

# CAPÍTULO 3

## ELECCIÓN DE ESCUELAS PÚBLICAS Y PRIVADAS, BACKGROUND FAMILIAR, SEGREGACIÓN Y CALIDAD DE LOS ENTORNOS EDUCATIVOS

### TABLA DE CONTENIDO

3.1. INTRODUCCIÓN.....	118
3.2. EL ROL DE LAS INTERACCIONES SOCIALES Y LA PROVISIÓN EFICIENTE DE BIENES PÚBLICOS .....	122
3.3. CALIDAD DE LOS ENTORNOS LOCALES, SEGREGACIÓN Y PREFERENCIAS POR LA EDUCACIÓN .....	131
3.3.1. <i>Índice de distribución del capital humano</i> .....	133
3.3.2. <i>Índice de sorting educativo en gastos</i> .....	136
3.3.3. <i>Segregación y elección de escuelas públicas o privadas</i> .....	139
3.3.3.1. <i>Estructura general de los modelos e hipótesis de comportamiento</i> .....	139
3.3.3.2. <i>Análisis de los resultados</i> .....	143
3.4. CALIDAD EDUCATIVA Y DIFERENCIAS EN LOGROS .....	150
3.4.3. <i>Rendimiento académico con datos del ICFES</i> .....	158
3.5. CALIDAD EDUCATIVA, ENTORNOS DE APRENDIZAJE Y COMPETENCIAS .....	162
3.5.1. <i>Categorías básicas y aspectos metodológicos</i> .....	163
3.5.2. <i>Análisis de los valores plausibles</i> .....	170
3.5.3. <i>Efectos endógenos, contextuales y correlacionados</i> .....	180
3.5.4. <i>Estimaciones</i> .....	187
3.6. CONCLUSIONES PARCIALES .....	205
3.7. BIBLIOGRAFÍA Y REFERENCIAS .....	210
ANEXOS.....	214
A.3.1. <i>Modelos Logit</i> .....	214
A.3.2. <i>Estadísticas descriptivas modelos de elección discreta</i> .....	215
A.3.3. <i>Notas metodológicas sobre la ECV 2003</i> .....	215
A3.4. <i>Resultados de las pruebas del ICFES</i> .....	217

<b>Gráficos</b>	
Gráfico 3.1.	Kernel de densidades: composición del capital humano intravecindad ..... 135
Gráfico 3.2	Distribución del índice según cuartiles..... 136
Gráfico 3.3.	Índice de sorting educativo..... 137
Gráfico 3.4.	Gráfico conjunto de distribución..... 138
Gráfico 3.5.	Elección de escuelas según habilidades e ingresos..... 154
Gráfico 3.6a.	Gráfico de dispersion pruebas del ICFES en Lenguaje ..... 160
Gráfico 3.6b.	Gráfico de dispersion pruebas del ICFES en Inglés... 160
Gráfico 3.6c.	Gráfico de dispersion pruebas del ICFES en Matemáticas..... 160
Gráfico 3.7.	Influencia del hogar, la escuela y el aula sobre las competencias en lectura..... 165
Gráfico 3.8.	Distribución de las experiencias tempranas en lectura..... 175
Gráfico 3.9.	Estimaciones empíricas de Bayes del intercepto aleatorio..... 195
Gráfico 3.10.	Distribución de valores plausibles según escala de ingresos..... 197
Gráfico 3.11.	Disponibilidad de recursos escolares..... 204

<b>Cuadros</b>	
Cuadro 3.1.	Modelos explorados en las estimaciones..... 142
Cuadro 3.2.	Total nacional modelos logit..... 149
Cuadro 3.3.	Clasificación de estudiantes según deciles..... 157
Cuadro 3.5.	Colombia: promedio y desviaciones en las pruebas del ICFES..... 160
Cuadro 3.5.	PIRLS comparaciones internacionales promedios generales..... 169
Cuadro 3.6.	Colombia: Valores Plausibles en la pruebas PIRLS 2001 según grupos y categorías de variables..... 173
Cuadro 3.7.	Valores Plausibles controlados por un conjunto de variables..... 179
Cuadro 3.8.	Colombia: modelos multinivel..... 198
Cuadro 3.9.	Modelos multinivel complementarios..... 202
Recuadro 3.1.	Características de las pruebas del ICFES..... 161

## 3

# ELECCIÓN DE ESCUELAS PÚBLICAS Y PRIVADAS, BACKGROUND FAMILIAR, SEGREGACIÓN Y CALIDAD DE LOS ENTORNOS EDUCATIVOS

### 3.1. Introducción

**L**a investigación que se desarrollará en este capítulo considera que la educación no es un bien público puro en la medida que tiene la característica de ser excluyente en el consumo y de operar como bien de club.

Las barreras de entrada vinculadas a la educación están constituidas por la estructura de costes y por los factores económicos y socio-demográficos relacionados con la calidad de los entornos familiares y locales que rodean las decisiones sobre la escolaridad de los niños y jóvenes en su ciclo educativo básico. Por tal razón, además de las restricciones de financiación, los estudiantes enfrentan barreras de acceso que están íntimamente ligadas a sus capacidades reales de funcionamiento.

Las diferencias en los recursos disponibles y en las tecnologías de producción de los servicios educativos entre escuelas públicas y privadas ha demostrado a lo largo de varias décadas de reformas que a pesar de los esfuerzos de financiación realizados por el sector público para proporcionar igualdad de oportunidades, persisten disparidades entre los planteles públicos y privados que tienden a ampliarse.

A diferencia del modelo de provisión predominantemente público que se aprecia en los países europeos, los países de América Latina cuentan con una amplia variedad de precios y calidades, un tamaño relativamente alto del sector privado en la oferta de la educación básica, así como resultados dispersos en las pruebas estandarizadas de logros académicos, de tal suerte que los dos tipos de planteles coexisten en mercados segmentados con un vasto menú de calidades y precios. Estos rasgos son más evidentes

en las zonas urbanas que en las rurales y, de manera particular, se pueden apreciar con mayor transparencia en el interior de las principales áreas metropolitanas.

En el caso colombiano, que puede considerarse como representativo de la mayoría de países de América Latina, el carácter de bien de club que se observa en la provisión de servicios educativos, además de su segmentación en precios y calidades, está asociado de manera ineludible a un arreglo social elitista y altamente estratificado. De acuerdo con la evidencia preliminar del cuadro 2.5 y los datos que se abordarán en un apartado posterior, es claro que en los estratos socio-económicos pobres la mixtura de educación pública y privada corrientemente está asociada con niveles bajos de calidad y precios.

En los estratos más altos esta relación calidad/precio es prácticamente unívoca, los colegios privados exhiben los mejores resultados en las pruebas de competencias académicas (matemáticas, ciencias naturales e idiomas foráneos), cuentan con una mayor disponibilidad de recursos, escogen a sus estudiantes a partir de criterios propios, los costes de acceso son casi prohibitivos para el resto de estratos socio-económicos y adicionalmente generan mecanismos soterrados de exclusión social que garantizan la persistencia intergeneracional del arreglo social establecido.

En estas condiciones, las predicciones que se derivan de los enfoques tradicionales en los que la existencia de mercados segmentados aumenta las posibilidades de consumo y puede llevar al óptimo social, se ven distorsionadas por la presencia de amplias desigualdades en la dotación inicial de recursos. De ahí, que el estudio haya insistido en la relevancia que tiene desigualdad inicial para explicar la inercia propia del sistema educativo colombiano.

Un argumento adicional que va en esta misma dirección, afirma que desde un punto de vista social y de acuerdo con el estado de desarrollo del país, si los mayores niveles de uniformidad en la provisión y la mejor calidad de la educación generan externalidades positivas, la creciente segmentación en el suministro del bien puede ser socialmente sub-óptima. Una garantía de tal aseveración se haya en la evidencia empírica proporcionada en los dos capítulos anteriores: cuando perduran diferencias importantes en las oportunidades de acceso a los servicios educativos básicos y predomina una

elevada disparidad en la distribución de los ingresos, los efectos dinámicos de auto-reforzamiento se magnifican y agravan los problemas distributivos.

Para abordar los aspectos anteriores el capítulo se estructura de la siguiente forma:

La sección 3.2 se ocupa del rol que juegan las interacciones sociales y la segregación socio-económica. Este apartado liga la cadena de razonamientos de la Tesis con los planteamientos del diagrama 2.1 del capítulo anterior y realiza una revisión de las principales contribuciones relacionadas con la provisión eficiente de bienes públicos como la educación y las posibilidades del surgimiento de equilibrios estratificados o de “*sorting*” escolar.

La sección 3.3 presenta algunos hechos estilizados de la segregación a nivel de micro-localidad. Propone y calcula indicadores de distribución intra-metropolitana del capital humano y un índice de *sorting* educativo en gastos que aportan una aproximación empírica a la estructura de preferencias por la calidad educativa y la particularidad de los entornos locales. Estos índices, también relacionados con la calidad de las vecindades, constituyen insumos necesarios para los ejercicios econométricos del apartado 3.4.

La sección 3.4 realiza algunos contrastes empíricos de los factores que inciden en la elección de escuelas y colegios públicos y privados. Se contrastan varios modelos de elección discreta que auscultan la importancia relativa del background familiar y los entornos locales. A partir de las estimaciones se realizan algunos ejercicios de estática comparada que aportan información sobre las probabilidades condicionadas y los cambios en probabilidad que se derivan de los ajustes obtenidos.

La muestra utilizada en esta última sección corresponde a 2381 escolares extraídos de la muestra total de niños y jóvenes entre los 5 y 17 años de la Encuesta de Calidad de Vida del año 2003 (ECV 2003). La selección solamente considera los escolares urbanos localizados en las principales áreas metropolitanas del país con información completa sobre el background familiar y las características socio-demográficas del entorno local.

La sección 3.5 profundiza en el estudio de las implicaciones de los entornos de aprendizaje sobre la calidad educativa, los logros y las competencias académicas de los estudiantes a partir de la base estadística de las pruebas PIRLS. Esta prueba permitirá en principio efectuar algunas comparaciones internacionales y luego profundizar en la mirada interna a los factores que inciden en el rendimiento de los escolares de 4° grado de primaria en las pruebas de proficiencia en lectura.

En este mismo apartado se realizarán algunas contrastes econométricos con una muestra de 5131 escolares distribuidos en 144 planteles educativos. Se explorarán algunas especificaciones que permitirán dilucidar el papel que juegan los efectos contextuales, correlacionados y endógenos que genera la interacción social de los estudiantes con sus compañeros de clase. Las técnicas utilizadas en esta sección 3.5 corresponden a estructuras multinivel que intentan capturar la variabilidad existente entre escuelas y extraer información relevante sobre el peso efectivo que tienen los perfiles económicos y educativos de los entornos familiares de los estudiantes.

### 3.2. El rol de las interacciones sociales y la provisión eficiente de bienes públicos

El modelo analítico planteado en el capítulo anterior esbozó la relevancia de las interacciones sociales y sus consecuencias sobre las trayectorias de acumulación del capital humano. La complejidad de nexos que pueden observarse en el diagrama 2.1 entre el background familiar, las características de la comunidad local, los perfiles de calidad de las escuelas públicas o privadas elegidas y las dotaciones innatas de los hijos, sugieren de inmediato algunos interrogantes sobre las trayectorias que seguirá el proceso de acumulación de capital humano y la estructura de ingresos, a través de las generaciones.

Bajo que condiciones *¿Persistirán las desigualdades intra-generacionales?*  
*¿Predominará un arreglo social estratificado o integrado?* *¿Prevalecerá el sorting o alguna modalidad de tracking en el suministro de los servicios educativos?* *¿Bajo que circunstancias se puede introducir alguna ruptura en estas trayectorias?*

Las respuestas a estas indagaciones no son tan fáciles como parece a primera vista. Diversos estudios empíricos han mostrado el nexo entre la segregación de las familias por niveles de ingreso, los fondos de financiación de la calidad educativa, los logros y el background familiar y los resultados que han obtenido no son unívocos. Trabajos como los de [Mayer \(2002\)](#) y el de [Bayer y Mc Millan\(2004\)](#) para los Estados Unidos, el de [Burgess, S., et. al., \(2004\)](#) para el caso de Inglaterra y algunas comparaciones internacionales en el trabajo de [Jenkins \(2006\)](#) son apenas una pequeña muestra del interés que despierta el interrogante sobre los nexos entre esta segregación socio-económica y las ganancias o pérdidas que generan.

Los perfiles educativos de los padres, los tamaños medios de los hogares, la calidad de los entornos locales (efectos de vecindad), la riqueza inicial y la pertenencia a redes sociales, étnicas, religiosas y la fortaleza de las instituciones, entre otros factores, que configuran en conjunto la noción de *capital social*, entran en juego con las condiciones iniciales y generan pérdidas y ganancias en su proceso de interacción.

Un primer argumento afirma que la influencia positiva o negativa que puedan ejercer los efectos de la vecindad o los lugares en los que habitan los individuos, los efectos de los compañeros de clase en las escuelas y los de la familia, se reflejarán en las condiciones y posibilidades futuras de los hijos, en sus logros y sus realizaciones. Si la trayectoria dinámica que describe el proceso de acumulación lleva a mayores grados de segregación entre ricos y pobres, evidentemente la distribución de talentos, habilidades y capacidades productivas reflejará este comportamiento en los espacios efectivos de realizaciones de las diferentes dinastías, en las condiciones objetivas y en los sentimientos de privación relativa entre los individuos de una misma generación. Este es un aspecto ampliamente tratado en la literatura desde los trabajos de [Merton y Kitt \(1950\)](#), así como en los trabajos de [Sen, A. \(1982, 1992\)](#).

Un argumento adicional, ligado a este razonamiento, es que los hijos que provienen de hogares pobres --con bajo nivel educativo o con calidad deficiente de su educación— cuando compiten con los hijos de los ricos, con alto nivel y buena calidad educativa, además de las redes sociales que los respaldan, tienen una alta probabilidad de quedar excluidos en una economía de mercado. En tal sentido la noción implícita es la de condiciones efectivas de privación relativa. Ahora bien, si los pobres se comparan con los ricos y con sus logros surgen sentimientos de privación relativa que aumentan la desdicha, el resentimiento, el aislamiento y que finalmente afectan sus expectativas de movilidad social. La combinación de estas dos situaciones (las condiciones y los sentimientos de privación relativas), cuando además se ven reforzadas por los efectos de los pares del grupo puede dar lugar al surgimiento de *ghettos*, amenazar la cohesión social y afectar el bienestar social. De ahí la relevancia del interrogante de si existe un umbral crítico a partir del cual los altos niveles de segregación inciden negativamente en el funcionamiento del sistema.

Un segundo argumento, cuyo punto de vista es opuesto al anterior, afirma que la segregación induce la confluencia de talentos por medio del sorting educativo y socio-económico, llevando así al aumento de los logros educativos de algunos grupos, a la homogenización de las preferencias de su descendencia y a la obtención de beneficios para el conjunto de la economía. En este sentido, el surgimiento de un arreglo social integrado o estratificado surge de manera endógena como el resultado natural de agentes



que optimizan sus recursos y que ponen en juego el capital social con el que cuentan en el proceso de acumulación de capital humano.

El resultado neto para la sociedad cuando el bien público analizado es la educación depende de las características específicas de los diferentes grupos, de la disponibilidad de fondos privados, de la formación y hábitos de los compañeros de clase y de la calidad de la oferta educativa pública o privada, tal y como se pudo apreciar de manera esquemática en el diagrama 2.1.

Es evidente que los altos niveles de segregación<sup>1</sup> revelan y refuerzan, en cualquiera de los casos, a un amplio espectro de variación de los talentos y de los logros educativos e inclusive, en muchas circunstancias, dependiendo del balance entre los beneficios y las pérdidas que genera, puede dar lugar a la disminución del nivel medio de estos logros cuando se hacen comparaciones entre regiones o países.

Cuando la dispersión en la calidad de los entornos locales es demasiado amplia el resultado esperado en términos de logros es también disperso y presumiblemente el proceso de auto reforzamiento conduzca a mayores grados de desigualdad de las generaciones futuras. Las implicaciones inter-generacionales que se derivan son obvias, los hijos de las familias con mayor dotación de capital social asisten a escuelas con mayores recursos, acumulan más capital humano, ganan más y reproducen un arreglo social segregado.

Sin embargo, de nuevo el resultado depende del balance entre los beneficios y las pérdidas forjadas por esta segregación y que será objeto de simulación en el último capítulo de esta Tesis. Cuando el balance es positivo (los beneficios de la segregación superan las pérdidas que genera) el desempeño educativo entra en una espiral ascendente, aumenta la acumulación agregada de capital humano, el crecimiento económico y la prosperidad, de tal suerte que tanto pobres como ricos obtienen beneficios. Cuando el balance es negativo el desempeño educativo medio puede verse seriamente afectado y así los pobres pierden más de lo que puedan ganar los ricos.

---

<sup>1</sup> Medidos a través de índices de densidad de capital humano de alta o baja educación en cada vecindad, barrio o comuna. También a través de los índices tradicionales de segregación como los de Duncan, las desviaciones relativas, las ratios entre varianzas intra-grupo y total o mediante índices de entropía.

Las posibles explicaciones a lo último son diversas y están relacionados, en primer lugar, con las distancias en los estándares de la calidad educativa; en segundo lugar con los mecanismos de financiación adoptados y, finalmente con el papel que desempeña la desigualdad inicial en la trayectoria dinámica del proceso de acumulación de capital humano. Los dos primeros aspectos están estrechamente ligados y por tal razón el acervo de literatura sobre calidad educativa, financiación y logros de los últimos años ofrece un menú amplio y variado.

Los diversos trabajos sobre segregación socio-económica, desde la perspectiva sociológica, enfatizan más en los efectos inmediatos de los entornos locales y familiares que en los costes o beneficios que desde una perspectiva dinámica puede generar la asimetría en la distribución de la calidad educativa. La literatura sobre funciones de producción de capital humano, en cambio, dirige su interés a la identificación de los beneficios que generan las interacciones con vecinos ricos o con alta composición de capital humano (*role-models*) o, en otros casos, los efectos de freno o de retraso inducidos por comunidades en las que predominan familias pobres o con bajos niveles educativos. Trabajos como los de [Benabou, 1996a](#) , [1996b](#) realizan una importante contribución sobre estos aspectos y lleva las consecuencias de su análisis a las implicaciones dinámicas de largo plazo.

### **Segregación y provisión eficiente de bienes públicos**

Los razonamientos anteriores llevan entonces al estudio de la provisión de bienes públicos. En este campo se identifican estudios ya clásicos que abordan el problema relacionándolo con los logros educativos, el *sorting* y la segregación escolar.

El trabajo de [Tiebout \(1956\)](#) inauguró una larga tradición que destaca los impactos de la descentralización fiscal sobre la eficiencia en la provisión local de bienes públicos como la educación. El planteamiento de manera simplificada es el siguiente: los agentes con estructura similar de preferencias se localizan en áreas metropolitanas, ciudades o micro-localidades (distritos, barrios o comunas) con perfiles específicos en la oferta de bienes públicos y, a la vez, escogen los niveles deseados de inversión según su valoración de la calidad. El aspecto crucial en los modelos que siguen esta tradición consiste en el ordenamiento espacial de las familias por niveles de ingreso y el

surgimiento de algún equilibrio estratificado en el que los individuos “*votan con los pies*”.

En esta misma línea se desprenden otros trabajos que se ocupan de la distribución espacial de las familias y el suministro de bienes locales cuando los hogares difieren en su estructura de rentas. [Ross y Yinger \(1999\)](#) realizan una revisión exhaustiva de esta literatura y destacan las investigaciones de [Ellickson \(1971\)](#) y [Westhoff \(1977\)](#) como trabajos de gran originalidad en este campo.

En la literatura de la década de los noventa aparecen los aportes de [Fernandez y Rogerson \(1997, 1998\)](#) que avanza en el estudio de los modelos de multi-comunidad y en el análisis de la incidencia de diversos mecanismos de financiación del gasto público en educación. Los autores abordan directamente los nexos entre la distribución del ingreso, la calidad de la educación pública y sus efectos redistributivos.

La hipótesis básica de esta literatura apunta al surgimiento de algún equilibrio estratificado en el que las familias se organizan en comunidades con un orden jerárquico de calidades educativas e ingresos. La segregación surge así de manera endógena del conjunto de decisiones secuenciales de las familias. En primer lugar, la generación de los padres escoge un lugar de residencia y, en segundo lugar, transforman gasto en educación mediante decisiones estrictamente privadas (provisión privada del bien público) o con algún mecanismo de votación para la provisión pública del bien.

El modelo identifica un conjunto de individuos con diferentes niveles de ingresos, pero con preferencias idénticas en las combinaciones de consumo y calidad de la educación para sus hijos. El mecanismo de decisión define finalmente la tasa óptima de inversión que garantizará los niveles de calidad educativa deseados para la descendencia. En efecto, el ejercicio de escogencia racional lleva a que cada comunidad maximice la utilidad de un individuo representativo bajo alguna restricción.

Un aspecto clave que es preciso retener es que la mayoría de estos modelos definen la calidad educativa desde la perspectiva del gasto, de tal manera que a mayor gasto per cápita mejor calidad. Los modelos de *sorting* escolar como los de [Fernandez y Rogerson, 1997](#), el de [Benabou, 1996b](#) y alguna literatura sobre crecimiento endógeno

como el trabajo de [Glomm y Ravikumar, 1992](#), así como algunas variantes recientes a los modelos de multi-comunidad, suponen que la calidad de la educación depende del gasto por estudiante y que su provisión pública o privada en ámbitos nacionales o locales exhibe rendimientos constantes a escala.

Los planteamientos anteriores llevan de inmediato a una nueva pregunta **¿cuál es el tipo de educación (pública o privada) que efectivamente prefieren los agentes y qué factores determinan su escogencia?** La sección 3.3 intentará dar cuenta del interrogante a través de algunos ejercicios empíricos que exploran el rol de la segregación<sup>2</sup>.

Desde una perspectiva más amplia, si los individuos distribuyen su ingreso entre consumo y proporción de gasto en la calidad educativa de sus hijos, en conjunción con su estructura de preferencias, entonces una clave que diferencia a los regímenes públicos y privados estriba en que la elección privada define de manera endógena en cada instante la calidad de su educación, otorgándole así un carácter específico a la decisión de cada agente. En la elección pública, en cambio, los individuos reciben de manera exógena los perfiles de calidad determinados por la disponibilidad de fondos públicos y por la ponderación que el Estado le otorgue a las inversiones en educación pública. Así, la elección privada cuenta con un rango amplio de posibilidades de producción de los servicios educativos, lo que se refleja en el campo de variación de las calidades; en la elección pública, las posibilidades de producción son estrechas y el rango de variación de la calidad educativa es limitado y restringido en sus estándares de competencia con el sector privado.

---

<sup>2</sup> Ahora bien, en determinadas condiciones la provisión pública puede implicar la redistribución de los pobres a los ricos en las escalas superiores de la educación. Tema que no se abordará en esta Tesis pero que puede señalar una agenda de investigación para el futuro. La explicación radica en que los individuos de bajos ingresos tienen obstáculos para acceder a la educación superior debido a las restricciones que impone su entorno familiar y local y que se revelan en las pruebas estándar de competencias o de selectividad. Este pobre desempeño se asocia a la baja calidad en la primaria, la educación secundaria básica y el bachillerato y, en consecuencia, la criba opera por la vía de las pruebas de competencias para ingresar a los planteles de educación superior, técnica y tecnológica. El corolario que se desprende del planteamiento es inmediato, si la calidad de la educación pública en los niveles medio y vocacional es deficiente y, a su vez, es consumida fundamentalmente por hogares pobres, la trayectoria dinámica de la acumulación de capital humano predice procesos sostenidos de segregación socio-económica o en el peor de los casos presagia el surgimiento de trampas de pobreza.

No obstante, también es preciso aclarar que en un régimen público completamente descentralizado el sistema de financiación debería reflejar la internalización que hacen los votantes de los retornos sociales y los spillovers esperados de la calidad educativa. Esto es aún más evidente cuando los fondos de financiación provienen de impuestos a las rentas estrictamente locales. Pero, tal y como se discutió en un acápite anterior, aunque los gobiernos locales capturen estas decisiones y asignen el gasto de mejor manera --obedeciendo a criterios de eficiencia en la asignación y en la producción de los servicios educativos-- la persistencia de asimetrías en las *valoraciones* de las familias pobres respecto al bien público o, inclusive, su fuerte aversión al riesgo, pueden inducir efectos negativos que pueden distorsionar finalmente la asignación del gasto público local en educación. En tal sentido, el espacio real de deseos estaría limitado por la capacidad efectiva de *desear* y de *valorar* la educación de buena calidad.

Cuando se trata de sistemas de financiación estrictamente local --tal y como sucede en los distritos educativos de algunos países industrializados en los que los fondos asignados para la educación pública provienen de los impuestos locales a la renta o a la propiedad— el modelo puede conducir a resultados muy cercanos o casi idénticos al sistema de asignación del gasto por la vía estrictamente privada en la que las familias asignan directamente sus gastos.

Cuándo conviven los dos tipos de regímenes --como es el caso de la mayoría de países de América Latina— con un tamaño relativamente alto del sector educativo privado, el sistema de asignación de gasto por la vía estrictamente privada que realizan los agentes desde los entornos locales, logra emular los mecanismos de sorting (en este caso una modalidad de sorting socio-económico) que se observan en los modelos de multi-comunidad.

Desde una perspectiva teórica, estas mixturas resultan difíciles de modelar debido a que en la práctica, a pesar de que los padres o los decisores de las inversiones en capital humano de las familias pagan tasas impositivas nacionales o locales, de todos modos se enfrentan a la decisión de enviar a sus hijos al sector público o a los colegios privados. Esta decisión depende de su estructura de preferencias por el servicio, generando así un proceso de “sorting” socio-económico en el que no votan en estricto con los pies en el sentido que le confiere el modelo inicialmente planteado por Tiebout, sino que revelan

sus preferencias enviando a sus hijos a escuelas o colegios de alta o baja calidad, sin necesidad de configurar en estricto un equilibrio espacial perfectamente estratificado, pero sí un equilibrio socio-económico estratificado.

La revelación de estas preferencias implica la elección de una tecnología de aprendizaje o de producción de los servicios educativos de sus hijos y que finalmente se expresa en la combinación particular entre calidad y precios de los planteles escogidos.

Ahora, tal y como se planteó en la introducción, además del sector público, el tamaño relativamente alto de la educación privada básica que se refleja en la proliferación de escuelas y colegios privados, como es el caso Colombiano, determinan por el lado de la oferta una amplia variedad de precios y calidades que configura un mercado de alta segmentación. El fallo de mercado al que conduce este proceso de elección y la potenciación explícita de espacios asimétricos de posibilidades que implica para las familias con diferentes niveles de ingresos lleva de nuevo a la reproducción de las desigualdades sociales. Por tal razón, desde otras perspectivas teóricas que tienen implicaciones inmediatas de política educativa surgen propuestas de financiación por el lado de la demanda como una alternativa para corregir los fallos del modelo.

Estos enfoques se desprenden fundamentalmente de los trabajos de [Hoxby\(1996a, 1996b\)](#) para el caso norteamericano. El punto de partida es que el sector público es ineficiente en el suministro de educación de buena calidad, no logra competir con los estándares que impone el sector privado –de acuerdo con esta concepción, más flexible en el suministro del servicio y en la combinación de tecnologías de producción—y que es más eficaz en la obtención de logros académicos.

La fórmula de política es, entonces, que el gobierno central o local suministre directamente los recursos de financiación educativa a las familias para que, de acuerdo con su estructura de preferencias sobre la calidad educativa, asignen estos bonos o “*vouchers*” entre las escuelas públicas y privadas de primaria o de secundaria de su predilección. El propósito del esquema es mejorar la eficiencia asignativa del sistema y generar competencia en la calidad entre los planteles educativos públicos y privados. Así, el modelo busca generar condiciones favorables para la redistribución del capital

humano de alta calificación entre los diferentes grupos sociales y brindar oportunidades de movilidad intergeneracional ascendente.

No obstante, el resultado de este modelo no es único y depende de su diseño y aplicación tal como lo sugiere [Nechyba \(1999\)](#). La evidencia empírica que discute este autor muestra que cuando se adopta este esquema es probable que se auto-refuerce la segregación socio-económica en detrimento de los pobres, se acelere la reducción del tamaño del Estado por la vía de la transferencia creciente de alumnos desde escuelas públicas a las privadas o, incluso, que aparezcan mecanismos sofisticados de discriminación contra individuos de baja habilidad a favor de la “*crema y nata*”.

Esto último puede surgir por varias razones:

(i) Sesgo de selección por el lado de las familias. Las que cuentan con mejor background familiar y mejor calidad de los entornos locales (efectos de vecindad) toman ventaja respecto a las demás. Valoran más el bien y no son aversos a los riesgos de inversión. De manera opuesta, los que no cuentan con background familiar, ni calidad de los ambientes locales favorables tienden a valorar menos el bien educación y son aversos a los riesgos de inversión.

(ii) Los oferentes de los servicios escolares atraen y escogen individuos de alta habilidad y con perfiles socio-económicos específicos a la manera de un club. Esta escogencia garantiza mejores resultados académicos, más prestigio y mayor demanda futura.

(iii) Los oferentes prefieren escoger mercados rentables, fundamentalmente localizados en las áreas metropolitanas y ciudades de tamaño relativo mayor, a la vez que desprecian la oferta en centros poblados de baja demanda o en las zonas rurales<sup>3</sup>. Esto puede llevar hacia una mayor polarización entre regiones.

---

<sup>3</sup> Para llevar el argumento a un extremo se formula el siguiente interrogante ¿Qué tipo de escuelas privadas ofrecerían los servicios educativos en el Departamento del Chocó (Colombia), localizado en la costa pacífica colombiana y con un alto porcentaje de población afro-colombiana en condiciones de miseria?

(iv) De acuerdo con (i), (ii) y (iii) las colas de la distribución del capital humano se ven afectadas y, en consecuencia, se genera mayor estratificación escolar y segregación socio-económica. De este modo mejora la eficiencia pero empeora la distribución de los ingresos.

(v) Finalmente, el esquema promueve el aumento del tamaño relativo del sector educativo privado, tal y como lo afirman [Nechyba \(1999\)](#) y [Epple and Romano \(2000\)](#).

A manera de inferencia de lo anterior, las diferencias en los logros educativos en términos de calidad se distribuyen de manera asimétrica entre los diferentes grupos sociales y afectan el potencial de inserción en el mercado laboral, lo que finalmente hace que se refuerce la estratificación.

