78	4,80%
V	1814	55	3,00%
VI	2828	50	1,80%
I	369	30	8,10%

Región	Total Estudiantes	Viajes sobre Umbral	Índice de Exceso de Viaje
III	157	5	3,20%
XI	202	3	1,50%
II	26	0	0,00%
XII	53	0	0,00%
TOTAL	25974	1427	5,50%

Fuente: Elaboración propia en Base Cuestionario Padres SIMCE 2004

Las regiones que tienen los índices más altos son justamente las que concentran la mayor cantidad de establecimientos rurales, lo que hace suponer que podrían haber zonas donde existe una sobreoferta y otras donde no se puede acceder a establecimientos a una corta distancia o tiempo de viaje.

II. Grupos socioeconómicos

La metodología consiste en clasificar en 5 clusters según educación de los padres, ingreso y el IVE. Al tratarse sólo de las familias en este caso se incluyen las variables de educación e ingresos.

Cuadro 5: Clasificación de ingresos

Tramo	Ingreso	Valor asignado
1	bajo \$100.000	100.000
2	entre \$100.000 y 200.000	200.000
3	entre \$201.000 y 300.000	300.000
4	entre \$301.000 y 400.000	400.000
5	entre \$401.000 y 500.000	500.000
6	entre \$501.000 y 600.000	600.000
7	entre \$601.000 y 800.000	800.000
8	entre \$801.000 y 1.000.000	1.000.000
9	entre \$1.001.000 y 1.200.000	1.200.000
10	entre \$1.201.000 y 1.400.000	1.400.000
11	entre \$1.401.000 y 1.600.000	1.600.000
12	entre \$1.601.000 y 1.800.000	1.800.000
13	entre \$1.801.000 y \$2.000.000	2.000.000
14	entre \$2.001.000 y \$2.200.000	2.200.000
15	Más de \$2.200.000	2.400.000

Fuente: Manual de Uso SIMCE 2006 para 4° básico

Cuadro 6: Clasificación del Nivel Educativo

Nivel Educativo	Años de estudios
Sin años de estudio	0 años
1° año de Educación Básica	1 años
2° año de Educación Básica	2 años
3° año de Educación Básica	3 años
4° año de Educación Básica	4 años
5° año de Educación Básica	5 años

Nivel Educativo	Años de estudios
6° año de Educación Básica	6 años
7° año de Educación Básica	7 años
8° año de Educación Básica	8 años
1° año de Educación Media	9 años
2° año de Educación Media	10 años
3° año de Educación Media	11 años
4° año de Educación Media	12 años
4° o 5° año de Educación Media Técnico Profesional	12 años
Educación Incompleta en un Centro de Formación Técnica o Instituto Profesional	14 años
Titulado de un Centro de Formación Técnica o Instituto Profesional	16 años
Educación Incompleta en una Universidad	15 años
Titulado de una Universidad	17 años
Grado de Magíster Universitario	19 años
Grado de Doctor Universitario	22 años

Fuente: Manual de Uso SIMCE 2006 para 4° básico

Luego de clasificar los ingresos y años de estudio, se limpia y estandariza la base de datos para realizar la segmentación. El resultado de ésta arroja los grupos socioeconómicos.

ANEXO II: Estadísticos descriptivos rendimiento

Cuadro 1: estadísticos descriptivos por establecimiento

	Mínimo	Máximo	Media	Desv. Típ.
Grados Combinados	1	6	1,77	1,193
Tamaño curso	5	47	21,93	8,025
Matrícula Establecimiento	12	1119	172,81	176,252
SIMCE establecimiento	127	316	228,96	28,705
Razón Alumno-profesor	4,2	52	16,79	7,05
Género profesor	0	1	0,73	0,442
Antigüedad profesor	0	49	20,96	11,815
Expectativas profesor	0,14	1	0,51	0,18
Selección	0	1	0,24	0,429
Indígenas curso	0	1,571	0,168	0,293
Uso recursos	0	1	0,454	0,152
IVE	0	95,6	63,391	13,465
Municipal	0	1	0,84	0,369
Subvencionado	0	1	0,15	0,362
GSE curso	1	5	1,36	0,563
Centro de Padres	0	1	0,93	0,258
Horas jornada escolar	4:00	9:30	7:08	1:06
Educación secundaria	0	1	0,08	0,266
JEC	0	1	0,55	0,497
Disponibilidad recursos	0	1	0,48	0,18

