



CENTRO DE
POLÍTICAS
PÚBLICAS UC

TEMAS DE LA AGENDA PÚBLICA

Clasificación de escuelas en la nueva institucionalidad educativa: contribución de modelos de valor agregado para una responsabilización justa

ERNESTO SAN MARTÍN

Facultad de Educación y Facultad de Matemáticas

ALEJANDRO CARRASCO

Facultad de Educación



Clasificación de escuelas en la nueva institucionalidad educativa: contribución de modelos de valor agregado para una responsabilización justa*¹

ERNESTO SAN MARTÍN

Facultad de Educación y Facultad de Matemáticas

ALEJANDRO CARRASCO

Facultad de Educación

Introducción

Chile ha dado paso a una nueva generación de reformas mediante cambios en la institucionalidad del sistema escolar que empleará mecanismos de rendición de cuentas y presión por resultados. La reciente aprobación por parte del Congreso Nacional de la Ley N° 20.529 que introduce un Sistema Nacional de Aseguramiento de la Calidad de la Educación tiene en su estructura fundamental la creación de dos nuevos organismos: la Superintendencia de Educación² y la Agencia de Calidad de la Educación. Esta última constituye una estrategia de mejora y cambio educacional que apunta a transformar la relación existente entre Estado, familias y establecimientos educacionales, mediante el uso intensivo de examinación censal de estudiantes (SIMCE), difusión pública de resultados y categorización de escuelas. En teoría, el incumplimiento de las metas de desempeño tendrá consecuencias no solo ante las medidas que el Estado establezca sobre las escuelas, sino también relacionadas al ejercicio de libertad de elección de las familias, que

se espera robustecer ante la mayor disponibilidad de información sobre el desempeño de los establecimientos³.

Todo lo anterior será posible dado que la nueva Agencia de Calidad de la Educación deberá clasificar a los establecimientos educacionales en términos de su efectividad escolar. A partir de dicha clasificación, un conjunto de medidas y consecuencias podrían recaer sobre los establecimientos que no cumplan los estándares educacionales instaurados por la autoridad. Sin embargo, más allá de este propósito general, la implementación específica de la ley supone una serie de aspectos conceptuales y metodológicos que requieren detenida discusión y tratamiento. Dadas las múltiples alternativas disponibles, la clasificación que realice la Agencia podría potencialmente juzgar y responsabilizar injustamente a algunos establecimientos escolares, ya sea subestimando su desempeño, o bien sobrestimando el de aquellas escuelas que solo en apariencia obtienen buenos resultados.

La ley aprobada ha indicado un camino en esta dirección al establecer que las escuelas serán clasificadas dentro

* Una versión preliminar de este documento fue presentada en una mesa de discusión organizada por el Centro de Políticas Públicas UC y realizada el 20 de junio de 2012. En la ocasión, comentaron el artículo José Weinstein, gerente del Centro de Innovación en Educación de Fundación Chile; Francisco Lagos, jefe del Centro de Estudios del MINEDUC y Claudia Peirano, directora de Grupo Educativo. Se agradece la participación de los comentaristas así como del público asistente, cuyas observaciones aportaron a la versión definitiva. La responsabilidad del texto final es total y exclusivamente de los autores.

1 Este artículo se enmarca en el Proyecto FONDECYT N° 1110315 *School Effectiveness and Value Added Models: From Quantitative Analysis to Qualitative Outcomes*, de E. San Martín y A. Carrasco.

2 Sus tareas se relacionan con una mayor fiscalización de las prácticas administrativas y contables de los establecimientos que reciben aportes del Estado. A modo de ejemplo, inéditamente, en abril de 2012 el Ministerio de Educación revocó el reconocimiento oficial a 37 establecimientos particulares subvencionados que transgredieron un conjunto de normativas previsionales, administrativas, laborales y de infraestructura. Se espera que la Superintendencia actúe en una dirección semejante.

3 Beyer *et al.* (2000), argumentaron durante los noventa que una de las falencias más graves de nuestro sistema escolar era que los establecimientos escolares no enfrentaban consecuencias por sus desempeños. La nueva institucionalidad educacional recoge ese diagnóstico y lo materializa mediante esta política pública.

de cuatro grupos según sus niveles de desempeño, para lo cual los estándares de aprendizaje alcanzados por los establecimientos medidos en pruebas censales (SIMCE) tendrán una ponderación no inferior al 67% del total (Ley N° 20.529, artículo 18). Además, la ley plantea que el procedimiento para realizar la clasificación podría eventualmente incluir modelos de valor agregado: “Para llevar a cabo esta ordenación, la Agencia deberá considerar los resultados de aprendizaje de todas las áreas evaluadas censalmente en las mediciones nacionales, la distribución de los resultados de los alumnos en relación con los estándares de aprendizaje y el grado de cumplimiento de los otros indicadores de calidad educativa. También deberá considerar las características de los alumnos del establecimiento educacional, incluida, entre otras, su vulnerabilidad, y, cuando proceda, indicadores de progreso o de valor agregado” (artículo 17).

En consecuencia, el proceso de clasificación de escuelas específico que defina la Agencia es contingente a las definiciones de modelización y estimación estadística que adopte, así como a la disponibilidad de datos, pero especialmente, importa recalcar que tales definiciones tendrán claras consecuencias para el sistema escolar y para las escuelas, las que no son neutrales en términos éticos. En este marco, este trabajo intenta anticipar alternativas técnicas e implicancias de política pública. Particularmente, porque el papel de clasificación de desempeños escolares a nivel establecimiento que tendrá la Agencia corresponde a lo que en la literatura sobre efectividad escolar se ha denominado durante décadas “efecto-escuela”. En consideración a la experiencia internacional, la Agencia deberá, en primer lugar, adoptar y ofrecer una conceptualización en base a la cual se estimará el efecto-escuela. En segundo término, deberá emplear metodologías de estimación que repercutan en clasificaciones justas, en el sentido que reflejen válidamente la contribución neta de los establecimientos al progreso escolar de los estudiantes.

El fundamento último de los mecanismos de rendición de cuentas supone que las escuelas son juzgadas respecto de aquellas acciones específicas que influyen en los aprendizajes y que están bajo su control e influencia. La investigación educacional acumulada por décadas demuestra que el aprendizaje escolar es resultado de una combinación de factores tanto endógenos (atribuibles a la acción de la escuela) como a variantes exógenas

(familia y entorno social). Al juzgar el desempeño de las escuelas, entonces, se está haciendo referencia a las dimensiones endógenas de su práctica a nivel de aula y escuela⁴. Si, por el contrario, la sociedad juzga el desempeño de una escuela a partir de aquellas dimensiones que están más allá de su control (pero que igualmente influyen en el nivel de aprendizaje de los niños), entonces se estaría incurriendo no solo en una estimación estadística sesgada del efecto-escuela, sino también en una responsabilización injusta respecto del real desempeño de las instituciones escolares.

Estimar el efecto-escuela constituye uno de los desafíos metodológicos de mayor complejidad de la investigación educacional contemporánea. La Agencia de Calidad, que hará estimaciones del efecto-escuela a gran escala, deberá enfrentar y resolver tales desafíos técnicos, pues su accionar tendrá repercusiones significativas sobre el funcionamiento y futuro de los establecimientos escolares. Ello, dado que las medidas o consecuencias van desde la intervención organizacional hasta el cierre definitivo de establecimientos. Importa subrayar nuevamente que algunas variantes metodológicas consideradas para clasificar escuelas podrían tener repercusiones éticas una vez que estén en régimen, si subestiman o sobrestiman injustamente el accionar de los establecimientos.

El presente artículo aspira a contribuir a este debate mediante una ilustración empírica de las marcadas diferencias en la posición que obtienen las escuelas en un potencial esquema de clasificación que resultan de la utilización de un modelo de valor agregado, respecto de la posición obtenida si se emplean promedios SIMCE a nivel de escuelas.

Visión comparada de sistemas de clasificación de escuelas

La experiencia internacional indica que la implementación de políticas educativas basadas en examinación intensiva del currículum nacional (*testing*), estándares nacionales y *accountability* o rendición de cuentas, como las que Chile está iniciando, trae consigo desafíos metodológicos de enorme envergadura relacionados con la estimación estadística del efecto-escuela que está a la base de todo sistema de comparación y clasificación de establecimientos educacionales.

Chile no es pionero en la implementación de sistemas

⁴ La complejidad adicional es que, al mismo tiempo, esas dimensiones endógenas no parecen operar de manera totalmente independiente a las dimensiones exógenas o contextuales (Thrupp, 1995, 1999; Thrupp *et al.*, 2002).

de rendición de cuentas y existe bastante experiencia acumulada internacionalmente, tanto a nivel de política pública como en la evidencia científica. Es probable que el sistema educacional más emblemático en esta materia sea el de Reino Unido, que por primera vez en 1986 estableció las bases del sistema de *rankings* o clasificación de escuelas, denominado *League Tables*, mediante la publicación de los resultados promedio de establecimientos secundarios reportados como indicativos de la calidad de los establecimientos británicos. A ello le siguió, a inicios de los noventa, la institucionalización de esta lógica de clasificación en la *Office for Standards in Education* (Ofsted), similar de la actual Agencia de Calidad en Chile. La Ofsted intensificó, extendió y profundizó estos mecanismos, introduciendo mediciones estandarizadas en los estudiantes de 7, 11 y 14 años (*Key Stage Tests*), mediante lo cual se clasificaron escuelas, divulgaron resultados y establecieron consecuencias.