Finalmente, un aspecto que no recibe la atención adecuada en la literatura y que sería de gran interés para una línea futura de investigación, apunta a que el aumento del tamaño del sector privado promueve la mayor segmentación de la calidad educativa en su interior, proliferan escuelas y colegios privados de dudosa calidad y obliga al sector estatal a diseñar mecanismos de filtro o sistemas complejos y costosos de acreditación de la calidad que terminan financiándose con recursos del mismo sector público.

### **3.3. Calidad de los entornos locales, segregación y preferencias por la educación**

La aproximación empírica a la segregación identifica una amplia variedad de métricas e índices<sup>4</sup> y el cumplimiento de las propiedades básicas da lugar a clasificaciones exhaustivas que no es del caso discutir en detalle aquí<sup>5</sup>.

---

<sup>4</sup> Entre los que se destacan i) la proporción de varianza intra-grupal en la varianza total ii) los denominados índices de Duncan que miden las desviaciones medias relativas ponderadas por el peso de cada grupo o atributo estudiado, iii) los índices de entropía tipo Theil fundamentados en las probabilidades a priori y a posteriori de pertenecer a determinado grupo, iv) los índices de exposición que miden las probabilidades de que los miembros de un determinado grupo social se localicen en vecindades con individuos pertenecientes a otros grupos y v) los índices de similaridad o disimilaridad que intentan capturar el grado de homogeneidad o de heterogeneidad intra-grupal e inter-grupal.

<sup>5</sup> Las propiedades necesarias que garantizan la consistencia de los índice responden a los siguientes interrogantes: a) ¿Cuál es el valor del índice cuando la segregación es nula o máxima? (*límites*)  
b) ¿Qué valores toma cuando se intercambian las categorías o atributos en la distribución? (*simetría*)  
c) ¿Las medidas son sensibles a cambios en el número total de individuos en cada grupo? (*invarianza al tamaño grupal*)



La clasificación de individuos o familias en grupos o localizaciones es el prerrequisito fundamental para la estimación de estos índices. A partir de aquí los ejercicios de descomposición de la varianza permiten computar la proporción del componente intra-grupal en la varianza total de una determinada área metropolitana.

Con el fin de estilizar los hechos se han medido dos índices simples que permiten apreciar, en primer lugar, las diferencias intra-metropolitanas en la concentración de capital humano calificado y no calificado y, en segundo lugar, las diferencias intra-metropolitanas en la estructura del gasto educativo de los hogares.

Mientras que el primer índice captura la distribución del capital humano de la población activa y está relacionado con la calidad de los entornos locales (niveles educativos medios de las micro-localidades), el segundo propone la aproximación a la estructura de preferencias de las familias hacia la educación por la vía del gasto.

La distribución de estos dos índices en el interior de las áreas metropolitanas proporcionará información que será utilizada en los modelos econométricos de la siguiente sección. Los cálculos se realizan con una submuestra de 6445 individuos, jefes de hogar, de la población activa y 3447 escolares (entre los 6 y 17 años de edad) pertenecientes a estos mismos hogares en el interior de las principales áreas metropolitanas del país. Los casos seleccionados corresponden a los que tienen información completa o casi completa y se descartaron aquellos casos con demasiados valores omitidos en la mayoría de variables relevantes. Estos individuos se distribuyen en 370 vecindades o micro-localidades seleccionados finalmente.

Debido a la ausencia de datos a nivel de barrios o comunas en Colombia, que correspondería a una noción más estándar de vecindad, el procedimiento utilizado para

---

d) ¿Si los individuos con los mismos atributos se reagrupan en otros grupos o categorías que ocurre con el índice? (*equivaloración*)

e) ¿Los integrantes de cada grupo son anónimos? (*imparcialidad*)

f) ¿Qué ocurre con el índice si un individuo es transferido desde un grupo de categoría más alta? (*principio de transferencia*).

aproximar el análisis a los entornos micro-locales siguió la estrategia que se describe en la nota al pie de página<sup>6</sup>.

### 3.3.1. Índice de distribución del capital humano

Este índice, que captura la composición de capital humano, establece la razón entre los trabajadores calificados y no calificados de las micro-localidades en el interior de las áreas metropolitanas de Colombia. De acuerdo con el arreglo educativo del país la definición de grupos de calificación suele utilizar 4 categorías. El primer grupo está constituido por quienes tienen menos de 6 años de educación o que solo han obtenido primaria completa. El segundo grupo entre 6-10 años, que constituye la meta constitucional de educación básica (5 años de primaria, al menos uno de preescolar y 4 años de educación secundaria) y en algunos casos se extiende hasta aquellos que llegan a 11 años sin validación en las pruebas del Estado (ICFES=Instituto Colombiano para el Fomento de la Educación Superior).

El tercer grupo incluye los que tienen bachillerato completo (11 años de educación y que hayan presentado las pruebas de estado del ICFES) hasta 14 años de estudios y que por lo general se pueden asociar a los que realizan formación intermedia o algunas carreras técnicas y tecnológicas profesionales.

Finalmente el cuarto grupo está conformado por quienes tienen educación superior.

---

<sup>6</sup> El concepto de “vecindades” utilizado corresponden a los conglomerados y segmentos muestrales que identifican la localización de los individuos en la Encuesta de Calidad de Vida 2003 (ECV2003) a partir del módulo de identificación y control. Estos segmentos corresponden a los mismos que utiliza del Departamento Nacional de Estadística (institución oficial encargada de los diseños metodológicos, operativos de recolección, procesamiento, administración y publicación de las estadísticas en Colombia) para el diseño muestral de los Censos de población y configuran la unidad básica de muestreo.

Teniendo en cuenta que la encuesta realiza un muestreo probabilístico, estratificado de conglomerados y polietápico, que luego permite identificar cada hogar en la base primaria con un número o llave de 20 dígitos y una variable adicional de identificación de la vivienda que consta de 13 dígitos, a partir de estos identificadores consecutivos el ejercicio consistió en desagregar la “llave” e identificar los puntos de quiebre de los segmentos (anónimos) en el interior de las áreas metropolitanas más grandes (A.M. Bogotá, A.M Cali, A.M Medellín) originando así un conjunto de unidades que constan de un mínimo de 10 hogares. Finalmente se seleccionaron 370 de estas unidades, descartando el resto por efectos de tamaño. La selección probabilística controlada de estas unidades muestrales, denominadas en esta Tesis como “vecindades” las define el DANE a a partir de criterios de estratificación socio-económica. De esta manera, garantiza hasta donde sea posible la homogeneidad de cada unidad. Los detalles del diseño muestral los proporciona el Departamento Nacional de Estadística, DANE, a través de la Dirección de metodología y producción estadística.

Para efectos del análisis empírico la población activa urbana de este estudio se clasifica en dos (2) grupos: calificados (los de tercer y cuarto grupo antes descrito) y los trabajadores no calificados (el primer grupo más los del segundo grupo).

La notación utilizada para el cálculo del primer índice es la siguiente:

El subíndice  $i$  se refiere al individuo; y el subíndice  $j$  alude a cada microlocalidad en el interior de las áreas metropolitanas.

$hhlh_{ij}$ : corresponde a la razón entre el porcentaje de trabajadores calificados y no calificados según la definición anterior

$hhlh_{ij}^p$ : es el valor del índice en la escala 0-100

$hhlh_{\min}$ : es el valor mínimo observado en las frecuencias de la ratio  $hhlh_{ij}$  para cada área metropolitana

$hhlh_{\max}$ : es el valor máximo observado en las frecuencias de la ratio  $hhlh_{ij}$  para cada área metropolitana

$CA_{hhlh_{ij}}$ : coeficientes de ajuste

El índice se normaliza en la escala [0, 100 ] de acuerdo con las distancias entre los valores máximos y mínimos observados para cada una de las áreas metropolitanas:

$$(3.1) \quad hhlh_{ij}^p = \left\{ \frac{hhlh_{ij} - hhlh_{\min}}{CA_{hhlh_{ij}}} \right\}.$$

El coeficiente de ajuste se calcula mediante la siguiente expresión:

$$(3.2) \quad CA_{hhlh_{ij}} = \frac{hhlh_{\max} - hhlh_{\min}}{100}$$

El indicador, representado en el [gráfico 3.1](#), informa sobre el peso relativo de trabajadores calificados y de su distribución en el interior de las principales áreas metropolitanas. En este sentido es una *proxy* de la calidad de las interacciones sociales o

del entorno en el que viven los escolares de la muestra. Valores alejados de cero denotan mayor densidad de población activa de alta educación en relación con la de educación baja. A medida que nos desplazamos hacia la derecha esta proporción entre calificados y no calificados aumenta y la curva de densidad proporciona las probabilidades correspondientes a los valores específicos de la ratio. El desplazamiento hacia la izquierda, indica que la ratio es cada vez más pequeña y que en la composición de capital humano son menos frecuentes los individuos calificados o predominan los individuos no calificados.

Gráfico 3.1

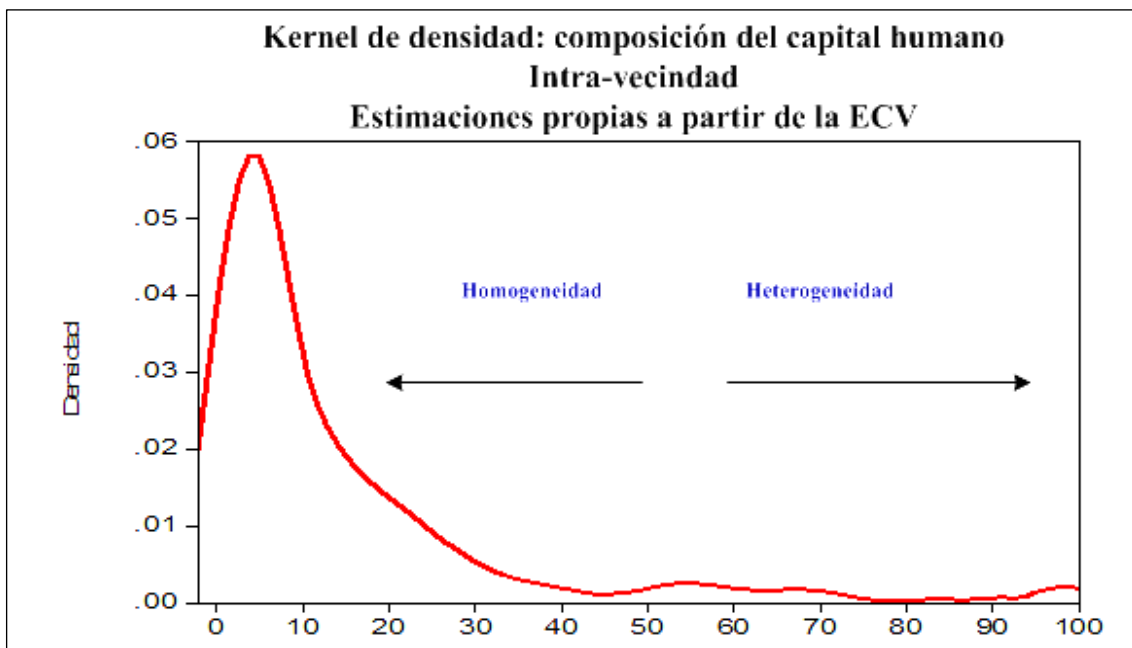


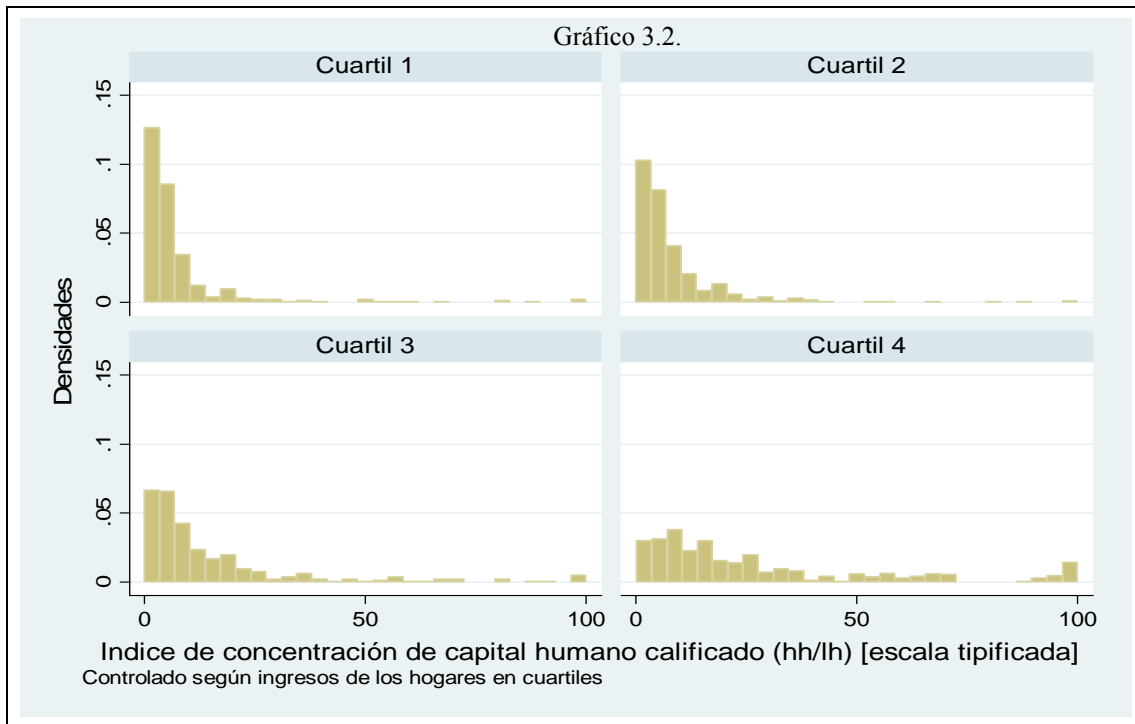
Gráfico 3.2 Distribución del índice según cuartiles de ingresos

El [gráfico 3.2](#) muestra el comportamiento del índice en cada cuartil de ingresos medios de los hogares. Las probabilidades de encontrar valores bajos del indicador son relativamente altas en los cuartiles 1 y 2 de ingresos y casi despreciables en los cuartiles 3 y 4.

Es preciso anotar que la distribución suavizada del indicador para las principales áreas metropolitanas del país se ajustó mediante una función de densidad Gaussiana

$f(x) = \frac{1}{N_c \Phi} \sum_{i=1}^{N_c} k\left(\frac{x - X_i}{\Phi}\right)$  donde  $N_c$  corresponde al número de observaciones para cada

una de las áreas metropolitanas,  $\phi$  es el parámetro de suavización y  $k$  un parámetro de ponderación.



### 3.3.2. Índice de sorting educativo en gastos

A partir de la información de gastos educativos,  $g_{ij}$ , que incluye pagos de matrículas, uniformes, útiles escolares y transporte, para cada uno de los escolares se estimó su proporción en los ingresos laborales del hogar,  $y_{ij}$ , obteniendo la varianza intravecindad,  $\text{var}_{\text{int}}$ , de la ecuación (3.3).

$$(3.3) \quad \text{var}_{\text{int}} = \text{var}[g_{ij} / y_{ij}].$$

$$(3.4) \quad \text{var}_{\text{tot}} = \frac{1}{N_c} \sum_{i=1}^T [(g_{ij} / y_{ij}) - (g_{ij} / y_{ij})_{\text{med}}]^2$$

La ecuación (3.4) expresa el cálculo de la varianza total,  $\text{var}_{\text{tot}}$  para cada una de las áreas metropolitanas, en donde  $(g_{ij} / y_{ij})_{\text{med}}$  corresponde al valor medio de la proporción de gasto educativo en el ingreso laboral<sup>7</sup>.

<sup>7</sup> La proporción media de gasto en el ingreso total es de 12.6% para la muestra utilizada con una desviación estándar de 8.96.

La razón entre esta varianza intra-vecindad y la total, para cada área metropolitana, proporciona una aproximación sencilla a lo que en este estudio se denomina como el índice de sorting educativo identificado por el lado del gasto.

$$(3.5) \quad S_{ij} = \frac{\text{var}_{\text{int}}}{\text{var}_{\text{tot}}}$$

La estandarización del índice se realiza mediante un coeficiente de ajuste ( $CA_j$ ) que tiene en cuenta los valores máximos ( $S_{\text{max}}$ ) y mínimos ( $S_{\text{min}}$ ) de la ratio observados en cada una de las áreas metropolitanas consideradas.

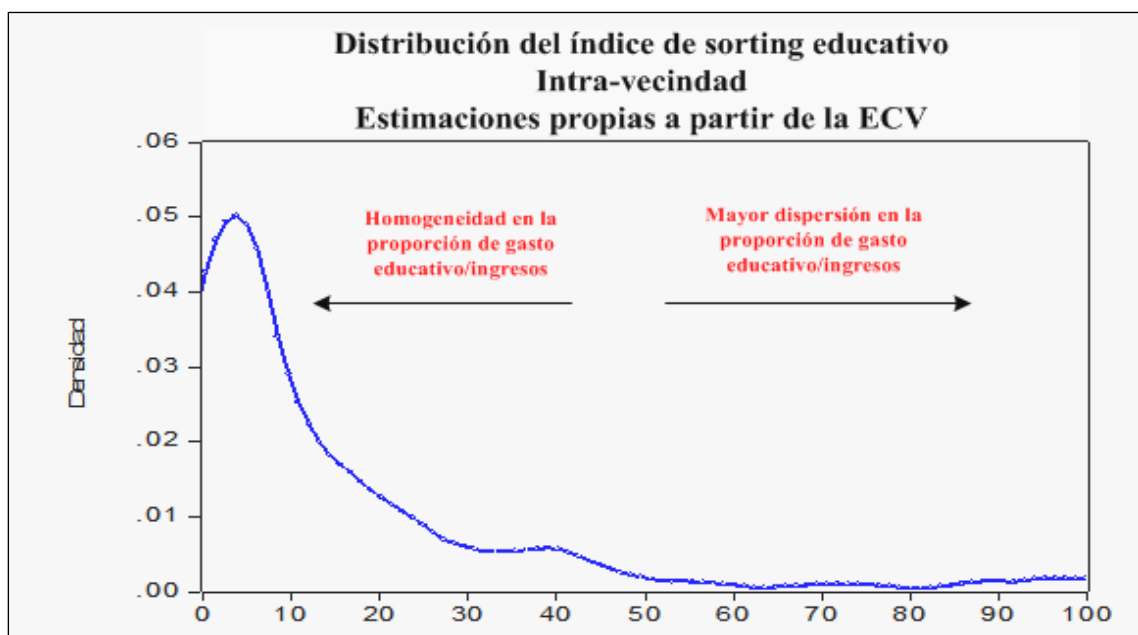
$$(3.6) \quad CA_j = \frac{S_{\text{max}} - S_{\text{min}}}{100}$$

El índice transformado para cada una de las micro-localidades queda expresado en la escala [0, 100] indicando mayor heterogeneidad grupal en los gastos educativos a medida que se aproxima a 100 y homogeneidad cuando se acerca a 0.

$$(3.7) \quad S'_{ij} = \left\{ \frac{S_{ij} - S_{\text{min}}}{CA_j} \right\}$$

El [gráfico 3.3](#) representa la distribución del índice para el conjunto de las áreas metropolitanas.

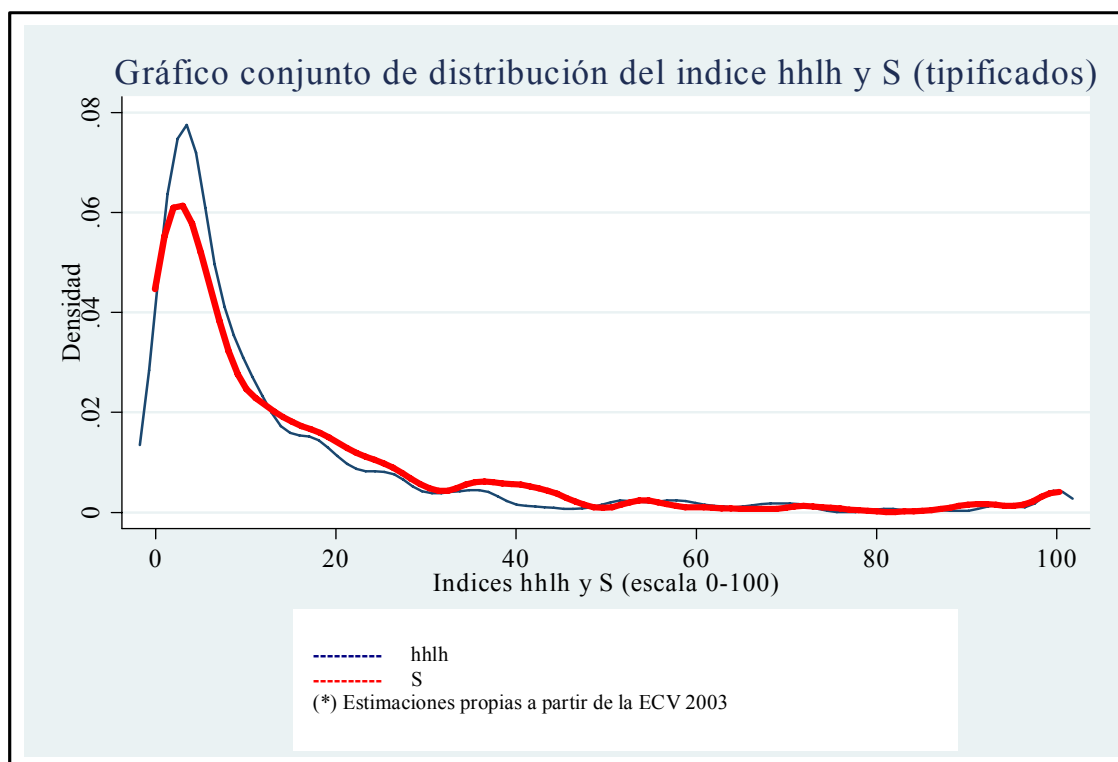
**Gráfico 3.3**



El indicador proporciona información sobre el grado de dispersión relativa de cada micro-localidad a la vez que nos aproxima de manera indirecta a la distribución de preferencias por la calidad educativa. Por tal razón se trata de un índice de sorting educativo propuesto aquí desde la perspectiva del gasto. Valores cercanos a cien indican que la variabilidad en el interior de la vecindad es superior a la encontrada para el área metropolitana o el agregado nacional como conjunto y, en efecto, indica que la micro-localidad es dispersa en la proporción de gasto educativo respecto a los ingresos laborales. Valores cercanos a cero indican que la variabilidad relativa es menor y que se trata de una micro-localidad con un mayor grado de homogeneidad en las asignaciones del gasto educativo.

La distribución conjunta de los dos índices ([gráfico 3.4](#)) permite apreciar la similaridad en la distribución de la población activa calificada, respecto a la no calificada  $h^k/h$ , y el índice  $S$  calculado a partir del gasto educativo.

**Gráfico 3.4**



### 3.3.3. Segregación y elección de escuelas públicas o privadas

Los modelos de elección discreta que se proponen y estiman a continuación intentan cuantificar los factores que afectan las decisiones de asistir a escuelas públicas o privadas.

El modelo econométrico planteado incluye variables del background familiar, del entorno local y algunas características individuales. La muestra utilizada en las estimaciones proviene de la base de escolares urbanos con información completa para el agregado nacional, extraídos de la ECV del año 2003.

En la línea de los planteamientos realizados en las secciones anteriores, se espera que las variables relacionadas con el background familiar y la calidad de los entornos locales sean significativas en la explicación de la estructura de decisiones.

#### 3.3.3.1. Estructura general de los modelos e hipótesis de comportamiento

La estructura general del modelo puede expresarse en las siguientes ecuaciones.

$$\begin{aligned}
 P(y_{ij} = \text{público}) &= F(\mathbf{x}, \beta) = \beta' \mathbf{x} = \frac{e^{\beta' \mathbf{x}}}{1 + e^{\beta' \mathbf{x}}} \\
 (3.8) \quad E[y_{ij} | \mathbf{x}] &= F(\mathbf{x}, \beta) \\
 \text{Ln} \frac{P_{ij}}{1 - P_{ij}} &= \beta' \mathbf{x} + \varepsilon_{ij}
 \end{aligned}$$

Los subíndices denotan el individuo  $i$  localizado en la microlocalidad  $j$  de cada área metropolitana.

La probabilidad de que los individuos asistan a escuelas o colegios públicos,  $P(y_{ij} = \text{público} = 1)$ , depende de un vector  $\mathbf{x}$  de atributos que incluye los sub-vectores de variables individuales, el background familiar y las variables proxy relacionadas con los entornos locales en los que los individuos toman sus decisiones. Tal y como se puede apreciar la especificación es lineal en el logaritmo de la factibilidad,

$\text{Ln} \frac{P_{ij}}{1 - P_{ij}} = \beta' \mathbf{x} + \varepsilon_{ij}$  y supone una función de distribución acumulativa logística. El

cociente  $\frac{P_{ij}}{1 - P_{ij}}$  corresponde a las denominadas “odds ratio” o factibilidades.



Los parámetros  $\beta$  corresponden a las pseudos-elasticidades del modelo y no tienen una interpretación directa, por lo que es necesario realizar la retransformación y calcular las probabilidades a través de  $\hat{P}(y = \text{público} | \mathbf{x}) = \Lambda(\mathbf{x}\hat{\beta})$ . Algunos detalles adicionales se pueden observar en el [anexo A3.1](#).

A continuación se definen las variables consideradas (cuyas estadísticas descriptivas básicas se pueden apreciar en el [anexo A3.2](#)) y las hipótesis de comportamiento. El [anexo A3.3](#) presenta la ficha metodológica de la Encuesta de Calidad de Vida 2003.

- *Edad de los escolares ( $Edad_{ij}$ )*
- *El origen de los padres ( $Origen$ )*. Esta variable intenta complementar los perfiles socio-demográficos que configuran el background familiar de los estudiantes.

Resulta de gran relevancia en el contexto colombiano debido a que de manera adicional a los efectos de la transición demográfica que el país experimentó desde finales de los 60s y la inercia propia de los procesos de urbanización que ya venían desde años atrás, las migraciones internas en Colombia tienen un altísimo componente explicado por las diversas fases del conflicto socio-político que se viene librando desde el año 1948. La primera ola de violencia 1948-1963 generó los primeros efectos de desplazamiento masivo hacia las zonas urbanas del país y la segunda ola, que viene desde el año 1963 hasta hoy, genera desplazamientos todavía más drásticos inducidos por la confrontación territorial entre los grupos guerrilleros y los grupos paramilitares. Por tal razón la variable de origen urbano o rural de los padres resulta de gran relevancia si tenemos en cuenta, además, la brecha socio-económica urbano-rural descrita en capítulos anteriores.

La hipótesis es que existen diferencias significativas entre la elección de escuelas y colegios públicos o privados cuando los padres son de origen urbano, respecto a los que tienen padres de origen rural. Las diferencias estriban en la valoración que pueden asignar a los proyectos educativos de sus hijos y a los rezagos generados por el canal de transmisión de habilidades descrito en el diagrama 2.2 del capítulo anterior. La variable toma valor de 1 para los padres de origen urbano y 0 para los padres de origen campesino.

- *Nivel educativo de los padres* que opera como una variable en categorías ( $Espadre$ ).

De acuerdo con las explicaciones previas, esta variable intenta capturar los impactos diferenciales sobre el espacio de probabilidades, que se derivan de los niveles de escolaridad de los padres. La categoría de referencias corresponde a los padres con primaria incompleta y sin educación, las demás categorías son la primaria completa ( $Espadre_p$ ), secundaria incompleta ( $Espadre_{si}$ ) y finalmente bachiller o nivel superior ( $Espadre_{bs}$ ).

- *Variables del entorno local*. Los modelos incluyen las variables analizadas en la sección previa: el índice de densidad de la población activa calificada respecto a la no calificada ( $h^o/h_{ij}$ ) y también se consideran algunos modelos que incluyen los índices estandarizados de sorting educativo ( $S_{ij}^o$ ). Adicionalmente se explora la incidencia directa del logaritmo del gasto educativo medio de la micro-localidad ( $LnG_j$ ) y los ingresos per cápita ( $Y_j$ )

La hipótesis de comportamiento apunta a que los estudiantes que provienen de entornos favorables, en los que la proporción de individuos calificados de la población activa (calidad de los entornos locales) y las preferencias reveladas en la estructura del gasto educativo son altos, afecta positivamente la probabilidad de asistir a planteles privados, o lo que es lo mismo, afecta negativamente la probabilidad de asistir a planteles educativos públicos.

Desde las perspectivas descritas en el apartado 3.2, las dos primeras variables captan el efecto de la segregación socio-económica sobre la estructura de decisiones.

- *Pobreza*. Algunas especificaciones consideran el indicador de pobreza disponible en la base de datos y que clasifica a los hogares en pobres y no pobres, de acuerdo con la línea de pobreza explicada en el capítulo 2. Esta variable dicotómica toma valor de 1 para los pobres y 0 para los no pobres.

La hipótesis subyacente, suficientemente discutida en las secciones previas, es que la procedencia de hogares pobres incide positivamente en la probabilidad de escoger escuelas públicas o negativamente en la elección de escuelas privadas.

Denotando por  $L_{ij}$  el logaritmo de la razón de probabilidades, las diversas especificaciones exploradas se resumen en el [cuadro 3.1](#).

El modelo (1) solamente incluye las variables relacionadas con el origen de los padres y su escolaridad. La diferencia con la especificación (2) radica en que esta última incluye el indicador de composición de capital humano en el ámbito de microlocalidad.

La especificación (3), incluye además del bloque de variables anteriores el indicador de pobreza y la (4) el logaritmo del gasto educativo medio de cada segmento. El modelo (6) explora la incidencia de los ingresos per cápita en lugar del gasto educativo.

Los modelos (7) y (8) incluyen el índice de segregación en gastos educativos en lugar del índice de composición de capital humano.