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

Cuadro 2: Estadísticos descriptivos estudiantes

	Mínimo	Máximo	Media	Desv. Típ.
Rendimiento	74,27	359,43	235,1856	54,96173
Género Estudiante	0	1	0,49	0,5
GSE Familia	1	5	1,28	0,665
Educación padre	1	21	10,53	3,92
Educación madre	1	21	10,97	3,523
Ingresos	1	13	2,21	1,869
Hogar Incompleto	0	1	0,03	0,181
Tamaño familiar	1	10	3,96	1,599
Expectativas padres	1	8	5,02	1,694
Satisfacción calidad	0	9	2,76	1,167

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

ANEXO III: Matrices de correlaciones y análisis factorial rendimiento

Cuadro 1: Matriz de correlaciones establecimiento

	Matrícula	Disponibilidad Recursos	IVE	Municipal	GSE curso	Exceso de viaje	Cantidad de cursos	Tamaño Curso	Indígenas curso	Uso Recursos	% Retirados	Educación Secundaria	Centro de Padres	Horas jornada escolar	Selección	Pago mensual
Disponibilidad Recursos	0,202	1														
IVE	-0,516	-0,273	1													
Municipal	-0,344	-0,193	0,439	1												
GSE curso	0,453	0,272	-0,877	-0,446	1											
Exceso de viaje	-0,142	-0,084	0,201	-0,027	-0,157	1										
Cantidad de cursos	-0,295	-0,079	0,228	-0,086	-0,188	0,248	1									
Tamaño Curso alumnos	0,61	0,178	-0,355	-0,288	0,296	-0,165	-0,076	1								
indígenas curso	-0,092	-0,063	0,245	-0,183	-0,216	0,166	0,268	0,056	1							
Uso Recursos	-0,079	0,329	0,032	0,101	-0,039	0,002	-0,066	-0,077	-0,01	1						
% Retirados	-0,046	-0,1	0,095	0,089	-0,103	0,092	0,093	-0,004	0,096	-0,004	1					
Educación Secundaria	0,587	0,151	-0,449	-0,369	0,474	-0,036	-0,115	0,246	-0,075	-0,088	-0,04	1				
Centro de Padres	-0,008	-0,047	0,193	0,177	-0,187	0,059	0,037	-0,04	0,041	-0,043	0,072	-0,072	1			
Horas jornada escolar	-0,191	-0,047	0,139	0,131	-0,097	0,041	0,062	-0,127	0,058	0,053	-0,03	-0,047	0,041	1		
Selección	0,357	0,212	-0,505	-0,478	0,506	-0,068	-0,069	0,238	-0,082	-0,066	-0,123	0,381	-0,135	-0,148	1	
Pago mensual	0,089	0,15	-0,563	-0,382	0,597	-0,028	-0,02	-0,013	-0,073	-0,082	-0,094	0,386	-0,236	0,008	0,4	1

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

Cuadro 2: Matriz correlaciones estudiantes

	personas hogar	educación padre	educación madre	padres indígenas	hogar incompleto	Ingresos	Expectativas Padres	Satisfacción Calidad	Género Alumno
personas hogar	1	-0,014	-0,051	0,016	0,022	0,109	-0,027	0,023	-0,006
educación padre	-0,014	1	0,482	-0,094	-0,459	0,435	0,409	-0,013	0,041
educación madre	-0,051	0,482	1	-0,107	-0,10	0,421	0,454	0,001	0,042
padres indígenas	0,016	-0,094	-0,107	1	0,012	-0,086	-0,057	0,025	0,001
hogar incompleto	0,022	-0,459	-0,100	0,012	1	-0,063	-0,048	0,019	-0,015
Ingresos	0,109	0,435	0,421	-0,086	-0,063	1	0,333	0,038	0,001
Expectativas Padres	-0,027	0,409	0,454	-0,057	-0,048	0,333	1	0,065	0,088
Satisfacción Calidad	0,023	-0,013	0,001	0,025	0,019	0,038	0,065	1	0,068
Género Alumno	-0,006	0,041	0,042	0,001	-0,015	0,001	0,088	0,068	1

