La medición del aprendizaje de los alumnos

MEJORES PRÁCTICAS PARA EVALUAR EL VALOR AGREGADO DE LAS ESCUELAS



OCDE

La medición del aprendizaje de los alumnos

MEJORES PRÁCTICAS PARA EVALUAR EL VALOR AGREGADO DE LAS ESCUELAS



Este trabajo se publica bajo la responsabilidad del Secretario General de la OCDE. Las opiniones expresadas y las conclusiones formuladas en este informe no necesariamente corresponden a las de los gobiernos de los países miembros de la OCDE.

Por favor, cite esta publicación de la siguiente manera:
OCDE (2011), *La medición del aprendizaje de los alumnos: Mejores prácticas para evaluar el valor agregado de las escuelas*, OECD Publishing.
http://dx.doi.org/10.1787/9789264090170-es

ISBN 978-92-64-09016-3 (impreso)
ISBN 978-92-64-09017-0 (PDF)

Esta publicación es producto del acuerdo de cooperación celebrado entre el Gobierno de los Estados Unidos Mexicanos y la Organización para la Cooperación y el Desarrollo Económicos, que busca mejorar la calidad de la educación en México.

Ilustración de la portada: ©UNESCO/José Gabriel Ruíz Lembo.

Las erratas de las publicaciones de la OCDE se encuentran en línea en *www.oecd.org/publishing/corrigenda*.
© OCDE 2011

Usted puede copiar, descargar o imprimir los contenidos de la OCDE para su propio uso y puede incluir extractos de publicaciones, bases de datos y productos de multimedia de la OCDE en sus propios documentos, presentaciones, blogs, sitios web y materiales docentes, siempre y cuando se dé el adecuado reconocimiento a la OCDE como fuente y propietaria del copyright. Toda solicitud para uso público o comercial y derechos de traducción deberá dirigirse a rights@oecd.org. Las solicitudes de permisos para fotocopiar parte de este material con fines comerciales o de uso público deben dirigirse al Copyright Clearance Center (CCC), *info@copyright.com* o al Centre français d'exploitation du droit de copie (CFC), *contact@cfcopies.com*.

Prefacio

LA EVALUACIÓN DEL VALOR AGREGADO DE LAS ESCUELAS

La prosperidad de los países ahora se deriva, en gran medida, de su capital humano, y para alcanzar el éxito en un mundo cambiante, las personas necesitan desarrollar sus conocimientos y habilidades durante toda su vida, empezando por su paso por la educación primaria y secundaria. Conforme los países miembros y asociados de la OCDE buscan mejorar sus sistemas educativos en este sentido, se otorga más importancia a las medidas efectivas de desempeño y desarrollo de alumnos y, por ende, de desempeño escolar. Con este enfoque en los resultados de aprendizaje, las medidas confiables y válidas de desempeño escolar pueden servir como base para políticas y programas educativos para la rendición de cuentas, los procesos de mejora escolar y para que familias y la sociedad en su conjunto cuenten con mayor información sobre la efectividad de escuelas.

El éxito de estas iniciativas reside en la precisión de la medida de desempeño escolar. Varios países han dejado de emplear puntuaciones brutas de exámenes como medida única del desempeño escolar, pues estas puntuaciones reflejan en gran parte los factores contextuales y de antecedentes de los alumnos, y no representan del todo el desempeño de las escuelas. Esto resulta problemático: sin una medida precisa del desempeño, sin resultados equitativos y sin respuestas eficaces de políticas e intervenciones educativas, se corre el riesgo de que los recursos no se canalicen a los alumnos, escuelas y docentes que más los necesitan para mejorar sus resultados. Por otra parte, no se pueden mejorar políticas y prácticas educativas si no se conoce lo que ha demostrado ser eficaz en el contexto local del país o de la entidad (estatal, municipal o zona escolar, por ejemplo).

Medidas más precisas y confiables aumentan la confianza que los actores del sistema educativo, incluyendo a directores y docentes, depositan en evaluaciones y políticas educativas. Los indicadores de valor agregado también son una herramienta importante para identificar buenas prácticas en el sistema educativo. Relativamente pocos países miembros y asociados de la OCDE, sin embargo, cuentan actualmente con mecanismos para proporcionar alguna información de valor agregado en el ámbito escolar. Para los países que ya cuentan con exámenes a grande escala para medir el desempeño de los alumnos, pero que aún no utilizan modelos de valor agregado, los requisitos técnicos que a continuacíon se presentan podrán servir como elementos importantes en la planeación en el corto y mediano plazos.

En este contexto, este informe presenta los múltiples desafíos en el proceso de diseñar, aplicar y mejorar sistemas basados en modelos de valor agregado. Estos retos incluyen el desarrollo adecuado de los instrumentos para obtener datos de desempeño, la elaboración de las bases de datos adecuadas, el diseño de técnicas estadísticas apropiadas y el diseño y puesta en marcha de las políticas y programas correspondientes, incluyendo iniciativas de mejora escolar. Los retos técnicos pueden tener repercusiones sustanciales en las políticas e imponen limitaciones

a la aplicabilidad de modelos de valor agregado en la elaboración de políticas y programas educativos. Por tanto, este informe dedica una considerable atención a los aspectos técnicos de la elaboración de modelos de valor agregado debido a su importancia en el proceso de diseñar políticas eficaces en esta área.

Este informe ofrece a los responsables de formular políticas educativas, administradores, directores escolares, docentes, y a otros actores clave, información y orientación sobre cómo entender y diseñar modelos de valor agregado en beneficio del sistema educativo. El informe se divide en tres partes: la Parte I analiza los objetivos y el uso de un sistema basado en modelos de valor agregado. El enfoque de estos capítulos reside en las principales aplicaciones para políticas y programas que beneficien las iniciativas de mejora escolar, los sistemas de rendición de cuentas, y la elección de escuelas. Al final del Capítulo 3, se presenta un resumen titulado ***"Establecimiento de un sistema basado en modelos de valor agregado: Pasos básicos en la fase de implementación"*** dirigido a las autoridades educativas que esperan comprender con rapidez el proceso requerido para establecer un sistema de evaluación basado en modelos de valor agregado. La Parte II aborda aspectos técnicos propios de los modelos de valor agregado. Se centra en aspectos de su diseño, el tipo de modelos posibles y los aspectos estadísticos y metodológicos pertinentes. La Parte III se centra en la puesta en marcha de sistemas basados en modelos de valor agregado, y abarca tanto los aspectos políticos e institucionales que deben abordarse como las consideraciones técnicas que necesitan resolverse.

El informe original se elaboró bajo la coordinación de Ben Jensen del Secretariado de la OCDE y reunió a expertos e investigadores en modelos de valor agregado de 12 países: Bieke De Fraine y Eva Van de Gaer (Bélgica, Fl.), Sine Frederiksen y Poul Schjørring (Dinamarca), Gasper Cankar (Eslovenia), Rosario Martínez Arias (España), Claudie Pascal (Francia), Torbjørn Hægeland (Noruega), Wim van de Grift (Países Bajos), Maria Eugénia Ferrão (Portugal), Andrew Ray (Reino Unido), Radim Ryska (Rep. Checa), Anders Broberg (Suecia), Henry Braun (invitado), y Maciej Jakubowski de la OCDE. Daniel McCaffrey, reconocido experto en el análisis de valor agregado, realizó una revisión técnica e hizo una contribución sustancial al informe original. María Castro Morera (España) realizó la revisión técnica y la actualización de referencias y citas para la versión en español.

La presente edición en español y el trabajo sobre políticas de evaluación e incentivos llevado a cabo como parte del acuerdo con el gobierno mexicano se produjeron bajo los auspicios de la División de Indicadores y Análisis del Directorado de Educación de la OCDE, bajo la responsabilidad de Andreas Schleicher, Michael Davidson, Alejandro Gómez Palma, Marlene Gras, Elisabeth Villoutreix, Niccolina Clements, Marika Boiron, Isabelle Moulherat, Fionnuala Canning, Juliet Evans y Shayne MacLachlan. Laura Milena Valencia E. realizó el trabajo de traducción y Fung Kwan Tam realizó la diagramación.