**Cuadro 3.1. Modelos explorados en las estimaciones**

Especificación	$L_{ij} = \ln \frac{P_{ij}}{1 - P_{ij}}, \text{ con } P_{ij} = P(y_{ij} = \text{público})$
(1)	$L_{ij} = \beta_0 + \beta_1 \text{Edad}_{ij} + \theta_1 \text{Origen} + \delta_1 \text{Espadre}_p + \delta_2 \text{Espadre}_{si} + \delta_3 \text{Espadre}_{bs} + \varepsilon_{ij}$
(2)	$L_{ij} = \beta_0 + \beta_1 \text{Edad}_{ij} + \theta_1 \text{Origen} + \delta_1 \text{Espadre}_p + \delta_2 \text{Espadre}_{si} + \delta_3 \text{Espadre}_{bs} + \tau_1 h_j^{\text{PH}} + \varepsilon_{ij}$
(3)	$L_{ij} = \beta_0 + \beta_1 \text{Edad}_{ij} + \theta_1 \text{Origen} + \theta_2 \text{Pobre} + \delta_1 \text{Espadre}_p + \delta_2 \text{Espadre}_{si} + \delta_3 \text{Espadre}_{bs} + \tau_1 h_j^{\text{PH}} + \varepsilon_{ij}$
(4)	$\ln y_{ij} = \beta_0 + \beta_1 \text{Edad}_{ij} + \theta_1 \text{Origen} + \theta_2 \text{Pobre} + \delta_1 \text{Espadre}_p + \delta_2 \text{Espadre}_{si} + \delta_3 \text{Espadre}_{bs} + \tau_1 h_j^{\text{PH}} + \tau_4 \ln G_j + \varepsilon_{ij}$

(5)	$\begin{aligned} Ln_{ij} = & \beta_0 + \beta_1 Edad_{ij} + \theta_1 Origen + \delta_1 Espadre_p + \delta_2 Espadre_{si} + \delta_3 Espadre_{bs} + \\ & \tau_1 h'_{ij} + \tau_4 LnG_j + \varepsilon_{ij} \end{aligned}$
(6)	$\begin{aligned} L_{ij} = & \beta_0 + \beta_1 Edad_{ij} + \theta_1 Origen + \theta_2 Pobre + \delta_1 Espadre_p + \delta_2 Espadre_{si} + \delta_3 Espadre_{bs} + \\ & \tau_1 h'_{ij} + \tau_3 LnY_j + \varepsilon_{ij} \end{aligned}$
(7)	$\begin{aligned} L_{ij} = & \beta_0 + \beta_1 Edad_{ij} + \theta_1 Origen + \delta_1 Espadre_p + \delta_2 Espadre_{si} + \delta_3 Espadre_{bs} + \\ & \tau_2 S'_j + \varepsilon_{ij} \end{aligned}$
(8)	$\begin{aligned} L_{ij} = & \beta_0 + \beta_1 Edad_{ij} + \theta_1 Origen + \theta_2 Pobre + \delta_1 Espadre_p + \delta_2 Espadre_{si} + \delta_3 Espadre_{bs} + \\ & \tau_2 S'_j + \varepsilon_{ij} \end{aligned}$

### 3.3.3.2. Análisis de los resultados

Los resultados de las estimaciones se pueden apreciar en el [cuadro 3.2](#) en el que aparecen las pseudo-elasticidades, los estadísticos relevantes y una selección de medidas de bondad del ajuste.

En general, todos los signos corresponden con los esperados en las hipótesis de comportamiento y los valores obtenidos fuertemente significativos. La bondad del ajuste medido por los coeficientes de McFadden<sup>8</sup>, McKelvey-Zavoina y el Cragg-Uhler reportados por el Stata en la utilidad Spot es el adecuado para este tipo de modelos.

#### *La importancia del background familiar*

El origen urbano o rural de los padres de los educandos, así como los coeficientes de la escolaridad de los padres aportan evidencia sobre la importancia del background familiar en la elección de escuelas públicas o privadas.

<sup>8</sup> El coeficiente de McFadden compara el modelo completo y solamente con intercepto, en este sentido es un índice de razón de versosimilitud que toma la siguiente forma:

$$R^2_{McF} = 1 - \frac{\text{Ln}(\hat{L}_{full})}{\text{Ln}(\hat{L}_{intercepto})}$$

El coeficiente de McKelvey-Zavoina supone que la decisión opera con una variable latente y estima como grado de bondad del ajuste la siguiente razón de varianzas:

$$R^2_{M\&Z} = \frac{V\hat{a}r(\hat{y}^*)}{V\hat{a}r(y^*)} = \frac{V\hat{a}r(\hat{y}^*)}{V\hat{a}r(\hat{y}^*) + V\hat{a}r(\varepsilon)}$$

La estimación (1) que solamente incluye este bloque de variables registra importantes impactos diferenciales en la probabilidad de asistir a escuelas públicas, de acuerdo con el nivel máximo de escolaridad, respecto al nivel de referencia, alcanzado por los progenitores. El modelo permite predecir, mediante algunos cómputos condicionados a padres de origen campesino y con primaria incompleta, que la probabilidad que tienen sus hijos de asistir a planteles públicos es de 0.79 y de asistir a establecimientos privados de 0.21.

Al realizar el mismo ejercicio, condicionando ahora con padres de origen urbano y primaria incompleta, las probabilidades cambian a 0.69 (públicos) y 0.31 (privados). Casi 10 puntos que puede explicarse por las diferencias en la valoración de las inversiones educativas en sus hijos, la aversión al riesgo o por las diferencias en la condición de pobreza, entre otros factores.

Para intentar dilucidar esta última hipótesis, controlamos los dos mismos ejercicios con la condición de pobreza hallando lo siguiente:

- (a) Si los padres son de origen rural y pobres, además de contar con primaria incompleta o ningún nivel de escolaridad, la probabilidad de elegir escuelas públicas es de 0.86
- (b) Si se trata de padres urbanos, también pobres, la probabilidad es ahora de 0.78
- (c) Si son urbanos y no pobres la probabilidad de enviar a sus hijos a los planteles públicos disminuyen a 0.598
- (d) Si estos padres son de origen campesino y no pobres la probabilidad es de 0.72.

Una sencilla matriz que se deriva de los ejercicios queda consignada en el siguiente cuadro:

<b>Origen\Pobreza</b>	<b>No pobre</b>	<b>Pobre</b>	<b>Pobre - No pobre</b>
<b>Urbano</b>	0.60	0.78	0.18
<b>Rural</b>	0.72	0.86	0.14
<b>Rural-Urbano</b>	0.12	0.8	

De este modo el origen de los padres y la condición de pobreza del hogar inciden en estas diferencias en las magnitudes obtenidas en la matriz.

Para evaluar los impactos diferenciales de los niveles educativos de los padres, respecto al de referencia, se evaluaron en el mismo modelo (1) los cambios discretos en la probabilidad obteniendo los siguientes resultados: los padres con primaria completa disminuyen la probabilidad de que sus hijos asistan a escuelas o colegios públicos en 0.13, los padres con secundaria incompleta en 0.24 y los que tienen bachillerato o niveles superiores en 0.37. Estos cálculos corresponden aproximadamente a los efectos marginales de cada variable. De acuerdo con estos resultados, la probabilidad que tiene un niño de asistir a una escuela pública cuando su padre es de origen urbano y bachiller o con educación superior es de 0.33, esto es, 77 de cada 100 con este perfil asisten probablemente a planteles privados. Si el padre es de origen rural, pero con alto nivel educativo la probabilidad de matricular a sus hijos en el sector público es 0.46 y si es urbano y pobre (controlando por esta variable) la probabilidad es 0.50.

### Calidad de los entornos locales

Los modelos (2)-(6) exploran la incidencia que tiene el entorno educativo de los niños y jóvenes mediante el índice  $h_{ij}^p/h_{ij}$ . El índice de composición de capital humano en la escala [0-100] resultó fuertemente significativo y con el signo esperado: a mayor proporción de individuos con nivel educativo alto, la probabilidad de matricular a los hijos en escuelas públicas es menor.

El efecto marginal de esta variable en el modelo (2) fue de -0.007 por cada punto de aumento. Si el indicador aumenta 10 puntos, manteniendo todo lo demás constante, la probabilidad de asistir a planteles públicos disminuye en 0.07.

El ejercicio de cambio discreto del valor mínimo del indicador, próximo a cero, al valor máximo próximo a 100, disminuye la probabilidad en 0.58.

La probabilidad general predicha por el modelo (2) cuando el indicador toma su valor medio (14.25) es de 0.57 (escuelas o colegios públicos) y 0.43 (privados).

El perfil<sup>9</sup> de un niño o joven que tiene un padre de origen urbano, con alto nivel educativo y que vive en un entorno socio-demográfico en el que prolifera gente de alta

<sup>9</sup> prvalue, x(origen=1 hhlh<sub>2</sub>= 100 espadre<sub>p</sub>=0 espadre<sub>si</sub>=0 espadre<sub>bs</sub>=1)

calificación, arroja en los ejercicios de estática, una probabilidad de matricular a los hijos en establecimientos privados de 0.95, esto es, solo 5 de cada 100 probablemente asistan a planteles públicos.

El perfil contrario, entornos con padres de origen rural, bajos niveles educativos y rodeados de vecinos con escolaridad precaria, dan lugar a una probabilidad de enviar a sus hijos a escuelas públicas de 0.81 (19 de cada 100 probablemente asistan a escuelas privadas). Este resultado es interesante porque muestra la ausencia de simetría en los dos experimentos.

Si los ejercicios anteriores se realizan tomando el valor medio del indicador  $h^0/h_{ij}$ , los resultados obtenidos en las dos experiencias son 0.42 y 0.73, respectivamente.

El modelo (3) incluye la dicotómica de pobreza y permite controlar los ejercicios de estática. El efecto del cambio en la calidad de pobre a no pobre implica un cambio discreto en la probabilidad de 0.17 en este modelo, el cual, a su vez predice una probabilidad general de 0.86 cuando el entorno familiar y local es desventajoso. Si el entorno familiar y local es muy favorable (como en el experimento anterior) esta probabilidad predicha es de 0.06.

Experimentando con los dos casos extremos de arriba, pero controlando ahora por la condición de pobreza y condicionando al valor medio del indicador  $h^0/h_{ij}$ , se encuentra que la condición de ser pobre, provenir de hogares con padres de origen campesino y sin educación, implica una probabilidad de 0.81.

La condición de no ser pobre y contar con padres educados y urbanos, en cambio, da lugar a una probabilidad de 0.38.

Cuando se incluye el logaritmo del gasto educativo medio de la microlocalidad la estimaciones (4) y (5) registran un efecto marginal de -0.21 que indica una disminución en la probabilidad de 2.1 puntos porcentuales por cada 10% de aumento en el gasto medio local, manteniendo todo lo demás constante o en sus valores medios. Esta variable resultó significativa en los dos modelos y tal como se puede apreciar absorbe

parte del impacto marginal del índice  $h_{ij}^p$ , lo que puede sugerir la presencia de correlación entre las dos variables. Algo análogo sucede con los ingresos medios de la microlocalidad en la estimación 6<sup>10</sup>.

Al estimar el modelo, excluyendo el indicador de la composición del capital humano e incluyendo el logaritmo del gasto educativo los resultados son los siguientes:

$$L_{ij}=12.2.-0.03Edad-0.50Origen-0.43Espadre_p-0.53Espadre_{si}-0.70Espadre_{bs}-0.93LnG$$

[18.1]**	[2.6]*	[4.4]**	[3.6]**	[3.6]**	[5.1]**	[16.1]**
<i>Ef. Mg.</i>	{-0.006}	{-0.12}	{-0.10}	{-0.13}	{-0.17}	{-0.23}

*McFadden's Adj R2: 0.175; Cragg-Uhler(Nagelkerke) R2: 0.292; McKelvey & Zavoina's R2: 0.299; N= 2379. (\*\*) significativas al 1% y (\*) al 5%. Valores absolutos del estadístico entre corchetes. Los valores entre llaves {} corresponden a los efectos marginales en la probabilidad*

El mismo ejercicio, pero ahora con los ingresos medios de la micro-localidad se resume a continuación:

$$L_{ij}=14.8.-0.04Edad-0.51Origen-0.29Espadre_p-0.45Espadre_{si}-0.68Espadre_{bs}-1.11LnY$$

[15.8]**	[3.6]*	[4.5]**	[2.4]**	[3.1]**	[4.9]**	[14.1]**
<i>Ef. Mg.</i>	{-0.009}	{-0.12}	{-0.07}	{-0.11}	{-0.17}	{-0.27}

*McFadden's Adj R2: 0.156; Cragg-Uhler(Nagelkerke) R2: 0.256; McKelvey & Zavoina's R2: 0.267; N= 2381. (\*\*) significativas al 1% y (\*) al 5%. Valores absolutos del estadístico entre corchetes. Los valores entre llaves {} corresponden a los efectos marginales en la probabilidad*

El efecto marginal en las probabilidades es ligeramente superior al encontrado en la regresión anterior.

Las regresiones (7) y (8) del [cuadro 3.2](#) exploran la incidencia del índice  $S_{ij}^g$ . Este indicador, como se explicó en secciones previas, intenta una aproximación a la estructura de preferencias relativas por educación de calidad desde la perspectiva del gasto. En la regresión (7) el efecto marginal sobre la probabilidad de que los niños y adolescentes estén matriculados en escuelas públicas es de -0.0012 por cada punto de

<sup>10</sup> Las pruebas de Pearson arrojaron una correlación alrededor de 0.5, significativa al 1%. Como se esperaba, las microlocalidades con alta proporción de personas calificadas también tienen ingresos elevados y mayores gastos educativos. Las de baja proporción de calificados cuentan con ingresos bajos y menores gastos educativos. Este resultado corrobora la presencia de segregación socio-económica en ámbitos espaciales.



aumento en el índice. El impacto es ligeramente inferior en la regresión (8) que incluye el indicador de pobreza.

El balance general de los resultados de los modelos explorados en este apartado permite concluir de manera parcial que efectivamente las variables asociadas al entorno familiar y local juegan un importante rol en la elección de escuelas públicas y privadas.

El conjunto de resultados refuerza la hipótesis de la presencia de segregación socioeconómica en esta estructura de decisiones. Los pobres asisten fundamentalmente a escuelas y colegios públicos, mientras que los que no tienen esta condición —de acuerdo con la línea de pobreza— asisten a planteles privados.

La variable relacionada con el origen de los padres también resultó significativa y con efectos marginales relativamente altos sobre la probabilidad de asistir a planteles públicos en el caso de padres de origen rural.

Tal y como se pudo apreciar, los resultados son plenamente consistentes con los planteamientos teóricos de secciones previas y corroboran la presencia de una modalidad de sorting socio-económico.

Las secciones que prosiguen se ocuparán de evaluar las diferencias en los logros académicos de los estudiantes y de auscultar los efectos que tienen el background familiar y la calidad de los entornos familiares, locales y de las escuelas, sobre los scores en las pruebas de competencias académicas.

Cuadro 3.2. Total Nacional Modelos Logit/Variable dependiente 1=Escuela Pública 0=Escuela Privada								
VARIABLES / Modelos	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Edad estudiante	-.059 (-5.83)**	-.0485 (-4.61)**	-.0429 (-4.03)**	-.023 (-2.07)*	-.027 (-2.37)*	-.036 (-3.27)**	-.059 (-5.68)**	-.051 (-4.82)**
<b>Background familiar</b>								
Origen de los padres(a)	-.567 (-5.32)**	-.475 (-4.37)**	-.452 (-4.11)**	-.451 (-3.91)**	-.468 (-4.08)**	-.472 (-4.16)**	-.552 (-5.10)**	-.499 (-4.56)**
espadre <sub>p</sub> (primaria completa)	-.523 (-4.61)**	-.369 (-3.20)**	-.332 (-2.84)**	-.344 (-2.86)**	-.373 (-3.12)**	-.246 (0.039)*	-.509 (-4.41)**	-.441 (-3.75)**
espadre <sub>si</sub> (secundaria incompleta)	-.980 (-7.26)**	-.567 (-3.94)**	-.452 (-3.09)**	-.326 (-2.13)*	-.402 (-2.67)**	-.316 (-2.12)*	-.992 (-7.27)**	-.807 (-5.76)**
espadre <sub>bs</sub> (bachilleres y superior)	-1.499 (-12.44)**	-.864 (-6.36)**	-.721 (-5.19)**	-.410 (-2.72)**	-.499 (-3.36)**	-.457 (-3.11)**	-1.532 (-12.50)**	-1.285 (-10.13)**
Pobreza según LP(b)			.727 (7.17)**	.503 (4.73)**				.883 (8.75)**
<b>Entorno local</b>								
Índice $h_{ij}^p/h_{ij}$ (c)		-.0299 (-9.33)**	-.0261 (-8.25)**	-.011 (-3.36)**	-.012 (-3.78)**	-.014 (-4.49)**		
Índice S de sorting(d)							-.005 (-2.31)*	-.004 (-1.93)*
Ln(Y per cápita) vecindad(e)						-.956 (-11.25)**		
Ln(gasto educ) vecindades(f)				-.802 (-12.67)**	-.849 (-13.54)**			
Constante	2.038 (12.72)**	2.002 (12.31)**	1.565 (9.04)**	10.360 (14.25)**	11.184 (15.71)**	12.993 (12.98)**	2.087 (12.56)**	1.547 (8.75)**
<b>McKelvey-Zavoina R2</b>	0.138	0.115	0.228	0.309	0.302	0.271	0.142	0.183
<b>McFadden's R2</b>	0.079	0.111	0.127	0.183	0.176	0.155	0.081	0.106
<b>Cragg-Uhler(Nagelkerke)R2</b>	0.144	0.196	0.221	0.305	0.295	0.264	0.149	0.189
<b>Observations (N)</b>	2381	2368	2368	2366	2366	2368	2321	2321

• Estimaciones propias a partir de la ECV de Colombia 2003

• \* Significativa al 5%; \*\* Significativa al 1% (Valores del estadístico z entre paréntesis)

• (a) Urbano=1, Rural=0; (b) Según Línea de Pobreza (DANE), Pobre=1, No pobre=0; (c) Índice de composición intra-vecindad de población activa calificada/no calificada; (d) Índice relacionado con la estructura del gasto educativo. Los índices (c) y (d) se expresan en la escala estandarizada [0,100], (ver los cálculos previos en la sección 3.3.1); (e) Ingreso per cápita a nivel de microlocalidad; (f) Gasto educativo medio de la micro-localidad en Logaritmos.

• La escolaridad de los padres utiliza como referencia “ningún nivel educativo o primaria incompleta”.

### 3.4. Calidad educativa y diferencias en logros

En la misma línea argumental trazada en la introducción de este capítulo se hizo alusión a que la educación no es un bien público puro y que más bien opera como un bien de club que puede ser excluyente en el consumo. Este argumento es aún más fuerte en entornos socio-económicos de alta segregación en los que los preceptos constitucionales o legales no logran garantizarse en la práctica, como es el caso de la gratuidad de la educación básica en Colombia suficientemente ilustrado en los capítulos 1 y 2.

Tal y como se pudo apreciar en la sección anterior los factores económicos y socio-demográficos relacionados con la calidad de los entornos familiares y locales que rodean las decisiones sobre la escolaridad de los niños y jóvenes, restringe el espacio de elecciones de los decisores en un mercado educativo en el que prolifera una amplia variedad de calidades.

En tal sentido se aludía que además de las restricciones estrictas de financiación, los estudiantes que provienen de entornos poco favorables enfrentan a menudo barreras de acceso al consumo de servicios educativo de buena calidad.

La variedad de calidades, tal y como se esbozaba ya en el modelo analítico del capítulo 2 (diagrama 2.1) se explica por las diferencias en los recursos disponibles y en las tecnologías de producción que utilizan los proveedores de los servicios educativos.

Es común encontrar en algunas investigaciones que comparan los logros escolares entre escuelas y colegios de primaria y secundaria en Colombia la siguiente afirmación: “*los planteles públicos son menos eficientes que los privados*”. Tal aseveración se apoya en el análisis de los resultados de las pruebas estándar de competencias en primaria y secundaria, denominadas pruebas SABER, o en las pruebas del Instituto Colombiano para el Fomento de la Educación Superior, ICFES, aplicadas al final del bachillerato y que, a su vez, se utilizan como exámenes de selectividad para ingresar a la educación superior.

Trabajos como los de [Barrera y Gaviria \(2003\)](#) y el de [Gaviria y Barrientos \(2001\)](#) afirman que las diferencias en la calidad educativa entre planteles públicos y privados

en Colombia se explican por las fallas en los mecanismos de incentivos y por los problemas asociados a la estructura organizacional del sector público, más que por los problemas de disponibilidad de recursos.

Así mismo, [Nuñez, Steiner, Cadena y Pardo \(2002\)](#) en su investigación sobre las diferencias entre colegios públicos y privados con información del ICFES del año 1999, hallan que los alumnos de los segundos superan en su desempeño a los primeros, especialmente en los niveles de ingresos medios y altos.

Estos resultados son contradictorios con los hallazgos de una investigación previa, [Piñeros y Rodríguez \(1998\)](#), en la que abordaron el estudio de los factores individuales y escolares que determinan el rendimiento académico de los estudiantes de secundaria para el año 1997 mediante técnicas multinivel. Los autores encontraron que el nivel socioeconómico afecta positivamente el rendimiento académico, que el tiempo de desplazamiento hacia la escuela tiene un efecto inverso sobre el rendimiento y al controlar la incidencia del nivel socioeconómico sobre el rendimiento académico los puntajes de los colegios oficiales en las diferentes áreas superan a los de los colegios privados.

Esta misma investigación detectó que los factores asociados a las escuelas tienen un efecto pequeño y significativo sobre el rendimiento, explican entre el 15% y el 18% de la varianza del rendimiento de los estudiantes de colegios privados y entre el 12% y el 16% de la varianza del rendimiento de los estudiantes de colegios oficiales.

De acuerdo con el marco analítico utilizado hasta el momento en esta Tesis, la interacción entre los factores familiares, locales y las restricciones de liquidez, determina la estructura de decisiones de enrolar a los estudiantes en establecimientos públicos o privados y que en consecuencia la probabilidad que tienen los estratos socioeconómicos medios y altos de matricular a sus hijos en escuelas privadas de buena calidad es relativamente alto.

Aunque el trabajo de [Gaviria y Barrientos \(2001\)](#) antes mencionado reconoce que la educación de los padres afecta el rendimiento académico de manera significativa y que, en el caso de los planteles privados, la educación promedio de los profesores y el

número de docentes por alumno están asociados positivamente con la calidad, insisten en que los aumentos del gasto público en educación impulsado por las reformas de descentralización de los servicios educativos no han repercutido positivamente sobre el desempeño debido a las rigideces de la estructura organizacional y de incentivos de la educación pública.

Las prescripciones de política que rápidamente se desprenden de estos enfoques apuntan a que para eliminar estas diferencias es preciso revertir los mecanismos de financiación y avanzar hacia un modelo de *vouchers* o de financiación directa por el lado de la demanda que fomente la competencia entre escuelas públicas y privadas y que brinden amplitud de escogencia a las familias<sup>11</sup>. Previamente se anotaron los riesgos a los que puede conducir un esquema de financiación como el propuesto y las repercusiones que puede tener en la profundización de la segregación socio-económica.

En sociedades tan heterogéneas como la colombiana, afirmar que las escuelas privadas son más eficientes que las públicas o que, literalmente, “*los colegios públicos son relativamente más eficientes para los alumnos que provienen de los estratos socio-económicos más bajos de la población*”<sup>12</sup> (Nuñez, Steiner, Cadena y Pardo (2002, pag. 34), no deja de generar algún tipo de desencanto en relación con los hallazgos que se han obtenido hasta el momento en esta Tesis.

Algunas consideraciones adicionales podrían ayudar a dilucidar tal aseveración.

En el sentido de [De Fraja \(2001\)](#), los costes de producción del bien educación dependen de las características de los consumidores: los *clubes* que se configuran proveen

---

<sup>11</sup> Valga anotar que en el sector salud, la Ley 100 promulgada a principios de la década de los noventa ya adoptó un esquema de financiación por el lado de la demanda, prácticamente idéntico al sistema de *vouchers* propuesto para el sector educativo. El balance después de más de una década permite afirmar que a pesar de la ampliación de la cobertura en los servicios de salud, los impactos sobre la calidad del servicio (suministrado por las denominadas Entidades Promotoras de Salud, EPS de carácter privado y que compiten con el Instituto Colombiano de los Seguros Sociales, ISS que hasta ese momento ostentaba el monopolio del servicio) han sido nefastos y adicionalmente, el ISS se encuentra en una fase casi terminal por efecto del desplazamiento de los usuarios hacia las EPS privadas. Tomando esta ilustración como simple analogía, los argumentos iniciales de que la participación del sector privado garantizaría la mayor cobertura y calidad en los servicios de salud es ahora dudosa para el sector educativo de acuerdo con esta experiencia.

<sup>12</sup> Lo que parece sugerir por deducción que los colegios privados son relativamente más eficientes para los estudiantes que provienen de estratos socio-económicos medios o altos.

servicios con características de un bien público que genera externalidades excluyentes con magnitud y signo no anónimas (“*club with non-anonymous crowding*”). La internalización de tal externalidad se realiza a través del mecanismo de precios.

Una manera elegante de estilizar estos hechos, bajo el supuesto de que las escuelas públicas son de menor calidad que las privadas, se desprende del modelo de [Epple y Romano \(1998\)](#) y de la exposición de [De Fraja \(2001\)](#).

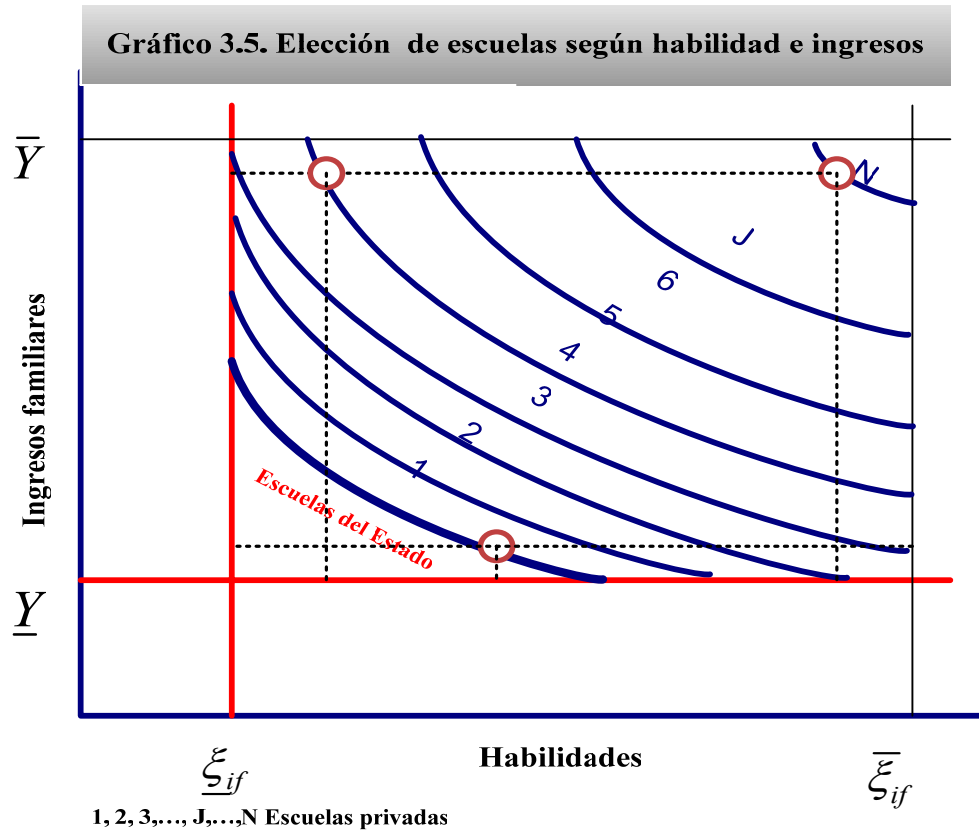
Cada individuo  $i$  que pertenecen a la familia  $f$  se diferencia por sus habilidades innatas ( $\xi_{if}$ ) y los ingresos familiares ( $Y_{if}$ ). Estos dos atributos toman valores en el rango  $[\underline{\xi}_{if}, \bar{\xi}_{if}]$  y  $[\underline{Y}, \bar{Y}]$ , las barras inferiores y superiores corresponden a las cotas inferiores y superiores, respectivamente.

El equilibrio que resulta del ejercicio de optimización conduce a un ordenamiento jerárquico en el que la combinación  $[\underline{\xi}_{if}, \bar{\xi}_{if}].[\underline{Y}, \bar{Y}]$  representa las características de cada estudiante con habilidad e ingreso familiar y cada punto en el espacio cartesiano  $(\xi_{if}, Y_{if})$  indica los perfiles de las escuelas a las que asisten: la calidad de la escuela es mayor en dirección nor-este y menor en la dirección contraria como lo muestra el [gráfico 3.5](#).

El modelo implícito en el diagrama admite la existencia de restricciones de liquidez.

Debido a que los gastos educativos,  $E_{it+1}^*$ , dependen positivamente de las habilidades innatas,  $\xi_{it+1}$ , y de los ingresos familiares,  $Y_{it}$ , cuando las familias presentan limitaciones de recursos para invertir en la educación de sus hijos, la única oportunidad para poder asistir a escuelas de alta calidad educativa es que sus habilidades excedan determinados umbrales.

$$(3.9) \quad \begin{aligned} E_{it+1}^* &= f(\xi_{it+1}, Y_{it}) \\ \partial E / \partial \xi &> 0 \\ \partial E / \partial Y &\geq 0 \end{aligned}$$



En cambio, si las familias son ricas aunque sus hijos sean menos habilidosos, probablemente obtengan certificación en escuelas de alta calidad en razón de que la dotación de recursos familiares podría compensar su ausencia de habilidades suficientes, además de que se beneficiarían de las externalidades generadas por algunos compañeros de curso que no necesariamente pertenecen a su clase social.

Este último argumento es el que sustenta el probable sesgo de sobre-representación de estudiantes provenientes de familias ricas en los colegios de mayor calidad. Se aprecia que sin restricciones de liquidez el ordenamiento sería estrictamente jerárquico, entonces las escuelas públicas concentrarían a los estudiantes de más baja habilidad innata y las privadas a los más capaces. Esto es, si no existiesen barreras de entrada, a cualquier nivel de ingreso los individuos más capaces irían a los colegios privados de mejor calidad y los menos capaces tenderían a ser atraídos por los colegios del Estado. En este caso no existiría ningún problema de sobre-representatividad de los individuos ricos y poco hábiles en escuelas privadas y tampoco se excluiría a individuos pobres y capaces de asistir a los colegios de mejor calidad.

Si existiese algún mecanismo perfecto de subvención (becas, subsidios, ayudas de manutención, o tarifas preferenciales según destrezas, etc.) a los individuos habilidosos provenientes de hogares pobres, tampoco se debería presentar problemas de sobre-representatividad en los extremos de la distribución.

De este modo, los resultados de las pruebas estándar deberían ser también estrictamente jerárquicos y capturarían de manera límpida la proficiencia real de los individuos, además de que los estudiantes más habilidosos generarían fuertes externalidades sobre el resto de compañeros.