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

Cuadro 3: Matriz correlaciones estudiantes

	Género Profesor	Antigüedad	Expectativas Profesor	Uso Recursos
Género Profesor	1	0,019	-0,015	0,057
Antigüedad	0,019	1	-0,118	0,074
Expectativas Profesor	-0,015	-0,118	1	0,100
Uso Recursos	0,057	0,074	0,100	1

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

Cuadro 4: KMO y prueba de Bartlett

Prueba		Estadístico
Medida de adecuación muestral de Kaiser-Meyer-Olkin.		0,794
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	113570,909
	g.l.	496
	Sig.	0,000

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

Cuadro 5: Matriz de componentes principales rendimiento

	Componente									
	1	2	3	4	5	6	7	8	9	10
Disponibilidad Recursos	,200	,128	-,034	,727	,019	,127	,038	-,079	-,169	,028
Índice de Vulnerabilidad Municipal	-,744	-,369	,181	-,094	-,036	-,001	-,128	,127	-,019	-,031
GSE Curso	-,569	-,252	-,271	,071	-,008	-,240	,004	,089	,349	-,004
Exceso de Viaje cantidad de cursos	,788	,294	-,157	,091	,054	,007	,168	-,088	,008	,046
Alumnos curso	-,012	-,192	,166	-,054	-,010	,120	-,038	,560	-,135	-,017
Alumnos indígenas curso	-,043	-,309	,192	-,142	2,05E-005	,139	-,058	,315	-,534	-,072
Género Profesor	,112	,711	,068	-,005	,035	,118	,043	-,056	-,121	,114
Antigüedad	-,079	-,031	,904	-,035	-,012	,004	-,022	,106	-,077	-,008
Expectativas Profesor	-,014	,105	,008	,039	-,005	,066	-,026	,026	,133	,722
Uso Recursos	-,111	-,083	,000	-,056	-,021	,131	,051	,033	,649	,100
Biblioteca	,429	-,070	,084	,313	-,046	,025	,174	-,109	-,072	,000
Retirados	-,120	-,143	,031	,667	-,052	,127	,112	-,003	,116	,228
Educación Secundaria	,075	,197	-,071	,609	,053	-,129	-,109	-,039	,074	-,190
	-,094	,060	,017	-,129	-,106	-,076	,054	,594	,019	,120
	,556	,441	-,016	,051	,008	-,027	-,023	,141	-,008	-,138

	Componente									
	1	2	3	4	5	6	7	8	9	10
Centro Padres	-,295	,109	-,061	,044	,059	-,052	,016	,456	,030	-,310
cantidad computadores	-,030	,810	,032	,084	,016	-,013	,046	-,051	,094	,046
Jornada Escolar	,022	-,309	,131	,190	,035	,050	-,103	,338	,158	,064
Matrícula	,267	,892	-,008	,078	,021	-,012	,015	-,001	,029	-,022
Matrícula indígena	-,033	,238	,825	,031	-,009	-,099	,021	,111	,095	-,009
Género Alumno	-,062	,037	,086	-,025	-,161	,171	,552	-,157	,040	-,382
educación padre	,381	,081	-,045	,036	,709	-,055	,337	-,002	,018	,053
educación madre	,425	,062	-,124	,042	,248	-,141	,535	,061	,025	,099
Ingresos	,834	-,128	-,057	,015	,124	-,049	,109	,000	,090	-,005
personas hogar	,159	-,043	,098	-,016	-,004	,091	-,349	,027	,364	-,347
hogar incompleto	,041	-,005	-,004	,015	-,891	-,043	,031	,081	,034	,027
padres indígenas	-,053	-,101	,700	-,025	-,013	,019	-,037	-,003	-,061	,001
Selección	,597	,259	-,025	-,001	,003	,117	,023	-,129	-,184	-,005
Expectativas Padres	,314	,073	-,026	,053	,157	-,033	,649	,070	,054	,047
Satisfacción Calidad	,043	-,021	,009	-,026	-,028	,779	,058	-,034	,038	,055
Satisfacción										
Infraestructura	,019	,082	-,067	,139	,037	,741	-,084	,064	,048	-,016
Pago Mensual	,856	-,166	-,020	-,009	,033	-,006	-,044	-,046	,030	-,034