Índice

INTRODUCCIÓN **9**

PARTE I – OBJETIVOS Y USO DE MODELOS DE VALOR AGREGADO **17**

CAPÍTULO 1 **OBJETIVOS DE POLÍTICAS PARA EL DESARROLLO DE UN SISTEMA BASADO EN MODELOS DE VALOR AGREGADO** **19**

Uso de modelos de valor agregado para procesos de mejora escolar **22**

- Toma de decisiones basada en datos **22**
- Precisión de las mediciones de desempeño **24**
- Uso de las características socioeconómicas en los modelos de valor agregado **27**

Rendición de cuentas de las escuelas mediante los modelos de valor agregado **30**

- Incentivos y resultados indeseados **35**

Elección de escuela con información de valor agregado **37**

Conclusión **40**

CAPÍTULO 2 **PRESENTACIÓN E INTERPRETACIÓN DE LOS MODELOS DE VALOR AGREGADO** **43**

Presentación de la información de valor agregado **45**

- Estudio de caso: Reino Unido **46**

Identificación de cambios significativos del desempeño escolar **53**

Creación de estándares y puntos de referencia con información de valor agregado **54**

Presentación en los medios **57**

CAPÍTULO 3 **APLICACIONES DE LOS MODELOS DE VALOR AGREGADO PARA PROCESOS DE MEJORA ESCOLAR** **63**

Escuelas como organizaciones de aprendizaje **64**

- Análisis de los perfiles de valor agregado de las escuelas **66**
- Aplicación de modelos de valor agregado para ayudar a las iniciativas de mejora escolar **68**
- Aplicación de modelos de valor agregado para proyecciones de desempeño **71**

Uso dirigido de los modelos de valor agregado y equidad **74**

Aplicación de modelos de valor agregado para mejorar el sistema de evaluación escolar **77**

Apéndice **Establecimiento de un sistema basado en modelos de valor agregado: Pasos básicos en la fase de implementación** **83**

PARTE II – EL DISEÑO DE MODELOS DE VALOR AGREGADO 91
CAPÍTULO 4 **CONSIDERACIONES DE DISEÑO** 97
Datos de la evaluación de alumnos 99
- Concepto de validez 100
- Error de medición 102
- Escalas de puntuaciones 103
- Resultados de evaluación en una escala ordinal 104
- Estructura de evaluaciones de los alumnos en países participantes 106

Filosofía del ajuste y características contextuales 110
- Importancia de las características contextuales 111
- ¿Qué características contextuales socioeconómicas? 114
- Datos del nivel de escuela 119

CAPÍTULO 5 **MODELOS ILUSTRATIVOS DE VALOR AGREGADO** 121
Modelos de valor agregado: Breve reseña histórica 122
Modelos de valor agregado de regresión lineal 124
Modelos de componentes de varianza o de efecto aleatorio 125
Modelos de efectos fijos de valor agregado 126
El modelo Dallas, TX 127
Modelos de respuesta de efecto aleatorio multivariado 129
Análisis de la curva de crecimiento 131
Conclusión 133

CAPÍTULO 6 **ELECCIÓN DE MODELO: ASPECTOS ESTADÍSTICOS Y METODOLÓGICOS** 135
Criterio estadístico: Varianza y estabilidad intertemporal 137
Criterio estadístico: Sesgo 140
Criterio estadístico: Error cuadrático de la media 142
- Datos perdidos 142

Elección de modelo en análisis de valor agregado 144
Conclusión 147

PARTE III – IMPLEMENTACIÓN DE UN SISTEMA BASADO EN MODELOS DE VALOR AGREGADO 151
CAPÍTULO 7 **ESTABLECER OBJETIVOS DE POLÍTICAS Y ELEGIR EL MODELO ADECUADO DE VALOR AGREGADO** 153
Determinación de la variable con la cual medir el valor agregado 155
- Medidas categóricas y continuas 156

Identificación del modelo adecuado de valor agregado que aborde mejor los objetivos de políticas 157
Elaboración de una base de datos eficaz 160
Importancia del programa piloto para los modelos de valor agregado 165

CAPÍTULO 8 **DESARROLLO Y USO DE LOS MODELOS DE VALOR AGREGADO** 169
Uso de resultados de una media móvil de tres años 173
Comunicación y compromiso con actores clave 174
Desarrollo de un programa de formación 178
Presentación y uso de la información de valor agregado 182

BIBLIOGRAFÍA 185

Lista de cuadros, gráficas y recuadros

CUADROS

Cuadro 2.1 Cantidad de escuelas por año con cambios anuales consecutivos significativos en puntuaciones de valor agregado contextualizado entre 2005 y 2006 ..... **56**

Cuadro 4.1 Exámenes de alumnos en países participantes que en potencia se usarían para modelos de valor agregado **107**

Cuadro 4.2 Datos contextuales recopilados en países participantes que en potencia se usarían en modelos de valor agregado **118**

Cuadro 6.1 Cambios absolutos en valor agregado contextualizado (VAC), valor agregado (VA) y resultados brutos (APS): Resumen de estadísticas, Etapa 4, 2005-06 (RU) **138**

GRÁFICAS

Gráfica 2.1 Ejemplo de cálculo de la línea de mediana del valor agregado en el Reino Unido **47**

Gráfica 2.2 Ejemplo del sitio de Internet de los Cuadros de Desempeño Escolar del valor agregado **48**

Gráfica 2.3 Ejemplo de información de valor agregado contextualizado proporcionada para una escuela secundaria del Reino Unido **50**

Gráfica 2.4 Cambios de valor agregado contextualizado (VAC) comparados con cambios en los resultados brutos de logros **53**

Gráfica 2.5 Extracto del periódico *The Guardian* (19 de enero de 2006) que presenta datos de valor agregado y de otros tipos **57**

Gráfica 2.6 Extracto de *The Times* (19 de enero de 2006) que presenta una "tabla de liga" de valor agregado **58**

Gráfica 2.7 Resultados de valor agregado de una escuela del Reino Unido disponibles en el sitio de Internet de la BBC, 2008 ..... **60**

Gráfica 3.1 Ejemplo del perfil de matemáticas de una escuela en el TVAAS **67**

Gráfica 3.2 Ejemplo de un Informe Escolar TVAAS, 2006 ..... **70**

Gráfica 3.3 Ejemplo de un Informe de Proyección TVAAS **72**

Gráfica 3.4 Ejemplo del Informe Escolar del TVAAS (Álgebra) **73**

Gráfica 3.5 Ejemplo de trayectorias de desempeño comparativas de TVAAS **75**

Gráfica 4.1 Ilustración gráfica del proceso de ajuste estadístico **110**

RECUADROS

Recuadro 2.1 Descripción del valor agregado contextualizado en los medios del Reino Unido **59**

Recuadro 3.1 Interés en procesos de enseñanza y aprendizaje específicos de inspecciones escolares en los Países Bajos **79**

Recuadro 8.1 Programas de formación en Polonia **179**

Introducción

Sistemas educativos de países miembros y asociados de la OCDE distinguen una creciente presión para mejorar los resultados del aprendizaje de los alumnos. Estudios recientes han demostrado que el logro de mejores resultados educativos brinda beneficios considerables individuales y para la sociedad en su conjunto (OCDE 2010a, 2010c). Conjuntamente con el reto de elevar *la calidad* de la educación, sistemas educativos de varios países enfrentan el reto de asegurar una mayor *equidad* en las oportunidades educativas ofrecidas para todos los alumnos, independientemente de su nivel socioeconómico, lengua materna o pertenencia a un grupo étnico. Una parte considerable del presupuesto público de varios países se destina a la educación para enfrentar estos retos. Sin embargo, incrementar el presupuesto educativo global o por estudiante no implica necesariamente que los países obtendrán mejores resultados; el caso de **Estados Unidos de América** es un buen ejemplo de ello. Actualmente, frente a la crisis económica global y recortes en los presupuestos públicos, que la inversión en la educación demuestre ser eficaz y eficiente cobra aún mayor importancia. Los sistemas educativos en distintos países enfrentan, por ende, el triple reto de mejorar los resultados de aprendizaje, asegurar mayor equidad, e incrementar su eficacia y eficiencia. En este contexto, medidas que logren identificar con mayor precisión y confiabilidad el desempeño de escuelas resultan ser sumamente relevantes. A pesar de que existe una gran variedad de medidas de desempeño de alumnos y de escuelas en distintos países, para todos resulta importante que las contribuciones de las escuelas individuales sean claramente identificadas, en lugar de que las mediciones reflejen simplemente las diversas condiciones socioeconómicas en las que enseñan los maestros y operan las escuelas. De no ser así, se corre el riesgo de que los recursos se asignen mal y que se establezcan incentivos no deseados, en lugar de que la inversión sea canalizada hacia los estudiantes y las escuelas de mayor necesidad ó específicamente, para mejorar las prácticas docentes.