Sin embargo, contundente y convergente evidencia, incluyendo un informe de la *Royal Statistical Society* y un estudio del Instituto de Educación de la Universidad de Londres (Goldstein *et al.*, 2000) comisionado por la misma Ofsted, demostró que las metodologías de clasificación empleadas por este organismo se basaban en estimaciones desajustadas, siendo catalogados como injustas, sesgadas y estadísticamente inapropiadas (Goldstein y Spiegelhalter, 1996; Fitz-Gibbon, 1997). En su informe (Goldstein *et al.*, 2000) y en Goldstein (2001), al comparar la clasificación que obtienen las escuelas según el método de estimación que se use, los expertos demuestran que la posición de las escuelas en los esquemas de clasificación es modificada, corregida y mejor explicada al usar modelos de valor agregado⁵.

Inglaterra ha continuado hasta hoy usando *League Tables*, incorporando por primera vez en el año 2002, para estudiantes de secundaria, metodologías de valor agregado contextualizadas tras un consenso producido respecto a las ventajas de estos modelos para capturar y controlar con mayor fuerza los efectos exógenos en las prácticas internas de las escuelas, que explican igualmente los rendimientos de los estudiantes (Sammons, 2012; Lyuten y Sammons, 2010; van de Grift, 2009; Goldstein, 2008; Gray *et al.*, 2001)⁶.

En el resto de Europa, un reciente informe de la OECD

(2008) detalla la manera en que países como Holanda, Bélgica y Polonia han incorporado metodologías de valor agregado en sus sistemas educacionales para fortalecer la validez de la implementación de estrategias de políticas relativas a elección parental de escuelas, sistemas de *accountability* y la estrategias de mejoramiento escolar. En Estados Unidos, un reciente informe de la *National Research Council* y la *National Academy of Education* (Braun *et al.*, 2010) ofrece evidencia y argumentos detallados tanto sobre las ventajas metodológicas y éticas del uso de estos modelos, especialmente en sistemas educacionales que poseen políticas de rendición de cuentas, así como respecto al conjunto de dificultades técnicas de estimación pendientes que deben abordar los expertos y las consideraciones que el uso e interpretación de esta metodología demanda a las autoridades educacionales.

De este modo, la experiencia internacional parece recomendar a los países que eligen estrategias de reforma educacional basadas en rendición de cuentas, decisiones de política pública que aseguren una responsabilización justa en sus metodologías de modelización, estimación y clasificación del desempeño de los establecimientos. Chile inicia una nueva fase en sus estrategias de cambio educacional y ha optado por un tipo de reforma educativa global que demanda alta sofisticación técnica y consideración ética al implementar nuevos dispositivos de presión, control y monitoreo.

Evolución de la investigación en efectividad escolar y difusión de resultados

La investigación en efectividad escolar emerge en países como Holanda, Inglaterra, Bélgica, Estados Unidos y Australia como campo de estudio en respuesta a dos décadas (1960 y 1970) de investigación sociológica cuya evidencia indicaba que el desempeño escolar era explicado en mayor proporción tanto por los antecedentes socioeconómicos de los alumnos, así como por la composición social de las escuelas (Coleman *et al.*, 1966). En reacción a estos estudios iniciales, emergió la evidencia fundacional de la investigación en efectividad escolar.

En Estados Unidos (Edmonds, 1979) y el Reino Unido (Rutter *et al.*, 1979), los investigadores concluyeron que una vez que los antecedentes socioeconómicos eran controlados, factores intraescuela explicaban una con-

5 Como consecuencia, los gobiernos de Irlanda, Escocia y Gales decidieron hace más de una década ir más lejos y eliminar no solo la clasificación de escuelas basada en exámenes de desempeño, sino también la examinación de aprendizajes estandarizada y censal, remplazándola por mediciones muestrales cuyos resultados no son públicos ni asociados a consecuencias para las escuelas.

6 Ver www.education.gov.uk/performance/tables/pilotks4_05/aboutcva.shtml

siderable porción de las diferencias de efectividad entre establecimientos: liderazgo, visión y metas, ambiente de aprendizaje, foco enseñanza y aprendizaje, altas expectativas en la capacidad de aprender de los estudiantes, vinculación escuela y familia, entre otras (Mortimore *et al.*, 1988; Sammons *et al.*, 1995; Teddlie y Reynolds, 2000). Este cuerpo de evidencia contribuyó a expandir internacionalmente los supuestos esenciales de la investigación en efectividad escolar: que las escuelas pueden hacer la diferencia con independencia de los factores extra escuela. La capacidad persuasiva de estos hallazgos es notoria en la decisiva influencia que han tenido en una serie de políticas y reformas globales de cambio y mejora escolar (inspección, estándares, rendición de cuentas, medición con consecuencias)⁷. Recientemente, sin embargo, la evidencia en este campo converge en que el desempeño escolar se explica en una importante proporción por una serie de influencias exógenas, lo que ha repercutido en la sugerencia de renombrar este campo de análisis como *Educational Effectiveness Research* (en lugar de *school*), donde la escuela aparece como una entre muchas otras influencias que explican los resultados escolares —junto a otras como capitales de familias nucleares, privación barrial, políticas nacionales, sistemas de apoyo locales, autoridades educativas intermedias, fundaciones educacionales, redes pedagógicas interesuelas, entre otras (Chapman *et al.*, 2012).

En la avanzada de este campo de estudio hay consenso en que las metodologías de valor agregado (VA) permiten en mayor medida estimar el efecto-escuela, contribuyendo a una medición refinada del desempeño de las escuelas en términos de la influencia que tiene su accionar sobre el progreso académico de sus estudiantes en un período de tiempo⁸. Desde un punto de vista estadístico, estos indicadores son típicamente estimados usando modelos HLM (*Hierarchical Linear Mixed*) (Raudenbush y Willms, 1995; Goldstein y Spiegelhalter, 1996; Goldstein

y Thomas, 1996; Goldstein, 2002; Raudenbush y Bryk, 2002; Hutchison *et al.*, 2005)⁹. En estas aplicaciones, el puntaje que un estudiante alcanza en un test puede ser explicado por un rango de factores o covariables¹⁰. Crucialmente, los modelos de valor agregado se caracterizan por incluir el puntaje obtenido por los estudiantes en un test previo (idealmente, cuando el estudiante ingresó a la escuela o a un ciclo educacional) y han sido implementados bajo estos términos en diversos sistemas escolares con mecanismos de rendición de cuentas.

Sin embargo, existe una diversidad de estrategias de modelamiento de VA. Recientemente, Timmermans *et al.* (2011) realizaron un estudio empírico con datos observacionales de escuelas secundarias de Holanda, identificando cinco tipos de estimación de valor agregado, cuya diferencia sustantiva está en los tipos de variables incluidas. Según los autores, el modelo más robusto para fines de rendición de cuentas incluyó covariables como el puntaje previo, indicadores composicionales y contextuales.

Similarmente, Braun *et al.* (2010), usando datos generados en Estados Unidos, concluyeron que uno de los problemas fundamentales que tienen por delante los métodos de valor agregado es la demanda por estimar el efecto-escuela con modelos que consideren y controlen la asignación no aleatoria que caracteriza la distribución de estudiantes en los colegios. La consecuencia de no hacerlo es incurrir en estimaciones que adolecen de sesgos de endogeneidad. En específico, la selección de estudiantes que realizan las escuelas es probablemente una de las covariables que en mayor medida tiene consecuencias sobre las características relativamente no aleatorias de la composición social y académica de las escuelas. En consecuencia, como se discutirá en la siguiente sección, la selección escolar en nuestro país constituye uno de los desafíos metodológicos de mayor envergadura que enfrentará la Agencia en la estimación del efecto-escuela.

7 Sin embargo, un conjunto de hallazgos, marcos conceptuales y procedimientos metodológicos a la base de la investigación en efectividad escolar, han sido desafiados (cf. Angus, 1993; Slee *et al.*, 1998; Thrupp, 1999; Goldstein y Woodhouse, 2000; Mortimore y Whitty, 2000; Gewirtz, 2002; Luyten *et al.*, 2005; Thrupp *et al.*, 2007; Wrigley, 2004 y 2006; Lupton, 2005; Carrasco, 2010; Thrupp y Lupton, 2010).

8 Sin embargo, existen dos objeciones fundamentales sobre el real valor de su uso. Primero, especialmente para el caso inglés, que las correlaciones entre mediciones desajustadas del desempeño escolar e indicadores de valor agregado son altas (Gorard, 2006, 2008). Por ejemplo, no existirían escuelas con bajo desempeño bruto pero con un puntaje alto en términos de valor agregado. Aunque nuestra evidencia es incipiente, Chile exhibe exactamente lo contrario (San Martín y Carrasco, próximamente). De ahí la importancia de los modelos de valor agregado para el caso chileno. Segundo, desde una perspectiva sociológica, el valor agregado oculta la operación de los efectos composicionales en las prácticas y procesos de la escuela que repercuten en sus desempeños. Los indicadores de valor agregado no pueden a la larga en términos empíricos ajustar o aislar el efecto de la composición social de la escuela sobre el proceso escolar, puesto que no capturan los microefectos cotidianos que tiene esta última sobre la práctica educacional a nivel organizacional y de aula (Thrupp, 1995, 1999).