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

ANEXO IV: : Matrices de correlaciones y análisis factorial elección

Cuadro 1: Correlaciones variables establecimiento elección

	JEC	SIMCE	Recursos Índice de Vulnerabilidad	Municipal	Exceso de Viaje	Cantidad de Cursos	Alumnos curso	% indígenas curso	género profesor	Ed. Secundaria	Matrícula	
JEC	1	-0,063	-0,042	0,162	0,087	0,125	0,134	-0,14	0,124	0,017	-0,085	-0,214
SIMCE	-0,063	1	0,178	-0,478	-0,26	-0,008	-0,153	0,044	-0,181	0,06	0,182	0,109
Recursos Índice de Vulnerabilidad	-0,042	0,178	1	-0,279	-0,198	-0,083	-0,08	0,164	-0,076	0,046	0,156	0,225
Municipal	0,162	-0,478	-0,279	1	0,445	0,203	0,228	-0,359	0,238	-0,008	-0,456	-0,526
Exceso de Viaje	0,087	-0,26	-0,198	0,445	1	-0,015	-0,091	-0,278	-0,172	0,017	-0,353	-0,323
Cantidad de Cursos	0,125	-0,008	-0,083	0,203	-0,015	1	0,251	-0,168	0,166	-0,02	-0,05	-0,171
Alumnos curso	0,134	-0,153	-0,08	0,228	-0,091	0,251	1	-0,073	0,278	-0,091	-0,113	-0,302
% indígenas curso	-0,14	0,044	0,164	-0,359	-0,278	-0,168	-0,073	1	0,054	0,089	0,252	0,633
género profesor	0,124	-0,181	-0,076	0,238	-0,172	0,166	0,278	0,054	1	-0,037	-0,096	-0,094
Ed. Secundaria	0,017	0,06	0,046	-0,008	0,017	-0,02	-0,091	0,089	-0,037	1	-0,023	0,068
Matrícula	-0,085	0,182	0,156	-0,456	-0,353	-0,05	-0,113	0,252	-0,096	-0,023	1	0,578
	-0,214	0,109	0,225	-0,526	-0,323	-0,171	-0,302	0,633	-0,094	0,068	0,578	1

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

Cuadro 2: Correlaciones variables familiares elección

	Género Alumno	Tamaño Familiar	Hogar Incompleto	Cercanía Familiaridad	Prestigio	Resultados SIMCE	Entrega de Valores	Costo	formación profesional	único en la comuna	GSE Alumno	
JEC	-0,014	0,009	0,015	0,009	-0,018	-0,025	-0,017	-0,035	-0,014	-0,002	0,209	-0,053
SIMCE	0,026	-0,021	-0,032	-0,116	-0,034	0,214	0,087	0,181	-0,072	0,095	-0,064	0,392
Recursos	0,022	0,002	-0,018	-0,052	-0,012	0,099	0,013	0,096	-0,01	0,04	-0,037	0,185