Este informe presenta los mejores métodos, comúnmente denominados ***modelos de valor agregado***, que permiten a los usuarios separar las contribuciones de las escuelas al desempeño de los alumnos de los factores contextuales ajenos al control de las aulas y las escuelas. Los modelos de valor agregado han generado un interés creciente debido a la mayor precisión que proporcionan en la medida del desempeño escolar y el papel que representan en la elaboración y establecimiento de políticas educativas e iniciativas de mejoras escolares. Varios estudios revelan que los modelos de valor agregado ofrecen estimadores más precisos del desempeño escolar que las comparaciones de puntuaciones brutas de exámenes o que los ***modelos de aprovechamiento contextualizado transversales*** con que a menudo se obtienen estimadores de desempeño escolar (Doran e Izumi, 2004). Los modelos de valor agregado ofrecen una base cuantitativa fundamentalmente más precisa y valiosa que las puntuaciones brutas de exámenes y estudios transversales para mejorar la planificación escolar, elaborar políticas y poner en marcha mecanismos eficaces de rendición de cuentas.

Los modelos de valor agregado son análisis estadísticos que ofrecen medidas de desempeño escolar (p. ej., una puntuación escolar de valor agregado) para supervisar, evaluar y desarrollar los procesos escolares y de otros elementos del sistema educativo. En este sentido, un sistema basado en modelos de valor agregado debe verse como un medio para un fin, y no como un fin en sí. Las condiciones, prioridades y objetivos de políticas educativas de cada sistema educativo determinarán en gran parte la forma de emplear las medidas de valor agregado. Por tanto, el desarrollo del sistema debe moldearse y adaptarse según el uso que tendrán las puntuaciones de valor agregado de las escuelas para alcanzar objetivos específicos de políticas educativas.

Con base en ejemplos de 13 países, este informe presenta tres objetivos amplios de políticas educativas que se benefician de los modelos de valor agregado: ***procesos de mejora escolar, sistemas de rendición de cuentas de las escuelas*** y mayor información para ***la elección de escuela***. La eficacia de los datos de desempeño en la toma de decisiones referentes a estos objetivos de políticas reside en la precisión de las medidas de desempeño empleadas. Sin embargo, el enfoque en ***la toma de decisiones basada en datos*** para alcanzar objetivos educativos enfrenta el problema de la falta de datos precisos de desempeño escolar, esencial para las mejoras educativas (Raudenbush, 2004; Vignoles *et al.,* 2000). Las puntuaciones brutas en los exámenes estandarizados son una medida del desempeño de los alumnos pero hay claros problemas para inferir sobre el desempeño escolar a partir de estos datos. Los modelos de aprovechamiento contextualizado transversales toman en cuenta características contextuales como los antecedentes de los alumnos, pero son menos útiles para aislar los efectos de las escuelas individuales en la educación de los alumnos. En este contexto, las medidas de valor agregado son un avance significativo, pues ofrecen una medida precisa del desempeño escolar sobre la cual basar decisiones dirigidas a los objetivos de políticas y elevar el desempeño escolar. Este informe ilustra cómo emplear la información de valor agregado para procesos de mejora escolar, para dar seguimiento a las políticas y programas específicos, y para la toma de decisiones en el ámbito escolar o bien, para el sistema educativo en su conjunto.

Para todos los ***procesos de mejora escolar*** es importante reconocer que las mejoras en una actividad o conjunto de actividades dadas requiere primero una evaluación precisa de la situación actual, lo que, a su vez, requiere una medida precisa del desempeño (Sammons *et al.,* 1994). Es difícil desarrollar programas eficaces para el futuro si no es posible analizar de manera precisa la situación presente (OCDE, 2010b). Para un sistema educativo en su conjunto, la información de valor agregado sirve para determinar las áreas y escuelas del sistema educativo que agregan más valor y las áreas y escuelas en las que se requiere más atención para mejorar resultados. En el ámbito escolar, la información de valor agregado puede servir para identificar las materias, grados y grupos de alumnos en donde la escuela agrega más valor y en donde se necesita mejorar. En este sentido, las puntuaciones e información de valor agregado son más valiosas si no sólo documentan el estado actual de las escuelas, sino también generan información que apoye un proceso de mejora escolar. Los análisis estadísticos de las relaciones entre las condiciones de entradas e insumos de las escuelas y los indicadores de desempeño escolar pueden señalar las estrategias que funcionan y las que no lo hacen, para guiar ajustes de políticas y la reasignación de recursos.

Los modelos de valor agregado sirven asimismo para crear proyecciones de desempeño escolar que ayuden en la planificación, asignación de recursos y toma de decisiones. Con las proyecciones se pueden identificar resultados futuros, por ejemplo, si continuasen las trayectorias de desempeño actuales, y también para establecer objetivos de desempeño. Estos objetivos dan información para tomar decisiones en el ámbito escolar sobre cómo utilizar mejor los recursos y estructurar la educación ofrecida para cumplir con objetivos de desempeño específicos (Hill *et al.,* 2005; Doran e Izumi, 2004). Si se combinan con información adicional recopilada dentro de las escuelas, las proyecciones del desempeño esperado de alumnos (basadas en estimadores de valor agregado) ofrecen una panorámica completa del desempeño de una escuela. El personal escolar tendrá así, a su disposición, información básica útil para planear y actuar.

Los modelos de valor agregado pueden beneficiar en gran medida los ***sistemas de rendición de cuentas de las escuelas*** cuando éstos están enfocados al aprendizaje de los alumnos. Los sistemas de rendición de cuentas establecen claramente qué entidades son responsables de rendir cuentas a qué organismos con base en las prácticas o resultados específicos esperados (McKewen, 1995). Estos sistemas también pueden ofrecer información al público en general: se puede informar a los contribuyentes si el recurso público se aplica con eficacia, y los usuarios estarían en posibilidades de elegir instituciones educativas de manera más informada. Sin embargo, el punto clave resulta ser si la evaluación de procesos y desempeño es ***precisa*** y ***justa*** para las escuelas individuales. Este informe ilustra que los modelos de valor agregado proporcionan una medida más precisa, y por ende más justa, del desempeño escolar (en tanto que se miden los incrementos en el desempeño de los alumnos), que al mismo tiempo sirve para mejorar la evaluación de los procesos escolares. Los resultados de los modelos de valor agregado (es decir, las puntuaciones de valor agregado de las escuelas) ofrecen medidas del grado al cual las escuelas lograron elevar el desempeño de los alumnos. Cuando se emplean en sistemas de rendición de cuentas de las escuelas, estas mediciones sirven de manera eficaz en las evaluaciones escolares, con consecuencias más justas y equitativas para las escuelas y personal escolar.