9 Además de los artículos del número especial de *Journal of Educational and Behavioral Statistics*, 2004, Vol. 29 (1).

10 Los modelos estadísticos utilizados en estos artículos van desde modelos de efecto fijo a otros más complejos, multivariados y longitudinales. Estos incorporan parámetros para efectos de la escuela y los profesores, a nivel de estudiante, a nivel de sala de clase, y efectos covariados a nivel de escuela. Igualmente, incorporan parámetros para permitir correlaciones residuales entre estudiantes en un mismo curso.

En el caso de Chile, la investigación en efectividad escolar ha estado principalmente focalizada en identificar factores intra escuela mediante estudio de casos (Bellei *et al.*, 2004; Eyzaguirre, 2004; Eyzaguirre y Fontaine, 2008). Sin embargo, a la luz de la investigación internacional, los trabajos existentes poseen importantes limitaciones metodológicas en sus estrategias para identificar escuelas efectivas, ya sea porque se basan en correlaciones simples ajustadas por nivel socioeconómico (NSE) que no expresan efectividad escolar, o bien porque emplean series de puntajes SIMCE promedio pero comparando los resultados obtenidos por diferentes cohortes de alumnos. Como consecuencia, estos estudios describen y definen como efectivas o exitosas las prácticas de escuelas que obtienen altos resultados SIMCE, los que seguramente se explican por un conjunto de factores exógenos a ellas –o por prácticas sofisticadas de selectividad– y las atribuyen sesgadamente a factores únicamente endógenos. En contraste, recientemente la investigación en esta área está introduciendo metodologías mixtas, que incluyen estimaciones de valor agregado y métodos cualitativos longitudinales y genuinamente inductivos (Muijis, 2012; San Martín y Carrasco, 2010).

Las limitaciones mencionadas han repercutido probablemente en la manera en que la política pública compara el desempeño de los establecimientos. El argumento para controlar únicamente por NSE es que se está comparando el desempeño de las escuelas en relación a poblaciones similares de alumnos en términos socioeconómicos, pero que asisten a establecimientos diferentes. Si bien este procedimiento en apariencia es correcto, su alcance es apenas aceptable si se emplea con el propósito de clasificar las escuelas en función de su efectividad. Una metodología como esta desestima varios requisitos que la literatura especializada ha identificado como mínimos para estimar la efectividad de una escuela. Primero, no se analiza el progreso de los estudiantes contrastando dos mediciones, de modo que se incluya el puntaje previo obtenido. Aunque se controle por NSE, solo se observa el desempeño en un instante del tiempo, y por tanto, no se expresa la contribución de la escuela al aprendizaje de los niños en términos de su progreso, componente esencial de la práctica educacional. Segundo, tampoco se hace ninguna predicción sobre el desempeño de la escuela, ni se estima su desempeño en relación a otras (o una escuela promedio). Tercero, los controles por NSE,

si bien apuntan en una dirección correcta en términos de considerar el impacto de otros factores extra escuela en el aprendizaje de los niños, son insuficientes para capturar todos los efectos composicionales que afectan dicho aprendizaje.

Pese a ello, en Chile los resultados de las pruebas SIMCE son públicos y se difunden tanto como *rankings* con promedios absolutos sin ajuste, como también ajustados por NSE. En base a ello, el Ministerio de Educación, desde la década anterior, ha elaborado cartillas de difusión para los padres, controlando por NSE. Recientemente, se distribuyó por correspondencia a todas las familias del país con hijos en el sistema escolar, un mapa semaforizado del desempeño de las escuelas cercanas a las de sus hijos, empleando promedios absolutos, sin ningún tipo de control de variables. Como podrá colegirse a partir de lo planteado, todas estas medidas de difusión han entregado información sesgada a las familias respecto de la real contribución de las escuelas a la educación de sus hijos. Por contraste, los modelos de VA resuelven en parte estos problemas elementales de estimación de la efectividad escolar, con el propósito de responsabilizar a las escuelas únicamente respecto de aquello que está en sus posibilidades revertir.

Definición y uso empírico de indicadores de valor agregado

Los componentes del valor agregado

Típicamente se define el valor agregado de una escuela como la contribución que hace la misma al desempeño escolar de los alumnos en relación al desempeño que hubiesen alcanzado en una escuela promedio. Es necesario tener información acerca de la condición inicial de los estudiantes, de manera de “descontarla” del aprendizaje esperado y, de esta manera, realizar una evaluación justa de lo que efectivamente agregó la escuela. Es por ello que, en términos empíricos, un requisito fundamental para desarrollar un análisis de valor agregado es contar con dos mediciones en el tiempo que capturen el desempeño de una misma cohorte (o mismos alumnos).

En Chile, a partir del año 2005, se decidió aplicar anualmente el SIMCE a 4° básico y alternar la aplicación de 8° básico y 2° medio. Con esto, se aseguró que ciertas cohortes sean medidas con el SIMCE una segunda vez¹¹,

11 Así, por ejemplo, los estudiantes que el 2005 cursaban 4° básico, el 2009 cursaban 8° básico y, dado el calendario de aplicaciones de las pruebas SIMCE, rindieron dicha prueba nuevamente. Similarmente, los estudiantes que rindieron el SIMCE en el 2007 cuando cursaban 4° básico, volvieron a ser medidos con el SIMCE, aplicado el 2011 a los estudiantes que cursaban 8° básico.

proporcionando la información necesaria para poder estimar el VA de un establecimiento educacional. Esta información es relevante para poder determinar cuán predictiva es la primera medición en relación a la segunda. Heurísticamente hablando, predictivo puede entenderse como un grado de estabilidad en el comportamiento de los puntajes del SIMCE de una medición a otra. Así, por ejemplo, suele decirse que un estudiante que obtuvo un buen puntaje en el 2005, tiende a tener un buen puntaje en el 2009. Sin embargo, puede ocurrir que un estudiante que en el 2005 obtuvo un puntaje bajo, obtenga en el 2009 un buen puntaje. En esta situación, decimos que el puntaje SIMCE 2005 de dicho estudiante no resultó un buen predictor de su puntaje en el 2009 y, por tanto, cabe preguntarse si ese aumento se debe a la escuela a la cual asiste o a otros factores externos a esta.

Además del puntaje SIMCE de la primera medición temporal, otros predictores que deben considerarse son los factores socioeconómicos, particularmente los que permiten controlar los procesos implícitos o explícitos de selectividad de estudiantes por parte de establecimientos educacionales. Estos factores están, por lo general, definidos a nivel del establecimiento: por ejemplo, el grupo socioeconómico de una escuela o el porcentaje de apoderados que declaran que sus hijos fueron seleccionados con alguna suerte de prueba de conocimientos.

Por otra parte, lo que un establecimiento hace está lejos de ser completamente capturado por los factores o covariables que se miden a nivel de escuela. A fin de representar la contribución que, además del puntaje inicial y factores como el grupo socioeconómico y la proporción de selectividad, hace la escuela a los puntajes medidos en una segunda ocasión, se introduce una componente no observable o no medida directamente, llamada efecto-escuela. Los puntajes (finales) de los estudiantes que pertenecen a una misma escuela están relacionados entre sí solo por medio del efecto-escuela. Operacionalmente, esto significa que, dadas las covariables o factores tanto individuales como aquellos medidos a nivel de escuela, los puntajes de los estudiantes están positivamente correlacionados entre ellos.

Así, son tres los componentes fundamentales en la metodología de valor agregado: (a) el puntaje inicial o pre-

vio, por ejemplo el SIMCE 2005; (b) el puntaje final, por ejemplo el SIMCE 2009; y (c) la contribución de la escuela al puntaje final por medio del efecto-escuela.

¿Qué es el valor agregado de un establecimiento educacional?

Usando los tres componentes ya descritos, nos hacemos la siguiente pregunta. Para predecir el rendimiento educacional de los estudiantes medido con el SIMCE 2009, ¿es suficiente la medición SIMCE 2005, junto a otros factores explicativos como el nivel socioeconómico de la escuela y la proporción de selectividad declarada por padres y apoderados?

Esta pregunta debe responderse para cada escuela, para lo que se construye una predicción promedio del SIMCE 2009 con el SIMCE 2005, el grupo socioeconómico y la proporción de selectividad, a nivel de escuela. Esta componente se conoce como la predicción promedio de los puntajes SIMCE 2009 como si los estudiantes hubiesen sido tratados por una escuela promedio (Manzi *et al.*, 2012; Carrasco y San Martín, 2012). Por otra parte, se predice el puntaje SIMCE 2009 de cada estudiante de dicha escuela por su SIMCE en el 2005, el grupo socioeconómico de esa escuela y su proporción de selectividad, además de la contribución debida al efecto-escuela.

El valor agregado de un establecimiento educacional corresponderá entonces a la diferencia entre la predicción en base a una escuela promedio y a la predicción promedio para esa escuela; esto es, lo que la escuela agrega para predecir los puntajes SIMCE 2009 más allá de lo que se puede pronosticar considerando que los estudiantes estén en una escuela promedio. Por lo tanto, una escuela que agrega valor es aquella que lleva a los estudiantes más allá de donde hubiesen llegado de haber sido educados por una escuela promedio¹².