Ahora bien, una consideración complementaria con lo anterior se deriva del modelo de [Stiglitz \(1974\)](#) en relación con la propensión al voto por alta o baja financiación pública.

Entre las posibilidades de múltiples equilibrios en el mercado educativo que se desprende de su modelo ya identificaba la posibilidad de que las familias ricas propugnaran por niveles bajos de financiación estatal (propensión por tasas impositivas bajas) debido a que preferirían enviar a sus hijos a escuelas y colegios privados de alta calidad.

Se infiere que las clases medias, en cambio, se inclinarían hacia la provisión mixta de los servicios educativos con un amplio rango de escogencia de calidades y precios, de acuerdo con su estructura de preferencias. La tensión que se deriva de este enfoque quedó cuidadosamente consignada en los trabajos posteriores de [Epple y Romano \(1998\)](#) y [Fernandez y Rogerson \(1997, 1998\)](#) en sus modelos de multicomunidad.

Un corolario importante que surge desde estas diferentes perspectivas apunta a las asimetrías en la provisión de calidades educativas públicas y privadas. El argumento es que el sector privado posee una mayor capacidad y habilidad para ofrecer las cantidades y calidades deseadas del bien educación, en contraste con la rigidez del sector estatal (ya sea centralizado o descentralizado) en el que el menú de calidades es limitado.

No obstante, cuando la oferta privada es demasiado amplia y variada esta singularidad en la provisión no conduce necesariamente en todos los estratos socioeconómicos a que la educación privada sea estrictamente mejor que la educación pública debido a que las



familias podrían escoger —en particular las clases medias—una combinación calidad/precio que las satisfaga pero que no necesariamente proporcione mejor calidad.

Tal y como se anotó en la introducción, en el caso colombiano es claro que en los estratos socio-económicos pobres la mixtura de educación pública y privada corrientemente está asociada con niveles bajos de calidad/precio. En los estratos medios la mixtura es todavía más compleja y la relación calidad/precio no es unívoca. En los estratos más altos, en cambio la relación sí es prácticamente unívoca, los colegios privados exhiben los mejores resultados en las pruebas de competencias en lengua materna, matemáticas, ciencias naturales e idiomas foráneos, cuentan con amplios flujos de recursos, los costes de acceso son casi prohibitivos para el resto de estratos socio-económicos y adicionalmente generan mecanismos soterrados de cribaje mediante los denominados “*bonos voluntarios*”<sup>13</sup> o prácticas directas de discriminación social por origen, color de piel o apellido que garantizan la persistencia intergeneracional de un arreglo social elitista.

Una mirada a los datos de la ECV 2003 sugiere que en el interior de cada estrato socio-económico y vecindad intra-urbana coexiste esta mixtura entre escuelas públicas y privadas, el peso de estas últimas, como se esperaba, aumenta con los deciles de ingresos medios de los hogares.

El [cuadro 3.3](#) clasifica a los estudiantes de acuerdo con las variables promedio de las vecindades de cada área metropolitana en la que se localizan.

Los hogares del primer decil, en cada sección muestral del área metropolitana de Bogotá, gastan en promedio un 10.7% de sus ingresos en educación. El 68.1% de los estudiantes asisten a escuelas y colegios públicos.

Para el último decil esta relación se invierte, el 67.7% asisten a escuelas y colegios privados, presumiblemente de mejor calidad.

---

<sup>13</sup> Que en la práctica constituyen cláusulas de obligatoriedad de pagar sumas millonarias para el acceso.

Cuadro 3.3. Clasificación de estudiantes según deciles						
Deciles(*)	A	B	C	A	B	C
	A.M. BOGOTÁ			A.M. CALI		
D1	10,7	68,1	31,9	8,4	72,5	27,5
D2	13,4	59,1	40,9	10,3	64,3	35,7
D3	9,8	62,7	37,3	13,3	63,7	36,3
D4	10,9	70,1	29,9	14,5	59,7	40,4
D5	10,1	60,4	39,6	11,3	52,6	47,4
D6	11,9	56,6	43,5	13,8	56,9	43,1
D7	11,1	51,7	48,3	12,9	61,1	38,9
D8	11,1	48,9	51,2	12,1	52,3	47,8
D9	14,6	39,6	60,4	12,6	46,6	53,4
D10	13,3	32,3	67,7	10,8	54,4	45,6
	A.M. MEDELLIN			CARIBE		
D1	22,2	78,5	21,6	10,6	63,2	36,9
D2	10,5	84,4	15,6	13,2	65,4	34,6
D3	12,8	81,8	18,2	8,9	61,8	38,3
D4	17,5	68,9	31,1	12,2	59,2	40,8
D5	13,0	85,1	14,9	9,9	57,9	42,1
D6	14,7	55,1	44,9	11,4	54,0	46,0
D7	14,0	67,0	33,0	11,2	48,0	52,0
D8	14,6	60,0	40,0	13,2	48,3	51,7
D9	15,9	49,4	50,6	13,7	40,6	59,4
D10	18,3	32,9	67,1	14,4	33,9	66,1
	TOTAL MUESTRAL					
	12,6	60,8	39,2			

(\*) Estimaciones propias con la base microdatos de la ECV 2003 (Tamaño muestral 3447 estudiantes)

D: Deciles de ingresos de los hogares a los que pertenecen los estudiantes

A= (Gasto educativo/Ingreso total hogares)%

B= %Estudiantes en escuelas y colegios publicos por micro-localidad

C= %Estudiantes en escuelas y colegios privados por micro-localidad

Es preciso advertir que el último decil puede presentar algún problema de censura estadística, debido a que la cola más alta es frecuentemente depurada por razones de reserva estadística o algunos valores reciben el tratamiento de *outliers*.

Resultados análogos al encontrado para Bogotá se pueden apreciar para las demás áreas metropolitanas. En la A.M. de Medellín el decil más bajo envía a sus hijos a escuelas públicas, en Cali el 72.5 y en la costa Caribe el 63.2. El último decil prácticamente invierte la relación.

### 3.4.3. Rendimiento académico con datos del ICFES

Infortunadamente la base de microdatos de la ECV 2003 utilizada en varios apartados de esta Tesis, por su naturaleza no cuenta con variables de rendimiento, lo que impide contrastar directamente un modelo de logros educativos según estratos socio-económicos.

Sin embargo, los reportes estadísticos agregados<sup>14</sup> de las pruebas del ICFES brindan algunos indicios sobre las hipótesis centrales de este apartado. La descripción de esta fuente se puede apreciar en el [recuadro 3.1](#).

El examen de los promedios históricos muestra que entre el 2000 y el 2006 los resultados globales mejoraron respecto a los años anteriores. Pero, más de la mitad de las instituciones públicas y privadas se ubican en la categoría media.

En Bogotá se identifican la mayoría de instituciones ubicadas en las categorías altas de competencias, mientras que regiones con menor grado de desarrollo relativo, Chocó, Amazonas y Vaupés, presentan el mayor número de instituciones educativas en los niveles inferiores de competencias.

La información disponible en el portal <http://www.icfesinteractivo.gov.co/><sup>15</sup> muestra algunos progresos al evaluar los resultados de las pruebas de matemáticas, lenguaje e inglés para los años 2000, 2003 y 2006 por planteles educativos públicos y privados en Bogotá, Cali y Medellín. Pero, la mayoría de instituciones todavía mantienen niveles relativamente modestos tal y como se puede apreciar en el [cuadro 3.4](#) de los promedios y desviaciones según percentiles.

---

<sup>14</sup> Las bases de microdatos SABER y la del ICFES no pudieron utilizarse debido que el Instituto Colombiano para el Fomento de la Educación Superior no respondió a las diversas solicitudes que se realizaron a lo largo de este proceso de investigación. Así mismo, los microdatos de los formularios C-600 que proporcionan información del Censo a Establecimientos de Educación en Colombia tampoco pudieron utilizarse pese a las múltiples solicitudes enviadas a la unidad metodológica y de producción estadística del Departamento Administrativo Nacional de Estadística DANE sin obtener respuesta alguna. Llama la atención que estos microdatos que aparecen referenciados y utilizados en diversos estudios de entidades privadas de investigación en Colombia y en varios estudios realizados por investigadores de universidades privadas de Santafé de Bogotá no sean de dominio público.

<sup>15</sup> La muestra disponible para estas tres ciudades consta de 2447 planteles públicos (847) y privados (1600). La información utilizada en esta sección corresponde al promedio por plantel en las pruebas de matemáticas, lenguaje e inglés para los años 2000, 2003 y 2006. También se dispone de la desviación estándar y del número de estudiantes evaluados en cada prueba.

Los [gráficos 3.6](#) y los [anexos A3.4](#) muestran las matrices de dispersión de la muestra total de colegios para las tres pruebas en los años seleccionados.

Los mayores avances se observan en las pruebas de competencias en lengua inglesa obtenidos por colegios privados, aunque con elevados niveles de dispersión. Los colegios que logran superar la banda superior en el año 2006 (puntajes promedio mayores a 70 en el eje vertical, [anexos A3.4](#)) son pocos y corresponden a colegios bilingües orientados a estudiantes de estratos socio-económicos altos<sup>16</sup>.

Las pruebas de matemáticas en Bogotá, para el 2006 muestran una alta concentración de colegios públicos y privados alrededor del promedio teórico (50 puntos).

Al examinar las cifras del [cuadro 3.4](#) queda claro que las instituciones privadas ofrecen la mejor educación del país, aunque la mayoría sigue presentando resultados muy deficientes.

En los últimos años cerca del 70% de estos establecimientos se ubicó en las categorías media, baja e inferior, lo cual sugiere que el suministro de la educación por parte del sector privado no siempre es garantía de calidad. Estos resultados ponen en entredicho el reclamo creciente de una política de privatización de los servicios educativos en Colombia en la medida que no necesariamente aseguraría mejores resultados.

Los diagramas cruzados de los promedios y las desviaciones por asignatura evaluada y ciudad de los anexos A3.4<sup>a</sup>-A3.4<sup>c</sup> identifican algunos colegios privados que se acercan a la banda superior y otros, también privados, que presentan promedios muy bajos, incluso inferiores a los colegios oficiales y con elevados niveles de dispersión.

El caso de Cali es diferente en estas mismas pruebas, las desviaciones estándar son mayores que en Bogotá y se identifican colegios privados que se alejan de los centroides de la distribución.

---

<sup>16</sup> Los estudiantes que provienen de colegios que imparten lengua alemana y francesa tienen la prerrogativa adicional de poder presentar en el componente flexible las pruebas de alemán o francés en lugar de la prueba de inglés.

Cuadro 3.4. Colombia: promedios y desviaciones en las pruebas del ICFES

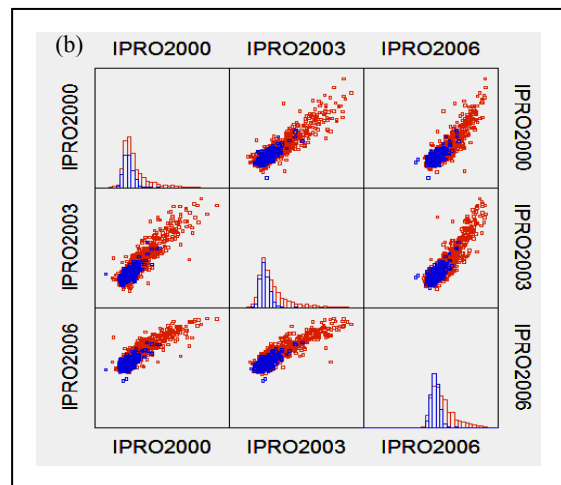
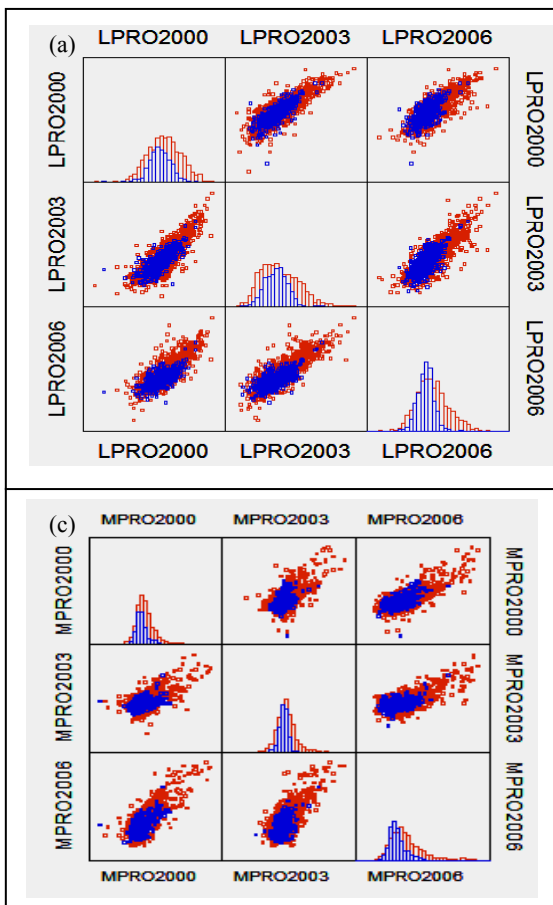
Percentiles	PRIVADO					PUBLICO					PRIVADO					PUBLICO					
	P5	P25	P75	P95	P99	P5	P25	P75	P95	P99	P5	P25	P75	P95	P99	P5	P25	P75	P95	P99	
<b>BOGOTA</b>																					
Promedios											Desviaciones Estándar										
MATEMAT	2000	40,7	42,1	43,8	46,2	48,9	41,1	41,9	43,1	44,4	46,0	3,7	4,5	5,3	6,2	8,2	4,1	4,6	5,2	5,7	6,3
	2003	40,0	41,5	43,5	45,9	48,9	40,5	41,5	42,5	43,3	44,7	3,7	4,6	5,7	7,0	8,6	4,0	4,6	5,3	6,0	6,8
	2006	41,2	44,0	48,9	55,0	61,1	41,9	43,5	45,8	48,4	52,1	5,2	6,9	8,6	10,0	12,1	6,4	7,1	8,1	9,2	10,1
LENGUAJE	2000	44,8	47,0	51,1	54,2	56,1	45,3	46,8	49,4	51,4	52,8	4,0	4,8	5,9	7,0	8,0	4,6	5,1	5,8	6,2	6,7
	2003	46,3	49,2	54,9	60,2	64,1	46,8	49,2	52,4	54,4	56,5	5,0	6,2	7,5	8,8	11,6	5,6	6,1	7,2	7,8	8,4
	2006	46,1	48,4	53,1	57,6	60,7	45,9	47,6	49,8	51,8	53,9	4,3	5,5	6,7	8,0	9,7	5,2	5,7	6,5	7,1	7,8
INGLES	2000	38,6	40,2	44,9	55,0	61,3	38,6	39,9	42,1	44,4	49,7	3,8	4,6	6,8	9,5	10,8	4,0	4,6	5,4	6,7	8,6
	2003	39,0	40,6	45,9	57,5	65,2	39,1	40,2	42,2	44,7	50,4	3,4	4,5	7,6	10,6	13,4	3,7	4,3	5,6	6,8	8,3
	2006	39,7	42,8	51,1	63,6	70,1	39,0	41,2	44,4	47,6	52,4	4,9	6,7	9,5	12,2	14,8	5,4	6,2	7,7	9,0	11,5
<b>CALI</b>																					
MATEMAT	2000	41,4	43,1	46,0	51,8	55,5	41,3	43,4	45,7	47,3		4,0	4,9	6,5	8,4	15,0	4,5	5,1	6,0	8,0	
	2003	38,6	40,4	43,1	47,0	50,7	40,3	41,2	42,4	43,2		3,4	4,4	5,7	7,5	9,3	3,9	4,5	5,4	6,3	
	2006	44,8	47,3	53,1	65,2	68,2	46,3	47,8	50,3	52,1	55,3	4,9	6,3	8,7	11,6	13,8	6,2	6,8	8,3	9,7	11,0
LENGUAJE	2000	41,0	43,2	46,9	50,8	53,5	41,5	44,1	46,4	47,7		4,1	5,1	6,3	7,9	12,7	4,8	5,3	6,0	7,2	
	2003	43,0	45,1	48,6	54,2	57,5	44,1	46,0	48,1	49,6		3,5	4,6	6,0	7,0	9,0	4,2	5,0	5,8	6,5	
	2006	42,8	45,1	50,5	57,0	60,7	42,5	45,3	48,4	51,3	52,6	4,4	6,0	7,5	9,3	11,8	5,5	6,6	7,5	8,3	9,4
INGLES	2000	37,4	38,9	42,3	58,8	68,4	37,4	39,0	41,2	43,3		3,6	4,6	7,2	10,3	15,1	4,2	4,8	6,1	7,9	
	2003	37,2	39,3	43,5	55,4	62,2	37,4	39,7	41,8	43,7		3,3	4,5	6,7	8,1	9,3	4,1	4,8	6,0	6,9	
	2006	36,5	38,7	47,7	65,5	75,3	36,4	38,2	41,1	45,2	48,9	4,4	6,2	9,6	11,9	14,5	4,6	6,2	8,1	8,9	12,7
<b>MEDELLIN</b>																					
MATEMAT	2000	41,1	42,2	44,3	46,8	49,6	41,2	41,9	42,7	43,7	44,8	3,7	4,7	5,5	6,6	11,0	4,2	4,6	5,1	5,8	8,2
	2003	39,4	41,0	43,5	47,0	49,8	39,4	40,7	41,7	42,8	43,3	4,0	4,7	5,8	7,2	10,6	3,9	4,4	5,3	5,9	7,9
	2006	41,3	43,7	49,3	57,7	62,7	41,6	43,2	45,6	48,4	51,2	5,0	6,9	8,9	10,7	13,1	5,8	6,9	8,2	9,4	10,8
LENGUAJE	2000	44,0	46,2	51,3	54,0	55,6	43,5	45,1	47,2	49,8	51,0	4,1	4,8	5,8	7,0	10,0	4,7	5,3	6,0	6,6	7,1
	2003	44,7	47,1	55,9	59,4	62,7	44,2	46,9	50,4	52,6	54,6	4,9	6,1	7,6	9,3	11,6	5,3	6,2	7,2	8,1	10,6
	2006	44,8	47,4	52,8	57,5	59,4	45,5	47,3	49,7	51,6	53,1	4,5	5,6	6,8	7,5	8,1	4,9	5,6	6,5	7,2	8,6
INGLES	2000	37,4	39,8	46,6	57,3	67,2	37,6	38,6	40,4	43,0	45,1	4,3	5,5	8,4	10,5	13,1	3,9	4,7	5,9	7,4	9,4
	2003	38,5	39,9	47,9	59,7	68,3	38,1	39,2	40,6	42,8	45,4	3,7	5,0	8,8	10,9	16,7	3,8	4,4	5,5	6,5	8,3
	2006	37,5	40,6	53,4	66,7	72,0	37,5	39,7	43,3	46,7	51,6	5,2	6,8	10,1	12,4	13,1	5,3	6,2	7,8	9,8	19,7

Fuente: cálculos propios a partir de la información del ICFES, <http://www.icfesinteractivo.gov.co/>

De acuerdo con el ICFES los rangos de puntajes son Bajo<30, Medio 31-70 y Alto >70

La escala es 0-100 con desviación 10

El ICFES define como valor teórico 50 y una desviación de 10



Gráficos 3.6(a) Lenguaje; 3.6(b) Inglés y 3.6(c) Matemáticas

-----Públicos  
-----Privados

LPRO = promedio en las pruebas de lenguaje  
IPRO = promedio en las pruebas de inglés  
MPRO= promedio en las pruebas de matemáticas  
(Escala ICFES 0-100)

Fuente: elaboración propia con la muestra total proveniente de <http://www.icfesinteractivo.gov.co/>

De acuerdo con los apartados precedentes de este capítulo la dispersión de los resultados aporta evidencia sobre la amplia variedad de calidades educativas en Colombia. La revisión de las cifras por plantel muestra, por ejemplo, que 36 colegios obtuvieron puntajes de matemáticas para el año 2006 en el rango más alto observado 60-70 puntos: 35 corresponden a colegios privados de coste elevado y solamente 1 colegio público localizado en Bogotá logró ubicarse en este rango. Los colegios privados a pesar de que en promedio exhiben resultados superiores a los establecimientos públicos presentan scores medios que no son competitivos a nivel internacional tal y como se pudo apreciar en la revisión preliminar realizada en el capítulo 1 de este estudio.

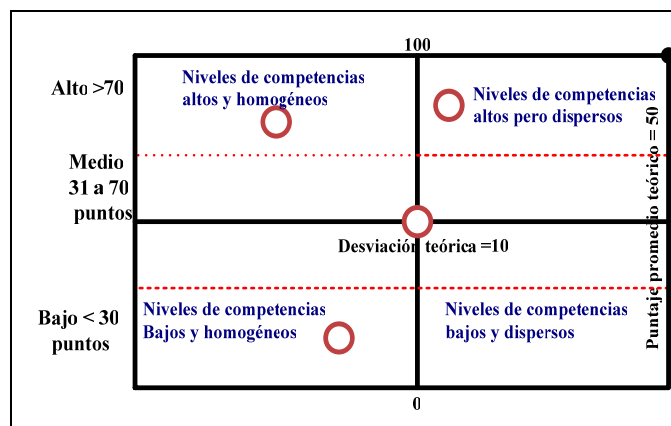
### Recuadro 3.1. Características de las pruebas del ICFES

El examen del Estado se aplica a todos los estudiantes del último grado de bachillerato y lo realiza el Instituto Colombiano para el Fomento de la Educación Superior ICFES. La prueba evalúa las competencias en las diferentes áreas (matemáticas, lenguaje, ciencias naturales, idiomas foráneos, biología, etc.) y configura uno de los requisitos para acceder a la educación superior. Su utilidad adicional es la de permitir realizar procesos de evaluación de la calidad en las instituciones educativas de todo el país.

El enfoque metodológico de las pruebas implica un dominio significativo del saber impartido en todo el ciclo educativo básico y medio, se hace especial hincapié en la comprensión, la inferencia y la deducción, la capacidad de análisis y la utilización oportuna y pertinente de conceptos. Las pruebas se aplican cada año en los calendarios académicos A (enero-noviembre) y B (septiembre-junio) adoptado en cada región del país.

Las pruebas se dividen en un núcleo común y en un componente flexible. Los puntaje se presentan en los reportes de resultados en la escala 0 a 100 representando así el nivel de competencia en cada prueba. Estos reportes proporcionan información sobre los niveles promedios por planteles, municipios y departamentos y definen los siguientes rangos en la escala nacional:

Rango de puntaje	Puntaje de la prueba
• Bajo	<30
• Medio	31 a 70
• Alto	>70



El ICFES define como valores teóricos un valor medio de 50 puntos y una desviación estándar de 10. Los niveles de competencia se clasifican a su vez en (A, B, C), siendo C el mayor nivel de competencia y A el menor. Los valores históricos de los reportes se pueden consultar en la página web <http://www.icfesinteractivo.gov.co/historicos/>

### 3.5. Calidad educativa, entornos de aprendizaje y competencias

Esta sección profundiza en el estudio de las implicaciones de los entornos de aprendizaje sobre la calidad educativa, los logros y las competencias académicas de los estudiantes a partir de la base estadística de las pruebas PIRLS. La primera parte introduce algunas consideraciones conceptuales y luego se analiza el rendimiento de los estudiantes y los factores que lo determinan.

La controversia suscitada desde los primeros reportes del denominado informe [Coleman \(1966\)](#) para los Estados Unidos generó inquietudes de suma relevancia a la sociología y a la economía de la educación sobre el rol que juegan los entornos familiares y locales en el rendimiento escolar de los niños y jóvenes. El énfasis inicial del reporte en la segregación racial y en el contexto socioeconómico, además de la disponibilidad de recursos oficiales para los planteles educativos, implicaba atribuir una responsabilidad importante al background familiar en la explicación de las diferencias en el rendimiento. La controversia siguió su curso y después de varias décadas de revisiones, críticas y esfuerzos por desentrañar los posibles errores metodológicos del informe del 66 --tal y como lo señala Barbara Heyns<sup>17</sup>-- lo cierto es que este trabajo alteró el curso de los enfoques sociológicos y logró permear el discurso económico predominante hasta entonces en el campo de la economía de la educación. En tal sentido y en palabras de [Albright y Conley \(2004\)](#), la ignición que el reporte inició todavía continúa y ha sido difícil de extinguir<sup>18</sup>.

El marco analítico introducido en el capítulo 2 de esta memoria de Tesis llamaba la atención sobre estos aspectos mediante un enfoque integrado en el que se destacaba la importancia de los entornos escolares para explicar los logros y realizaciones de los estudiantes (grados alcanzados, el desempeño en las pruebas estándar nacionales e internacionales y en el rendimiento académico en general). Estas características interactúan con la calidad educativa de los entornos locales y los perfiles de instrucción de los padres, así como con la esfera gubernamental que traza los lineamientos básicos de la política educativa, diseñando mecanismos compensatorios a los grupos sociales de

---

<sup>17</sup> B. Heyns, "Educational effects: issues in conceptualization and measurement". En J. C. Richardson, *Handbook of Theory and Research in the Sociology of Education*, New York: Greenwood, 305-340; Hallinan, citado por *Karen Albright and Dalton Conley (2004)* "American educational policy in historical perspective". En [Albright y Conley \(Eds\) \(2004\)](#)

<sup>18</sup> Ibid, *Introduction*, p.11.

mayor vulnerabilidad e impulsa programas para el mejoramiento de la calidad. Todos estos componentes, de manera conjunta con las destrezas innatas de los escolares, sus hábitos de estudio, disciplina y sus experiencias tempranas en lecto-escritura antes de la iniciación de su ciclo formal de escolaridad inciden en su desempeño.

Para lograr el propósito trazado en este apartado y dilucidar los factores de mayor relevancia en el rendimiento, la mirada ahora se dirige hacia los estudiantes, a sus scores o los resultados de proficiencia y la interacción con los factores antes mencionados. La base estadística utilizada en este apartado corresponde a las pruebas PIRLS (Progress in International Reading Literacy Study) de la IEA<sup>19</sup> aplicada en el año 2001 a 35 países entre los que salió seleccionada Colombia. La muestra consta de 5131 escolares de grado 4° de primaria, distribuidos en 147 planteles educativos y 196 docentes.

### 3.5.1. Categorías básicas y aspectos metodológicos

#### Definiciones previas

Las pruebas PIRLS intentan esclarecer cuáles son los factores que inciden en las **competencias lectoras** de los niños, considerándola como destreza esencial que configura la base de su futuro procesos de aprendizaje, disfrute e interacción social.

De acuerdo con el marco conceptual y la especificación de la prueba<sup>20</sup>, la competencia lectora corresponde a la habilidad adquirida por los individuos para comprender y utilizar formas lingüísticas en el proceso de interacción con los demás y con los textos que proporcionan información, conocimientos y disfrute.

La prueba evalúa así las destrezas de los niños en los *procesos de comprensión*, en los *propósitos de la lectura*, en sus *conductas y actitudes* hacia el acto de leer. Los *procesos de comprensión* incluyen la identificación y obtención de información, la capacidad para realizar inferencias directas, así como el trabajo de interpretación e integración de ideas. Los *propósitos de la lectura* consideran la capacidad para desentrañar

---

<sup>19</sup> International Association for the Evaluation of Educational Achievement.

<sup>20</sup> Martin, Michael O, Ina V.S. Mullis, y Eugene J. Gonzalez (2004). *International Achievements in the Processes of Reading Comprehension*. International Study Center, Lynch School of Education, Boston College, USA.



información y la experiencia en el manejo de textos literarios o la relación con estructuras narrativas. En las *conductas y actitudes ante la lectura* se explora el ejercicio de la lectura como disfrute u obligación.

Los diversos cuestionarios que configuran la prueba se dirigen a dilucidar cada uno de estos propósitos desde la perspectiva del estudiante, de su entorno escolar y de los profesores.

La elección de niños de 4º de primaria obedece al estadio de transición en el que se encuentran. Estadio que supone que ya han aprendido a leer y la lectura se convierte en el principal vehículo de aprendizaje en todas sus asignaturas. El denominado por PIRLS como el contexto para el desarrollo de la competencia lectora es el resultado de la interacción entre el *hogar*, la *escuela* y el *aula* (ver [gráfico 3.7](#)). Como se puede observar este enfoque conceptual presenta gran proximidad con el modelo analítico planteado desde el capítulo 2 de esta Tesis.

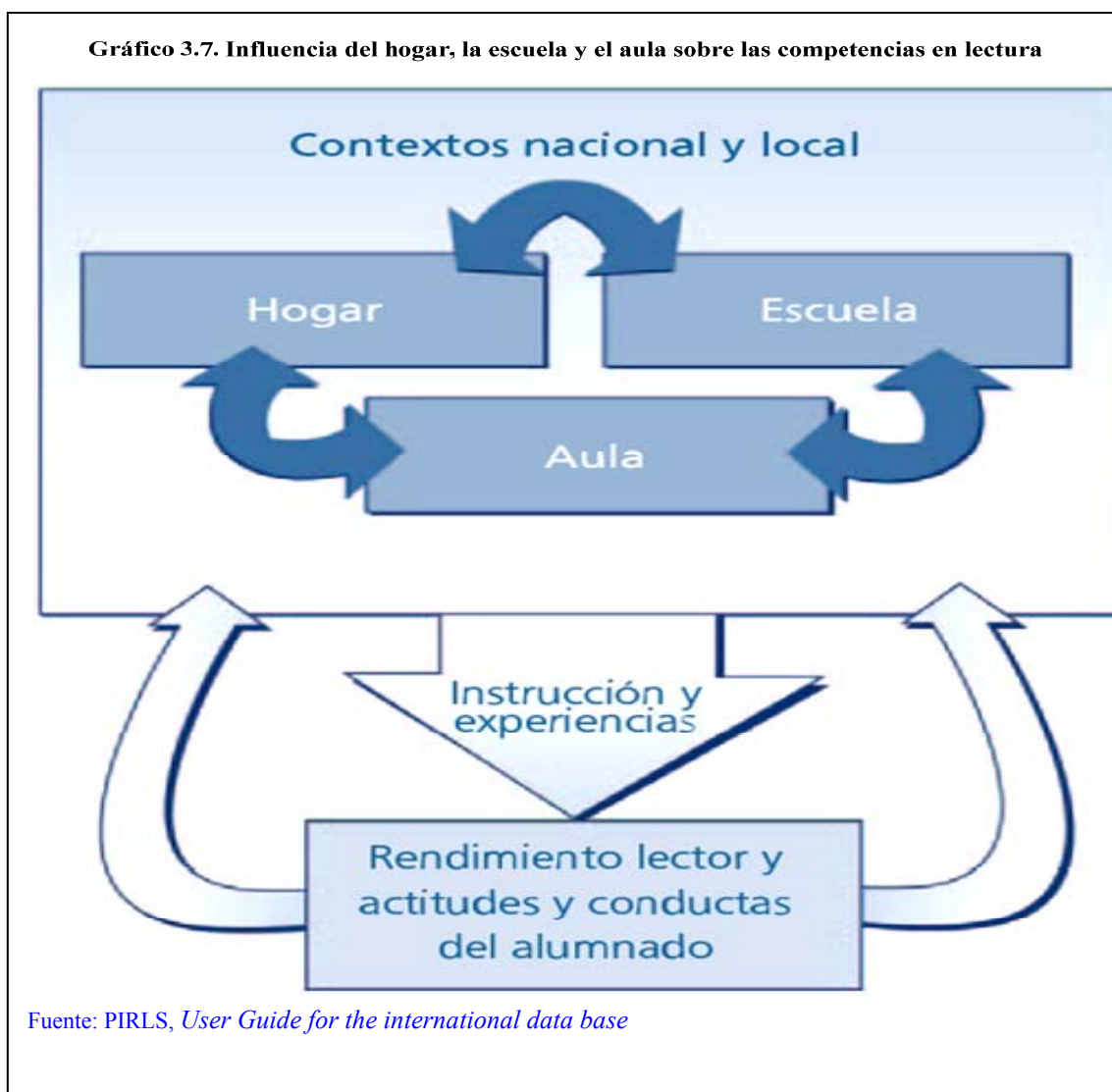
Los progresos en las competencias de lectura están afectados por los siguientes factores:

**Nacionales:** entre los que se destacan las acciones del gobierno y la organización del sistema educativo, el diseño y las políticas curriculares.