La ***elección de escuela*** es otro objetivo posible que se presenta en este informe y que se beneficia de los modelos de valor agregado. Estos datos pretenden informar a las familias sobre el desempeño de diferentes escuelas para contribuir a su elección de escuela. Esto requiere publicar los datos de los resultados escolares (Gorard, Fitz y Taylor, 2001). Si bien esto no sucede en todos los países, es una tendencia creciente entre los países miembros de la OCDE (OCDE, 2007a). Como se presenta en la Parte I de este informe, los beneficios de mejorar los niveles de elección escolar en un sistema educativo pueden ser múltiples. Los padres de alumnos pueden elegir las escuelas más adecuadas a sus necesidades, y los recursos fluyen de esta manera a las escuelas que satisfacen mejor estas necesidades (Hoxby, 2003). Sin embargo, estos beneficios dependen de una medida precisa del desempeño escolar, pues de otro modo las elecciones de las familias estarán mal informadas y los recursos por ende se podrán asignar de manera equivocada. Es esencial la mayor precisión de los modelos de valor agregado para la eficacia de un sistema de elección escolar. Ofrece a los padres una medida más precisa del desempeño escolar sobre la cual basar sus decisiones, y a las escuelas, una oportunidad más justa de mejorar su desempeño con base en una evaluación más equitativa.

El alto grado de precisión propia de los modelos de valor agregado genera más confianza en las medidas de desempeño para promover objetivos de políticas educativas como los que se describen aquí. Esta mayor confianza proviene de las mejoras que se han hecho a estos modelos con el tiempo y de las ventajas, en comparación con otros métodos, de estimar el desempeño escolar. Una discusión de los principales modelos de valor agregado, así como una breve reseña histórica, se presentan en el Capítulo 5.

VENTAJAS DE LOS MODELOS DE VALOR AGREGADO

Los modelos de valor agregado son una mejora sustancial en comparación con muchas medidas de desempeño escolar que actualmente se utilizan. Las comparaciones de puntuaciones brutas de pruebas proporcionan cierta información importante, pero son medidas deficientes del desempeño escolar. No toman en cuenta el rendimiento previo y generan resultados que en gran medida reflejan diferencias en las características contextuales como los antecedentes socioeconómicos de los alumnos. Los modelos de aprovechamiento contextualizado intentan abordar estos problemas al medir el impacto de las características contextuales sobre una medida de desempeño específica, pero son menos útiles para separar los efectos escolares en el avance estudiantil de otras características contextuales, y por ende son menos útiles para medir el desempeño escolar. Los modelos de valor agregado pretenden resolver estos problemas al incorporar medidas de aprovechamiento previo y, en algunos casos, características contextuales de los alumnos. Esto permite un análisis más refinado del progreso del desempeño de alumnos y resulta más eficaz para desentrañar los efectos de diversos factores que afectan al avance de los alumnos. Estas ventajas permiten una mayor precisión al medir el desempeño, lo que genera más confianza en la interpretación de las mediciones de desempeño escolar.

Para aquellos países y sistemas educativos que se encuentran en procesos de reformas e innovación, este informe sostiene que los modelos de valor agregado pueden contribuir a los siguientes esfuerzos:

- aprendizaje en todos los niveles del sistema educativo al identificar con precisión los aspectos de mayor y menor desempeño;
- identificación y análisis de "lo que funciona" para procesos de mejora escolar;
- establecimiento de sistemas de rendición de cuentas más transparentes y más equitativos que pueden servir como base para el diseño de incentivos para elevar el desempeño escolar;
- desarrollo de sistemas de información que permitan a las escuelas analizar y evaluar su desempeño y fortalecer el sistema de evaluación;
- mayor eficacia y equidad en la asignación de recursos (humanos, materiales y económicos) a las áreas críticas que presenten mayor necesidad; y
- atención a desigualdades socioeconómicas arraigadas que puedan estar ocultas en el ámbito escolar mediante medidas de desempeño indiscriminadas, imprecisas e inequitativas.

MODELOS DE VALOR AGREGADO: DEFINICIONES

En distintos países hay una gran variación en el uso de modelos de valor agregado y en los análisis estadísticos utilizados para analizar el desempeño escolar. Dadas las ventajas que

ofrece el uso de los modelos de valor agregado, resulta importante distinguir los modelos de valor agregado de otras aproximaciones estadísticas. En este informe, se utilizan las siguientes definiciones:

Valor agregado	La contribución de una escuela al crecimiento de los alumnos hacia objetivos educativos establecidos o prescritos (p. ej., el aprovechamiento cognitivo o niveles mínimos de alfabetización). Esta contribución está aislada de otros factores que contribuyen al progreso educativo de los alumnos.
Modelos de valor agregado	Una clase de modelos estadísticos que estiman las contribuciones de las escuelas al avance de los alumnos en objetivos educativos establecidos o prescritos (p. ej., el aprovechamiento cognitivo o niveles mínimos de alfabetización) medido en al menos ***dos*** momentos temporales.

Los modelos específicos de valor agregado pueden ajustarse a una definición más precisa de la estimación del desempeño escolar, pero esta definición general es aplicable a diversas especificaciones de valor agregado y aún distingue con claridad los modelos de valor agregado de otros tipos de análisis estadísticos. Algunos análisis estadísticos que se aplican en varios países para supervisar el desempeño escolar no se considerarían modelos de valor agregado según estas definiciones. En este informe se define a estos análisis transversales como ***modelos de aprovechamiento contextualizado***. Estos análisis suelen excluir al menos dos mediciones de desempeño estudiantil tomadas en distintos momentos, lo cual se considera la base de los modelos de valor agregado.

Una característica distintiva de los modelos de valor agregado es la inclusión de medidas anteriores de desempeño, que permiten una estimación más precisa de la contribución de la escuela al progreso de los alumnos. Doran e Izumi (2004) destacaron las ventajas de los modelos de valor agregado al registrar a los alumnos a través del tiempo en comparación con modelos transversales (o de aprovechamiento contextualizado), que ofrecen una imagen "estática" del desempeño escolar. Los modelos de valor agregado facilitan un análisis más detallado de mejora escolar al estimar la contribución de la escuela a la mejora del desempeño de los alumnos en un periodo determinado. Además, los modelos de valor agregado pueden registrar de mejor manera aquellos factores no observados que influyen en la medición de desempeño inicial, como la capacidad del alumno, lo cual representa un problema sistémico en la mayoría de los modelos de aprovechamiento contextualizado (Raudenbush, 2004).

La inclusión de una medición de rendimiento previo permite estimar el valor agregado de la escuela. El valor agregado debe interpretarse como la contribución de la escuela al desempeño de los alumnos entre dos mediciones de desempeño. Esto es un punto importante, y es posible emplear diversas evaluaciones de alumnos en diferentes intervalos. Es necesario reconocer estas diferencias en la interpretación de la contribución de escuelas individuales (es decir, la puntuación de valor agregado de una escuela). Una distinción básica es la materia de las evaluaciones de los alumnos, pues el valor agregado de la escuela se estima sólo con la materia incluida en las evaluaciones (se analiza esto con más detalle en el Capítulo 1). Otra consideración son los momentos temporales de las evaluaciones. Varias estimaciones de valor agregado calculan la contribución de la escuela en un año dado. Sin embargo,

diversos sistemas educativos no realizan evaluaciones anuales ni cuentan con una estructura de evaluaciones que permita la estimación de una puntuación de valor agregado para un único año. Esto no implica que no sea posible estimar el valor agregado en un periodo de varios años. Al contrario, estas estimaciones se efectúan en varios sistemas educativos. Sin embargo, es importante reconocer que difieren de las puntuaciones de valor agregado de un año único, pues el análisis de las puntuaciones del valor agregado de las escuelas ha de precisar la materia y el periodo en que se mide el valor agregado.