Es importante mencionar que si la efectividad de una escuela se representa solo por el SIMCE promedio 2009, entonces lo que se hace es entremezclar diferentes componentes del logro de los estudiantes, atribuyéndole todo el mérito o bien a los mismos estudiantes, como si no fuesen tratados por dicha escuela, o a los factores sociodemográficos, tanto a nivel de estudiantes como de escuela¹³. Una responsabilización justa solo puede ase-

12 Es importante enfatizar que la contribución del efecto-escuela al puntaje SIMCE 2009 es la misma para todos los estudiantes que pertenecen a una misma escuela. Por lo tanto, el posible fenómeno de puntajes techo SIMCE 2009 y/o SIMCE 2005 no sería un problema para la estimación del VA, pues empíricamente se constata que no hay escuelas donde todos los estudiantes obtienen puntajes techo en ambas mediciones. También es importante enfatizar que los modelos de valor agregado no explican el progreso individual representado por la diferencia de puntajes del *pre* y *post test*. El *pre test* (SIMCE 2005 en nuestro ejemplo) es un predictor y, como cualquier otro predictor, no necesita estar en la misma escala que la variable dependiente (en nuestro ejemplo, el SIMCE 2009). Para detalles, ver Carrasco y San Martín (2012).

13 Una discusión rigurosa de esta descomposición se puede encontrar en Manzi *et al.*, 2012 (sección 2.2).

gurarse cuando se atribuye al establecimiento el aporte que hace al logro de los estudiantes después de descontar las contribuciones exógenas, tanto aquellas que se logra capturar por factores explicativos, como aquellas que quedan en el error idiosincrático, también llamado error de medición, que corresponde a lo que no es explicado ni por los factores individuales o escolares, ni por el valor agregado.

¿Por qué es necesario incluir en los modelos de valor agregado el nivel de selectividad de una escuela?

Dada la particularidad de nuestro sistema escolar en relación a la selección de estudiantes (OCDE, 2010), los modelos de valor agregado aquí construidos incorporan covariables o factores que capturan prácticas selectivas de las escuelas. La selectividad tiene efectos composicionales que repercuten en los desempeños escolares, por lo que esta decisión metodológica tiene un fundamento educacional importante. Asimismo, la Ley N° 20.529 incluyó cierto resguardo en esta dirección, que otorga bases para incluir la selectividad en los modelos de clasificación¹⁴.

Un supuesto elemental que permite comparar la efectividad de las escuelas es que sus alumnos han sido asignados aleatoriamente en ellas. De ese modo, podría afirmarse que la efectividad de una escuela es independiente del tipo de estudiantes que atiende. O dicho de otro modo, uno podría predecir el desempeño de cualquier estudiante, al margen de su condición inicial, si sabemos el grado de efectividad de un establecimiento educacional. Sin embargo, para el caso chileno este supuesto en la práctica no se cumple, lo que introduce una complejidad adicional a la comparación de desempeño entre escuelas. Y no se cumple en parte, debido a que nuestro sistema escolar se organiza en base a la elección abierta de escuelas por parte de las familias, pero, especialmente, porque hasta hace poco el país permitía la selección abierta de estudiantes por parte de todos los establecimientos.

Es paradójal que en el país la investigación sobre selección escolar sea aún limitada. La legislación introducida que la prohíbe, se basa más en la constatación de sus efectos (segregación), que en la presencia de un conocimiento de base sobre la magnitud, estrategias y características de los establecimientos que la practican. Un

ejemplo es el estudio de Valenzuela *et al.* (2009) que, aunque constata la existencia de segregación educacional, no ofrece una explicación sobre los mecanismos que la producen. Asimismo, como se discutirá, Contreras *et al.* (2011) infiere la presencia de prácticas selectivas pero a partir de datos secundarios autoreportados que adolecen de confiabilidad. Con todo, la selección escolar parece constituir una práctica extendida en todo el sistema –antecedente que explica su reciente regulación– cuyas consecuencias educacionales y sociales se mencionan a continuación.

Primero, la selección puede transformarse en una estrategia de mejoramiento escolar en sí misma (mejorar mediante la captura de los estudiantes más aptos), quitando presión a los establecimientos para que mejoren sus políticas y prácticas internas. Segundo, esta práctica rompe el supuesto sobre el que funciona un sistema de *vouchers* como el chileno. Las escuelas no compiten por capturar estudiantes valiéndose del mejoramiento de su prácticas educacionales, si no que, en su lugar, la competencia se restringe a la captura de los estudiantes más fáciles de educar; o de aceptar a aquellos estudiantes que el resto del sistema excluye. La evidencia en Chile concluye que el esquema de cuasi mercados permite a las escuelas competir más por la vía de la selección de estudiantes que mediante la generación de mejores prácticas pedagógicas y mayores aprendizajes en sus alumnos (Contreras *et al.*, 2011). Tercero, la selección por criterios adscriptivos (socioeconómicos, étnicos o religiosos) constituye una práctica discriminatoria y que por tanto está reñida con el fundamento que organiza un sistema escolar inclusivo y equitativo. Cuarto, la selección temprana, ya sea por rendimiento u otra condición, en la práctica lo que considera son las oportunidades de origen que los niños traen consigo y no sus aptitudes académicas potenciales. Si en algo hay acuerdo a nivel internacional (OECD, 2010), es en la idea de retrasar lo más posible la selección de manera que ella no se valga ni refleje anticipadamente en la experiencia de los niños la posesión desigual de capitales transmitidos y distribuidos diferencialmente en las sociedades. Si bien la selección es parte de la naturaleza de todo sistema escolar, globalmente se emplea tardía y horizontalmente. Chile, por contraste, la ha usado tempranamente (pre básica) y en todas sus modalidades. Por último, otro de los as-

14 En el artículo 17, párrafo 3°, dice así: “La Agencia determinará la metodología de ordenación, previo informe del Consejo Nacional de Educación, la que deberá ser aprobada por decreto supremo del Ministerio de Educación. Esta metodología podrá considerar resguardos que aseguren que las modificaciones de categoría no ocurran por un cambio en la composición del alumnado y se actualizará en un plazo no inferior a cuatro y que no exceda de ocho años”.

pectos discutidos para evaluar la conveniencia de seleccionar alumnos es su impacto en el desempeño escolar como resultado del denominado “efecto-pares”. Los resultados de estas estimaciones varían en la magnitud y efectos encontrados. Mientras algunos afirman que la investigación internacional no es para nada concluyente, lo que desaconseja adoptar medidas que apuntan a una mayor heterogeneidad composicional (Beyer, 2007), otros plantean que la evidencia comparada otorga bases para impedir la selección (Manzi, 2007). Recientemente en Chile, Taut y Escobar (2012), mediante análisis multinivel de datos longitudinales del SIMCE 2004-2006, analizaron el efecto de la diversidad social y rendimiento de los pares en estudiantes de enseñanza media. El estudio concluye que especialmente la diversidad social del aula tiene un impacto positivo en el aprendizaje, particularmente en matemática, tanto en establecimientos particulares subvencionados como municipales.

De este modo, al no considerarse estadísticamente la selectividad, la efectividad encontrada de una escuela podría estar más asociada a sus prácticas selectivas (o falta de ellas), que a sus prácticas pedagógicas o de gestión. De este modo, lo que espuriamente se estaría juzgando sería la capacidad de la escuela de constituir una composición social y académica de estudiantes en mejores condiciones para aprender, y, menos, el efecto neto de las prácticas educativas de la escuela sobre el aprendizaje de los estudiantes que atiende. Y en el otro extremo, para aquellas escuelas que no seleccionan, dado que reciben estudiantes que las escuelas restantes excluyen, el juicio sobre el accionar de la escuela omitiría el hecho que, en términos relativos, su capacidad de controlar la composición de la escuela es reducida, estando su desempeño neto afectado por ese hecho que los análisis, sin embargo, no consideran.

Para revertir la selección, Chile ha introducido legislación para prohibir la selección hasta 6° básico en el caso de la Ley General de Educación¹⁵ y en todo nivel para quienes suscriban convenios de Igualdad de Oportunidades en el caso de la Ley de Subvención Escolar Preferencial. Ambos cuerpos de ley buscan dotar de mayor heterogeneidad a la composición de las escuelas, no solo para cautelar principios de cohesión social y formación

cívica, sino también para trasladar a los establecimientos escolares la responsabilidad de educar académicamente a los niños, al margen de su situación de origen, ante lo cual la sociedad crecientemente pedirá cuentas, por ejemplo, mediante la nueva Agencia de Calidad de la Educación.

Sin embargo, el impacto buscado en el sistema escolar de ambas leyes no está necesariamente asegurado. Algunos especialistas han expresado dudas especialmente respecto de la LGE. En específico, la prohibición de seleccionar que introduce ha sido criticada por su laxitud y calificada como una ‘selección oficializada’ (Contreras, 2010), o también criticada por su ambigüedad, dado que acto seguido de restringir la admisión por rendimiento, afirma que las pruebas de admisión deben ser públicas (Mena y Corbalán, 2010). A estas críticas se suma el hecho de que es probable que los dispositivos de fiscalización de la selección que pudiese contemplar la nueva institucionalidad educacional tengan diversos límites prácticos para supervigilarla.

Por tanto, en el mediano plazo, incluir covariables de selectividad resulta aconsejable e indispensable si se espera distinguir qué escuelas se desempeñan mejor que otras en términos netos. Los indicadores de selectividad que aquí se emplean provienen de los cuestionarios a padres que administra el SIMCE. Dado su carácter autorreportado, son en general de mediana confiabilidad, lo que agrega un desafío metodológico de envergadura para capturar con medidas confiables el nivel de selectividad de una escuela, especialmente puesto que es muy probable que ocurran mediante mecanismos tanto explícitos como implícitos.