	Género Alumno	Tamaño Familiar	Hogar Incompleto	Cercanía	Familiaridad	Prestigio	Resultados SIMCE	Entrega de Valores	Costo	formación profesional	único en la comuna	GSE Alumno
Índice de Vulnerabilidad Municipal	-0,04	0,006	0,04	0,138	0,052	-0,181	0,014	-0,194	0,037	-0,115	0,111	-0,557
Exceso de Viaje	-0,036	-0,032	0,018	0,246	0,046	-0,2	0,003	-0,253	0,039	-0,163	0,038	-0,316
Cantidad de Cursos	-0,012	0,012	0	-0,044	0,012	-0,009	0,015	-0,002	-0,063	0,012	0,066	-0,09
Alumnos curso	0,001	0,012	0,018	-0,025	-0,018	-0,013	0,031	-0,005	-0,04	0,017	0,067	-0,101
% indígenas curso	0,034	-0,006	-0,019	-0,086	-0,011	0,117	0,008	0,095	0,045	0,063	-0,064	0,122
género profesor Ed. Secundaria	0,01	0,031	0,032	-0,067	-0,024	-0,022	0,004	-0,01	-0,031	0,011	0,084	-0,138
Matrícula	-0,007	-0,013	-0,008	-0,002	0,011	-0,001	-0,006	0,028	0,015	0,017	-0,001	0,011
Género Alumno	0,028	0,01	-0,026	-0,095	-0,008	0,084	-0,048	0,124	-0,017	0,058	-0,071	0,352
Tamaño Familiar	0,043	-0,003	-0,027	-0,072	-0,017	0,082	-0,045	0,079	0,046	0,059	-0,097	0,211
Hogar Incompleto	1	-0,004	-0,013	0,001	-0,007	0,002	-0,007	0,002	-0,02	-0,005	0,001	0,024
Cercanía	-0,004	1	0,037	-0,022	0,046	0,001	0,003	0,017	-0,004	-0,013	-0,017	-0,022
Familiaridad	-0,013	0,037	1	0,01	0	-0,019	-0,015	-0,023	0,001	-0,001	0,016	-0,259
Prestigio	0,001	-0,022	0,01	1	0,007	-0,236	-0,091	-0,214	0,078	-0,176	-0,022	-0,091
Resultados SIMCE	-0,007	0,046	0	0,007	1	-0,01	0	-0,03	0,059	-0,085	-0,042	-0,047
Entrega de Valores	0,002	0,001	-0,019	-0,236	-0,01	1	0,137	0,268	-0,116	0,052	-0,061	0,147
Costo	-0,007	0,003	-0,015	-0,091	0	0,137	1	0,081	-0,05	0,022	0,024	-0,004
formación profesional	0,002	0,017	-0,023	-0,214	-0,03	0,268	0,081	1	-0,078	0,066	-0,005	0,153
único en la comuna	-0,02	-0,004	0,001	0,078	0,059	-0,116	-0,05	-0,078	1	-0,021	0,091	-0,075
GSE Alumno	-0,005	-0,013	-0,001	-0,176	-0,085	0,052	0,022	0,066	-0,021	1	-0,006	0,106
	0,001	-0,017	0,016	-0,022	-0,042	-0,061	0,024	-0,005	0,091	-0,006	1	-0,121
	0,024	-0,022	-0,259	-0,091	-0,047	0,147	-0,004	0,153	-0,075	0,106	-0,121	1

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa

Cuadro 3: KMO y prueba de Bartlett

Prueba		Estadístico
Medida de adecuación muestral de Kaiser-Meyer-Olkin.		0,732
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	40822,639
	g.l.	276
	Sig.	0,000

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa

Cuadro 4: Matriz de componentes principales elección

	Componente								
	1	2	3	4	5	6	7	8	9
JEC	-,262	,277	,010	-,338	,414	,216	-,096	,148	,070
SIMCE	,516	,165	-,471	-,111	,090	,093	-,164	,059	,056
Recursos	,398	-,009	-,013	-,032	,150	,135	-,058	,187	,086
Índice de Vulnerabilidad Municipal	-,832	,062	,095	,141	-,018	-,026	,129	-,019	,051
Exceso de Viaje	-,593	-,401	-,275	,109	,085	-,037	,063	,042	,004
Cantidad de Cursos	-,237	,445	,068	-,226	-,139	,155	-,037	-,020	,224
Alumnos curso	-,296	,525	,258	-,190	-,218	,010	,069	-,001	-,002
% indígenas curso	,537	-,138	,530	,191	,067	-,059	,227	,047	,041
género profesor	-,210	,442	,537	-,046	-,103	-,029	,153	,042	,085
Ed. Secundaria	,067	-,104	-,025	,142	,416	,076	,241	,051	,774
Matrícula	,636	-,063	,228	-,218	-,095	,102	-,123	-,007	-,055
Género Alumno	,714	-,302	,430	,061	,028	-,013	,052	,029	,005
Tamaño Familiar	,061	2,69E-005	,053	-,088	-,083	-,059	,105	,750	-,102
Hogar Incompleto	-,007	,054	,093	,137	-,174	,457	-,370	-,053	,071
	-,115	-,017	,205	,374	,110	-,059	-,725	,186	,047