La importancia de múltiples medidas de aprovechamiento genera la cuestión de lo que debe considerarse una medición adecuada de desempeño anterior a partir de la cual medir el crecimiento. Hay un acalorado debate sobre la comparabilidad de las puntuaciones de exámenes y la conversión de puntuaciones en escalas significativas y comparables (Braun, 2000; Dorans *et al.,* 2007; Patz, 2007; Kolen y Brennan, 2004). Desde luego, muchos modelos de valor agregado no requieren en realidad que las puntuaciones de los exámenes se escalen de manera vertical. Tan sólo requieren que las puntuaciones en grados sucesivos tengan una relación aproximadamente lineal y, en la mayoría de los casos, que sea una medición razonable (Doran y Cohen, 2005). Este informe no analiza el desarrollo de instrumentos de evaluación de alumnos en sí mismo: una revisión de la abundante literatura que analiza los problemas de evaluación cae fuera de su alcance. Sin embargo, la definición de valor agregado en este informe se centra ***en el crecimiento hacia objetivos establecidos o prescritos*** (p. ej., aprovechamiento cognitivo). Esto excluye a algunos modelos de aprovechamiento contextualizado que incluyen medidas de inteligencia, como puntuaciones de CI, que tal vez se considere una medición de capacidad general pero es menos adecuada como medida de rendimiento previo sobre la cual medir crecimiento. En el análisis de las puntuaciones de valor agregado de las escuelas siempre debe quedar claro lo que verdaderamente representan las medidas de aprovechamiento anteriores y actuales, y cómo deben por tanto afectar acciones de políticas educativas y de escuelas.

Aún con la mayor precisión de los modelos de valor agregado, persisten ciertas dificultades para medir el desempeño escolar. La interpretación de las puntuaciones del valor agregado de las escuelas debe incluir varias advertencias y precauciones para que sea correcta. Estos retos se abordan en la Parte II de este informe. Si bien el análisis pretende ilustrar los diversos problemas de medición al diseñar y emplear modelos de valor agregado, no es la intención negar su considerable potencial. Al contrario, las estimaciones precisas de valor agregado tienen gran potencial en la elaboración de políticas e iniciativas de mejoras escolares, y son un avance sustancial respecto a mediciones alternativas. Por ejemplo, el Capítulo 6 analiza los problemas estadísticos y metodológicos que deben abordarse en la elaboración y uso de modelos de valor agregado. Se destacan estos problemas no para desalentar el empleo de modelos de valor agregado en los sistemas educativos, sino para alentar su elaboración eficaz para avanzar hacia objetivos específicos de políticas. De hecho, una razón básica para promover los modelos de valor agregado es que los problemas estadísticos y metodológicos a menudo crean problemas mayores de falta de especificación cuando otras estrategias estadísticas y mediciones de desempeño escolar son utilizadas. Estas estrategias distintas a los modelos de valor agregado suelen proporcionar medidas menos precisas del desempeño escolar y son, por ende, menos útiles para un desarrollo eficaz del sistema educativo y de las escuelas. Por tanto, la atención que en este informe se dedica a los problemas estadísticos y

metodológicos pretende destacar la necesidad de elaborar y ofrecer medidas precisas de valor agregado del desempeño escolar para dar información útil para el desarrollo de políticas e iniciativas de mejora escolar, y ganar la confianza de los actores clave del sector educativo, incluyendo a los administradores y los docentes.

ORGANIZACIÓN DE ESTE INFORME

El informe se divide en tres partes para facilitar su lectura y uso por públicos distintos. La Parte I analiza los objetivos y uso de los modelos de valor agregado. Esto incluye un análisis de los objetivos de políticas (Capítulo 1) que se mejoran con los modelos de valor agregado. De igual manera, se presenta un análisis de cómo presentar la información de valor agregado y puntuaciones escolares a distintos actores del sistema educativo, al distinguir entre la presentación de la información de valor agregado con propósitos internos, para uso público y presentación a los medios. Se ofrecen varios ejemplos de métodos de presentación eficaces por países en el Capítulo 2. El análisis de una presentación de información de valor agregado con propósitos internos se centra en la aplicación de modelos de valor agregado con propósitos de mejora escolar en el Capítulo 3. El análisis muestra que la información juega un papel clave en promover decisiones basadas en datos con escuelas que utilizan mediciones de desempeño precisas para desarrollar y supervisar iniciativas de mejora escolar. Este análisis considera a las escuelas como organizaciones de aprendizaje que aplican y se benefician de la discusión de diversos aspectos del desempeño de las escuelas y de los alumnos. Se dedica atención al uso deseable de los modelos de valor agregado: para subgrupos específicos de la población estudiantil y aspectos específicos de las escuelas; para establecer objetivos de desempeño y proyecciones de desempeño; para identificar a los alumnos que necesitan ayuda especial e intervenciones oportunas; y para mejorar el sistema de evaluación, particularmente lo que se refiere a evaluaciones escolares y de desempeño. Al final del Capítulo 3, se presenta un resumen como referencia general titulado "***Establecimiento de un sistema basado en modelos de valor agregado: Pasos básicos en la fase de implementación***". El resumen no pretende ofrecer una lista exhaustiva de todas las actividades necesarias, sino ayudar a las autoridades educativas que esperan comprender con rapidez el proceso requerido para establecer un sistema de evaluación basado en modelos de valor agregado.

La Parte II analiza el diseño de modelos de valor agregado y se centra en los aspectos técnicos. El Capítulo 4 analiza las principales consideraciones de diseño al elaborar un sistema de modelos de valor agregado e identifica los retos básicos que necesitan abordarse. El Capítulo 5 presenta ejemplos de los principales tipos de modelos de valor agregado para ilustrar sus diversos requerimientos, y cómo deben adaptarse a contextos particulares. El Capítulo 6 presenta las consideraciones estadísticas y metodológicas básicas en la elaboración de modelos de valor agregado. Se destaca la identificación de los criterios básicos para ayudar a elegir un modelo o modelos de valor agregado que sean más apropiados para un sistema educativo. Se señala que un aspecto clave para las autoridades educativas es decidir cuál es el modelo de valor agregado más adecuado para cumplir con los objetivos y el uso planeado que se les dará a los modelos de valor agregado.

La Parte III del informe analiza la implementación de sistemas basados en modelos de valor agregado en los sistemas educativos. Este análisis guía a los responsables de formular políticas y a las autoridades en el establecimiento de un sistema que mejor responda a sus necesidades.

Experiencias de distintos países ilustran los principales problemas y estrategias potenciales que se pueden emplear. El Capítulo 7 se centra en los pasos que necesitan darse en la elaboración del sistema y que incluyen la fase ***piloto*** de implantación. El Capítulo 8 examina la puesta en marcha, con considerable atención al desarrollo de una política de comunicación y compromiso de los actores clave del sistema educativo. Esta política de compromiso debe acompañar la introducción de un sistema de modelos de valor agregado e incluir la formación de los actores relevantes. Será necesario expresar con claridad las acciones y consecuencias para directores escolares, maestros y otros actores, no sólo para generar confianza en el nuevo sistema sino también para disipar temores por la introducción de un sistema que puede percibirse como potencialmente carente de justicia y transparencia. Será necesario elaborar estrategias específicas que expliquen el sistema e informen a los actores sobre aspectos básicos, como el cálculo de las puntuaciones de valor agregado, por ejemplo. En este proceso de comunicación, información y formación, es importante destacar los beneficios que ofrecen los modelos de valor agregado en comparación con otras medidas de desempeño. En varios países que se toman como ejemplos, representantes de distintos sectores y actores clave recibieron favorablemente la elaboración y uso de modelos de valor agregado una vez que se logró comunicar aspectos importantes: su mayor precisión ofrece una medida más justa del desempeño escolar. Sobre ello, se pueden construir sistemas más equitativos de rendición de cuentas de las escuelas y elección, así como iniciativas más precisas, y por ende más eficaces, de mejora escolar.