Diferencias en la posición y clasificación de las escuelas

¿Es la posición de los establecimientos sensible al método utilizado para clasificarlos respecto a su desempeño en términos de resultados académicos? Para responder a esta pregunta, a continuación se presentan algunos resultados del análisis de datos efectuado usando dos definiciones distintas de efectividad escolar: un modelo de valor agregado (VA) y un puntaje SIMCE. El

15 La LGE fue promulgada en agosto de 2009 y determinó que los establecimientos que reciben subvención estatal están impedidos de considerar el rendimiento escolar pasado o potencial de los postulantes, así como establecer como requisito de admisión la presentación de antecedentes socioeconómicos de la familia de los postulantes entre el primer nivel de transición y el sexto año de educación básica (Ley N° 20.370 artículos 12°, 13° y 14°). Esta Ley –que posee un rango constitucional en la mayor parte de sus artículos–, dado su carácter universal, extiende el criterio de no selección de alumnos sobre todos los establecimientos con financiamiento público en los niveles mencionados. Así, el cambio en la legislación significó, por una parte, la restricción en la selección académica o rendimiento de alumnos durante la educación básica y, por otra, la validación de esta práctica a partir de séptimo básico (LGE, 2010; especialmente artículos 12°, 13° y 14°).

propósito fue explorar si se advertían diferencias en la posición de las escuelas una vez clasificadas mediante el puntaje SIMCE 2009 o a través de un indicador de valor agregado.

Datos utilizados en el análisis

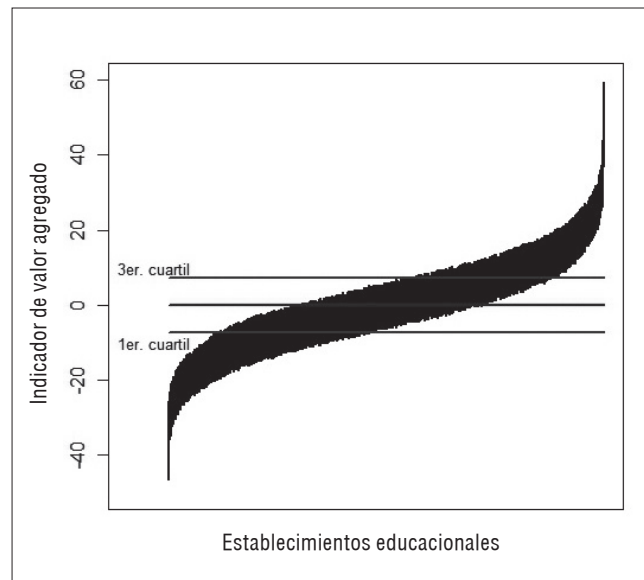
En el año 2005, el SIMCE de matemáticas fue aplicado a 252.481 estudiantes que cursaban 4° año de enseñanza básica y que estaban repartidos en 7.517 establecimientos educacionales. La gran mayoría de estos estudiantes (a saber, 227.993 estudiantes de 5.842 establecimientos educacionales) fueron nuevamente medidos el 2009 con el SIMCE de matemáticas. A partir de estas dos mediciones, se construyó una base de datos para realizar un análisis de valor agregado de establecimientos educacionales municipalizados y particulares subvencionados, obteniendo una muestra que incluye 153.689 estudiantes de 3.203 establecimientos educacionales. Los establecimientos de la muestra son municipalizados pertenecientes a corporaciones municipales (MC), municipalizados de administración directa (MD) y particulares subvencionados (PS), y de acuerdo al SIMCE se clasifican en cuatro grupos socioeconómicos (A, B, C y D, siendo este último el más alto).

La escuela promedio con respecto a la cual se comparan las contribuciones de cada establecimiento educacional de la muestra está caracterizada, además de la medición individual SIMCE 2005, por tres variables que representen los procesos explícitos e implícitos de selección de los estudiantes por parte de los establecimientos educacionales: el grupo socioeconómico de la escuela según la clasificación SIMCE; la proporción de padres y apoderados que en el cuestionario de padres y apoderados declara que su hijo fue seleccionado por la escuela por medio de una prueba de habilidad; y el promedio de puntajes SIMCE 2005 para los estudiantes que están en un mismo establecimiento educacional en el 2009.

La Figura 1 muestra el indicador de valor agregado que obtiene cada establecimiento. Para fines de ordenación, se emplean intervalos que permiten ilustrar su posición en términos de efectividad. Dos escuelas tienen indicadores de valor agregado diferentes cuando los intervalos correspondientes no se superponen uno al otro. Como se observa en la Figura 1, es posible hacer distinciones

entre establecimientos educacionales en términos de indicadores de valor agregado: los establecimientos cuyos indicadores de VA están por debajo del primer cuartil son menos efectivos que los establecimientos cuyos indicadores están por encima del tercer cuartil. También se puede observar que los establecimientos cuyos indicadores de VA están entre el primer cuartil y la media, son menos efectivos que los colegios cuyos indicadores están sobre el tercer cuartil.

Figura 1 | Ordenación de cada establecimiento según indicador de valor agregado



Fuente: Elaboración propia.

Los resultados anteriores son estadísticamente válidos si se asegura que el puntaje inicial no está correlacionado con el efecto-escuela. En relación a los datos utilizados en este estudio, es relevante mencionar que el SIMCE 2005 fue aplicado cuando los estudiantes cursaban 4° básico, en efecto, el SIMCE se aplica hacia el último tercio del año escolar. Esta aplicación coincide con el término del segundo ciclo académico. A partir de 5° básico y hasta 8°, los estudiantes cursan el tercer ciclo académico, cuyas características de organización de la enseñanza son muy diferentes a las del segundo ciclo¹⁶. Por lo tanto, es probable que los establecimientos educacionales traten de distinta forma a los alumnos en estos ciclos. De esta manera, la medición inicial SIMCE 2005 proporciona

¹⁶ Por ejemplo, en el segundo ciclo los estudiantes tienen un docente para todas las asignaturas mientras que, a partir de 5° básico, hay prácticamente un docente por asignatura. En el segundo ciclo, el mismo docente a cargo de impartir las asignaturas, es el profesor jefe, a saber, el docente que guía aspectos

Nota 16 continúa en pág. siguiente

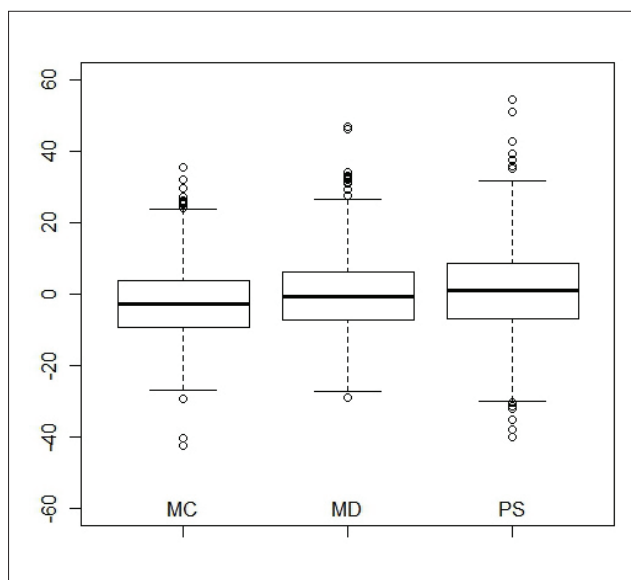
información del nivel de logro alcanzado por cada estudiante al finalizar el segundo ciclo, y la medición final SIMCE 2009 coincide con la finalización del tercer ciclo. Por lo tanto, bajo estas consideraciones, el valor agregado de un establecimiento educacional corresponde a lo que la escuela añade para que los estudiantes alcancen cierto nivel de logro al terminar el tercer ciclo, teniendo en cuenta el nivel alcanzado al finalizar el segundo ciclo. Así, hay escuelas que ciertamente añaden valor en el sentido de permitir que sus estudiantes alcancen un nivel de logro al finalizar el tercer ciclo, superior que el que hubiesen alcanzado considerando, además de factores de selectividad, su nivel de logro al terminar el segundo ciclo¹⁷.

Efectividad escolar por dependencia y su relación con puntaje SIMCE promedio

Una de las preguntas centrales que ha ocupado el debate sobre el impacto de la provisión mixta de nuestro sistema escolar se relaciona con la comparación del desempeño entre establecimientos municipales o particulares subvencionados (Drago y Paredes, 2011). Sin embargo, como indica Carrasco y San Martín (2011), la evidencia acumulada en el país al respecto no ha empleado metodologías de valor agregado para concluir sobre el desempeño relativo por dependencia (esto es, no incluyen dos mediciones).

En la Figura 2 se aprecian los *box-plots* de los indicadores de valor agregado para cada dependencia educacional. El 75% de los establecimientos de cada dependencia tienen efectividades comparables, lo cual significa que en general la efectividad escolar no está asociada a la dependencia educacional.

Figura 2 | Indicadores de VA, incluyendo puntaje previo, NSE y control de selectividad de la escuela



Fuente: Elaboración propia.

Por otro lado, en la Figura 3 se aprecia que, para cada grupo socioeconómico, hay una gama de establecimientos con altos puntajes SIMCE promedio que, una vez analizados con indicadores de VA, se descomponen en grupos homogéneos y emergen establecimientos de bajísima agregación de valor, así como de agregación media y alta. Similarmente, hay establecimientos con bajo promedio SIMCE que tienen indicadores de valor agregado altos, bajos y medios. Por lo tanto, es necesario concluir que los puntajes SIMCE promedio están lejos de capturar la efectividad de una escuela.

sociales y relacionales de los alumnos a nivel de escuela. Además, típicamente las escuelas ponen a los estudiantes en el segundo ciclo a cargo de un mismo docente para asegurar cierta homogeneidad de enseñanza y así enfrentar de la mejor manera el SIMCE. En el tercer ciclo, estos hábitos desaparecen y son remplazados por otras dinámicas organizacionales internas.