	Componente								
	1	2	3	4	5	6	7	8	9
Cercanía	-,264	-,478	-,020	-,286	-,116	,032	-,012	,196	,057
Familiaridad	-,059	-,118	,011	,179	-,192	,680	,144	-,139	,042
Prestigio	,329	,389	-,204	,391	,011	,065	,107	,087	-,077
Resultados SIMCE	,025	,238	-,198	,396	,105	,019	,289	,095	-,293
Entrega de Valores	,336	,394	-,128	,269	,090	,097	,038	,023	-,047
Costo	-,063	-,283	,223	-,080	,328	,238	,086	-,316	-,321
formación profesional	,193	,258	,019	-,028	,192	-,419	-,153	-,393	,095
único en la comuna	-,191	,163	,149	-,199	,640	,156	-,026	,100	-,312
GSE Alumno	,617	,094	-,313	-,414	-,099	,059	,099	-,070	,012

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

ANEXO V: Estimación de los parámetros del modelo de elección

Tipo de Establecimiento(a)		B	Error típ.	Wald	g.l.	Sig.	Exp(B)	Intervalo de confianza al 95% para Exp(B)	
								Límite inf.	Límite sup.
Bajo	Intersección	-30,12	1,31	531,32	1	0			
	SIMCE establ.	0,04	0	114,37	1	0	1,04	1,03	1,05
	Cantidad de Cursos	22,98	0,61	1420,12	1	0	9,56+E09	2,89+E09	3,16+E10
	[Cercanía=0]	-0,01	0,15	0,04	1	0,95	0,99	0,73	1,34
	[Cercanía=1]	0(b)			0				
	[Ed. Secund.=0]	6,53	0,23	782,98	1	0	683,68	432,8	1079,99
	[Ed. Secund.=1]	0(b)			0				
	Exceso de Viaje	0,01	1,92	0	1	1	1,01	0,02	43,76
	[Prestigio=0]	-0,58	0,18	10,01	1	0	0,56	0,39	0,8
	[Prestigio=1]	0(b)			0				
	[Resultados SIMCE=0]	-0,67	0,23	8,22	1	0	0,51	0,33	0,81
	[Resultados SIMCE=1]	0(b)			0				
	[Entrega Valórica=0]	-0,79	0,18	20,39	1	0	0,46	0,32	0,64
	[Entrega Valórica=1]	0(b)			0				
	[JEC=0]	-0,39	0,16	6,07	1	0,01	0,68	0,5	0,92
	[JEC=1]	0(b)			0				
	[Único=0]	0,27	0,15	2,96	1	0,09	1,3	0,96	1,76
	[Único=1]	0(b)			0				
IVE.	0,16	0,01	555,01	1	0	1,18	1,16	1,19	
GSE alumno	1,44	0,12	134,59	1	0	4,2	3,3	5,36	
Tamaño Curso	-0,49	0,01	1632,51	1	0	0,62	0,6	0,63	
Indígenas Curso	-0,04	0,02	4,6	1	0,03	0,96	0,92	1	
Medio - Bajo	Intersección	-24,7	1,01	592,58	1	0			
	SIMCE establ.	0,03	0	99,71	1	0	1,03	1,03	1,04
	Cantidad de Cursos	16,05	0	.	1	.	9355665	9355665	9355665
	[Cercanía=0]	0,06	0,13	0,21	1	0,65	1,06	0,82	1,38
	[Cercanía=1]	0(b)			0				
	[Ed. Secund.=0]	3,21	0,15	488,59	1	0	24,74	18,61	32,88
	[Ed. Secund.=1]	0(b)			0				
	Exceso de Viaje	-0,91	1,87	0,24	1	0,63	0,4	0,01	15,59