Parte I

Objetivos y uso de modelos de valor agregado

CAPÍTULO 1

Objetivos de políticas para el desarrollo de un sistema basado en modelos de valor agregado

Uso de modelos de valor agregado para procesos de mejora escolar **22**
- Toma de decisiones basada en datos **22**
- Precisión de las mediciones de desempeño **24**
- Uso de las características socioeconómicas en los modelos de valor agregado **27**

Rendición de cuentas de las escuelas mediante los modelos de valor agregado **30**
- Incentivos y resultados indeseados **35**

Elección de la escuela con información de valor agregado **37**

Conclusión **40**

El interés actual en el desempeño escolar en muchos países gira en torno a temas sobre la eficacia de inversiones en las escuelas, junto con la preocupación extendida sobre la competitividad económica nacional. Dado el papel central del capital humano en la economía moderna (Friedman, 2005; OCDE, 1994, 1996, 2001, 2010), las escuelas de una nación se ven como fuente potencial de ventaja competitiva. Una preocupación relacionada es que la existencia de niveles sustanciales de heterogeneidad en el desempeño escolar junto con diferencias significativas en resultados educativos para grupos reconocibles en la población generen tensiones sociales y creen ineficiencias económicas (OCDE, 2008a; Lucas, 1988; Romer, 1994). Para abordar de forma adecuada estos temas se necesitan métodos para medir con exactitud el desempeño escolar y evaluar así con eficacia las inversiones en escuelas, identificar las mejores prácticas y destacar las áreas donde se necesitan mejoras. Un sistema así debe transmitir adecuadamente esta información para ilustrar cómo efectuar dichas mejoras y aumentar el desempeño de todas las escuelas.

Un análisis de valor agregado se diseña para evaluar a las escuelas con base en lo que sus alumnos aprendieron mientras estaban inscritos en ellas. Las puntuaciones de valor agregado de la escuela son agregaciones de las trayectorias del desempeño estudiantil individual que pueden recibir influencia de diversos factores, además de la influencia propia de la escuela. Las estrategias de valor agregado pretenden aislar la contribución de la escuela al aprendizaje estudiantil de otros factores asociados (en sentido estadístico) al aprendizaje estudiantil, como los antecedentes socioeconómicos de los alumnos. Sea cual fuere la ambición que impulse el desarrollo de un sistema basado en modelos de valor agregado, existe la necesidad de elaborar mediciones precisas del desempeño escolar; mediciones que reflejen el desempeño real de una escuela y no factores que más o menos sean ajenos al control de la escuela, como diferencias en la composición estudiantil o el "ruido aleatorio" estadístico. Los modelos de valor agregado ofrecen mediciones de desempeño escolar que, para la mayoría de los sistemas educativos, mejorarán en gran medida los datos e información usada hoy en día para dar información en la toma de decisiones. De manera importante, las mediciones de valor agregado proporcionan medidas precisas de la contribución de la escuela al desempeño estudiantil, lo que resuelve muchos problemas con las mediciones de desempeño escolar actuales. La información más precisa sobre el desempeño escolar y estudiantil facilita iniciativas más dirigidas y bien definidas que rindan mejoras sostenidas.

Los modelos de valor agregado sirven para centrar la atención en programas educativos o grupos de estudiantes particulares con desempeño bajo o alto. Esta información es útil para que las autoridades educativas, administradores y directores escolares y maestros identifiquen mejor los problemas de desempeño, y guíen el desarrollo y evaluación de programas escolares. Las políticas y programas diseñados para aumentar el desempeño requieren una forma de evaluación que identifique tanto las áreas de alto desempeño como las que necesitan mejoras. Los modelos de valor agregado ofrecen indicadores cuantitativos precisos de desempeño que facilitan la identificación de áreas para mejorarse al interior de las escuelas y sistemas escolares, permiten la creación de puntos de referencia de desempeño y facilitan el aprendizaje dentro y entre las escuelas. Los modelos de valor agregado también sirven para incrementar la eficacia de instituciones existentes, como la inspección escolar, y permiten juicios más informados sobre las escuelas.

Con base en ejemplos de 13 países, este informe se centra en tres objetivos amplios de políticas que impulsan el desarrollo de modelos de valor agregado en un sistema educativo: ***iniciativas para procesos de mejora escolar, rendición de cuentas de las escuelas y elección de escuela***. En este capítulo se delinean estos temas y se continua su análisis en capítulos subsecuentes mediante la ilustración de diversas aplicaciones –en los ámbitos del sistema educativo y de las escuelas– de modelos de valor agregado. Aunque estos tres objetivos amplios de políticas difieren en su interés central y en el desarrollo de programas para cumplirlos, todos tienen el objetivo general de mejorar los resultados de los sistemas educativos. El análisis de estos objetivos de políticas es relativamente breve, en vista de su amplitud y complejidad. Se dedica más atención a la presentación y aplicación de modelos de valor agregado que promuevan estos objetivos de políticas en los capítulos 2 y 3, pues se considera de mayor relevancia e interés para las autoridades educativas.

Es importante subrayar que, para la mayoría de los sistemas educativos, el establecimiento de un sistema basado en modelos de valor agregado tiene múltiples objetivos. Por ejemplo, en el **Reino Unido**, los modelos de valor agregado ahora se usan:

- En Cuadros de Desempeño escolares[1] para ofrecer información a padres y considerar escuelas.
- En sistemas para mejoras Escolar, donde los datos sirven para autoevaluaciones y establecimiento de objetivos.
- Para informar inspecciones escolares incorporadas al proceso más amplio de mejoras escolares.
- Para contribuir a seleccionar escuelas para iniciativas particulares.
- Para ofrecer información sobre la eficacia de tipos particulares de escuela o iniciativas de políticas.

Estos múltiples objetivos ilustran la importancia de obtener mediciones de desempeño precisas en varias áreas de sistemas educativos. Asimismo, debe reconocerse que los programas específicos pueden atender múltiples objetivos de políticas. Muchos programas e iniciativas agrupados en una política de mejoras escolares también serían aplicables a fines de rendición de cuentas de las escuelas. En algunos casos, las iniciativas para promover elecciones informadas de escuelas impondrían una forma de rendición de cuentas a las escuelas, maestros y funcionarios escolares.

Surgen más complicaciones en un informe como éste, dirigido a las autoridades correspondientes y educadores que operan en diferentes contextos políticos y culturales. Estas diferencias pueden generar interpretaciones divergentes de los objetivos de programas particulares. La historia de un sistema educativo, la interacción entre instituciones educativas y el clima actual del desarrollo del sistema afectan la forma de apreciar una política o iniciativa particulares. Un ejemplo es el desarrollo de un sistema que aplique modelos de valor agregado a la mejora de la elección de escuelas. En esta situación, los resultados de valor agregado de las escuelas estarían disponibles para los padres al publicarse en un sitio de Internet centralizado que permita a padres y estudiantes conocer mejor el desempeño de la escuela. El grado en el que esta medida también se considere una forma de rendición de cuentas de las escuelas difiere en función del contexto y del desarrollo histórico del sistema educativo. Si esta política

se desarrollara en un sistema donde previamente se hubiera analizado poca información de desempeño, habría una mayor probabilidad de que se percibiera como el establecimiento de una forma de rendición de cuentas escolar, a diferencia del caso de un sistema en donde la información sobre el desempeño de escuelas y estudiantes estuviera disponible para el público de forma habitual. Lo que se considera común en un sistema educativo puede representar un cambio significativo en otro. El impacto de las políticas en escuelas, administradores escolares y maestros varía de forma correspondiente. El análisis del uso de modelos de valor agregado para promover objetivos de políticas presentado en este informe no incluye estimaciones sobre el impacto en diversas políticas. El análisis se centra en emplear la información y puntuaciones escolares de valor agregado como base para la acción: el desarrollo y supervisión de iniciativas y prácticas que se pueden establecer para múltiples objetivos de políticas.