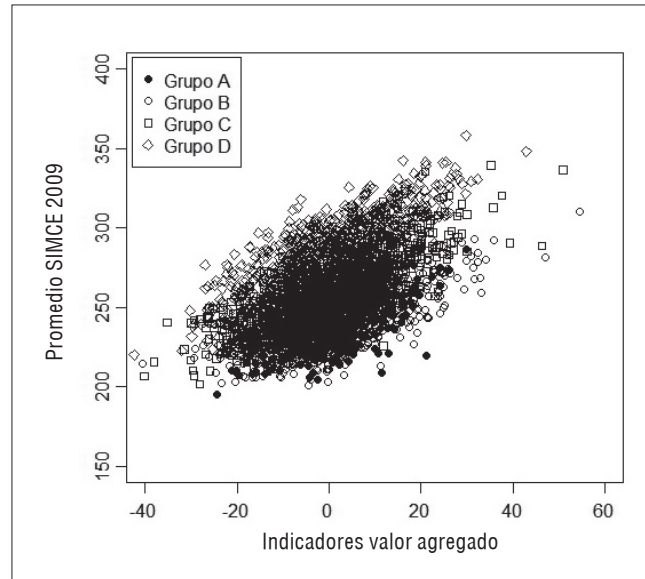
17 Dos observaciones adicionales deben mencionarse en relación a la estimación de los indicadores de valor agregado. En primer lugar, los resultados anteriores han sido obtenidos usando una muestra de escuelas, aquellas que tienen al menos 20 estudiantes medidos con el SIMCE 2005 y el SIMCE 2009. Esta muestra contiene el 56% de los establecimientos educacionales que tienen estudiantes que fueron medidos con ambos SIMCE. Sin embargo, cabe preguntarse qué ocurriría con el ordenamiento de este 56% de establecimientos educacionales si los indicadores de valor agregado se estimasen con una muestra de colegios tales que haya al menos 15 estudiantes medidos con ambos SIMCE (lo que equivale al 66% de las escuelas), o al menos 10 estudiantes (que equivale al 77% de las escuelas), o al menos 5 estudiantes (equivalente al 86% de ellas). Las correlaciones de Spearman (que miden cómo se mantiene el ranking) entre los indicadores de valor agregado del 56% inicial de escuelas, cuando estos se calculan con las tres diferentes muestras, es igual a 0.99. Una segunda observación refiere a la precisión con que los puntajes SIMCE individuales son calculados. Baste decir que en principio la Oficina del SIMCE utiliza un modelo de Teoría Respuesta al Ítem que permite estimar parámetros de adivinanza de cada pregunta. Sin embargo, recientemente se ha establecido que dicho modelo tiene 'problemas de identificación', lo cual afecta a la estimación no solo de las características de los ítems, sino también a la estimación de los puntajes individuales; ver San Martín y Rolin, J. M. (próximamente).

La Tabla 1 muestra el cruce de las clasificaciones de escuelas, considerando puntajes SIMCE 2009 y VA¹⁸. El número de establecimientos en la diagonal (en negrita) expresa aquellos establecimientos que no cambian de posición una vez reclasificados (tomando en consideración toda la muestra). Dicho de otro modo, su posición en términos de efectividad se mantiene independiente de la metodología de clasificación empleada. En contraste, todos los establecimientos que están fuera de la diagonal han cambiado su posición y son por tanto sensibles al método de clasificación utilizado. En específico, las escuelas que mantienen su clasificación corresponden al 39,2% del total, mientras que el 60,8% de las escuelas cambia su clasificación.

Más específicamente, el 27,95% de las escuelas clasificadas usando valor agregado aumenta su categoría si se las reclasifica usando SIMCE promedio; y el 32,8% de las escuelas clasificadas usando SIMCE aumenta su categoría al ser reclasificadas usando valor agregado. Estas últimas corresponden a aquellas que han sido clasificadas injustamente si solo se tiene en cuenta el resultado promedio SIMCE al mejorar su posición si se emplea VA para juzgar su desempeño.

La descripción anterior puede repetirse controlando por dependencia escolar. En la Tabla 2 aparecen las clasificaciones de las escuelas municipalizadas (MC o MD): el 38,4% de las escuelas municipalizadas mantiene su clasificación, mientras que el 61,6% no tiene la misma

Figura 3 | Relación entre SIMCE 2009 promedio y VA según NSE



Fuente: Elaboración propia.

clasificación usando SIMCE promedio y valor agregado. De hecho, el 13,2% de las escuelas clasificadas usando valor agregado aumenta su categoría si se las reclasifica usando SIMCE promedio; y el 48,3% de las escuelas clasificadas usando SIMCE aumenta su categoría al ser reclasificadas usando valor agregado¹⁹.

Tabla 1 | Clasificación de escuelas usando VA y SIMCE promedio

VA \ SIMCE	Primer cuartil	Segundo cuartil	Tercer cuartil	Cuarto cuartil	Total
Primer cuartil	411	181	140	68	800
Segundo cuartil	295	196	182	129	802
Tercer cuartil	182	182	242	195	801
Cuarto cuartil	42	112	238	408	800
Total	930	671	802	800	3203

Fuente: Elaboración propia

18 Se definieron cuatro grupos de escuelas de acuerdo a los cuartiles de las distribuciones correspondientes a los indicadores de valor agregado y al puntaje SIMCE promedio.

19 En relación a la nota 17, es importante mencionar que si se comparan las clasificaciones de las escuelas usando valor agregado, donde este último ha sido calculado con la muestra inicial y las muestras de escuelas tales que al menos haya 15, al menos haya 10 y al menos haya 5 estudiantes medidos con ambos SIMCE; se encuentra que solo un 2,4%, un 6,3% y un 8,2%, respectivamente, de las escuelas de la muestra inicial, son reclasificadas.

Tabla 2 | **Clasificación de establecimientos Municipales**

VA / SIMCE	Primer cuartil	Segundo cuartil	Tercer cuartil	Cuarto cuartil	Total
Primer cuartil	295	88	28	1	412
Segundo cuartil	239	123	61	8	431
Tercer cuartil	147	118	89	17	371
Cuarto cuartil	30	81	126	82	319
Total	711	410	304	108	1533

Fuente: Elaboración propia

Si analizamos estas clasificaciones para las escuelas particulares subvencionadas, la Tabla 3 muestra que el 40% de las escuelas mantiene su clasificación, mientras que el 60% no tiene la misma clasificación usando SIMCE promedio y valor agregado. De hecho, el 41,4% de las escuelas clasificadas usando valor agregado aumenta su categoría si se las reclasifica usando SIMCE promedio, y el 18,6% de las escuelas clasificadas usando SIMCE aumenta su categoría al ser reclasificadas

usando valor agregado. Para extender el ejemplo, las escuelas en cursiva empeorarían su posición: aunque obtienen un alto SIMCE, poseen un bajo puntaje en los indicadores de VA, es decir, se sobrestima su efectividad.

En suma, es notable la sensibilidad del cambio de posición de las escuelas al reclasificarse con SIMCE promedio o con valor agregado. Esta última apreciación se puede resumir de la siguiente manera en la Tabla 4.

Tabla 3 | **Clasificación de establecimientos Particulares Subvencionados**

VA / SIMCE	Primer cuartil	Segundo cuartil	Tercer cuartil	Cuarto cuartil	Total
Primer cuartil	116	93	112	67	388
Segundo cuartil	56	73	121	121	371
Tercer cuartil	35	64	153	178	430
Cuarto cuartil	12	31	112	326	481
Total	219	261	498	692	1670

Fuente: Elaboración propia

Tabla 4 | **Cambio de posición al reclasificar desde indicador VA a SIMCE**

Tipo de establecimiento	VA ↓ SIMCE	VA ↑ SIMCE
Municipalizado	48.3%	13.2%
Particular Subvencionado	18.6%	41.4%

Fuente: Elaboración propia

La flecha indica, según se trate, si aumenta o disminuye la posición de las escuelas una vez que se reclasifican, desde el uso de un indicador de VA, a una clasificación efectuada con el promedio SIMCE. Así, por ejemplo, el 41,4% de los particulares subvencionados mejoró su posición si es reclasificado con puntajes SIMCE en lugar de emplear para estimar su nivel de efectividad y posición el indicador de VA. Por contraste, el 48% de los establecimientos municipalizados que son reclasificados usando puntajes SIMCE en lugar del indicador de VA, descienden en su posición de efectividad.

En consecuencia, estos resultados indican que la posición en que las escuelas son clasificadas para fines de rendición de cuentas es muy sensible al tipo de metodología que se utilice. Adicionalmente, estos resultados también indican que a cada modelo de estimación subyace una definición conceptual propia respecto al significado de la efectividad escolar atribuible a la escuela.