**Locales:** los recursos sociales y culturales en el entorno inmediato.

**Familiares:** las actividades alrededor de la lectura en el entorno próximo al estudiante, las lenguas habladas en casa, los recursos y el status socioeconómico, la disponibilidad de libros y material impreso en casa, las relaciones entre la familia y la escuela, las actividades extraescolares en el hogar.

**Escuela:** las prácticas curriculares, el entorno escolar y de recursos, la calidad de los recursos disponibles, la calidad docente, la existencia de espacios propicios para las labores académicas, aulas, dotación de laboratorios y salas de informática, biblioteca, material didáctico, deberes escolares y los procesos de evaluación.



En la misma línea trazada hasta el momento y sin el ánimo de profundizar en otros aspectos relacionados con los procesos cognitivos de los infantes, es necesario señalar desde una perspectiva psicológica que la experiencia temprana de los niños determina en gran medida la fortaleza o debilidad de su desarrollo futuro y sienta las bases de lo que en la literatura anglosajona se denomina como el “*school readiness*”, toda la gama de actitudes, la presteza y soltura que exhiben los estudiantes para realizar tareas académicas y avanzar en el proceso de formación y de adquisición de capacidades. Según [Linver, Sidle y Bruks-Gunn \(2006\)](#), los desarrollos o dificultades observados en la proficiencia lingüística y en las competencias en lectura en los primeros años es un excelente predictor del grado de alfabetismo o analfabetismo funcional en la adolescencia o en la vida laboral; así mismo, es un predictor de la fluidez con la que los individuos actuarán en su vida adulta y de la seguridad con la que trazan metas y emprenden sus proyectos. Los pobres desarrollos en el manejo del lenguaje en la edad

temprana aumentan el riesgo del fracaso escolar en las fases académicas que prosiguen e impiden alcanzar los progresos necesarios para operar con pericia en la vida social.

De ahí la necesidad de que la asistencia a la infancia para la creación de capacidades mínimas constituya una de las piezas de mayor prioridad en el diseño de las políticas educativas hoy en día<sup>21</sup>.

### **Precisiones metodológicas**

A continuación se introducen algunas precisiones metodológicas sobre la muestra utilizada, los procedimientos y los productos que proporciona la base de microdatos.

No se trata de una descripción o análisis minucioso de los diseños muestrales, sino más bien de una explicación escueta del marco de interpretación de los resultados que es posible derivar de esta base estadística.

La metodología utilizada por PIRLS tiene un grado relativamente alto de complejidad y requiere de un manejo cuidadoso de los resultados. Oportunamente se hará referencia al soporte metodológico de la base en el que se pueden consultar aspectos técnicos de mayor finura.

- Las muestras son estratificadas, bietápicas y de conglomerados. El primer estadio selecciona las escuelas y en la segunda etapa selecciona una o más clases de 4º de primaria por plantel y a todos los estudiantes que las conforman. El diseño define 75 estratos muestrales dentro de cada país y en cada uno de ellos se seleccionan las escuelas de manera independiente.

Debido a que no se trata de muestreo aleatorio simple, el manejo de la base demanda la utilización de técnica de replicación que proporcionen errores muestrales consistentes

---

<sup>21</sup> Desde una perspectiva más amplia, la nueva economía del ciclo vital postula que las diferencias tempranas en las estructuras cognitivas de los infantes persisten a lo largo del ciclo vital de los individuos y repercuten en la edad adulta. De acuerdo con el trabajo pionero de [Ben-Porath](#) (1967) y la amplia revisión de la literatura y los contrastes empíricos realizados recientemente por [Cunha, Heckman, et., al.](#) (2005), las brechas en la dotación de habilidades observadas en la edad adulta emergen antes de la iniciación del ciclo escolar. La disponibilidad de ingresos familiares y el background del entorno son factores decisivos, de tal modo que las “*habilidades engendran habilidades*” a través de un proceso de múltiples estados en el que las inversiones iniciales tienen implicaciones en la inversiones posteriores (multiplicador de habilidades).

(como las técnicas jackknife, JRR). La base suministra las variables Jkzone y jkcrep que identifican las zonas muestrales y que sirven de soporte en el proceso de estimación de los errores.

- Las variables disponibles, además de los resultados de los *tests* de proficiencia, provienen de 4 cuestionarios que se describen a continuación:

(a) Un cuestionario de estudiantes que indaga sobre sus experiencias en el hogar y en la escuela, la autopercepción y las actitudes hacia la lectura, los hábitos de estudio y los recursos bibliográficos disponibles en el hogar, etc.

(b) El primer cuestionario se complementa con uno diligenciado por los padres o acudientes de los estudiantes y que pregunta también sobre los recursos bibliográficos disponibles en el hogar, sus hábitos y actitudes hacia la lectura, sus conexiones con la escuela y algunos indicadores socioeconómicos básicos (educación, estatus ocupacional y percepción relativa de su riqueza).

(c) El cuestionario a los docentes indaga sobre las características generales de la clase, los niveles de lectura observados en el grupo, habilidades en lenguaje, los tiempos de instrucción y de preparación de clases, materiales didácticos, recursos disponibles y conexiones con los hogares de los educandos, nivel educativo de los mismos profesores, entre otros.

(d) El cuestionario a las escuelas se ocupa de las características demográficas de los estudiantes, su localización y el entorno escolar, algunos indicadores asociados al perfil económico de los estudiantes, políticas curriculares, tiempo total de instrucción en un año escolar, disponibilidad de materiales y plantilla, percepción del clima educativo, cercanía de bibliotecas públicas, museos, teatros y librerías, etc.

- Adicionalmente a las variables originales la base cuenta con una batería de 10 indicadores sintéticos que combinan diversos atributos y proporcionan una visión amplia de diversos aspectos.

Estos indicadores dan una idea compacta en tres categorías (alto, medio o bajo) de conceptos más amplios, como por ejemplo, los “recursos educativos del hogar” que se obtiene de la combinación de variables como el número de libros en el hogar y disponibilidad de soportes (ordenador, mesa de trabajo, libros propios, etc.).

- Scores. En razón de que la proficiencia, destreza o habilidad de un estudiante es una variable no observable directamente, el proceso de escalamiento que hace PIRLS permite estimar modelos de variables latentes que describen la probabilidad de respuesta a una determinada pregunta de acuerdo con la proficiencia<sup>22</sup>.

De este modo, según el número de parámetros considerados, inicialmente se estima la probabilidad de que una persona, cuya proficiencia sobre una escala predefinida y un conjunto de atributo observables, responderá correctamente a un determinado ítem. Automáticamente queda determinada la probabilidad de una respuesta incorrecta y con base en la distribución posterior para cada respuesta se obtiene un conjunto de cinco **valores plausibles (PV)** cuya variación revela la incertidumbre relacionada con la proficiencia estimada para cada individuo.

Ahora bien, debido a la magnitud de las pruebas completas ningún estudiante fue evaluado en todos los componentes sino en una fracción, así entonces los valores plausibles estiman mediante técnicas de imputación cómo el estudiante podría haber respondido al test completo controlado por un conjunto de atributos<sup>23</sup>. Por tal razón, los scores o valores reportados en la base de PIRLS son 5 por cada ítem y estudiante, y no un único valor puntual.

Adicionalmente, los errores provienen de dos fuentes: del muestreo y del proceso de imputación del indicador de proficiencia

El proceso de estandarización de estos valores plausibles está en una escala media de 500 y una desviación de 100.

---

<sup>22</sup> Los desarrollos estadísticos pueden apreciarse en PIRLS 2001 (2003). *User Guide for the international data base*. Editado por Eugenio González y Ann M. Kennedy, Boston College e IEA.

<sup>23</sup>Ver *ibid.*, p. 6-11.

Cuadro 3.5. PIRLS, comparaciones internacionales promedios generales

Países	Promedio General Tot	Años de Escol formal	Edad promedio	Promedio Niñas	Promedio Niños	Diferencia
Suecia	561 [2.2]	4	10.8	572 [2.6]	550 [2.5]	22 [2.6]
Holanda	554 [2.5]	4	10.3	562 [2.7]	547 [2.8]	15 [2.2]
Inglaterra	553 [3.4]	5	10.2	564 [3.9]	541 [3.7]	22 [3.3]
Bulgaria	550 [3.8]	4	10.9	562 [3.7]	538 [4.7]	24 [3.6]
Latvia	545 [2.3]	4	11.0	556 [3.1]	534 [2.6]	22 [3.4]
Canada	544 [2.4]	4	10.0	553 [2.6]	536 [2.6]	17 [2.1]
Lituania	543 [2.6]	4	10.9	552 [3.0]	535 [2.7]	17 [2.7]
Hungría	543 [2.2]	4	10.7	550 [2.4]	536 [2.5]	14 [2.1]
EEUU	542 [3.8]	4	10.2	551 [3.8]	533 [4.9]	18 [4.1]
Italia	541 [2.4]	4	9.8	545 [2.6]	537 [2.7]	8 [2.5]
Alemania	539 [1.9]	4	10.5	545 [2.2]	533 [2.5]	13 [2.7]
R. Checa	537 [2.3]	4	10.5	543 [2.8]	531 [2.6]	12 [2.8]
N.Zelanda	529 [3.6]	5	10.1	542 [4.7]	516 [4.2]	27 [5.4]
Escocia	528 [3.6]	5	9.8	537 [3.9]	519 [4.2]	17 [4.0]
Singapur	528 [5.2]	4	10.1	540 [5.3]	516 [5.7]	24 [4.1]
Rusia	528 [4.4]	3 or 4	10.3	534 [4.3]	522 [4.8]	12 [2.3]
Hong Kong	528 [3.1]	4	10.2	538 [3.0]	519 [3.5]	19 [2.9]
Francia	525 [2.4]	4	10.1	531 [2.7]	520 [3.0]	11 [3.3]
Grecia	524 [3.5]	4	9.9	535 [3.8]	514 [4.0]	21 [3.9]
Slovakia	518 [2.8]	4	10.3	526 [3.0]	510 [3.3]	16 [3.0]
Islandia	512 [1.2]	4	9.7	522 [1.9]	503 [1.5]	19 [2.4]
Rumanía	512 [4.6]	4	11.1	519 [4.2]	504 [5.7]	14 [3.8]
Israel	509 [2.8]	4	10.0	520 [3.4]	498 [3.7]	22 [4.3]
Eslovenia	502 [2.0]	3	9.8	512 [2.5]	491 [2.4]	22 [2.8]
<b>Promedio muestral</b>	<b>500 [0.6]</b>	<b>4</b>	<b>10.3</b>	<b>510 [0.7]</b>	<b>490 [0.7]</b>	<b>20 [0.7]</b>
Noruega	499 [2.9]	4	10.0	510 [3.5]	489 [3.4]	21 [3.9]
Chipre	494 [3.0]	4	9.7	506 [3.3]	482 [3.6]	24 [3.5]
Moldova	492 [4.0]	4	10.8	504 [4.7]	479 [4.0]	25 [4.0]
Turquía	449 [3.5]	4	10.2	459 [4.0]	440 [3.7]	19 [3.1]
Macedonia	442 [4.6]	4	10.7	452 [5.1]	431 [4.8]	21 [3.6]
<b>Colombia</b>	<b>422 [4.4]</b>	<b>4</b>	<b>10.5</b>	<b>428 [5.1]</b>	<b>416 [4.7]</b>	<b>12 [4.3]</b>
Aregentina	420 [5.9]	4	10.2	428 [6.2]	410 [6.5]	18 [4.7]
Irán	414 [4.2]	4	10.4	426 [5.7]	399 [5.6]	27 [8.1]
Kuwait	396 [4.3]	4	9.9	422 [5.6]	373 [6.3]	48 [8.4]
Marruecos	350 [9.6]	4	11.2	361 [9.6]	341 [10.9]	20 [6.8]
Belize	327 [4.7]	4	9.8	341 [5.3]	314 [5.2]	27 [4.8]

Fuente: Elaboración propia a partir de Almanacs pruebas PIRLS, base de microdatos y documentación

Desviaciones estándar entre corchetes

Media 500

### 3.5.2. Análisis de los valores plausibles

Esta sección presenta los resultados más importantes que se derivan del procesamiento de la base. La primera parte realiza brevemente algunas comparaciones internacionales, la segunda avanza hacia el análisis cruzado de algunas variables de relevancia en este capítulo y finalmente se estiman algunos modelos econométricos que buscan determinar cuales son los factores de mayor peso relativo en la explicación de la proficiencia.

Un primer examen de los resultados globales de la prueba ubica a Colombia en el lugar 30 de un total de 35 países (ver [cuadro 3.5](#)). El promedio general fue de 422 con una desviación estándar de 4.4. Los únicos países que obtuvieron promedios inferiores fueron Argentina (420), Irán (414), Kuwait (396), Marruecos (350) y Belize (327).

Las niñas registraron puntajes superiores a los de los niños con una diferencia que resultó estadísticamente significativa después de realizar una prueba de diferencia de medias de las dos distribuciones.

El país que obtuvo los mayores scores fue Suecia con un promedio de 561 y una desviación de 2.2, seguido por Holanda (554) e Inglaterra (553).

Las mayores dificultades de los niños colombianos se evidenciaron en la comprensión de los textos literarios (con un puntaje medio en todos los items de 425 y desviación 4.2) y en los textos informativos (con un puntaje medio de 424 y una desviación de 4.3). De manera particular, en los procesos inferenciales y de extracción de información el valor plausible promedio fue 429 y en la interpretación, integración y evaluación de textos solamente llegó a 417. La correlación observada en estos dos componentes fue relativamente alta para todos los países. Colombia obtuvo un coeficiente de correlación de Pearson de 0.92 equivalente a la mediana internacional de la muestra de países.

En la comparación internacional el 48.5% de los estudiantes colombianos está por debajo del percentil 25 de la distribución de los puntajes totales, el 34.2% entre el 25-50, el 13.3 entre el 50-75, el 3.1% entre el 75-90 y solamente el 0.8% por encima del percentil 90 (43 estudiantes de un total de 5131 se ubicaron en el nivel más alto en la escala internacional).

La misma comparación, pero ahora utilizando la distribución de los puntajes en el ámbito doméstico, muestra que el 12.8% se localiza en el nivel más alto (por encima del percentil 90), 17.8% entre 75-90, el 26.2% entre el 50-75, el 22.9% entre el percentil 25-50 y el 20.4% en el nivel más bajo (<25 percentil).

### **¿Cuáles son los factores específicos que están asociados con este pobre desempeño de los estudiantes colombiano en las pruebas?**

A continuación se realizan algunos ejercicios que exploran la distribución de los puntajes de acuerdo con un conjunto de variables de control y que proporcionarán una primera respuesta al interrogante.

El [cuadro 3.6](#) presenta un resumen de los resultados utilizando todos los valores plausibles. Los valores corresponden a las medias, las desviaciones y las proporciones calculadas después de los ejercicios de replicación mediante técnicas Jackknife<sup>24</sup>. También aparecen los errores estándar que capturan los efectos de muestreo e imputación en las aproximaciones a la proficiencia. Los indicadores y las variables de clasificación de los puntajes calculados aparecen en cinco grandes grupos para facilitar su lectura. Estos grupos corresponden a las características individuales, el background familiar, el entorno en el que se localiza la escuela, algunas variables seleccionadas sobre el perfil de la escuela y finalmente dos variables relacionadas con los docentes.

#### **Características individuales**

La variable de clasificación en percentiles muestra que del total de 5131 niños, el 10% obtuvo un puntaje promedio de 549.52 con un error estándar de 7.98 que se deriva del proceso de iteración de los cinco valores plausibles. De acuerdo con el diseño de la muestra este 10% (656 niños) representa 97517 niños del universo muestral. El percentil más bajo (<25) obtuvo un puntaje medio de 329.58 y una desviación de 7.53.

---

<sup>24</sup> Tal y como lo sugieren las notas técnicas de utilización de la base. Ver PIRLS 2001, *User Guide for the international data base*, p. 7.3 y sgtes. Obsérvese que el método de replicación Jackknife utiliza toda la información disponible (5 valores plausibles) para estimar de manera insesgada los errores muestrales. La variación entre la muestra original y las estimaciones de las pseudo-réplicas corresponde al error muestral del estadístico correspondiente. Las replicas utilizadas en los cálculos fueron 75 por cada ejercicio controlado con la respectiva variable de clasificación.



Las columnas C(5) y C(6) muestran los valores correspondientes a la primera réplica y se presenta para observar su diferencia respecto a valor medio finalmente obtenido en la C(3).

Las columnas C(7) y C(8) presentan los cálculos de las proporciones y su desviaciones estándar. Como se puede observar, este cálculo coincide con la distribución en percentiles.

En este mismo bloque de indicadores también está el índice que capta la experiencia temprana en lectura. Este indicador, relevante para la tesis previamente planteada sobre la importancia de los primeros años de los infantes, hace referencia a los procesos de estimulación realizados por los padres antes de que sus hijos iniciaran su ciclo de educación formal (leían libros, contaban historias, cantaban estribillos, jugaban con el alfabeto, realizaban juegos de palabras, enseñaban el sonido de los signos alfabéticos).

La distribución del indicador está dada en 3 categorías (alta, media y baja) y los valores observados en la primera columna indican el número de casos en la muestra para cada una de las categorías y el valor contiguo la población a la que representa esta muestra. Como se puede apreciar los valores altos del índice coinciden con los mayores puntajes medios obtenidos en la replicación con todos los valores plausibles: aquellos niños con experiencias tempranas de lectura “altas” o muy favorables obtuvieron un puntaje medio de 434,57, superior al que obtuvieron los niños con experiencias “medias” (425,84) y “bajas” (410,29). Las desviaciones estándar se pueden ver en la columna adyacente a los valores medios C(4).

Tal y como se anotó antes, para efectos de control las dos columnas que siguen a los valores medios y a la desviación de la media final presentan los puntajes y los errores estándar de la primera réplica (de un total de 75 y que corresponden al número total de estratos definidos en el diseño muestral). Como se puede ver la variabilidad no es demasiado elevada.

Cuadro 3.6. Colombia: Valores plausibles (scores) en las pruebas PIRLS 2001 según grupos y categorías de variables de clasificación

INDICADORES Y VARIABLES DE CLASIFICACIÓN	Casos en la Muestra	Peso poblacional	Media (***)		Media 1er (***)		%	
			VP=SCORE	Err.Std	VP=SCORE	Err.Std	Estimado	Err.Std
Categorías	C(1)	C(2)	C(3)	C(4)	C(5)	C(6)	C(7)	C(8)
<b>CARACTERÍSTICAS ESTUDIANTE</b>								
<b>PUNTAJES SEGÚN PERCENTILES</b>								
≤ 25 PERCENTIL	1049	243463	329.58	7.53	318.42	2.93	24.97	2.04
25-50	1173	243919	396.34	0.97	396.64	0.57	25.01	1.37
50-75	1342	243921	448.28	1.27	450.1	0.6	25.01	0.99
75-90	911	146349	492.58	3.96	498.82	0.65	15.01	0.91
≥ 90 PERCENTIL DE PUNTAJES ALTOS	656	97517	549.52	7.98	560.74	3.84	10.00	1.34
<b>INDICE DE EXPERIENCIAS TEMPRANAS EN LECTURA (c)</b>								
ALTA	1832	325578	434.57	6.13	433.79	6.06	33.39	1.69
MEDIA	1745	315882	425.84	4.85	425.23	4.77	32.39	1.27
BAJA	783	170495	410.29	5.01	410.69	5	17.48	1.21
<b>INDICE DE ACTITUD DEL ESTUDIANTE HACIA LA LECTURA(a)</b>								
ALTA	2301	436134	444.87	5.07	445.66	4.99	44.72	1.52
MEDIA	2426	469709	404.52	5.03	403.54	4.96	48.17	1.41
BAJA	162	24412	432.99	8.3	430.25	7.8	2.5	0.29
<b>INDICE DE AUTOPERCEPCIÓN DEL ESTUDIANTE HACIA LA LA LECTURA(b)</b>								
ALTA	1369	221457	466.96	6.17	466.82	5.98	22.71	1.02
MEDIA	3416	681979	410.78	4.33	410.04	4.23	69.93	1.09
BAJA	190	42125	412.35	9.8	415.02	9.27	4.32	0.53
<b>BACKGROUND FAMILIAR</b>								
<b>INDICE DE ACTITUD DE LOS PADRES HACIA LA LECTURA(d)</b>								
ALTA	1796	309799	442.63	5.45	443.02	5.43	31.77	1.26
MEDIA	2051	397940	413.56	4.98	413.28	4.96	40.81	1.61
BAJA	281	50536	427.5	7.41	426.57	7.22	5.18	0.5
<b>INDICE DE POSESIONES DEL HOGAR (e)</b>								
3-4	1287	208741	429.31	7.96	428.43	7.91	21.41	1.26
1-2	3730	744612	421.94	4.11	421.88	4.09	76.36	1.34
<b>INDICE DE POSESION DE LIBROS EN EL HOGAR (ESTUD)</b>								
0-11	1849	417649	400.63	4.81	400.49	4.7	42.83	1.69
11-25	1318	219283	437.83	4.74	439.47	4.35	22.49	0.95
26-100	996	163265	457.39	7.31	457.42	7.28	16.74	1.01
101-200	341	59220	455.14	8.98	452.46	8.61	6.07	0.51
≥ 200	264	44244	451.09	9.9	447.73	9.52	4.54	0.44
<b>LIBROS EN EL HOGAR (PADRES O ACUDIENES)</b>								
0-11	1927	401928	408.17	4.41	408.32	4.39	41.22	1.77
11-25	1197	214371	421.07	6.41	420.55	6.24	21.98	1.45
26-100	1092	174932	450.86	7.01	449.87	6.87	17.94	0.95
101-200	241	36539	471.33	11.74	471.78	11.04	3.75	0.4
> 200	169	27906	469.56	11.34	471.21	11.22	2.86	0.38
<b>POSESION DE ORDENADOR EN CASA</b>								
SI	1272	195620	442.68	9.27	442.22	9.15	20.06	1.28
NO	3726	755156	418.12	4.13	417.92	4.12	77.44	1.37
<b>INGRESO FAMILIAR ANUAL(*)</b>								
≤ 20000 US AÑO (BAJOS)	2032	385683	409.92	4.71	409.97	4.69	39.55	2.35
20000 a 59999 US AÑO (MEDIOS)	1259	215232	438.41	4.77	439.42	4.58	22.07	1.26
MÁS DE 60000 US AÑO (ALTOS)	1304	250857	436.34	9.67	434.25	9.45	25.72	2.19
<b>INGRESOS FAMILIARES x POSESIÓN ORDENADOR (**)</b>								
INGR BAJOS x SI	311	49495	396.99	8.13	397.31	7.73	12.83	1.31
INGR BAJOS x NO	1660	325349	413.13	4.7	413.07	4.69	84.36	1.48
INGR MEDIOS x SI	294	45467	440.14	6.88	439.53	6.31	21.12	1.57
INGR MEDIOS x NO	937	164954	438.78	5.16	440.05	5.03	76.64	1.71
INGR ALTOS x SI	508	74891	473.74	15.06	470.78	14.76	29.85	3.46
INGR ALTOS x NO	771	171612	421.09	8.31	419.44	7.96	68.41	3.44
<b>NIVEL EDUCATIVO MÁS ALTO DE LOS PADRES</b>								
UNIVERSITARIO FINALIZADO O SUPERIOR	701	122014	461.54	14.23	461.63	14.13	12.51	1.55
POSTSECUNDARIA SIN UNIVERSIDAD	215	33228	453.21	10.71	447.19	9.93	3.41	0.38
BACHILLERATO FINALIZADO	653	105524	442.57	5.59	442.25	5.31	10.82	0.85
SECUNDARIA MEDIA	646	108916	427.73	6.73	426.35	6.44	11.17	0.77
PRIMARIA O ALGO SEC INCOMPL O NINGUNO	1820	387021	402.65	3.78	402.81	3.69	39.69	1.55

(...)Cuadro 3.6. Continuación Valores plausibles (scores) en las pruebas PIRLS según grupos y categorías de variables de clasificación

INDICADORES Y VARIABLES DE CLASIFICACIÓN	Casos en		Media (***)	Err.Std	Media 1er (***)		% (***)		
	la Muestra	Peso poblacional	5 VP=SCORE		VP=SCORE	Err.Std	Estimado	Err.Std	
<b>LOCALIZACIÓN ESCUELA</b>									
<b>ZONA DE LOCALIZACIÓN ESCUELA</b>									
URBANA	4197	618520	435,41	4,77	435,32	4,76	63,43	3,18	
SUBURBANA	76	13680	402,05	17,54	394,49	16,56	1,4	0,98	
RURAL	557	252978	388,33	7,28	387,12	7,05	25,94	3,03	
<b>TAMAÑO ZONA DE LOCALIZACIÓN ESCUELA</b>									
< 3000 PERSONAS	385	99052,33	397,47	15,09	397,36	14,79	10,16	3,77	
3001 - 100000 PERS	1327	244834,1	429,53	11,35	430,34	11,32	25,11	4,35	
100001 - 500000 PERS	1059	159420,7	418,01	10,33	416,65	10,26	16,35	3,75	
> 500000 PERS	1617	230925,1	444,08	8,15	443,58	8,11	23,68	4,14	
<b>DISPONIBILIDAD DE LIBRERÍAS</b>									
30 MINUTOS EN COCHE	4205	651599	433,22	5,7	432,98	5,69	66,82	4,31	
2 HORAS EN COCHE	267	100644	398,16	17,01	397,91	16,88	10,32	2,83	
NO DISPONIBLE	484	147292	400,63	6,96	400,24	6,61	15,1	3,57	
<b>DISPONIBILIDAD DE MUSEOS</b>									
30 MINUTOS EN COCHE	2424	367300	444,74	8,15	443,89	8,08	37,67	4,53	
2 HORAS EN COCHE	1146	213452	429,49	9,59	429,65	9,5	21,89	3,96	
NO DISPONIBLE	1345	327329	395,09	6,6	395	6,38	33,57	4,83	
<b>DISPONIBILIDAD DE BIBLIOTECA PÚBLICA</b>									
30 MINUTOS EN COCHE	4239	706932	426,72	4,49	426,1	4,47	72,49	3,91	
2 HORAS EN COCHE	291	52026	438,55	12,99	440	12,84	5,34	1,88	
NO DISPONIBLE	435	145894	386,69	11,09	386,11	10,74	14,96	2,85	
<b>ESCUELA</b>									
<b>% DE ESTUDIANTES PROCEDENTES DE FAMILIAS CON DESVENTAJA ECONOMICA (ESCUELA)</b>									
0-10%	592	97633	479,52	19,24	477,92	19,19	10,01	2,55	
11-25%	258	42308	440,35	17,09	437,14	16,74	4,34	1,31	
26-50%	716	128078	448,85	8,56	449,33	8,5	13,13	3,62	
>50%	3404	669127	406,92	5,18	407,15	5,12	68,62	4,04	
<b>% DE ESTUDIANTES AL FINAL DEL AÑO ESCOLAR</b>									
98-100%	2915	534867	423,47	5,94	423,11	5,92	54,85	4,92	
95-97%	1206	215972	442,53	9,94	443,06	9,93	22,15	4,12	
90-94%	541	103386	405,12	15,39	404,04	14,82	10,6	3,6	
80-89%	267	67227	410,66	15,27	412,95	15,08	6,89	2,48	
> 80%	171	40249	359,65	27,53	355,3	26,4	4,13	1,81	
<b>PROFESORES</b>									
<b>NIVEL FORMAL DE EDUCACIÓN DOCENTES(f)</b>									
NIVEL 3 ISCED	413	92688	396,56	19,41	393,92	19,3	9,5	2,31	
NIVEL 4	210	42282	436,52	30,91	438,02	30,8	4,34	1,49	
NIVEL 5 Y SUP	4323	803220	424,18	5,02	424,26	5	82,37	3,02	
<b>SEXO DEL DOCENTE</b>									
FEMENINO	4068	776820	425,88	5	425,85	4,98	79,66	3,08	
MASCULINO	931	173329	403,82	10,78	403,13	10,72	17,77	2,93	

Fuente: estimaciones propias partir de las bases originales de PIRLS procesadas en SPSS.

Los valores corresponden a los promedios de los scores y replicaciones Jackknife para el cálculo de los errores estándar (número de replicaciones = 75=número de estratos muestrales).