USO DE MODELOS DE VALOR AGREGADO PARA PROCESOS DE MEJORA ESCOLAR

Los modelos de valor agregado ofrecen indicadores e información de desempeño precisos, útiles como base para la acción en favor de objetivos de mejoras escolares. Estas acciones diferirán entre sistemas educativos, e incluirán diversas iniciativas que variarán en tamaño y propósito específico. Son más valiosos si no sólo son capaces de documentar el estado actual del sistema, sino también de generar información que apoye mejoras continuas, en particular si se efectúan análisis posteriores más detallados. Por ejemplo, en el ámbito de las políticas, los modelos de valor agregado sirven para identificar escuelas con desempeño alto o bajo y dirigir la atención y financiación a donde más se necesite. Más aún, los análisis estadísticos de las relaciones entre las condiciones iniciales de las escuelas y el desempeño escolar pueden sugerir qué estrategias son más eficaces, lo que conduce a ajustes de las políticas en curso y a reasignaciones de recursos.

Al crear mediciones precisas de desempeño escolar, los modelos de valor agregado facultan a las escuelas y administradores para tomar decisiones más informadas para mejorar el desempeño escolar (Saunders, 2000). Esta información permite el desarrollo de iniciativas a menudo más dirigidas a mejoras escolares. Además, la información de valor agregado faculta a las escuelas y a las autoridades correspondientes a supervisar y evaluar estas iniciativas. Este manejo y toma de decisiones con base en datos resuelve muchas dificultades al evaluar si los recursos se emplean de manera eficaz y, por ende, deben permitir el desarrollo continuo de mejoras en eficiencia y calidad.

Toma de decisiones basada en datos

Avanzar en las mejoras escolares utilizando mediciones de valor agregado requiere prestar mayor atención a la toma de decisiones basada en datos dentro de las escuelas y de los sistemas educativos. En los últimos años, tomadores de decisiones educativas de distintos países vieron una oleada de interés en la toma de decisiones basada en datos, y se dieron varias iniciativas para crear modelos escalables que empleasen los datos como apoyo de mejoras escolares (Saunders, 2000). Los datos y las mediciones juegan un papel crucial en la determinación de estrategias y supervisión del avance hacia objetivos de políticas (Atkinson Review, 2005). En este contexto de toma de decisiones hay un interés explícito en el uso de datos comparativos para identificar, por ejemplo, áreas de mejoras potenciales y establecer metas significativas.

La estructura cambiante de los sistemas educativos también intensificó el interés en la asignación y uso de recursos y el desempeño de las escuelas. Se dio un giro hacia una mayor autonomía escolar con menos regulaciones centralizadas de insumos y procesos (OCDE, 2004, 2008b). Sin embargo, con un control menos centralizado, se precisa que un sistema mida el desempeño de las escuelas, y que esas mediciones estén disponibles de manera sistemática. Por ejemplo, Ryska (2006) analiza como en la **República Checa** antes de 1990 las características habituales del sistema escolar eran un gobierno centralizado, educación uniforme y una supervisión rígida de los maestros. El Ministerio de Educación de la República Checa representó un papel decisivo, utilizando instrumentos de gobierno y control, mientras que los insumos iniciales y los procesos se prescribían con todo detalle. Las inspecciones escolares eran el principal instrumento de supervisión en el ámbito escolar. Los directores escolares y los maestros tenían poca libertad para actuar respecto del cumplimiento del currículo, en términos tanto del contenido como de los métodos de instrucción. En el ámbito estudiantil, las evaluaciones de los maestros constituían el principal método de evaluación, y se centraba sobre todo en evaluar el conocimiento adquirido según el currículo prescrito. Los estándares y desempeño del sistema de evaluación en su conjunto no se supervisaban ni evaluaban. Cambios más recientes dieron lugar a una mayor descentralización en el sistema escolar, con un mayor énfasis en las medidas de desempeño y el uso eficiente de los recursos. Un sistema escolar más descentralizado presentó a todas las partes interesadas una situación nueva. Las escuelas se percataron de la necesidad de una evaluación estructurada y sistemática que realimentase "lo que funciona" en todos los niveles del sistema educativo (Ryska, 2006).

La toma de decisiones basada en datos no sólo debe ser asunto de las autoridades educativas: los participantes de todos los niveles del sistema educativo pueden aprovechar los datos. Los directores escolares y maestros pueden utilizar los datos sobre insumos, procesos y resultados para analizar las asignaciones de recursos y la eficacia de diversas políticas, programas y decisiones administrativas (Odden y Busch, 1998). Es importante ver las escuelas como organizaciones de aprendizaje de la misma manera que otras organizaciones del sector público o privado (Caldwell y Spinks, 1998). Sin embargo, los datos por sí solos no generan el éxito de estrategias más completas de toma de decisiones basada en datos. Debe haber una estrategia sistémica para mejorar el desempeño que emplee mediciones de desempeño precisas y alinee los esfuerzos hacia objetivos claramente establecidos.

Muchos países miembros de la OCDE intentaron, en los últimos años, desplazar el interés del sector público desde las entradas o insumos (*inputs*) al sistema educativo a los resultados, con el fin de mejorar el desempeño del sector público (Eurostat, 2001). En los sistemas educativos, los datos tradicionalmente se han dirigido a los recursos (Atkinson Review, 2005), y la información disponible en las estadísticas oficiales y en los sistemas administrativos oficiales se relaciona sobre todo con los recursos o insumos al sistema en lugar de hacerlo con los resultados. Los políticos tienen a menudo acceso a información muy detallada sobre los recursos e insumos canalizados al sistema educativo. La información financiera sobre el capital y gastos corrientes a menudo se divide en gastos en diferentes escuelas e insumos escolares, como edificios y mantenimiento, y salarios de profesores y demás personal escolar (OCDE, 2007a). Esta información en ocasiones se fragmenta para analizar los gastos en diferentes niveles, como gastos centralizados, regionales, de programas, en escuelas y en alumnos. Los

beneficios de la toma de decisiones basada en datos se optimizan con un análisis combinado de datos de entradas, procesos y resultados. Depender de un solo tipo de datos puede desviar las conclusiones y acciones. Combinar los datos de entradas con los procesos escolares permite un análisis más extenso tanto de la asignación como del uso de los recursos. En muchos países se ha recopilado información sobre procesos escolares durante muchos años. El método más común para recopilar esta información es mediante un marco evaluativo específico que, en la mayoría de los países miembros de la OCDE, incluye inspecciones escolares (OCDE, 2007a). El interés de estas evaluaciones y la información reunida varía por país, pero por lo general el interés se centra en los procesos del ámbito escolar y en asegurar la adhesión a las regulaciones y procedimientos escolares. La información suele recopilarse sobre la forma y estructura de la enseñanza y sobre problemas específicos dentro de las escuelas. En varios países, las inspecciones escolares evalúan el desempeño escolar respecto a criterios determinados en esas áreas (OCDE, 2007a).

Si bien estos análisis son más extensos que un sistema de supervisión centrado sólo en los insumos, la toma de decisiones en este contexto se ve restringida por la falta de datos de desempeño. Las decisiones sobre insumos y procesos no pueden, por tanto, analizarse en términos de sus efectos en el desempeño, y la asignación óptima de recursos y "mezcla" de políticas y programas que afectan los procesos escolares tampoco pueden analizarse de manera eficaz. Una vez obtenida la información de valor agregado, los tomadores de decisiones podrán analizar mejor cómo ajustar los recursos y poner en marcha los procesos escolares adecuados para mejorar el desempeño estudiantil. La inclusión de información de valor agregado permite el aprendizaje organizacional sobre lo que mejor contribuye a las mejoras de desempeño.