	[Prestigio=0]	-0,67	0,16	17,5	1	0	0,51	0,38	0,7
	[Prestigio=1]	0(b)			0				
	[Resultados SIMCE=0]	-0,17	0,2	0,72	1	0,4	0,84	0,57	1,25
	[Resultados SIMCE=1]	0(b)			0				
	[Entrega Valórica=0]	-0,69	0,15	20,21	1	0	0,5	0,37	0,68
	[Entrega Valórica=1]	0(b)			0				
	[JEC=0]	0,03	0,14	0,05	1	0,82	1,03	0,79	1,35
	[JEC=1]	0(b)			0				
	[Único=0]	0,12	0,14	0,7	1	0,4	1,12	0,86	1,47
	[Único=1]	0(b)			0				
	IVE	0,15	0,01	559,18	1	0	1,16	1,14	1,17
	GSE alumno	1,29	0,1	165,52	1	0	3,62	2,97	4,4
	Tamaño Curso	-0,19	0,01	414,13	1	0	0,83	0,81	0,84
	Indígenas Curso	-0,06	0,02	12,01	1	0	0,94	0,91	0,97
Medio	Intersección	1,77	0,97	3,32	1	0,07			
	SIMCE establ.	0	0	1,33	1	0,25	1	0,99	1
	Cantidad de Cursos	-1,2	0		1		0,3	0,3	0,3
	[Cercanía=0]	-0,01	0,13	0	1	0,96	0,99	0,77	1,29
	[Cercanía=1]	0(b)			0				
	[Ed. Secund.=0]	1,98	0,14	211,93	1	0	7,23	5,54	9,44
	[Ed. Secund.=1]	0(b)			0				
	Exceso de Viaje	6,46	1,85	12,19	1	0	637,42	16,99	23920,51
	[Prestigio=0]	-0,61	0,16	15,5	1	0	0,54	0,4	0,74
	[Prestigio=1]	0(b)			0				
	[Resultados SIMCE=0]	0,18	0,2	0,76	1	0,38	1,2	0,8	1,78
	[Resultados SIMCE=1]	0(b)			0				
	[Entrega Valórica=0]	-0,53	0,15	12,51	1	0	0,59	0,44	0,79
	[Entrega Valórica=1]	0(b)			0				
	[JEC=0]	0,89	0,13	46,59	1	0	2,44	1,89	3,15
	[JEC=1]	0(b)			0				
	[Único=0]	0,26	0,14	3,62	1	0,06	1,3	0,99	1,71
	[Único=1]	0(b)			0				
	IVE	0,06	0,01	106,39	1	0	1,06	1,05	1,07
	GSE alumno	0,91	0,1	87,2	1	0	2,47	2,05	2,99
Tamaño Curso	-0,11	0,01	140,29	1	0	0,9	0,88	0,91	

	Indígenas Curso	-0,03	0,02	3,44	1	0,06	0,97	0,93	1
	Intersección	15,34	5879,8	0	1	1			
	SIMCE establ.	-0,05	0,01	114,05	1	0	0,95	0,94	0,96
	Cantidad de Cursos	-1,35	5879,8	0	1	1	0,26	0	
	[Cercanía=0]	0,3	0,19	2,67	1	0,1	1,35	0,94	1,94
	[Cercanía=1]	0(b)			0				
	[Ed. Secund.=0]	-21,88	2830,3	0	1	0,99	0	0	
	[Ed. Secund.=1]	0(b)			0				
	Exceso de Viaje	-25,92	5,38	23,23	1	0	0	0	0
	[Prestigio=0]	-0,14	0,22	0,41	1	0,52	0,87	0,57	1,33
	[Prestigio=1]	0(b)			0				
Medio	[Resultados SIMCE=0]	0,79	0,39	4,15	1	0,04	2,21	1,03	4,73
-Alto	[Resultados SIMCE=1]	0(b)			0				
	[Entrega Valórica=0]	0,22	0,22	0,98	1	0,32	1,24	0,81	1,9
	[Entrega Valórica=1]	0(b)			0				
	[JEC=0]	2,46	0,25	93,38	1	0	11,64	7,08	19,16
	[JEC=1]				0				
	[Único=0]	0,86	0,23	14,05	1	0	2,36	1,51	3,69
	[Único=1]				0				
	IVE	-0,05	0,01	21,68	1	0	0,95	0,93	0,97
	GSE alumno	-0,31	0,15	4,37	1	0,04	0,73	0,55	0,98
	Tamaño Curso	-0,05	0,01	15,69	1	0	0,95	0,92	0,97
	Indígenas Curso	0,15	0,04	16,38	1	0	1,16	1,08	1,25