Los índices (a)-(e) extraídos de la base original con las siguientes definiciones: (a) Construido según las intensidades en las siguientes respuestas de los estudiantes: "leo porque tengo que hacerlo", "hablo con los demás sobre los libros", "me alegra si alguien me regala un libro", "leer es aburrido", "disfruto leyendo". La media se computa sobre la siguiente escala de 4 puntos o intensidad de la respuesta: 1=un poco en desacuerdo, 2=de acuerdo, 3=de acuerdo mucho. Alto nivel indica una media entre 3-4, medio 2-3 y bajo 1-2; (b) Capta la visión propia del estudiante hacia la lectura y se construye sobre una escala de respuesta a lo siguiente: "leer es muy fácil para mí", "no leo tan bien como otros estudiantes", "leer es muy difícil para mí". Las categorías alto, medio y bajo se construyen a partir de las respuestas 1=un poco en desacuerdo, 2=disenso, 3=un poco de acuerdo y 4=niego. Alto 3-4, medio 2-3 y bajo 1-2; (c) Este índice se basa en las respuestas de los padres en relación con las actividades realizadas en el hogar antes de iniciar el proceso formal de educación de sus hijos: leía libros, contaba historias, cantaba estribillos, jugaba con el alfabeto, juegos de palabras, enseñaba el sonido de signos alfabéticos. Según las respuestas (nunca o casi nunca, algunas veces, frecuentemente). El nivel Alto corresponde al valor entre 2.33-3; el medio 1.67-2.33 y el bajo entre 1-1.67 en las frecuencias de respuesta; (d) Obtenido a partir de las siguientes respuestas de los padres: "leo únicamente si tengo que hacerlo", "me gusta hablar con otras personas sobre libros", "reservo tiempo para leer", "leo únicamente si necesito información", "leer es una importante actividad en mi hogar". De la misma manera que en los demás índices se contruyen las categorías; (e) Este índice tiene en cuenta el número de libros en el hogar, soportes educativos como escritorío, ordenador propio, libros propios y acceso a diarios.

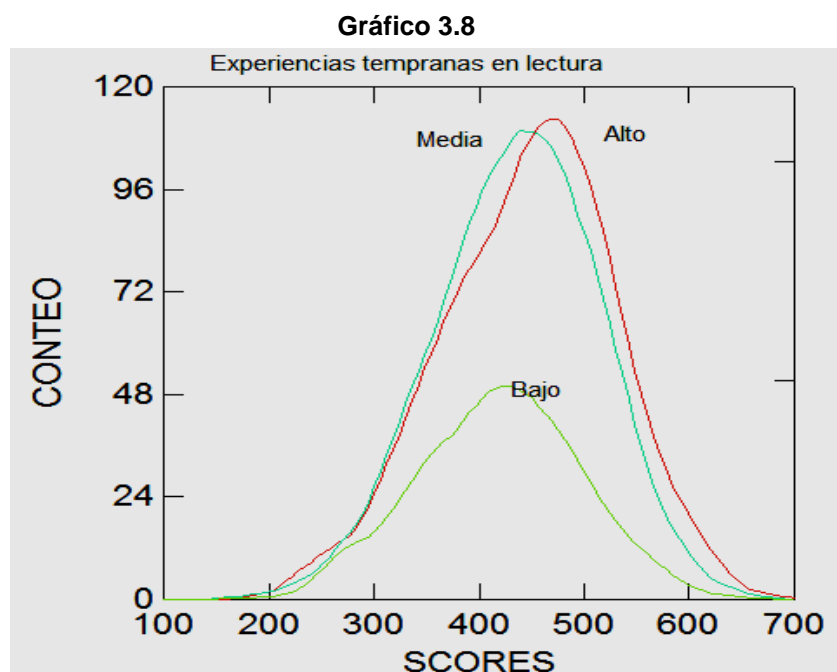
(f) De acuerdo con International Standard Classification of Education de la Unesco en los siguientes niveles: N3=Educación secundaria (bachillerato), N4=educación postsecundaria o terciaria, N5=primer estado de la educación terciaria, N6=segundo estado de educación terciaria (cualificación avanzada en investigación, ej. profesores universitarios).

(\*) Recodificada a partir de la variable original de 6 categorías; (\*\*) Replicaciones con variables en tablas cruzadas; (\*\*\*) Corresponden a la media de todos los valores plausibles y también aparece la media que resulta en la primera replicación; (\*\*\*\*) corresponde a la proporción estimada de cada categoría en la muestra y su respectivo error estándar.

El porcentaje o proporción de la población que representa cada una de las categorías (alta, media y baja) se calcula en la penúltima columna con su respectiva desviación.

Los resultados indican que el 33.4% de los estudiantes de la muestra recibieron una elevada estimulación hacia la lectura en los primeros años, el 32.4% una experiencia media y el 17.5% no contaron con un entorno adecuado en el hogar hacia la lectura, antes de iniciar su ciclo de educación primaria.

El [gráfico 3.8](#) muestra los histogramas de frecuencias suavizados de los puntajes para cada una de las categorías.



En este mismo bloque se incluyeron dos índices, un índice de auto-percepción de los estudiantes hacia la lectura y otro que intenta capturar su actitud. Aquellos niños con auto-percepción y actitud en la categoría alta muestran resultados consistentes con lo esperado.

Los que se clasificaron en las categorías medias y altas, sin embargo, presentan una diferencia respecto a los valores previstos: se esperaba que la auto-percepción y la actitud media coincidiera con valores más elevados del score que aquellos ubicados en la categoría baja. Una posible explicación radica en que la proporción de estudiantes en esta última categoría es solamente del 2.5% en el índice de auto-percepción y de 4.32 en

el de actitud hacia la lectura. Otra explicación puede atribuirse a los errores individuales de apreciación de los niños y que no es del caso intentar dilucidar en esta memoria de Tesis.

### **Background familiar**

El segundo bloque presenta una amplia variedad de indicadores que brindan una visión amplia del entorno familiar en el que viven los niños.

En general los valores computados proporcionan evidencia sobre la relación positiva entre las condiciones económicas de las familias, los niveles de escolaridad máximos alcanzados por los padres, la actitud de estos padres hacia la lectura y la posesión de bienes físicos que sirvan de soporte para el desarrollo de las actividades académicas de los niños en el interior de los hogares. De nuevo, esta vez desde una perspectiva más desagregada que la de los capítulos anteriores, se encontró que las restricciones financieras de las familias juegan un papel importante en el rendimiento de los escolares. Utilizando como variable de clasificación los ingresos anuales se puede apreciar que en el rango más bajo (< 20 mil dólares/año) sus hijos obtuvieron puntajes pobres (409.9 en promedio y una desviación de 4.7), mientras que los niños que provienen de hogares más solventes (ingresos medios y altos), obtuvieron valores plausibles mucho más elevados, 438.4 y 436.3 respectivamente. La diferencia entre el rango más bajo y estos últimos está por encima de los 26 puntos. Llama la atención que los de ingresos más altos obtuviesen un valor ligeramente inferior a los de ingresos medios<sup>25</sup>

En relación con la educación máxima alcanzada por los padres, los resultados son consistentes con las hipótesis formuladas a lo largo de este trabajo. Los niños con padres de bajo nivel de escolaridad obtuvieron puntajes inferiores, 402.65, mientras que los hijos de padres con nivel educativo universitario o superior alcanzaron una media de 461.54 en el score. Los datos sugieren una relación estrictamente creciente en los puntajes de los niños a medida que aumenta el grado de escolaridad de sus padres.

---

<sup>25</sup> Diferentes pruebas de diferencia de medias rechazaron la hipótesis nula de igualdad entre estos dos grupos.

La variable asociada a la disponibilidad de recursos bibliográficos en casa también proporcionó resultados interesantes y muestra, salvo el último valor de la escala, que a mayor número de libros en el hogar los valores factibles de la prueba son mayores. Este índice, conjuntamente con el de actitud de los padres hacia la lectura informa sobre la aprehensión hacia la actividad de estudiar.

Por razones presumiblemente muestrales los valores bajos del índice de actitud y el último rango en el de posesión de libros arrojaron valores medios imprevistos.

Otro indicador de este mismo bloque es el de posesión de ordenador en casa. La proporción de respuesta fue la siguiente: 77.4% no tenían ordenador y de 20.1% si contaba con este soporte en el año de aplicación de la prueba. La diferencia en los puntajes medios de los dos grupos, al controlar por esta variable, fue de casi 25 puntos.

### **Localización de la escuela**

Existe una amplia diferencia entre los puntajes de los niños urbanos, respecto a los rurales o suburbanos. Los niños de zonas urbanas (que representan el 63.4% del total de la muestra) alcanzaron una media de 435.4, frente a los 402.1 y 388.3 de los niños ubicados en zonas suburbanas y rurales, respectivamente.

En este mismo bloque también aparecen variables relacionadas con las oportunidades que brinda el entorno en la disponibilidad de bibliotecas públicas, museos y librerías.

### **Escuelas y profesores**

Los últimos dos bloques muestran que desde la perspectiva de los directores de los planteles educativos, el 68.6% de los niños provienen de hogares con deventajas económicas. Solamente un 10% proceden de hogares solventes. La diferencia entre estos dos rangos extremos es de más de 70 puntos en los valores plausibles.

En relación con las variables de los docentes, el 80% eran profesoras y solamente 18% profesores. La diferencia entre los puntajes es de casi 22 puntos, aunque los resultados de los niños con profesores hombres presentan una mayor variabilidad.

Finalmente, los puntajes obtenidos por los niños con profesores de bajo nivel (bachillerato) de educación formal según la escala ISCED<sup>26</sup> fueron precisamente los más bajos. Los otros dos niveles de educación formal de los docentes no muestran una asociación muy clara con los puntajes de la prueba.

### *¿Cuál fue el resultado de los compañeros de curso en las pruebas?*

Para responder a este interrogante a partir de los valores plausibles de cada estudiante se calculó el valor medio de los puntajes de los compañeros de clase ( $PEER_{ics}$ )<sup>27</sup>. Este valor corresponde a la siguiente expresión:

$$(3.10) \quad PEER_{ics} = \frac{\sum_{j \neq i}^{n_c} VP_{jcs} - VP_{ics}}{n_c - 1}$$

Los subíndices denotan el estudiante  $i$  de la clase  $c$  en la escuela  $s$ .

$VP$  corresponde a los valores plausibles obtenidos en la prueba y  $n_c$  el tamaño de la clase. Para cada uno de los estudiantes se obtiene así el puntaje promedio de sus compañeros en las pruebas y que será de utilidad para poder examinar los primeros indicios de los denominados *peer effects* de la literatura anglosajona.

Con estos valores y la clasificación de los puntajes en percentiles se realizó el remuestreo para los 5131 estudiantes (aplicado los mismos parámetros de los ejercicios de replicación jackknife realizados en los acápite anteriores) y se obtuvieron los resultados que aparecen en el [cuadro 3.7](#). Los niños con puntajes ubicados en el percentil más bajo de la distribución estudian con otros infantes también de bajo rendimiento en las pruebas (una media general de 373.9 y un error estándar de 6.2).

<sup>26</sup> Clasificación internacional de la Unesco.

<sup>27</sup> Utilizando la variable *idclass* como identificador se calculó el valor medio mediante el procedimiento *aggregate* del paquete estadístico SPSS. Para cada uno de los valores plausibles (scores) se obtuvo el correspondiente valor de la variable PEER.

Cuadro 3.7. Valores Plausibles controlados por un conjunto de variables

CLASIFICACIÓN EN PERCENTILES DE LOS VALORES PLAUSIBLES	Casos en	Peso	Media (***)	Media 1er (***)	%			
	la Muestra	poblacional	5 VP=SCORE	Err.Std	VP=SCORE	Err.Std	Estimado	Err.Std
	C(1)	C(2)	C(3)	C(4)	C(5)	C(6)	C(7)	C(8)
<b>VALOR PROMEDIO DE LOS PV COMPAÑEROS DE CLASE</b>								
< 25 PERCENTIL EN LOS VP	1049	243463	373,91	6,13	373,33	6,03	25,01	2,04
25-50	1173	243919	412,08	3,47	411,61	3,41	25,05	1,36
50-75	1341	243154	434,57	3,47	434,22	3,43	24,97	0,99
75-90	911	146349	456,06	4	456,44	3,95	15,03	0,92
>90 PERCENTIL DE PUNTAJES ALTOS EN LOS VP	655	96750	488,03	9,15	488,97	9,1	9,94	1,35
<b>PEER SCORES DE LOS PERCENTILES BAJOS Y ALTOS SEGÚN INGRESO FAMILIAR ANUAL</b>								
<b>RANGOS DE INGRESO FAMILIAR ANUAL PERCENTIL &lt; 25 (=25% MÁS BAJO EN LOS SCORES)</b>								
< 20000 US AÑO (BAJOS)	508	103219	372,4	7,44	373,16	7,4	42,4	4,52
20000 a 59999 US AÑO (MEDIOS)	179	36335	391,22	5,17	389,81	5	14,92	1,76
MÁS DE 60000 US AÑO (ALTOS)	216	62391	372,3	9,49	371,52	9,34	25,63	3,21
<b>PERCENTIL &gt; 90 (= 10% MÁS ALTO EN LOS SCORES)</b>								
< 20000 US AÑO (BAJOS)	131	21230	456,87	5,33	457,61	5,19	21,94	3,74
20000 a 59999 US AÑO (MEDIOS)	148	21004	472,98	6,45	473,6	6,36	21,71	2,95
MÁS DE 60000 US AÑO (ALTOS)	323	44140	504,95	13,1	506,07	13,06	45,62	5,79
<b>PEER SCORES DE LOS PERCENTILES BAJOS Y ALTOS SEGÚN NIVEL EDUCAT MÁS ALTO DE LOS PADRES</b>								
<b>NIVEL EDUCATIVO MÁS ALTO DE LOS PADRES PERCENTIL &lt; 25 (=25% MÁS BAJO EN LOS SCORES)</b>								
UNIVERSITARIO FINALIZADO O SUPERIOR	65	19514	371,53	10,08	371,64	9,43	8,02	2,6
POSTSECUNDARIA SIN UNIVERSIDAD	25	5546	403,56	16,88	402,09	16,36	2,28	0,64
BACHILLERATO FINALIZADO	104	18694	389,62	5,99	388,76	5,96	7,68	1,04
SECUNDARIA MEDIA	134	24844	393,23	7,01	392,53	6,96	10,2	1,47
PRIMARIA O ALGO SEC INCOMPL O NINGUNO	479	116238	369,6	6,63	369,93	6,53	47,74	2,94
<b>PERCENTIL &gt; 90 (= 10% MÁS ALTO EN LOS SCORES)</b>								
UNIVERSITARIO FINALIZADO O SUPERIOR	230	31312	510,13	15,49	510,53	15,45	32,36	5,37
POSTSECUNDARIA SIN UNIVERSIDAD	48	6419	481,96	11,04	482,45	10,96	6,63	1,14
BACHILLERATO FINALIZADO	90	11831	469,39	4,86	470,53	4,5	12,23	2,4
SECUNDARIA MEDIA	60	9226	462,97	7,22	460,13	6,83	9,54	2,17
PRIMARIA O ALGO SEC INCOMPL O NINGUNO	90	15883	455,93	6,69	460,65	5,93	16,42	3,27

Fuente: estimaciones propias partir de las bases de PIRLS.

El mismo comportamiento se aprecia en los demás percentiles, el 10% de los niños con puntajes más altos tienen compañeros de clase que alcanzaron los promedios más altos en la prueba, aunque con mayor variabilidad que en los otros percentiles.

### ¿Cuál es el background familiar de los compañeros de curso?

Si controlamos el cuadro anterior por las variables ingreso y máximo nivel educativo de los padres de los compañeros de clase los resultados permiten apreciar que los estudiantes con puntajes más bajos (percentil <25) tienen compañeros (42%) con puntajes también bajos y con los niveles de ingresos correspondientes a los más pobres (<20 mil US\$ al año). Así mismo, el nivel educativo de los padres (ningún nivel de escolaridad, primaria o secundaria incompleta) de sus compañeros es el inferior con una proporción del 47.7%.



Ahora, al examinar el percentil de puntajes más altos se puede concluir que los niños con logros superiores en los scores tienen compañeros de clase con padres predominantemente ricos (> 60 mil U\$ al año) y educados (educación universitaria finalizada o superior). Estos resultados, además de sugerir la presencia de efectos contextuales aportan indicios adicionales a la idea de que persiste un proceso de estratificación o sorting socio-económico.

### 3.5.3. Efectos endógenos, contextuales y correlacionados

La exploración que hasta el momento se ha realizado lleva de inmediato a uno los aspectos cruciales que la economía de la educación ha venido escudriñando en los últimos años y que podríamos resumir en el siguiente interrogante *¿Cuál es la responsabilidad relativa de cada uno de los factores en los perfiles de competencia de los estudiantes?*

Las posibles respuestas a este interrogante no son fáciles y algunos intentos de respuesta demandan algunas precisiones analíticas y metodológicas que se abordan a continuación.

Desde la brillante reflexión de [Manski \(1993, 1995\)](#) quedó claramente dilucidada la diferencia entre los efectos *endógenos*, *contextuales* y *correlacionados*, así como los problemas que se derivan de los ejercicios de modelamiento econométrico cuando persisten correlaciones altas a nivel individual o entre clases y escuelas que no son modeladas explícitamente. El primer tipo de efectos alude a la influencia que ejercen sus pares o compañeros de clase y cuya intensidad no depende estrictamente de su background familiar. En tal sentido, el individuo  $i$  que pertenece a la clase  $c$  en la escuela  $s$  recibe los influjos del rendimiento académico de los demás y dependiendo de la intensidad de este efecto su desempeño tiende a ser parecido al que predomina en el grupo al que pertenece. Los efectos contextuales hacen referencia al influjo exógeno en el desempeño de  $i$  que ejercen sus compañeros o el grupo. El canal de transmisión de los compañeros proviene de su vector de características socioeconómicas y demográficas (educación de los padres, pertenencia a redes sociales, niveles de ingreso familiar, etc). Dependiendo de la intensidad de este efecto el individuo  $i$  tiende a comportarse de acuerdo con la distribución del background predominante en el grupo.

Los efectos correlacionados, de su parte, predicen que los individuos tienden a comportarse de manera similar debido a que provienen de entornos similares, enfrentan ambientes institucionales parecidos, reciben los mismos beneficios de la calidad de la escuela o de la disponibilidad de recursos de ésta. A manera de ejemplo, entre estos efectos pueden citarse la influencia de un mismo docente o del canal de recursos de la escuela o también el que se desprende del proceso de elección de escuelas de sus padres los cuales revelaron la estructura de preferencias por determinados perfiles de calidad educativa según sus niveles educativos o de riqueza familiar.

Los dos primeros efectos, tal y como lo señala [Manski \(1995\)](#), resumen diferentes tipos de interacción social y es muy fácil confundirlos y todavía más difícil aislarlos. Si la clase se tomara como una “microsociedad”, entonces el conjunto de interacciones que se forja en su interior representaría los denominados *peer effects*. Si la habilidad o las destrezas de sus compañeros de curso afectan el rendimiento escolar de un estudiante  $i$ , de tal manera que al adoptar nuevos métodos de enseñanza o mejorar la calidad del staff de docentes de esta “microsociedad”, el efecto externo de los compañeros sobre los logros académicos de  $i$  aumenta y se trataría entonces de un efecto endógeno puro que actuaría a la manera de un multiplicador microsocial en este caso.

Una distinción clave entre los tres tipos de efectos es que los endógenos operan con un mecanismo multiplicador, mientras que los otros dos no. Si en la microsociedad definida previamente existiesen inicialmente estudiantes de alta y baja habilidad, al proveer de mayores recursos a los primeros (tutorías particulares, sesiones de refuerzo, optativas, etc) aumentaría la calidad y se generaría un impacto positivo sobre sus compañeros de clase que depende de la magnitud del componente endógeno en la transmisión de habilidades. Si por alguna razón el efecto endógeno es nulo o despreciable en su magnitud, entonces el rendimiento del individuo  $i$  sería independiente del de sus pares y las diferencias estarían explicadas por otros factores. Por el contrario, si la magnitud del efecto endógeno es infinita o muy grande, el comportamiento de los individuos tiende a ser idéntico y sus logros y realizaciones escolares serían casi las mismas.

Una manera elegante de formalizar las ideas anteriores puede apreciarse a través de la especificación de un modelo lineal a la manera de [Manski \(1995\)](#):

Sea  $R$  un escalar que denota el resultado en alguna prueba y que condensa la proficiencia del individuo  $i$ .

El vector  $G$  resume el conjunto de características asociadas al grupo de referencia de los individuos, por ejemplo grupos étnicos o clases.

$F$  indica el conjunto de variables relacionadas con los atributos socioeconómicos y corresponde al vector exógeno de atributos.

$U$  la habilidad no observable directamente.

El modelo lineal se puede escribir de la siguiente manera:

$$(3.11) \quad \begin{aligned} R &= \alpha + \beta E(R|G) + E(F|G)' \gamma + F' \lambda + U \\ E(U|G, F) &= G' \delta \end{aligned}$$

Donde  $E$ , operador de esperanza matemática, denota el valor esperado.

La regresión de  $R$  sobre  $(G, F)$  genera la ecuación fundamental de interacción social:

$$(3.12) \quad E(R|G, F) = \alpha + \beta E(R|G) + E(F|G)' \gamma + F' \lambda + G' \delta$$

Esta expresión corresponde al valor de proficiencia esperado, dados los vectores grupales y familiares.

$(R, G, F)$  se observan directamente pero no las realizaciones de la variable  $U$ .

$(\alpha, \beta, \gamma, \lambda, \delta)$  es el vector de parámetros.

Así, la especificación intenta capturar los efectos endógenos ( $\beta$ ), contextuales ( $\gamma$ ), los efectos correlacionados ( $\delta$ ) y los directos que genera el background familiar ( $\lambda$ ).

Obsérvese que si  $\delta$  es significativo los efectos correlacionados importan y en consecuencia los individuos que conforman el grupo tienden a comportarse o exhibir

resultados similares en el atributo no observado  $U$ . Si por el contrario,  $\delta=0$ , los resultados individuales en logros son independientes de  $G$ . Si  $\delta$  es muy grande o tiende a infinito, todos los miembros del grupo tienden a tener idéntica distribución del atributo no observado  $U$ .

Adicionalmente es preciso anotar que si  $E(R|G)$ ,  $E(F|G)$ ,  $F$  y  $G$  fuesen tratamientos exógenos entonces el vector de parámetros  $(\beta, \gamma, \lambda, \delta)$  sencillamente capturaría efectos puros de tratamiento; pero, tanto  $E(R|G)$  como  $E(F|G)$  dependen de  $G$  funcionalmente y por tal razón aparece un problema de identificación en el sentido que le confiere [Manski \(1995\)](#).

Al aplicar esperanza matemática respecto a  $F$  a la ecuación de interacción social anterior se obtiene que:

$$(3.13) \quad E(R|G) = \alpha + \beta E(R|G) + E(F|G)' \gamma + E(F|G)' \lambda + G' \delta$$

Y, en consecuencia,

$$(3.14) \quad E(R|G) = \frac{[\alpha + E(F|G)'(\gamma + \lambda) + G' \delta]}{(1 - \beta)},$$

Esto último, si y solo si  $\beta$  es diferente de la unidad.

Esta ecuación reducida es la solución única de equilibrio social que indica la dependencia funcional que tiene  $R$  de  $G$ , de tal modo que  $E(R|G)$  es perfectamente colineal con  $[I, E(F|G), G]$ .

Esto último implica que los parámetros no quedan identificados y así entonces los efectos endógenos puros no pueden distinguirse de los contextuales ni de los correlacionados.

Al sustituir la solución anterior en la ecuación de interacción social, la forma reducida de esta última puede escribirse de la siguiente manera:

$$(3.15) \quad E(R|G, F) = \frac{\alpha}{(1-\beta)} + E(F|G) \left[ \frac{\gamma + \lambda\beta}{(1-\beta)} \right] + G' \frac{\delta}{(1-\beta)} + F' \lambda$$

Aquí los parámetros quedan identificados cuando no hay colinealidad perfecta. La única dificultad es que no se puede distinguir entre efectos endógenos puros y contextuales, pero la virtud de la ecuación radica en que permite capturar la incidencia de los factores sociales como conjunto ([Manski, 1995, p. 132](#)).

Obsérvese que cuando  $\left[ \frac{\gamma}{(1-\beta)} + \frac{\lambda\beta}{(1-\beta)} \right]$  es significativo una implicación es que  $\gamma$  y  $\beta$  también deberían ser diferentes de cero.

Ahora bien, si se supone en la ecuación reducida que  $\delta = \gamma = 0$  (efectos contextuales y correlacionados nulos), el modelo que se deriva es endógeno puro:

$$(3.16) \quad E(R|G, F) = \frac{\alpha}{(1-\beta)} + E(F|G)' \frac{\lambda\beta}{(1-\beta)} + F' \lambda$$

Bajo  $\lambda$  diferente de cero (que captura el efecto directo del background familiar), entonces el modelo queda identificado.

Al identificar el valor del parámetro compuesto  $\frac{\lambda\beta}{(1-\beta)}$ , entonces se pueden obtener  $\beta$  y  $\alpha$ .

Bajo tal supuesto tendríamos una aproximación indirecta a la magnitud de los efectos endógenos ( $\beta$ ) y directa a los que genera el background familiar ( $\lambda$ ).

Una especificación compacta y adecuada para la realización de los ejercicios econométricos que se presentarán más adelante expresa la forma reducida (ecuación 3.15) como sigue:

$$(3.17) \quad R_{ics} = \frac{\alpha}{(1-\beta)} + \lambda F_{ics}' + \phi Sch_{cs} + \frac{\gamma + \lambda\beta}{(1-\beta)} \bar{F}_{(-i)cs} + \mu_{cs} + \varepsilon_{ics}$$

En donde  $R_{ics}$  corresponde a los puntajes, valores plausibles o scores de las pruebas y que, tal como se anotó con anterioridad, intentan capturar la proficiencia del estudiante  $i$  que pertenece a la clase  $c$  de la escuela  $s$ .

$F_{ics}$  : contiene el background familiar y las características individuales.

$\bar{R}_{(-i)cs}$  : representa el background medio de los compañeros de clase (excluyendo al individuo  $i$ ) o algún indicador que defina el perfil de los pares escolares.

$Sch_{cs}$  : corresponde al conjunto de atributos asociados a la escuela  $s$  (entorno o disponibilidad de recursos)

$\mu_{cs}$  y  $\varepsilon_{ics}$  son los errores a nivel de escuela o clase y los individuales, respectivamente.

$\lambda$ ,  $\phi$  y  $\frac{\gamma + \lambda\beta}{(1-\beta)}$  son los parámetros asociados al background familiar, el entorno escolar y los atributos de los compañeros de clase, respectivamente.

La ventaja de esta especificación respecto a la ecuación de interacción social explicada previamente es que elimina el problema de simultaneidad. Tal dificultad se generaría al incluir directamente los scores promedio ( $\bar{R}_{(-i)cs}$ ) de los compañeros de clase y por tanto  $\text{cov}(\bar{R}_{(-i)cs}, \varepsilon_{ics}) \neq 0$ .

De igual manera se elimina el *problema de reflejo* señalado por [Manski \(1993, 1995\)](#), salvo que se presentara un problema extremo de colinealidad perfecta en los regresores utilizados o por la manera como se haya diseñado la muestra.

Adicionalmente, es necesario tener en cuenta que el error,  $\mu_{cs}$ , captura efectos correlacionados que condensan la influencia común que reciben los estudiantes a nivel de escuela.

No obstante, si se supone que los estudiantes se asignan aleatoriamente a nivel de clase en el interior de cada escuela, pero no entre escuelas, entonces el componente

sistemático del background medio opera entre estas últimas. Así, al existir covarianza no nula a nivel de escuela ( $s$ ),  $\text{cov}(\bar{F}_{(-i),cs}, \mu_{cs}) \neq 0$ , se garantizaría una estimación consistente de los parámetros, siempre y cuando la estimación incluya los efectos grupales pertinentes y logre eliminar o al menos modelar el componente omitido. Ver [McEwan \(2003\)](#) y [Ammermueller y Pischke \(2006\)](#).

Tal como se discutió en capítulos anteriores y en los primeros apartados de éste, si las escuelas en Colombia claramente atraen a estudiantes con determinados perfiles socioeconómicos o provenientes de determinados estratos sociales es evidente entonces que las características no observadas que condensa el error  $\mu_{cs}$  presenten correlación sistemática con  $\bar{F}_{(-i),cs}$  al comparar escuelas. En tal sentido la hipótesis que subyace a esta última afirmación es que prevalece una modalidad de sorting socioeconómico entre los planteles educativos y por tal razón la especificación anterior podría reescribirse como sigue:

$$(3.18) \quad R_{ics} = \alpha_s^* + \lambda F_{ics} + \phi Sch_{cs} + \rho \bar{F}_{(-i)cs} + \mu_{cs} + \varepsilon_{ics}$$

En donde  $\alpha_s^* = \frac{\alpha}{(1-\beta)}$  captura los efectos específicos de cada escuela y  $\rho = \frac{\gamma + \lambda\beta}{(1-\beta)}$  la incidencia que tiene el background familiar de los compañeros de clase. Como  $\mu_{cs}$  varía entre escuelas, entonces el intercepto compuesto es aleatorio.

Se aprecia de inmediato que la última ecuación muestra una estructura multinivel en la que los estudiantes están anidados en escuelas. Esto implica que la estimación de algún modelo que no considere esta estructura de agrupamiento sería incorrecta en la medida que supondría que los individuos se distribuyen de manera independiente. Es bien conocido en la literatura de modelos jerárquicos y mixtos<sup>28</sup> generalizados que al no modelar la variación simultánea del intercepto con el componente de error asociado se generan sesgos de estimación que inflan la significancia de los predictores y violan uno de los supuestos básicos del teorema de Gauss-Markov.

<sup>28</sup> Aquellos que combinan efectos fijos y aleatorios de manera simultánea.

De acuerdo con la última especificación planteada, si  $\alpha_s^* = \alpha_0 + \mu_s$ , en donde  $\alpha_0$  es un efecto común a todas las escuelas y  $\mu_s$  el error asociado, al no modelar la correlación existente se genera el sesgo mencionado. Por tal razón la especificación incluye de manera explícita los efectos de grupo, capturados a través de  $\bar{F}_{(-i)cs}$ .<sup>29</sup>

De manera general esta configuración jerárquica implica que la proficiencia de los individuos dependa de su vector de características específicas en una primera fase y que, en una segunda fase, capture los efectos específicos relacionados con su entorno escolar o con el background de los compañeros de clase.

La técnica econométrica apropiada en este caso es entonces la de los modelos multinivel<sup>30</sup> ampliamente desarrollados en los trabajos de [Goldstein \(1999, 1993\)](#), [Hox\( 2002\)](#), [Rowe, Hill y Holmes-Smith \(1995\)](#), [Brik y Raudenbush \(1992\)](#) y en los algoritmos recientes de [Rabe-Hesketh, Skrondal y Pickles \(2004\)](#)<sup>31</sup>

Así mismo, la utilización de técnicas multinivel aporta información de gran relevancia sobre la variabilidad total e intra-grupal de la muestra.