Una política de responsabilización justa de la efectividad de las escuelas

Un sistema educativo que avanza en poner mayor presión a sus escuelas y que asocia consecuencias al desempeño, requiere en paralelo de dispositivos para una responsabilización justa que permita distinguir con mejor precisión aquellos establecimientos que avanzan de los que no lo hacen²⁰. Aunque en Chile la política de clasificar escuelas es celebrada como una estrategia adecuada de mejoramiento educacional (Brunner y Peña, 2007), algunos han advertido los desafíos metodológicos que trae consigo esta nueva institucionalidad clasificadora (Espínola y Claro, 2010; Contreras, 2010).

Para contribuir a la discusión pública, en este artículo se contrastó la clasificación de escuelas inducida por indicadores de valor agregado respecto de aquella inducida por los puntajes SIMCE a nivel de escuela. Se ha planteado que la literatura internacional sobre efectividad escolar concuerda en que hay diferencias en los distintos tipos de modelización de valor agregado, siendo los más robustos, aquellos que incorporan puntaje previo, nivel

socioeconómico, selectividad y variables composicionales (Timmermans *et al.*, 2011). Coincidentemente, se ha demostrado que la mayoría de las escuelas (60,8%) son reclasificadas una vez que su efectividad se analiza a partir de modelos de valor agregado.

Sin embargo, Chile no ha desarrollado sistemas de medición de valor agregado ni existen a nivel de Estado modelizaciones para estimar si las escuelas agregan o no valor en los aprendizajes de sus estudiantes. A nuestro juicio, esta es una pieza central para la puesta en marcha de una agencia aseguradora de la calidad que garantice una responsabilización justa del desempeño de los establecimientos educacionales²¹.

Sin embargo, es crucial remarcar que la puesta en práctica de modelos de valor agregado para clasificar escuelas demanda un horizonte de tiempo de mediano plazo. Pero esto no debiese convertirse en un argumento para sugerir la inconveniencia de su uso ante dicha restricción temporal. Especialmente porque otras metodologías alternativas de clasificación –distintas al valor agregado– adolecen de restricciones para conceptualizar, modelizar y estimar estadísticamente el efecto-escuela. Tales limitaciones podrían tener consecuencias éticas sobre los establecimientos educacionales, cuyo significado e impacto sobre las comunidades escolares resulta de mayor relevancia que la urgencia política de implementar aceleradamente la ordenación de escuelas. Así entonces, la puesta en funcionamiento de la Agencia de Calidad podría consistir más bien en la generación de las condiciones y ajustes técnicos que permitan, con bases científicas sólidas, implementar en un mediano plazo la clasificación de escuelas propiamente tal. En términos de política pública, parece ser que el camino recomendable a corto plazo es emprender un programa I+D que busque tanto prototipar el uso del VA para clasificar escuelas en Chile, así como mejorar la confiabilidad de los indicadores o covariables fundamentales que deben incluirse en las estimaciones. Para ese fin, a continuación se mencionan algunos problemas técnicos que demandan atención.

20 Es importante subrayar que este tipo particular de políticas que Chile viene adoptando a paso firme, basadas en competencia, presión y rendición, han sido abandonadas, sustituidas o desestimadas en países como Canadá, Escocia, Finlandia y Singapur. En su lugar, diversos especialistas –Robin Alexander, Andrew Hargreaves, Diane Ravitch y Pasi Sahlberg– han demostrado la eficacia de políticas educativas que promueven el profesionalismo, colegialidad, liderazgo para construir capacidad local, accountability interna, pedagogía dialógica, aprendizaje creativo, evaluación para monitorear progresos, confianza multinivel y estándares indicativos (Carrasco, 2012).

21 Una importante excepción iniciada pero posteriormente descontinuada, fueron los análisis exploratorios sobre valor agregado encargados por el MINEDUC en 2008 y 2009, respecto de los resultados SIMCE para 3°, 4°, 5°, 8° básico y 2° medio, entre los años 2004 y 2007 (G. del Pino, San Martín, E., Manzi, J., Taut, S., González, J. de la Cruz, R.).

En primer lugar, como se ha mencionado, los procesos de selectividad son gravitantes en las estimaciones y se requeriría la recolección de covariables que ayuden a tener información de mayor validez sobre estos procesos, ya sea mediante cuestionarios SIMCE u otros medios. También la Agencia debe dar una señal política clara al proponer las variables que se utilizarán en un modelo de VA. En efecto, esto consistiría en describir la escuela promedio con respecto a la cual las escuelas compararán sus efectividades.

En segundo lugar, los establecimientos con pocos alumnos medidos en el SIMCE no solo afectan la precisión de indicadores de VA, sino también del promedio SIMCE bruto: ¿cómo podrán clasificarse estas escuelas? ¿es posible cuestionar la existencia de establecimientos con menos de 20 estudiantes medidos por el SIMCE?

En tercer lugar, un tema que se ha discutido muy marginalmente es la calidad y confiabilidad de la estimación del grupo socioeconómico de las escuelas que efectúa el SIMCE. Esta es una covariable crucial para estimar el efecto-escuela, que requiere máxima precisión. Existen, sin embargo, indicaciones que en su estado actual no logran capturar con máxima sensibilidad las diferencias en las condiciones materiales de las escuelas y estudiantes que asisten. Estadísticamente hablando, hay que preguntarse si corresponde a una variable medida con error o no; en caso de serlo, se induce un problema de endogeneidad (Grilli y Rampichini, 2011) que debe ser considerado a la hora de calcular el VA.

En cuarto lugar, hay dos cuestiones que actúan en paralelo y deben tomarse en cuenta conjuntamente. Primero, los problemas de endogeneidad que impiden capturar el efecto neto de la escuela, sobre todo porque los alumnos medidos, por ejemplo en 4° básico, han sido previamente tratados por la misma escuela. En este trabajo hemos propuesto una interpretación de los datos que evita dicha dificultad; si esta interpretación se cuestiona, hay que considerar sistemáticamente el problema de endogeneidad²². Segundo, y relacionadamente, una agenda de mayor calendarización del SIMCE (2°, 6°, etc.), tam-

bién podría influir en términos endógenos en la explicación de los resultados.

En quinto lugar, cual sea el método de clasificación de escuelas utilizado, es de interés examinar la estabilidad de la efectividad de una escuela a lo largo del tiempo. Esto significa comparar la efectividad de una escuela calculada con una determinada cohorte, con aquella obtenida por una segunda. En el campo del VA, la problemática de la estabilidad de los indicadores de VA en el tiempo es aún objeto de desarrollo metodológico. Para visualizar los problemas, basta imaginar el significado de una escuela promedio cuando dos o más cohortes son empleadas.

En último término, hay que decidir qué mediciones estandarizadas deberían utilizarse para calcular la efectividad de una escuela: ¿solo el SIMCE de matemáticas? ¿el SIMCE de lenguaje? ¿una medida resumen de ambas? ¿o se consideraría la posibilidad de calcular dos indicadores de VA, uno usando el SIMCE de matemáticas, otro el SIMCE de lenguaje? De ser así, es importante recordar que dichos indicadores no se pueden combinar, pues el significado de una escuela promedio con una u otra medida estandarizada es completamente diferente.

Mirado en perspectiva, no contar con un sistema que permita estimar si los establecimientos educacionales mejoran o no por efecto de su propia agencia, podría exponer el fundamento de política pública que está a la base de un Sistema de Aseguramiento de la Calidad. El uso del SIMCE sin ajustes como indicativo de los estándares nacionales de aprendizaje es insuficiente para estos efectos, puesto que informa sobre los resultados absolutos obtenidos por las escuelas, o relativos en relación a sus pares, sin tomar en cuenta la estructura multivariada que explica el desempeño de las escuelas, ni descontar los efectos de factores ajenos a su control. La consecuencia probable será, ya sea sobrestimar o bien subestimar el desempeño real de los establecimientos escolares, comprometiendo la responsabilización justa de su ordenamiento público.

22 Para ese propósito, la tarea es estudiar teóricamente factores exógenos al efecto-escuela que no solo permitan explicar la contribución neta de las escuelas, sino que además permitan “romper” la endogeneidad (Manzi *et al.*, 2012). Estas consideraciones permitirían, además, dar cierta dirección sustantiva a una sistemática revisión de los cuestionarios complementarios del SIMCE.