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

ANEXO VI: Estadísticos de modelos de elección según GSE

I. GSE Bajo

Cuadro 1: Información del ajuste del modelo

Modelo	Criterio de ajuste del modelo	Contrastes de la razón de verosimilitud		
	-2 Log verosimilitud	Chi-cuadrado	g.l.	Sig.
Sólo la intersección	25505,707	12968,214	56	0,000
Final	12537,493			

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

Cuadro 2: Bondad de ajuste

Test	Chi-cuadrado	g.l.	Sig.
Pearson	21336,406	26940	1,000
Desviación	12537,493	26940	1,000

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

Cuadro 3: Pseudo R²

Pseudo R ²	
Cox y Snell	0,721
Nagelkerke	0,785
McFadden	0,508

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

II. GSE Medio-Bajo

Cuadro 4: Información del ajuste del modelo

Modelo	Criterio de ajuste del modelo	Contrastes de la razón de verosimilitud		
	-2 Log verosimilitud	Chi-cuadrado	g.l.	Sig.
Sólo la intersección	2427,728	1399,377	44	0
Final	1028,351			

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

Cuadro 5: Bondad de ajuste

Test	Chi-cuadrado	g.l.	Sig.
Pearson	2261,133	2740	1
Desviación	1028,351	2740	1

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

Cuadro 6: Pseudo R²

Pseudo R ²	
Cox y Snell	0,822
Nagelkerke	0,866
McFadden	0,576

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

III. GSE Medio

Cuadro 7: Información del ajuste del modelo

Modelo	Criterio de ajuste del modelo	Contrastes de la razón de verosimilitud		
	-2 Log verosimilitud	Chi-cuadrado	g.l.	Sig.
Sólo la intersección	476,108	337,609	36	0
Final	138,499			

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

Cuadro 8: Bondad de ajuste

Test	Chi-cuadrado	g.l.	Sig.
Pearson	154,648	588	1
Desviación	138,499	588	1

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

Cuadro 9: Pseudo R²

Pseudo R ²	
Cox y Snell	0,868
Nagelkerke	0,921
McFadden	0,709

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

IV. GSE Medio-Alto

Cuadro 10: Información del ajuste del modelo

Modelo	Criterio de ajuste del modelo	Contrastes de la razón de verosimilitud		
	-2 Log verosimilitud	Chi-cuadrado	g.l.	Sig.
Sólo la intersección	220,575	112,361	24	0
Final	108,214			

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

Cuadro 11: Bondad de ajuste

Test	Chi-cuadrado	g.l.	Sig.
Pearson	139,908	304	1
Desviación	108,214	304	1

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

Cuadro 12: Pseudo R²

Pseudo R ²	
Cox y Snell	0,709
Nagelkerke	0,778
McFadden	0,509

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

V. GSE Alto

Cuadro 13: Información del ajuste del modelo

Modelo	Criterio de ajuste del modelo	Contrastes de la razón de verosimilitud		
	-2 Log verosimilitud	Chi-cuadrado	g.l.	Sig.
Sólo la intersección	340,466	269,6	24	0
Final	70,865			

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

Cuadro 14: Bondad de ajuste

Test	Chi-cuadrado	g.l.	Sig.
Pearson	28277,541	396	0
Desviación	70,865	396	1

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.

Cuadro 15: Pseudo R²

Pseudo R ²	
Cox y Snell	0,79
Nagelkerke	0,918
McFadden	0,792

Fuente: Elaboración propia, resultados de los modelos arrojados por el programa SPSS.