#### 3.5.4. Estimaciones

La base de las estimaciones se fundamenta en el modelo previamente planteado y se exploran diversas alternativas en cada uno de los niveles.

$$(3.19) \quad R_{ics} = \alpha_s^* + \lambda F_{ics} + \phi Sch_{cs} + \rho \bar{F}_{(-i)cs} + \mu_{cs} + \varepsilon_{ics}$$

Las **variables** consideradas se definen a continuación:

<sup>29</sup> Se trata de un procedimiento remedial a un problema de variables omitidas.

<sup>30</sup> La bondad de este tratamiento reside en que los puntajes individuales no pueden ser tratados como observaciones aleatorias independientes, sino que dependen del cluster en el que se anidan, en este caso las escuelas y clases.

<sup>31</sup> Estos últimos, creadores del módulo Glamm (Generalized Linear Latent and Mixed Models) recientemente incorporado en el Stata y que tiene la ventaja de explorar diferentes configuraciones de la estructura de covarianzas.



- ❖  $R_{ics}$ : Corresponde a los valores plausibles VP de las pruebas PIRLS en todos los componentes de competencia lectora de los estudiantes de 4º grado de primaria.

Tal y como se anotó, la base estadística de PIRLS proporciona 5 VP (scores) de acuerdo con la incertidumbre asociada al proceso de imputación utilizado. Por tal razón cada bloque de regresiones se estima para cada uno de estos valores con el propósito de evaluar su variabilidad<sup>32</sup>.

La muestra consta de 5131 escolares distribuidos en 147 planteles educativos.

- ❖  $F_{ics}$ : Este vector incluye las características individuales y del background familiar. Aquí se considera el sexo de los estudiantes (*sex*), el indicador de actividades tempranas de alfabetismo en el hogar discutido previamente y que intenta capturar la incidencia de los procesos de estimulación temprana hacia la lectura, así como el nivel educativo de los padres.

El indicador de estimulación temprana es ordinal en la base original y se transformó en dos variables dicotómicas  $ET_m$  (1 actividades medias) y  $ET_b$  (actividades bajas o poca estimulación antes de que los niños iniciaran su ciclo formal de educación), tomando como referencia el nivel más alto de la variable ( $ET_a$ ).

La educación de los padres considera la educación universitaria o superior, postsecundaria, bachillerato completo, secundaria incompleta, primaria.

- ❖  $Sch_{cs}$  incluye al menos una variable relacionada con la calidad del entorno escolar como la disponibilidad de recursos escolares (*RE*) que proviene de la base original de PIRLS.

Este índice tiene en cuenta la disponibilidad de recursos de instrucción, quilificación del cuerpo docente, materiales y estado general de las instalaciones, disponibilidad de ordenadores para instrucción, etc. De acuerdo con la metodología de PIRLS en la elaboración del índice se tiene en cuenta una escala de 4 puntos para cada respuesta según el recurso: 1=nada (no disponible, escaso o inadecuado), 2=poco 3= más o menos

---

<sup>32</sup> Esto implica correr cinco veces cada bloque de regresiones.

(disponibilidad media) y 4=mucho (gran disponibilidad o adecuación de los recursos). El indicador compacto toma la media de respuesta de todas las variables consideradas y elabora 3 categorías que son las que aparecen finalmente en la base estadística: 1=disponibilidad alta de recursos ( $RE_a$ ), 2=disponibilidad media ( $RE_m$ ) y 3= baja disponibilidad ( $RE_b$ ). Los niveles medios y bajos se toman como variables dummy y la referencia corresponde al nivel más alto. De este modo el coeficiente  $\phi$  captura el impacto diferencial respecto al de referencia.

- ❖  $\bar{F}_{(-i)cs}$  es un vector de variables a nivel de grupo. Incluye indicadores relacionados con el background de los compañeros de clase como el nivel educativo de los padres o los niveles de ingreso familiar.

Teniendo en cuenta que las variables proporcionadas por PIRLS son ordinales, mediante procedimientos de agregación por clase y escuela se obtuvo el porcentaje de compañeros en cada categoría de educación de sus padres. El mismo procedimiento se hizo para los ingresos. Así, las variables (alternativas para evitar problemas de colinealidad perfecta) que entran como regresores son:

$PCEP_a$  = porcentaje de compañeros con padres ubicados en las escalas más altas de escolaridad (educación postsecundaria, universitaria o superior)

$PCEP_b$  = porcentaje de compañeros con padres ubicados en las escalas más bajas de escolaridad (secundaria incompleta, primaria o ninguno) y

$PCEP_m$  = porcentaje de compañeros con padres ubicados en la escala media de educación (bachillerato completo).

Un ejercicio análogo se realizó con los ingresos familiares de los compañeros:

$PCY_b$  =porcentaje con ingresos familiares bajos (< 20 mil U\$ año)

$PCY_m$  =porcentaje con ingresos medios (20-60 mil U\$ año) y

$PCY_a$  =porcentaje con ingresos altos (>60 mil U\$ anuales)

Debido a la alta correlación entre los niveles educativos de los padres y los niveles de ingresos, las dos variables entran en modelos distintos.

Los valores de los coeficientes  $\rho$  intentan capturar el impacto que tiene sobre el rendimiento cada punto porcentual de aumento en los atributos mencionados.

De acuerdo con las hipótesis sobre la composición de las varianzas, el intercepto aleatorio puede tomar distintas configuraciones.

Según las hipótesis formuladas anteriormente, como la covarianza es distinta de cero con las dos últimas variables (educación de los padres e ingreso familiar de los pares escolares) los resultados serán sensibles a la especificación exacta de la estructura de la matriz de covarianzas<sup>33</sup>.

Esto depende de la especificación en la estructura multinivel,  $\alpha_s^* = \alpha_0 + \theta \bar{F}_{(-i)cs} + \mu_{cs}$ .

Obsérvese que al suponer  $\theta = 0$  la variabilidad del intercepto depende estrictamente del error entre escuelas, mientras que al suponer que  $\theta \neq 0$  la variabilidad del intercepto aleatorio también dependería de las variables específicas a nivel de grupo lo que implica una especificación más compleja de la matriz de covarianzas.

Finalmente,  $\varepsilon_{ics}$  es el shock estocástico puro a nivel individual.

Las estimaciones se realizan con el módulo *multilevel mixed-effects regression(xtmixed)* de *Stata* incorporado a la interfaz de la versión 9 y en la 10 que tiene su equivalente en el módulo externo *Glamm*<sup>34</sup>. Algunos cálculos se contrastan con las salidas que

<sup>33</sup> Si la estructura de covarianzas es independiente, entonces se tendrían varianzas distintas por cada efecto aleatorio. La estructura de identidad supone que todas las varianzas son iguales y las covarianzas nulas. La no estructurada supone que todas las varianzas y covarianzas son distintas. La definición de una u otra estructura tiene implicaciones sobre la convergencia por máxima verosimilitud. En algunos casos, dependiendo del algoritmo utilizado es posible que los modelos no converjan o que requieran de la utilización de demasiados recursos de memoria. El módulo *xtmixed* models de *Stata 9* y *10* incluyen estas posibilidades de modelamiento con estructura distintas de las covarianzas.

<sup>34</sup> Una referencia completa en [Rabe-Hesketh, Skrondal y Pickles \(2004\)](#).

proporciona el *Systat* v.11 en su módulo avanzado *Mixed Regression: Hierarchical Data*.

Inicialmente se realiza una exploración parsimoniosa de los modelos.

A continuación se presentan las primeras especificaciones y se describen los resultados de las corridas con los primeros valores plausibles (ver [cuadro 3.8](#)).

Es preciso recordar que estos valores plausibles se distribuyen con media de 500 y desviación de 100.

Se estiman 8 modelos con los primeros valores plausibles como variable dependiente y las estimaciones aparecen agrupadas en columnas que exploran la incidencia de las variables previamente señaladas.

La variable sexo aparece en todas las estimaciones debido a que las diferencias entre niños y niñas son significativas en la prueba a favor de las últimas.

La variable edad no se incluyó en razón de que su variabilidad es mínima (niños y niñas de 4° grado de primaria con una media de 10.5 años) y no aporta a la explicación de los puntajes.

La parte inferior de cada bloque de regresiones presenta el número de observaciones individuales incluidas (estudiantes), el número de escuelas (clusters) y el tamaño medio (número de estudiantes por escuela). También aparece la proporción de varianza en cada nivel. Los regresores aparecen agrupados en categorías.

La línea de base o modelo no condicionado, como suele denominarse en la literatura de modelos mixtos, ecuación (3.20), muestra que en una estrategia jerárquica de dos niveles (estudiantes y escuelas) el 53% de la variabilidad (detectada mediante la correlación intra-cluster) está explicada por las características de los estudiantes no incluidos inicialmente, mientras que el 47% restante lo explica las diferencias entre escuelas, columna (1) [cuadro 3.8](#).

$$(3.20) \quad \begin{aligned} R_{ics} &= \alpha_s^* + \varepsilon_{ics} \\ \varepsilon_{ics} &: N(0, \sigma_\varepsilon^2) \\ \alpha_s^* &= \alpha_0^* + \mu_{is} \\ \mu_{is} &: N(0, \sigma_s^2) \end{aligned}$$

Si el modelo se amplía tres niveles, incluyendo la variabilidad en el interior de las clases, la proporción de varianza explicada por las diferencias entre escuelas es de 43%, la variabilidad en el interior de las clase explica el 4.5% y el resto de factores relacionados con las características de los individuos un 52.5%. Estos resultados muestran la importancia relativa que tienen las diferencias entre los planteles educativos en la explicación del rendimiento de los estudiantes.

Ahora bien, siguiendo con la estrategia parsimoniosa ahora se incluyen efectos fijos individuales. La variable sexo entra en el modelo y toma como referencia a los hombres, así mismo se incluye la variable relacionada con la presencia de estimulación temprana en lectura y el nivel educativo de los padres de los escolares.

$$(3.21) \quad \begin{aligned} R_{ics} &= \alpha_s^* + \alpha_c^* + \lambda_1 sex + \lambda_2 ET_m + \lambda_3 ET_b + \lambda_4 EDP_{m4} + EDP_b + \varepsilon_{ics} \\ \varepsilon_{ics} &: N(0, \sigma_\varepsilon^2) \\ \alpha_s^* &= \alpha_{0s}^* + \mu_{is} \\ \alpha_c^* &= \alpha_{0c}^* + \mu_{ic} \\ \mu_{is} &: N(0, \sigma_s^2) \\ \mu_{ic} &: N(0, \sigma_c^2) \end{aligned}$$

*Nivel 1= estudiantes; Nivel 2= escuelas y Nivel 3= clases.*

Los resultados, que pueden apreciarse en todas las columnas del [cuadro 3.8](#), muestran que efectivamente las niñas tienen un impacto diferencial en el rendimiento que oscila alrededor de los 6 puntos, respecto a los niños.

De otra parte, la ausencia o bajo nivel de estímulo a la lectura en los primeros años de infancia, antes de que los niños ingresaran al sistema de escolaridad formal, tiene un impacto negativo y significativo de casi 10 puntos en los scores, respecto al nivel más alto de esta variable ordinal. La intensidad *media* de este indicador proporcionado por

PIRLS no resultó significativa, indicando así que las diferencias respecto al nivel de referencia son nulas.

La educación de los padres muestra fuertes impactos diferenciales. Los niños y niñas que provienen de hogares con educación primaria de sus padres obtienen puntajes por debajo de 26.5 puntos respecto al nivel de referencia (educación universitaria o superior), los que tienen padres con educación secundaria incompleta se rezagan en (-23.7 puntos), los que cuentan con bachillerato en (-10.6 puntos) y los que tienen algún nivel de educación postsecundaria no presentan diferencias significativas, ver columna (3).

Ahora bien, si el modelo incluye un indicador relacionado con el número de libros en el hogar, columna (4), en lugar de la educación de los padres, a fin de explorar una variable que también representa una buena proxy del background familiar y de la presencia de un entorno adecuado que asigne una alta ponderación al ‘mundo de las ideas’ o que expresa de alguna manera la estructura de preferencias por el estudio, los resultados obtenidos son plenamente satisfactorios y significativos.

$$(3.22) \quad (3.23) R_{ics} = \alpha_s^* + \alpha_c^* + \lambda_1 sex + \lambda_2 ET_m + \lambda_3 ET_b + \lambda_4 Books_m + \lambda_4 Books_h + \varepsilon_{ics}$$

$$\varepsilon_{ics} : N(0, \sigma_\varepsilon^2)$$

$$\alpha_s^* = \alpha_{0s}^* + \mu_{is}$$

$$\alpha_c^* = \alpha_{0c}^* + \mu_{ic}$$

$$\mu_{is} : N(0, \sigma_s^2)$$

$$\mu_{ic} : N(0, \sigma_c^2)$$

Los hogares con más de 26 libros en casa, en los rangos definidos del indicador, muestran impactos diferenciales positivos y de gran magnitud, respecto al nivel de referencia (0-11 libros). El nivel 2 que corresponde a los que tienen entre 11 y 25 libros en la casa no resultó significativo. Esta variable que podría operar en algún modelo alternativo como instrumental de los niveles de ingresos familiares corrobora de nuevo el fuerte peso que tiene el background familiar en los logros escolares.

Ahora, la estrategia conduce a la inclusión del indicador que sintetiza la disponibilidad de recursos de las escuelas en las que se anidan los estudiantes, columna (4). Esta

variable ordinal suministrada por PIRLS, tal y como se indicó previamente, se configura en tres niveles alta, media y baja disponibilidad de recursos (materiales, instructores, dotación de ordenadores e instalaciones adecuadas) y brinda una idea sobre el tipo de plantel.

Regularmente, las instituciones públicas suelen presentar escasez o baja provisión de insumos, mientras que las privadas cuentan con mayor disponibilidad. Sin embargo, esta presunción debe tomarse con cautela debido a que infortunadamente los datos de PIRLS no distinguen entre escuelas según su naturaleza jurídica.

Esta especificación en la que la variable entra en la parte fija del modelo como predictor muestra un resultado contundente: los planteles educativos con baja disponibilidad de recursos tienen un impacto diferencial de (-53.5 puntos) respecto a los de mayor dotación que opera como referencia. Los planteles de dotación media en la escala ordinal tienen un impacto diferencial de (-32 puntos) en el indicador de proficiencia, columna (4).

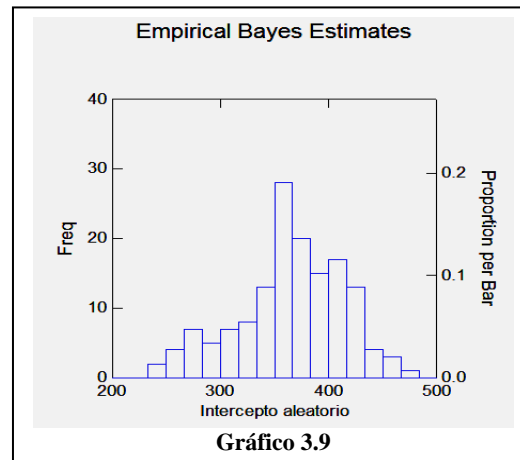
$$(3.24) \quad R_{ics} = \alpha_s^* + \alpha_c^* + \lambda_1 sex + \lambda_2 ET_m + \lambda_3 ET_b + \lambda_4 Books_m + \lambda_5 Books_h + \phi_m RE_m + \phi_b RE_b + \varepsilon_{ics}$$

Este resultado refuerza aún más la argumentación que se ha venido manejando en los diversos apartados de esta Tesis: las diferencias en la dotación media de recursos, en este caso de los planteles educativos, constituye un rasgo notable en la explicación de las diferencias en el rendimiento.

Si la presunción de que los dos extremos alto y bajo de la distribución representan escuelas privadas y públicas respectivamente la conclusión llevaría a reconocer por esta vía la existencia de fuertes inequidades en la provisión de los servicios educativos de calidad.

La misma especificación, pero ahora incluyendo la disponibilidad de recursos en la parte aleatoria del modelo a nivel de escuela arroja resultados similares en los coeficientes de los predictores y disminuye la correlación intra-cluster al 0.30. Esto indica la relevancia del indicador cuya variabilidad ahora queda subsumida en el

intercepto aleatorio del modelo y cuya distribución se puede apreciar en la distribución empírica de Bayes que se desprende de las estimaciones.



Manteniendo la última especificación, pero agregando ahora en la parte fija del modelo el porcentaje de compañeros de clase con padres educados al nivel de bachillerato completo y padres con bajo nivel educativo (primaria y secundaria incompleta), los resultados son atractivos (el nivel más alto se excluye de la corrida para evitar problemas de colinealidad perfecta). Ver columna (5).

$$(3.25) \quad R_{ics} = \alpha_s^* + \alpha_c^* + \lambda_1 sex + \lambda_2 ET_m + \lambda_3 ET_b + \lambda_4 Books_m + \lambda_5 Books_h + \phi_m RE_m + \phi_b RE_b + \rho_m PCEP_m + \rho_b PCEP_b + \varepsilon_{ics}$$

Las dos variables son significativas y con signo negativo: estudiar con compañeros cuyos padres son poco educados disminuye los scores en casi 6 puntos por cada 10 puntos porcentual de aumento en la frecuencia de este perfil de pares escolares. El resultado para el otro nivel es de (-0.473).

Al estimar el mismo modelo incorporando esta última variable en el tercer nivel (clase) la variabilidad intra-cluster de este nivel se acerca a cero sin mayores repercusiones en los coeficientes de los predictores. La importancia de este indicador reside en la aproximación a la magnitud de los efectos *contextuales* en los que desarrollan sus actividades los estudiantes. De acuerdo con estos resultados esta magnitud es alta y sugiere que la composición intra-clase e intra-escolar es relativamente homogénea.



De acuerdo con la estrategia parsimoniosa seguida hasta el momento se plantea una especificación que incluye el nivel educativo de los profesores y la zona en la que se localiza la escuela (urbana, rural o suburbana).

$$(3.26) \quad R_{ics} = \alpha_s^* + \alpha_c^* + \lambda_1 sex + \lambda_2 ET_m + \lambda_3 ET_b + \lambda_4 Books_m + \lambda_5 Books_h + \lambda_6 Subur + \lambda_7 Rur + \phi_m RE_m + \phi_b RE_b + \phi_m Eprof_m + \phi_h Eprof_h + \rho_m PCEP_m + \rho_b PCEP_b + \varepsilon_{ics}$$

Los estudiantes localizados en zonas rurales del país tienen una diferencia de (-29.3 puntos) respecto a los estudiantes de zonas urbanas. La dicotómica para zonas suburbanas no resultó estadísticamente significativa. Ver columna (6).

En relación con la educación de los docentes solamente los niveles más altos resultaron significativos con un impacto diferencial positivo en los puntajes de sus estudiantes de 43 puntos respecto al nivel más bajo. La correlación intra-cluster a nivel de escuela de esta especificación es de 0.28.

Un modelo que finalmente se explora incluye el rango de ingresos de los compañeros de clase, en lugar de la educación de sus padres como predictor. El porcentaje de varianza explicada a nivel de escuela disminuye al 21% y la compañía de niños y niñas provenientes de familias con ingresos bajos repercute negativamente en el rendimiento en la magnitud de (-0.6) respecto al nivel más alto de ingresos. Esto es, el aumento de un 10% en los compañeros pobres hace que el rendimiento disminuya en 6 puntos.

Al incorporar esta misma variable en el componente aleatorio del modelo, a nivel de escuela, la proporción de varianza de este nivel en la varianza total disminuye ostensiblemente, demostrando así que el anidamiento de los perfiles homogéneos de ingreso constituye una variable decisiva a la hora de explicar la variabilidad de los resultados.

Otra especificación que aporta resultados interesantes considera la presencia de una estructura de varianzas no escalares en los dos niveles. Esto es, a nivel de escuela incluye efectos aleatorios en las variables de disponibilidad de recursos y a nivel de clase introduce el porcentaje de compañeros con niveles medios y bajos de ingreso

también como efectos aleatorios. Esta especificación intenta así explicar la variabilidad intra-cluster en los dos niveles superiores (escuelas y clases).

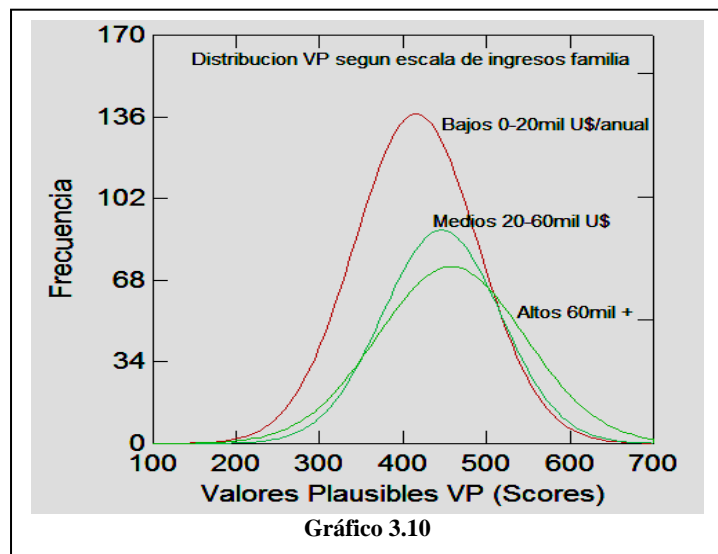
El modelo que incluye de nuevo la educación de los padres a nivel individual y la variable de status económico de los compañeros de clase, toma la siguiente especificación:

$$(3.27) \quad R_{ics} = \alpha_s^* + \alpha_c^* + \lambda_1 sex + \lambda_2 ET_m + \lambda_3 ET_b + \lambda_4 Books_m + \lambda_5 Books_h + \lambda_6 Subur + \lambda_7 Rur + \phi_m RE_m + \phi_b RE_b + \rho_m P Y F_m + \rho_b P Y F_b + \varepsilon_{ics}$$

La estructura de la matriz de covarianzas de este modelo<sup>35</sup> es no escalar debido a que el intercepto aleatorio a nivel de escuela depende de la disponibilidad de recursos escolares y el intercepto aleatorio a nivel de clase queda determinado por el porcentaje de pares escolares según ingresos familiares. La variabilidad a nivel de escuela explica ahora 14.7% de la variabilidad total y a nivel de clase el aporte se reduce ostensiblemente. Ver columna (8).

Estos hallazgos proporcionan una pieza adicional de evidencia a la existencia de segregación socioeconómica en el suministro de los servicios educativos y brindan soporte empírico a la identificación de los factores que determinan los diferenciales del rendimiento entre escuelas.

Finalmente, El [gráfico 3.10](#) muestra la distribución de los valores plausibles según la escala de ingresos familiares. Los grupos de ingresos medios y altos presentan promedios cercanos y los dos se diferencian claramente del grupo de ingresos bajos.



<sup>35</sup> Las estimaciones se realizaron con el módulo xtmixed en Stata versión 9.1.

Cuadro 3.8. Modelos multinivel Variable dependiente: Primeros Valores Plausibles PIRLS

VARIABLES e indicadores \ Regresiones	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sexo (1=femenino, 0=masculino) Sex			5.900** (1.980)	5.764** (1.849)	6.424** (2.004)	5.830** (2.085)	5.468** (2.030)	5.406** (2.035)
<b>Background Familiar</b>								
<b>Estimulación en lectura (Ref = nivel alto)</b>								
Estimulación temprana en lectura (nivel medio, ET <sub>m</sub> )			-1.336 (2.155)	-2.929 (2.002)	-2.256 (2.181)	-1.632 (2.267)	-2.725 (2.190)	-2.826 (2.188)
Estimulación temprana en lectura (nivel bajo, ET <sub>b</sub> )			-9.678*** (2.885)	-10.05*** (2.714)	-10.63*** (2.930)	-10.16*** (3.030)	-10.61*** (2.977)	-10.16** (3.460)
<b>Educación de los padres (Ref. Universitario o superior EDP<sub>1</sub>)</b>								
Post-secundaria (EDP <sub>2</sub> )			-8.723 (4.808)					
Bachillerato (EDP <sub>3</sub> )			-10.57** (3.547)					
Secundaria incompleta (EDP <sub>4</sub> )			-23.71*** (3.654)					
Primaria o ninguno (EDP <sub>5</sub> )			-26.53*** (3.292)					
<b>Posesión de libros en el hogar (Ref=0-11 libros)</b>								
Books (11-25)				1.394 (2.312)	2.382 (2.518)	1.103 (2.615)	0.339 (2.511)	0.637 (3.048)
Books (26-100)				9.298*** (2.517)	9.989*** (2.747)	8.661** (2.819)	8.488** (2.702)	9.045** (3.256)
Books (101-200)				16.44*** (4.333)	18.54*** (4.774)	16.55*** (4.912)	14.37** (4.774)	14.72** (5.315)
Books (>200)				18.27*** (5.031)	19.28*** (5.581)	15.99** (5.995)	15.98** (5.729)	16.13** (6.221)

(...) Continuación	Regresiones	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8) <sup>a</sup>
<b>Recursos de la Escuela (Referencia RE=alto<sub>h</sub>)</b>									
Disponibilidad <b>media</b> de recursos (RE <sub>m</sub> )					-32.10**	-18.25	-8.581	-9.314	-4.122
					(10.63)	(10.16)	(10.01)	(9.049)	(8.360)
Disponibilidad <b>baja</b> de recursos (RE <sub>b</sub> )					-53.48***	-33.69**	-28.89*	-28.84*	-22.63*
					(12.18)	(12.04)	(11.92)	(11.40)	(9.837)
<b>Entorno Intraclase</b>									
Compañeros de clase									
Porcentaje de niños con padres de educación Bachiller (PCEP <sub>m</sub> )						-0.473	-0.722*		
						(0.288)	(0.302)		
Porcentaje de niños con padres secundaria incompleta o primaria (PCEP <sub>b</sub> )						-0.656***	-0.555**		
						(0.163)	(0.173)		
<b>Zona (localización de la escuela, Ref=Urbana)</b>									
Suburbana (Suburb)							-7.531	30.03	30.70
							(32.10)	(39.38)	(27.74)
Rrural (Rur)							-29.31**	-44.25***	-33.71**
							(10.08)	(11.89)	(10.64)
<b>Docentes (Ref= nivel bajo= Eprof<sub>b</sub>)</b>									
Educación docentes (nivel alto=Eprof <sub>h</sub> )							43.69**	15.63	
							(16.91)	(19.19)	
Educación docentes (nivel medio=Eprof <sub>m</sub> )							16.72	5.118	
							(9.996)	(11.12)	
<b>Ingreso familiar de los compañeros de clase (Ref= PYF<sub>b</sub> =Bajo)</b>									
Porcentaje con ingreso familiar anual (30-60mil US) PYF <sub>m</sub>								0.621**	0.846***
								(0.210)	(0.202)
Porcentaje con ingreso familiar anual (>60mil US) PYF <sub>h</sub>								0.923***	0.830***
								(0.177)	(0.188)
<b>Intercepto</b>		421.8***	422.2***	441.4***	449.3***	485.8***	470.8***	395.6***	390.8***

(...) Continuación \ Regresiones	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	(4.801)	(4.788)	(5.332)	(8.842)	(13.79)	(15.84)	(15.80)	(11.41)
Desviaciones estimadas por niveles								
Nivel escuela (Log)	4.026*** (0.0631)	3.982*** (0.0697)	3.895*** (0.0734)	3.852*** (0.0755)	3.730*** (0.0856)	3.666*** (0.0933)	3.388*** (0.137)	2.713*** (0.116)
Residual (Log)	4.095*** (0.0100)	4.082*** (0.0101)	4.078*** (0.0114)	4.078*** (0.0106)	4.083*** (0.0115)	4.080*** (0.0120)	4.079*** (0.0117)	4.068*** (0.0120)
Nivel de clase (Log)		2.861*** (0.141)	2.748*** (0.161)	2.839*** (0.146)	2.811*** (0.161)	2.794*** (0.178)	3.013*** (0.162)	-0.546*** (0.150)
Varianza intracluster nivel de escuela/ Varianza total	0.466	0.429	0.394	0.370	0.314	0.289	0.183	0.147
Varianza intracluster nivel de clase/Varianza total	---	0.046	0.0397	0.049	0.049	0.050	0.087	0.000
Número de individuos <i>N</i>	5131	5131	4035	4626	3981	3677	3810	3902
Número de grupos nivel de escuela	147	147	144	144	141	131	98	99
Número de grupos nivel de clase	---	196	193	193	190	174	138	142
Tamaño medio del grupo nivel escuela	34.9	34.9	28.0	32.1	28.2	28.8	38.9	39.4
Tamaño medio del grupo nivel de clase		26.2	20.9	24.0	21.0	21.1	27.6	27.5

Errores estándar entre paréntesis; Niveles de significancia \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Fuente: Estimaciones propias con base en PIRLS 2001. La variable dependiente corresponde al primer valor plausible de los cinco proporcionados por PIRLS.

(<sup>a</sup>) La regresión (8) incluye en el componente aleatorio lo siguiente: (i) en el nivel de escuela el índice de disponibilidad de recursos escolares y (ii) a nivel de clase el porcentaje de compañeros según ingresos familiares.

### **Exploraciones complementarias con todos los valores plausibles**

Esta parte presenta algunas exploraciones adicionales con todos los valores plausibles de la prueba PIRLS. El [cuadro 3.9](#) presenta el resumen de los principales resultados.

El primer bloque incluye los ingresos familiares en los efectos de los compañeros de clase, y los otros tres grupos de regresiones la educación de los padres.

### **Efectos de la características individuales y del background familiar**

En general el coeficiente de la variable sexo oscila entre los 5 y 8 puntos de impacto diferencial de las niñas respecto a los niños, dependiendo del valor plausible utilizado y de los otros regresores.

El primer modelo del bloque 1, por ejemplo, muestra que las niñas tienen un efecto diferencial positivo de 6.5 puntos sobre VP1 y de 7.97 cuando se utiliza VP5 como variable dependiente. Esta variable es significativa en todos los modelos al 1%.

Los bloques tres y cuatro incluyen el indicador de estimulación temprana en lectura, además de la disponibilidad de recursos escolares y los perfiles educativos de los padres de los compañeros de clase, tal y como se hizo en las estimaciones previas.

La presencia de actividades de lectura en el hogar antes de que los niños iniciaran su ciclo escolar formal muestra que aquellos infantes con niveles bajos de estimulación mediante la lectura de libros, narración de historias, juegos de palabras, etc., presentan un efecto marginal negativo y estadísticamente significativo, respecto a los del nivel alto.