Referencias

- Bellei, C., Raczynski, D., Muñoz, G. y Pérez, L.,** 2004. *¿Quién dijo que no se puede?: Escuelas efectivas en contextos de pobreza.* Santiago de Chile, UNICEF.
- Beyer, H.,** 2007. Algunas notas sobre selección escolar. En: Brunner, J.J. y Peña, C., eds. *La reforma del sistema escolar: aportes para el debate.* Santiago, Universidad Diego Portales.
- Beyer, H., Eyzaguirre, B. y Fontaine, L.,** 2000. La reforma educacional: una apreciación crítica. Santiago: CEP, Documento de Trabajo N° 306.
- Braun, H., Chudowsky, N. and Koenig, J.,** 2010. *Getting value out of value-added.* Washington, D.C: The National Academies Press.
- Brunner, J.J. y Peña, C., eds.,** 2007. *La reforma del sistema escolar: aportes para el debate.* Santiago, Universidad Diego Portales.
- Carrasco, A.,** 2012. Evaluación retroactiva de los sistemas escolares anglófonos que adoptaron políticas de accountability, estandarización, examinación y privatización educacional. Book Review, manuscrito.
- Carrasco, A. y San Martín, E.,** 2012. Exploring new approaches in school choice research: a partial examination of educational market-driven hypothesis from value-added models. Conditionally accepted in *Estudios de Economía.*
- Carrasco, A. y San Martín, E.,** 2011. Are quasi-markets in education meeting their policy purposes in Chile? Re-examining empirical hypothesis from value-added models. *Informes Técnicos MIDE UC, IT104.* En: www.mideuc.cl/category/publicaciones.
- Chapman, C., Armstrong, P., Harris, A., Muijs, D., Reynolds, D. and Sammons, P., eds.,** 2012. *School effectiveness and improvement research, policy and practice: challenging the orthodoxy?* London: Routledge.
- Contreras, D., Sepúlveda, P. y Bustos, S.,** 2011. When schools are the ones that choose: the effects of screening in Chile. *Social Science Quarterly*, 91 (5), 1349-1368.
- Contreras, D.,** 2010. Derecho a la educación, inclusión y selección escolar. En: Bellei, C., Contreras, D. y Valenzuela, J.P., eds. *Ecós de la revolución pingüina. Avances, debates y silencios de la reforma educacional.* UNICEF-Universidad de Chile. Pehuén Editores.
- Coleman, J.S., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, R. and York, R.,** 1966. *Equality of educational opportunity.* Washington, DC: U.S. Government Printing Office.
- Drago, J. y Paredes, R.,** 2011. La brecha de la calidad de la educación. *CEPAL Review*, 104, 167-180.
- Edmonds, R. R.,** 1979. Effective schools for the urban poor. *Educational Leadership*, 37(10), 15-24.
- Espinola, V. y Claro, J. P.,** 2010. El sistema nacional de aseguramiento: una reforma basada en estándares. En: Bellei, C., Contreras, D., y Valenzuela, J.P., eds. *Ecós de la revolución pingüina. Avances, debates y silencios de la reforma educacional.* UNICEF-Universidad de Chile. Pehuén Editores.
- Eyzaguirre, B. y Fontaine, L.,** 2008. *Las escuelas que tenemos.* Santiago: Centro de Estudios Públicos.
- Eyzaguirre, B.,** 2004. Claves para la educación en pobreza. *Estudios Públicos*, 93, 249-277.
- Fitz-Gibbon, C. T.,** 1997. *The value added national project.* London: SCAA.
- Gorard, S.,** 2008. The value added of primary schools: what is it really measuring? *Educational Review*, 60 (2), 179-185.
- Gorard, S.,** 2006. Value-added is of little value. *Journal of Educational Policy*, 21 (2), 233-241.
- Goldstein, H.,** 2008. Evidence and education policy. Some reflections and allegations. *Cambridge Journal of Education*, 38 (3), 393-400.
- Goldstein, H.,** 2002. *Multilevel statistical models.* London: Kendall's Library of Statistics.
- Goldstein, H.,** 2001. Using pupil performance data for judging schools and teachers: scope and limitations. *British Educational Research Journal*, 27, 4433-4442.
- Goldstein, H., Huiqi, P., Rath, T. and Hill, N.,** 2000. *The use of value-added information in judging school performance. Perspectives on Education Policy.* Institute of Education, University of London.
- Goldstein, H., and Spiegelhalter, D.,** 1996. League tables and their limitations: statistical issues in comparison of institutional performance. *Journal of the Royal Statistical Society, Series A* 159, 385-443.
- Goldstein, H. and Thomas, S.,** 1996. Using examination results as indicators of school and college performance. *Journal of the Royal Statistical Society, Series A* 159, 407-442.
- Gray, J., Goldstein, H. and Thomas, S.,** 2003. Of trends and trajectories: searching for patterns in school improvement. *British Educational Research Journal*, 29 (1), 83-88.
- Gray, J., Goldstein, H. and Thomas, S.,** 2001. Predicting the future: the role of past performance in determining trends in institutional effectiveness at a level. *British Educational Research Journal*, 27 (4), 1-15.

- Grilli, L. and Rampichini, C.**, 2011. The role of sample cluster means in multilevel models: a view on endogeneity and measurement error issues. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 7, 121-133.
- Hutchison, D., Mifsud, C., Morrison, J., Grech, R., Rudd, P. and Hanson, J.**, 2005. The Malta primary literacy value added project: a template for value added in small island states? *Research Papers in Education*, 20, 303-345.
- Luyten, H. y Sammons, P.**, 2010. Multilevel modelling. En: Creemers, B., Kyriakides, L. and Sammon, P. *Methodological Advances in Educational Effectiveness*. London: Routledge.
- Manzi, J., San Martín, E. and Van Belleghem, S.**, 2012. *School system evaluation by value added analysis under endogeneity*. Una versión borrador está disponible en <http://mideuc.cl/wp-content/uploads/2011/09/it1102.pdf>.
- Manzi, J.**, 2007. Selección de estudiantes en el sistema escolar chileno: ¿La buena educación debe ser solo para los mejores? En: Brunner, J.J. y Peña, C., coords. *La reforma del sistema escolar: aportes para el debate*. Santiago, Universidad Diego Portales.
- Mena, P. y Corbalán, F.**, 2010. La selección escolar: una cuestión de libertad, competencia, igualdad e integración. En: Bellei, C., Contreras, D., y Valenzuela, J.P., eds. *Ecos de la revolución pingüina. Avances, debates y silencios de la reforma educacional*. UNICEF-Universidad de Chile. Pehuén Editores.
- Mortimore, P., Sammons, P., Stoll, L., Lewis, D. and Ecob, R.**, 1988. *School matters: the junior years*. Somerset, England: Open Books.
- Muijs, D.**, 2012. Methodological change in educational effectiveness research. En: Chapman, C., Armstrong, P., Harris, A., Muijs, D., Reynolds, D. and Sammons, P., eds. *School effectiveness and improvement research, policy and practice: challenging the orthodoxy?* London: Routledge.
- OCDE**, 2010. *PISA 2009: what makes schools successful? Resources, policies and practices*. Volume IV. Paris, Francia.
- OECD**, 2008. *Best practices to assess the value added of schools*. Paris, Francia.
- Raudenbush, S. W. and Bryk, A. S.**, 2002. *Hierarchical linear models: applications and data analysis methods*, 2nd edition. Newbury Park, CA: Sage.
- Raudenbush, S. W. and Willms, J.D.**, 1995. The estimation of school effects. *Journal of Educational and Behavioral Statistics* 20, 307-335.
- Rutter, M., Maughan, B., Mortimore, P. and Ouston, J.**, 1979. *Fifteen thousand hours: secondary schools and their effect on children*. Cambridge MA: Harvard University Press.
- Sammons, P.**, 2012. Methodological issues and new trends in educational effectiveness research. En: Chapman, C., Armstrong, P., Harris, A., Muijs, D., Reynolds, D. and Sammons, P., eds. *School effectiveness and improvement research, policy and practice: challenging the orthodoxy?* London: Routledge.
- Sammons, P., Hillman, J. and Mortimore, P.**, 1995. *Key characteristics of effective schools: a review of school effectiveness research*. London: DfSE.
- San Martín, E. and Rolin, J.M.**, 2012. Identification of the 1PL model with guessing parameter: parametric and semi-parametric results. *Psychometrika*, accepted to be published.
- San Martín, E. y Carrasco, A.**, 2010. *School effectiveness and value added models: from quantitative analysis to qualitative outcomes*. Manuscrito, Proyecto FONDECYT N° 1110315.
- Taut, S. y Escobar, J.**, 2012. *El efecto de las características de los pares en el aprendizaje de estudiantes chilenos de enseñanza media*, Informes Técnicos MIDE UC, IT1201, Junio 2012. En: www.mideuc.cl/category/publicaciones
- Teddlie, C. and Reynolds, D., eds.**, 2000. *International handbook of school effectiveness research*. London and New York: Falmer.
- Timmermans, A., Doolaard, S., and Wolf, I.**, 2011. Conceptual and empirical differences among various value added models for accountability. *School effectiveness and school improvement*, 22 (4), 393-413.
- Thrupp, M.**, 1999. *Schools making a difference: let's be realistic! School mix, school effectiveness and the social limits of reform*. Buckingham: Open University Press.
- Thrupp, M.**, 1995. The school mix effect: the history of an enduring problem in educational research, policy and practice. *British Journal of Sociology of Education*, 16, 183-203.
- Thrupp, M., Lauder, H., and Robinson, T.**, 2002. School composition and peer effects. *International Journal of Educational Research*, 37, 483-504.
- Valenzuela, J.P., Bellei, C. y De los Ríos, D.**, 2009. Evolución de la segregación socioeconómica de los estudiantes chilenos y su relación con el financiamiento compartido. En: *Evidencias para políticas públicas en educación*. FONIDE, Ministerio de Educación.
- van de Grift, W.**, 2009. Reliability and validity in measuring the value added of schools. *School effectiveness and school improvement*, 20 (2), 269-285.



CENTRO DE
**POLÍTICAS
PÚBLICAS UC**

www.politicaspUBLICAS.uc.cl
politicaspUBLICAS@uc.cl

SEDE CASA CENTRAL

Av. Libertador Bernardo O'Higgins 324, piso 3, Santiago.
Teléfono (56-2) 354 6637.

SEDE LO CONTADOR

El Comendador 1916, Providencia.
Teléfono (56-2) 354 5658.

CENTRO DE POLÍTICAS PÚBLICAS UC

- Facultad de Arquitectura, Diseño y Estudios Urbanos • Facultad de Ciencias Económicas y Administrativas
- Facultad de Ciencias Sociales • Facultad de Derecho • Facultad de Educación
- Facultad de Historia, Geografía y Ciencia Política • Facultad de Ingeniería • Facultad de Medicina