Los coeficientes del nivel medio de este indicador no resultaron significativamente diferentes de cero implicando así que su impacto marginal no es estadísticamente diferente a los del nivel alto.

Cuadro 3.9. Colombia: modelos multinivel complementarios / Variable independiente Valores Plausibles (Scores totales) en las pruebas PIRLS

## Bloque 1

VALORES PLAUSIBLES (SCORES)	VP1			VP2			VP3			VP4			VP5		
	Coef	Std	Sig	Coef	Std	Sig	Coef	Std	Sig	Coef	Std	Sig	Coef	Std	Sig
Sexo (1 =femenino 0= masculino) ( <i>Sex</i> )	5,16	1,78	**	6,57	1,78	**	6,08	1,79	**	6,08	1,78	**	5,47	1,79	**
Disponibilidad de recursos escolares ( <i>Ref: alta</i> )															
Disponibilidad media ( <i>REm</i> )	-23,46	10,45	*	-22,73	10,59	*	-23,40	10,19	*	-23,49	10,26	*	-21,59	10,69	*
Disponibilidad baja ( <i>REb</i> )	-38,85	12,27	**	-37,13	12,43	**	-35,80	11,98	**	-37,25	12,06	**	-37,08	12,55	**
Efectos de los compañeros (ingresos familiares)															
Porcentaje de ingresos familiares medios ( <i>PCYm</i> )	0,92	0,23	**	0,86	0,24	**	0,95	0,23	**	0,94	0,23	**	0,90	0,24	**
Porcentaje de ingresos familiares altos ( <i>PCYa</i> )	0,56	0,17	**	0,57	0,17	**	0,62	0,16	**	0,57	0,16	**	0,53	0,17	**
Intercepto	403,31	13,38	**	403,08	13,55	**	400,06	13,07	**	401,95	13,16	**	404,05	13,67	**
N observaciones individuales	5100			5100			5100			5100			5100		
Ng número de clusters (escuelas)	144			144			144			144			144		
Nm tamaño medio clusters	35,4			35,4			35,4			35,4			35,4		
Correlación intracluster	0,36			0,39			0,36			0,37			0,39		
Wald Chi2	56,07		**	57,32		**	63,37		**	61,13		**	51,63		**
Varianza (constante)	2006,83	299,20		2270,83	308,10		2090,30	284,80		2123,14	287,60		2316,67	314,30	
Varianza (residual)	3594,65	72,27		3595,53	72,29		3644,99	73,30		3602,60	72,40		3604,61	72,50	

## Bloque 2

VALORES PLAUSIBLES (SCORES)	VP1			VP2			VP3			VP4			VP5		
	Coef	Std	Sig	Coef	Std	Sig	Coef	Std	Sig	Coef	Std	Sig	Coef	Std	Sig
Sexo (1 =femenino 0= masculino) ( <i>Sex</i> )	6,64	2,01	**	8,24	2,02	**	7,98	2,03	**	7,98	2,03	**	8,07	2,02	**
Disponibilidad de recursos escolares ( <i>Ref: alta</i> )															
Disponibilidad media ( <i>REm</i> )	-23,58	10,36	*	-24,88	10,57	*	-24,48	10,07	*	-24,48	10,07	*	-22,60	10,66	*
Disponibilidad baja ( <i>REb</i> )	-41,48	12,15	**	-43,15	12,39	**	-37,80	11,82	**	-37,80	11,82	**	-40,85	12,49	**
Efectos de los compañeros (educación de los padres)															
Porcentaje con padres de educación postsecundaria, universitaria o superior ( <i>PCEPa</i> )	0,55	0,14	**	0,44	0,15	**	0,59	0,14	**	0,59	0,14	**	0,50	0,15	**
Porcentaje con padres de educación secundaria completa (Bachillerato) ( <i>PCEPm</i> )	0,26	0,19	(ns)	0,23	0,19	(ns)	0,30	0,19	(ns)	0,30	0,19	(ns)	0,33	0,19	+
Intercepto	425,74	10,10	**	429,89	10,28	**	424,43	9,87	**	424,43	9,87	**	425,88	10,35	**
N observaciones individuales	4032			4032			4032			4032			4032		
Ng número de clusters (escuelas)	141			141			141			141			141		
Nm tamaño medio clusters	28,6			28,6			28,6			28,6			28,6		
Estructura de covarianzas															
Correlación intracluster	0,37			0,37			0,35			0,35			0,38		
Wald Chi2	54,00			50,70			60,98			60,90			53,75		
Varianza (constante)	2102,17	302,90		2196,97	318,90		1963,02	283,60		1963,02	283,60		2238,54	324,40	
Varianza (residual)	3642,02	82,80		3676,29	83,60		3706,77	84,20		3706,77	84,20		3672,24	83,50	

\*\* Significativo al 1%; Significativo al 5%; + Significativo al 10%; (ns) no significativo

Bloque 3 (...)															
VALORES PLAUSIBLES (SCORES)															
Variables	VP1			VP2			VP3			VP4			VP5		
	Coef	Std	Sig	Coef	Std	Sig	Coef	Std	Sig	Coef	Std	Sig	Coef	Std	Sig
Sexo (1 =femenino 0= masculino) ( <i>Sex</i> )	6,54	2,01	**	8,15	2,02	**	7,89	2,03	**	6,95	2,01	**	7,97	2,02	**
Estimulación temprana en lectura (hogar)															
Nivel medio ( <i>ETm</i> )	-2,62	2,17	(ns)	-3,30	2,18	(ns)	-1,40	2,19	(ns)	-2,78	2,18	(ns)	-3,07	2,18	(ns)
Nivel bajo ( <i>ETb</i> )	-13,24	2,89	**	-12,15	2,91	**	-12,28	2,92	**	-13,67	2,90	**	-12,53	2,91	**
Disponibilidad de recursos escolares (Ref: alta)															
Disponibilidad media ( <i>REm</i> )	-23,50	10,38	*	-24,82	10,58	*	-24,39	10,07	*	-25,52	10,32	**	-22,54	10,69	*
Disponibilidad baja ( <i>REb</i> )	-40,11	12,18	**	-41,97	12,40	**	-36,43	11,83	**	-42,60	12,11	**	-39,62	12,52	**
Efectos de los compañeros (educación de los padres)															
Porcentaje con padres de educación postsecundaria, universitaria o superior ( <i>PCEPa</i> )	0,54	0,14	**	0,42	0,15	**	0,57	0,14	**	0,50	0,14	**	0,48	0,15	**
Porcentaje con padres de educación secundaria completa (Bachillerato) ( <i>PCEPm</i> )	0,22	0,19	(ns)	0,18	0,19	(ns)	0,25	0,19	(ns)	0,10	0,19	(ns)	0,28	0,19	(ns)
Intercepto	430,14	10,21	**	434,32	10,38	**	428,02	9,97	**	435,36	10,16	**	430,34	10,46	**
N observaciones individuales	4032			4032			4032			4032			4032		
Ng número de clusters (escuelas)	141,0			141,0			141,0			141,0			141,0		
Nm tamaño medio clusters	28,6			28,6			28,6			28,6			28,6		
Estructura de covarianzas															
Correlación intracluster	0,37			0,38			0,35			0,36			0,38		
Wald Chi2	75,00			68,06			79,60			75,72			72,12		
Varianza (constante)	2112,89	303,50		2203,14	318,90		1967,47	283,50		2083,01	301,00		2252,56	325,50	
Varianza (residual)	3623,58	82,40		3661,47	83,30		3690,71	83,90		3638,36	82,70		3655,95	83,10	
Bloque 4															
VALORES PLAUSIBLES (SCORES)															
Variables	VP1			VP2			VP3			VP4			VP5		
	Coef	Std	Sig	Coef	Std	Sig	Coef	Std	Sig	Coef	Std	Sig	Coef	Std	Sig
Sexo (1 =femenino 0= masculino) ( <i>Sex</i> )	6,60	2,01	**	8,19	2,02	**	7,95	2,02	**	6,97	2,01	**	8,04	2,02	**
Estimulación temprana en lectura (hogar)															
Nivel medio	-2,72	2,17	(ns)	-3,39	2,18	(ns)	-1,53	2,19	(ns)	-2,83	2,17	(ns)	-3,21	2,18	(ns)
Nivel bajo	-13,43	2,89	**	-12,31	2,90	**	-12,51	2,91	**	-13,76	2,89	**	-12,78	2,90	**
Disponibilidad de recursos escolares (Ref: alta)															
Disponibilidad media ( <i>REm</i> )	-23,89	10,47	*	-25,16	10,65	*	-24,83	10,17	*	-25,70	10,35	**	-23,08	10,81	*
Disponibilidad baja ( <i>REb</i> )	-41,51	12,23	**	-43,17	12,44	**	-38,04	11,89	**	-43,24	12,10	**	-41,48	12,61	**
Efectos de los compañeros (educación de los padres)															
Porcentaje con padres de educación postsecundaria, universitaria o superior ( <i>PCEPa</i> )	0,53	0,14	**	0,41	0,15	**	0,57	0,14	**	0,50	0,14	**	0,47	0,15	**
Intercepto	433,78	9,82	**	437,43	9,98	**	432,27	9,59	**	437,02	9,73	**	435,10	10,11	**
N observaciones individuales	4032			4032			4032			4032			4032		
Ng número de clusters (escuelas)	141			141			141			141			141		
Nm tamaño medio clusters	28,6			28,6			28,6			28,6			28,6		
Estructura de covarianzas															
Correlación intracluster	0,37			0,38			0,35			0,37			0,39		
Wald Chi2	72,77			66,49			76,65			75,06			68,84		
Varianza (constante)	2153,794	306,9		2238,378	321,2		2012,827	287,9		2100,209	300,7		2313,257	331,2	
Varianza (residual)	3621,69	82,3		3659,564	83,2		3688,84	83,8		3636,738	82,6		3653,921	83,1	

Estimaciones propias a partir de la base de datos PIRLS 2001

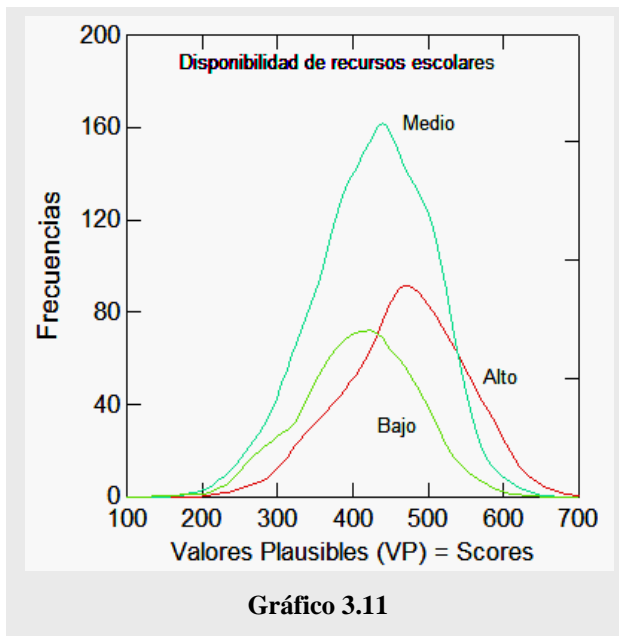
\*\* Significativo al 1%; \* Significativo al 5%; + Significativo al 10%; (ns) no significativo



### **Efectos de los recursos escolares**

Una de las variables de mayor impacto y significancia nuevamente es la disponibilidad de recursos escolares. Los valores de los coeficientes estimados muestran que aquellos planteles con carencias de recursos o con disponibilidad media tienen un impacto marginal negativo muy fuerte sobre los puntajes, tal y como se había encontrado en la estrategia de estimación anterior. Las diferencias observadas respecto al nivel alto (referencia) corroboran que la dotación de recursos (materiales, instructores, dotación de ordenadores e instalaciones adecuadas) repercute vigorosamente en el desempeño de los estudiantes. Los valores estimados son relativamente estables en todas las corridas y oscilan entre (-22.5) y (-24) para el nivel medio del indicador y entre (-36) y (-41.5) para los de nivel bajo, de acuerdo con el valor plausible utilizado.

El [gráfico 3.11](#) muestra la distribución de los puntajes para cada una de las categorías del indicador.



Como se anotó en las estimaciones anteriores, la muestra de PIRLS no distingue entre escuelas públicas y privadas, sin embargo, de acuerdo con los análisis realizados hasta el momento existen indicios de que aquellos planteles con la mayor dotación corresponden precisamente a los privados y que las diferencias pueden revelar de manera indirecta la diferencia entre los dos tipos de planteles.

### **Efectos del background de los compañeros de clase**

Tanto los ingresos familiares como el nivel educativo más alto alcanzado por alguno de los padres de los compañeros de clase resultaron significativos en todas las estimaciones.

El primer bloque de regresiones muestra que el coeficiente asociado a los ingresos medios oscila entre 0.86 y 0.95 según el valor plausible utilizado, mientras que para los ingresos altos se ubica entre 0.53 y 0.57. Esto indica que los niños o niñas con compañeros que pertenecen a familias de ingresos medios reciben un beneficio externo positivo entre 8.6 y 9.5 puntos adicionales por cada 10 puntos porcentuales de aumento en la frecuencia de este perfil de ingresos. La magnitud de la externalidad generada por los compañeros provenientes de familias de ingresos altos es sistemáticamente inferior aunque también positiva y significativa.

Los bloques 2, 3 y 4 incluyen el background educativo de los padres en lugar de los ingresos. Los coeficientes relacionados con el porcentaje de padres según perfiles educativos fluctúan entre 0.41 y 0.57 para los de educación alta (postsecundaria, universitaria o superior) y entre 0.19-0.33, de acuerdo con el score utilizado como variable dependiente.

### **3.6. Conclusiones parciales**

El capítulo avanzó en su primera parte en el estudio de los determinantes de la elección de escuelas (públicas y privadas) en Colombia y exploró entre sus determinantes el papel que juegan la distribución intra-metropolitana del capital humano, los niveles de escolaridad de los padres y un índice de sorting educativo desde la perspectiva del gasto educativo.

Entre los resultados de mayor relevancia se destacan los siguientes:

- ❖ Las estimaciones proporcionaron evidencia favorable sobre la repercusión que tiene el origen urbano o rural de los padres, así como su nivel educativo. Las probabilidades obtenidas en las corridas econométricas indican que 77 de cada 100 niños con padres urbanos y con bachillerato completo o superior asisten probablemente a planteles privados, mientras que los niños que provienen de hogares en los que el padre es de origen rural y con nivel educativo de secundaria completa tienen una probabilidad inferior de asistir a los planteles privados.

- ❖ Las variables relacionadas con la distribución del capital humano calificado y no calificado en los ámbitos intra-metropolitanos mostraron que los niños que viven en entornos socio-demográficos de alta composición de población activa calificada asisten fundamentalmente a escuelas y colegios privados. 95 de cada 100 niños con padres de origen urbano, nivel educativo alto y que vive en un entorno socio-demográfico en el que proliferan vecinos de alta calificación asisten fundamentalmente a planteles privados, mientras que 81 niños de cada 100 con padres de origen campesino dotados de niveles educativos bajos y rodeados de vecinos con escolaridad precaria asisten a planteles públicos.

Las estimaciones logran reforzar la hipótesis de la presencia de segregación socioeconómica en esta estructura de decisiones. Los más pobres asisten fundamentalmente a escuelas y colegios públicos, mientras que los que no tienen esta condición —de acuerdo con la línea de pobreza— asisten a planteles privados. Aunque los individuos no votan en estricto con los pies para alcanzar un equilibrio espacial perfectamente estratificado, tal y como lo predicen los modelos del tipo Tiebout, si existe evidencia a favor de un arreglo socioeconómico que emula los mecanismos de sorting que imperan en los modelos de multicomunidad. Las clases socioeconómicas medias, de su parte, establecen una mixtura entre planteles públicos y privados.

Estos resultados muestran que la educación de buena calidad opera como un bien de club que concentra a los más pudientes, los más pobres se ubican en escuelas públicas y la clase media distribuye sus elecciones entre escuelas públicas y privadas con una amplia variedad de precios y calidades.

- ❖ En relación con la calidad. Los colegios privados, a pesar de que en promedio exhiben resultados superiores en las pruebas del ICFES a los establecimientos públicos, presentan scores medios todavía deficientes. La mayoría de instituciones pública y privadas se ubicaron en el rango medio (30-70 en la escala 0-100 del ICFES) y solamente un pequeño puñado de planteles privados orientados hacia estratos de ingresos altos logró ubicarse en los niveles superiores de competencias en matemáticas, lenguaje e inglés.

La parte final del capítulo exploró el efecto de los entornos de aprendizaje sobre la calidad educativa, los logros y las competencias académicas de los estudiantes de grado 4° de primaria con los datos de la prueba PIRLS. La mirada integral de las características del entorno familiar, la calidad del entorno escolar, los perfiles socioeconómicos de los compañeros de clase y la formación de los profesores proporcionó evidencia sobre la existencia de efectos correlacionados, contextuales y endógenos. Entre los principales hallazgos de los ejercicios se pueden destacar los siguientes:

- ❖ Los ejercicios econométricos con estructuras multinivel (modelos mixtos de efectos fijos y aleatorios con datos anidados) mostraron que al estimar el modelo no condicionado (sin variables explicativas) la correlación intra-cluster en dos niveles (individuos y escuelas) fue de 0.47 y la residual de 0.53. Estos resultados revelaron en principio que la variabilidad en la proficiencia proviene en una alta proporción de las diferencias entre escuelas y de los factores asociados a los estudiantes y al entorno escolar.
- ❖ Las niñas presentan un impacto diferencial positivo en las pruebas que se ubica en aproximadamente 5 puntos respecto a los niños.

#### *Background familiar*

- ❖ Todos los modelos estimados mostraron evidencia sobre la importancia que tiene el trabajo de estimulación hacia la lectura en el hogar, antes de que los niños inicien su ciclo escolar formal. La presencia de actividades de lectura en el hogar antes de incorporar a los infantes a la escuela resultó significativa y muestra que los niños y niñas con niveles bajos de estimulación temprana (lectura de libros, narración de historias, juegos de palabras, etc.) presentan una fuerte desventaja respecto a los del nivel alto. Este resultado es consistente con algunos de los postulados de los nuevos enfoques del ciclo vital ([Cunha, et., al., 2005](#)).
- ❖ Adicionalmente, las variables de educación de los padres resultaron significativas y corroboran la presencia de un canal directo de transmisión de habilidades de los padres a los hijos. De igual manera, los perfiles de ingresos de las familias resultaron también significativos en las corridas econométricas

complementarias y muestran efectivamente que la dotación media de recursos juega un papel decisivo a la hora de explicar el desempeño académico de los estudiantes.

#### Efectos correlacionados

- ❖ En relación con la dotación media de insumos escolares, la disponibilidad de recursos en las escuelas (materiales, instructores, dotación de ordenadores e instalaciones adecuadas) fue precisamente una de las variables de mayor impacto en todas las corridas econométricas. Los valores de los coeficientes estimados muestran que aquellos planteles con carencias de recursos o con disponibilidad baja, tienen un impacto marginal negativo muy fuerte sobre los puntajes, respecto a las escuelas con niveles altos de recursos. Esta variable aporta evidencia sobre la magnitud de los efectos correlacionados por la vía del ambiente escolar o clima educativo de los planteles. Su magnitud depende del conjunto de regresores incluidos en las estimaciones y la especificación adoptada. En las corridas econométricas los impactos diferenciales de los planteles con dotación alta y media de recursos, respecto a los de menor disponibilidad resultaron significativos y con magnitudes relativamente altas: por ejemplo, en la regresión (4) del [cuadro 3.8](#) que incluye solamente variables del background familiar y la disponibilidad de recursos escolares, el impacto diferencial sobre los puntajes, por el hecho de estudiar en un plantel con baja dotación de recursos es de -53.5 puntos, respecto a los planteles de mayor disponibilidad de estos recursos; al incluir otras variables relacionadas el background de los compañeros de clase, el impacto diferencial de la disponibilidad baja de recursos es de -33.7 puntos y al incluir la zona en la que se localiza la escuela (urbana, rural o suburbana) el impacto diferencial respecto a los planteles más ricos es de -28.9 puntos.

#### Efectos contextuales y endógenos

- ❖ En relación con los *pares* de la clase, los resultados indican que los niños o niñas con compañeros que pertenecen a familias de ingresos medios reciben un beneficio externo positivo cercano a un punto adicional por cada punto porcentual de aumento en la frecuencia de este perfil de ingresos. Los mismo ejercicios, pero incluyendo la educación de los padres de los compañeros de

clase, en lugar de los rangos de ingreso de sus familias, también arrojó resultados consistentes con los planteamientos teóricos. La conclusión que se desprende es que compartir aula con compañeros de padres educados (educación postsecundaria, universitaria o superior) genera impactos positivos sobre los resultados. Por ejemplo, de acuerdo con las estimaciones de la regresión (7) -- [cuadro 3.8](#)—por cada punto porcentual de aumento en el perfil de ingresos altos de los compañeros, los scores del individuo  $i$  aumentan en 0.9 puntos<sup>36</sup> y el aumento que generan los perfiles de ingresos medios llega a 0.6 puntos. Estos resultados proporcionan evidencia sobre la magnitud de los efectos *contextuales*, asociados a la calidad de los compañeros de clase o al clima sociodemográfico medio, intra-clase, que enfrentan los educandos.

- ❖ Es preciso señalar finalmente que de acuerdo con la estrategia utilizada, los efectos *endógenos puros* no pueden identificarse de manera aislada. Así, la influencia neta que ejercen los compañeros de clase sobre los resultados de un individuo  $i$  en la prueba no pueden en estricto diferenciarse de los impactos exógenos de su background familiar<sup>37</sup>. Por tal razón, es necesario tener en cuenta que los efectos contextuales señalados previamente incluyen este componente endógeno y no se puede desligar de los perfiles sociodemográficos y económicos de los pares escolares (efectos *contextuales puros*) y que no sabemos a ciencia cierta cuál es su magnitud. Tal y como se desprende del planteamiento de la ecuación (3.15), aunque hubiese sido deseable calcular la contribución individual de cada uno de estos factores, la naturaleza “ordinal” de los regresores dificultó el cálculo indirecto sugerido en la ecuación (3.16) y, en consecuencia, los ejercicios econométricos capturan la incidencia de los factores sociales como conjunto, lo que constituía precisamente uno de los objetivos básicos trazado inicialmente en esta investigación.

---

<sup>36</sup> Esto es, por cada 10 puntos porcentuales de aumento en la proporción de compañeros de clase ricos, los puntajes obtenidos por el individuo  $i$  en la prueba PIRLS aumentan en 9 puntos.

<sup>37</sup> Valga anotar que la correlación de Pearson entre los puntajes de cada individuo y el valor medio de sus compañeros de clase se ubica en el entorno de 0.65 cuando se utilizan los cinco valores plausibles proporcionados por PIRLS.

### 3.7. Bibliografía y Referencias

Albright y Conley (Eds) (2004). *After the Bell – Family Background, Public Policy, and Educational Success*. Routledge advances in Sociology, U.S.A.

Ammermueller y Pischke Jörn-Steffen (2006). Peer effectst in european primary schools: evidence from PIRLS. National Bureau of Economic Research, Working paper series 12180. Cambridge, MA, April.

Barrera, F. y Gaviria, A. (2003). Efficiency of Colombian schools. *Fedesarrollo Documento de trabajo*, (mimeo), septiembre.

Bayer, Patrick, Robert McMillan, and Kim Rueben, (2004). An Equilibrium Model of Sorting in an Urban Housing Market. Mimeo, Yale University.

Bénabou, R., (1996a). Equity and Efficiency in Human Capital Investment: The Local Connection. *Review of Economic Studies*, 63, 2, 237-264.

Bénabou, R., (1996b). Heterogeneity, Stratification, and Growth: Macroeconomic Implications of Community Structure and School Finance. *American Economic Review*, 86, 3, 584-609.

Ben-Porath, Yoram. (1967). The Production of Human Capital and the Life Cycle Earnings, *Journal of Political Economy*, 75(4), pp. 352-365.

Brik A, Raudenbush, S. W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage Publications, California.

Burgess, S., et., al., (2004). Sorting and Choice in English Secondary Schools, CMPO, The University of Bristol. *Working Paper Series* No. 04/111

Coleman, J. S., E. Campbell, C. Hobson, J. McPartland, A. Mood, F. D. Weinfeld, and R. York (1966). *Equality of Educational Opportunity*, Washington, DC: Department of Health, Education and Welfare.

Checchi, D. (2006). *The Economics of Education*. (Reimprinted 2007). Cambridge. University Press. 278pp.

Cunha, F, Heckman, J., Lance Lochner y Dimitry Masterov (2005). Interpreting the evidence on life cycle skill formation. National Bureau of Economic Research, NBER, Working Paper 11331, Mayo.

De Fraja, G. (2001). Education policies: equity, efficiency and voting equilibria. *Economic Journal* 111, 104-119.

Ellickson, Bryan, (1971). Jurisdictional Fragmentation and Residential Choice. *American Economic Review*, 61(2), 334-339

Epple, D., and T. Romano (2000). Neighborhood Schools, Choice, and the Distribution of Educational Benefits, *NBER Working Paper* 7850.

Epple, D., y Romano, R. (1998). Competition between private and public schools, couchers, and peer-group effects. *American Economic Review* 88 (1), 33-62.

Fernandez, R. (2000). Sorting, Education and Inequality. *NBER Working Paper* 8101.

Fernandez, R., and R. Rogerson (1997). Keeping People Out: Income Distribution, Zoning and the Quality of Public Education. *International Economic Review*, 38, 23-42.

Fernandez, R., and R. Rogerson (1998). Public Education and Income Distribution: A Dynamic Quantitative Evaluation of Education Finance Reform. *American Economic Review*, 88(4), 813-833.

Gaviria, A. y Barrientos, J. (2001). Determinantes de la calidad de la educación en Colombia. Departamento Nacional de Planeación, *Archivos de Economía* 159, noviembre.

Glomm, G., and B. Ravikumar, (1992). Public versus Private Investment in Human Capital: Endogenous Growth and Income Inequality. *Journal of Political Economy*, 100(4), 818-834.

Goldstein, Harvey (1999). *Multilevel Statistical Models*. Institute of Education, Multilevel models project, London. <http://www.arnoldpublishers.com/support/goldstein.htm>.

Goldstein,H., Rasbash,J., Yang,M., Woodhouse,G., Pan,H., Nuttall,D., and Thomas,S. (1993). A multilevel analysis of school examination results. *Oxford review of education*, 19, 425-33.

Hox, Joop (2002). *Multilevel Analysis: Techniques and Applications*. Lawrence Erlbaum Associates, Publishers, London.

Hoxby, C. (1996a). How teachers' unions affect education production. *Quarterly Journal of Economics* 111 (3), 671–718.

Hoxby, C. (1996b). Are efficiency and equity in school finance substitutes or complements? *Journal of Economic Perspectives* 10 (4), 51–72.

Hoxby, C. (2000a). The effects of class size on student achievement: new evidence from population variation. *Quarterly Journal of Economics* 115 (4), 1239–85.

Hoxby, C. (2000b). *Peer Effect in the Classroom: Learning from Gender and Race Variation*. Working Paper no. 7867, National Bureau of Economic Research, Cambridge, MA.

Jenkins, S., et., al.,(2006), Social Segregation in Secondary Schools: How Does England Compare with Other Countries?, IZA Discussion Paper No. 1959 January



Linver, Miriam R., Sidle, A. y Jeanne Brooks-Gunn (2004). How do parents matter? En Albright y Conley (Eds)(2006). *After the Bell – Family Background, Public Policy, and Educational Success*, Routledge advances in Sociology, Cap. I.

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

Manski, Charles (1995). *Identification problems in the social science*. Harvard University Press, USA.

Manski, Charles (1993). Identification of endogenous social effects: The reflection problem, *The Review of Economic Studies* 60(3), 531-542.

Mayer, S. (2002). How Economics Segregation Affects Children's Educational Attainment. *Social Forces*, Vol. 81, No. 1, 153-176. Sept.

Merton R.K. y Kitt, A.S. (1950). Contributions to the theory of reference group behavior. En R.K. Merton y P. Lazarsfeld (Eds.), *Continuities in social research. Studies in the scope and method of "The American soldier"*. Glencoe: Free Press.

Nechyba, T. (2003). School Finance, Spatial Income Segregation and the Nature of Communities. *Journal of Urban Economics*, 54, 1, pp. 61-88

Nechyba, T.J. (2000). Mobility, Targeting, and Private-School Vouchers. *American Economic Review*, 90(1), 130-46.

Nechyba, T. (1999). School Finance Induced Migration Patterns: the Impact of Private School Vouchers. *Journal of Public Economic Theory*, 1, 1, pp.5-50

Núñez, J.; Steiner, R.; Cadena, X. y Pardo, R. (2002). ¿Cuáles colegios ofrecen mejor educación en Colombia? Departamento Nacional de Planeación, *Archivos de Economía* 193, junio.

Piñeros, L. y Rodríguez, A. (1998). Los insumos escolares en la educación secundaria y su efecto sobre el rendimiento académico de los estudiantes: Un estudio en Colombia. Banco Mundial, Departamento de Desarrollo Humano, *LCSHD Paper Series* No. 36, diciembre.

Rabe-Hesketh, Skrondal y Pickles (2004). *GLAMM Manual*. Working paper series 160, U.C. Berkeley Division of Biostatistics. <http://www.gllamm.org>.

Ross, S., and J. Yinger (1999). Sorting and Voting: A Review of the Literature on Urban Public Finance. En E.S. Mills and P. Cheshire, (eds.), *Handbook of Regional and Urban Economics*, Elsevier Science Publishers: North Holland.

Rowe K.J, Hill, P. W., Holmes-Smith, P. (1995). "Methodological Issues in educational Performance and School Effectiveness Research: A Discussion with Worked Examples". *Australian Journal of Education* V. 39 N° 3, pp. 217-248.

Sen, Amartya (1992). *Inequality Reexamined*. Cambridge, mass., Harvard University Press.

Sen, A. (1982). *Poverty and Famines : An Essay on Entitlements and Deprivation*. Oxford, Clarendon Press.

Stiglitz, J. (1974). The demand for education in public and private school system. *Journal of Public Economics* 55 (3), 349-385.

Tiebout, C. (1956). A Pure Theory of Local Expenditures. *Journal of Political Economy*, 64, 416-24.

Westhoff, Frank, (1977). Existence of Equilibria in Economies with a Local Public Good. *Journal of Economic Theory*, 14, 84-112.