Precisión de las mediciones de desempeño

Dada la creciente necesidad de análisis del desempeño escolar, se requiere una medición de logro precisa para medir el avance del desempeño estudiantil y el efecto de la asignación y uso de recursos en el sistema educativo. Sin duda, la precisión de estas mediciones es primordial si se van a emplear en la evaluación y desarrollo del sector educativo. En distintos países, las mediciones de desempeño escolar se concentran en puntuaciones de pruebas sin ajustar o en mediciones de aprovechamiento estudiantil; por ejemplo, las puntuaciones promedio de pruebas estandarizadas o el porcentaje de estudiantes en cada escuela que avanzan a niveles educativos más elevados. Sin embargo, cada vez se reconoce más que hay problemas con estas mediciones de desempeño escolar. Estas mediciones no suelen considerar otros factores que influyen en los logros educativos, como la capacidad innata de los alumnos, sus antecedentes socioeconómicos, aspectos de idioma para estudiantes indígenas, la influencia de sus compañeros e individuos dentro y fuera de la escuela, diversos acontecimientos y situaciones que ocurren fuera de la escuela y tal vez afecten el aprendizaje de los estudiantes, y la aleatoriedad general de las evaluaciones estudiantiles.

En países que descentralizaron la estructura del sistema educativo e introdujeron un nuevo interés central en la rendición de cuentas de las escuelas se reconoce que un énfasis en los datos de desempeño origina preocupaciones sobre la equidad en ausencia de mediciones de valor agregado (Jakubowski, 2008; Hægeland, 2006). Estas preocupaciones causaron consternación entre los interesados en la educación en varios países (Linn, 2004, 2005). Los directores escolares y los maestros perciben que su desempeño se juzga de manera

injusta con la imposición de mediciones de rendición de cuentas basadas en factores ajenos al control de las escuelas. Estas preocupaciones se extienden a las comunidades, familias, asociaciones de padres de familia y sindicatos de educadores (Bethell, 2005; OCDE, 2009). Una preocupación común es que la situación o aprovechamiento de un alumno en un momento dado corresponde a su desarrollo cognitivo previo a su ingreso a la escuela, así como a su crecimiento durante todos sus años en la escuela. De hecho, el desarrollo de un alumno recibe la influencia no sólo de su escolaridad anterior, sino también de sus experiencias fuera de la escuela y del apoyo de su familia y comunidad a lo largo del tiempo. Hacer a la escuela la única responsable de los resultados no es defendible ni justo. Por ejemplo, muchos estudios demuestran que el aprovechamiento estudiantil se correlaciona de manera estrecha con las características familiares y de la comunidad, lo que socava más la credibilidad de usar sólo datos sobre el aprovechamiento actual del estudiante como base para la rendición de cuentas de las escuelas. McCall, Kingsbury y Olson (2004) reportaron correlaciones entre las medias de puntuaciones de la escuela y el porcentaje de estudiantes candidatos a almuerzos gratuitos o con descuento (medida gruesa del nivel de pobreza de la población estudiantil en **EUA**). Los datos se obtuvieron de cientos de escuelas ubicadas en varios estados. El desempeño escolar utilizó puntuaciones de la prueba de Medida del Avance Académico administrado por la *Nortwest Evaluation Association* en 2002 y 2003. Al resumir los resultados de los Grados 3 a 8, las correlaciones en las puntuaciones de lectura oscilaron entre -0.54 a -0.66, y en matemáticas, las puntuaciones fueron de -0.51 a -0.59. Cuando se reemplazaron las medias escolares con una medida muy sencilla basada en los cambios de las puntuaciones de pruebas escolares, las correlaciones en las puntuaciones de lectura oscilaron entre -0.07 a -0.27, y en matemáticas variaron de -0.02 a -0.24. La correlación entre los cambios o avances de las puntuaciones de los alumnos y los almuerzos escolares gratuitos fue por tanto mucho más baja que las correlaciones entre esta medición del estado socioeconómico y las puntuaciones brutas. Esto indica que el desempeño escolar se aísla más fácilmente de otros factores en los análisis del progreso escolar en lugar de depender del desempeño estudiantil en un momento dado. Zvoch y Stevens (2006) ofrecen más evidencias para preferir indicadores basados en el crecimiento escolar respecto de los basados en el rendimiento escolar, con base en un análisis de datos de tres cohortes sucesivas en un distrito escolar grande de EUA. Estos hallazgos proporcionan un apoyo indirecto a las aproximaciones basadas en valor agregado escolar que emplean trayectorias de puntuaciones escolares como variables de entrada. Jakubowski (2008) ilustra que en **Polonia** existe una fuerte convicción de que las pruebas externas sin procesar resultan de poco valor para evaluar la calidad de la enseñanza o del desempeño escolar. Asimismo, en **Noruega** se ven con recelo los juicios sobre desempeño escolar basados en mediciones que no consideren la diversidad de factores ajenos al control de las escuelas y que afectan al desempeño estudiantil (Hægeland, 2006). Si el objetivo que se persigue es el de presentar indicadores que reflejen el desempeño escolar, parece claro que los promedios escolares sin ajustar del aprovechamiento escolar individual son mediciones insuficientes, pues reciben la influencia de muchos factores importantes ajenos al control de las escuelas o que se distribuyen de manera inequitativa entre ellas.

En el **Reino Unido**, a principios de la década de 1990, el nuevo acento en los datos de desempeño para hacer responsables a las escuelas originó inquietudes respeto a si se juzgaría con justicia a las escuelas en ausencia de mediciones de valor agregado. Al mismo tiempo, el desarrollo

de las pruebas de Etapa Clave (*Key Stage*) ofreció la posibilidad de calcular puntuaciones de valor agregado por escuela basado en el avance entre cada Etapa, una vez que estuvieran disponibles los datos nacionales para las cohortes correspondientes a los alumnos. Esto significó que se evaluaría a las escuelas según el desempeño de sus alumnos en pruebas nacionales en inglés, matemáticas y ciencias a los 11 y 14 años de edad, en exámenes nacionales de todas las materias a los 16 y 18, y el avance de los alumnos entre estas pruebas[2] (Ray, 2006; Ray, Evans and McCormack, 2009). Los modelos de valor agregado continuaron su desarrollo con el tiempo. Al comienzo del empleo de mediciones de desempeño escolar, se presentaron los resultados escolares en términos de la proporción de estudiantes que excedía el umbral pertinente en cada materia. En este sentido, las escuelas se compararon con un estándar al ordenarlas según la proporción de alumnos que cumplían con el estándar. Estas llamadas "tablas de liga" (*league tables*), que tanto interés provocan en el público, no incluyeron, cuando se publicaron por primera vez, consideraciones de crecimiento escolar individual. Como se describió anteriormente, las comparaciones basadas en puntuaciones brutas pueden ser contraproducentes si no toman en cuenta este aspecto o el contexto escolar. Como señaló Jane Davidson, ex ministra de Educación de Gales en 2002:

> "No necesito que una tabla de liga (que entonces se basaron en puntuaciones brutas) me diga que el desempeño será mejor en una de nuestras comunidades más ricas que en una de las más pobres" (Bethell, 2005, p. 8).

Por otra parte, un análisis de valor agregado ofrece una medición comparativa del desempeño escolar. Es decir, se compara a cada escuela con el promedio de todas las escuelas incluidas en el análisis, considerando los cambios en las puntuaciones de los estudiantes individuales en las pruebas a lo largo del tiempo. Un análisis de valor agregado se diseña para evaluar a las escuelas sobre lo que sus alumnos aprendieron mientras estaban inscritos en ellas y no en medir, sin pretenderlo, lo que los alumnos ya sabían antes de ingresar en la escuela. Esto se considera una base más justa para comparar escuelas que atienden a poblaciones escolares diferentes, con distintos niveles de habilidades y conocimientos. La retroalimentación que proporciona un programa de formación de profesores que acompañó a la introducción de los modelos de valor agregado en **Polonia** en 2006 ilustra que una mayor precisión se traduce en mayor justicia del sistema (Jakubowski, 2007; véase también el Capítulo 8, para un análisis más detallado). Se destacaron varias áreas en la retroalimentación recibida por parte de los participantes en la formación, que apoyó la introducción de los modelos de valor agregado. Los maestros subrayaron:

- Los beneficios de la objetividad de los resultados de valor agregado que hacen destacar a las buenas escuelas que trabajan con estudiantes en desventaja y combaten comparaciones inválidas basadas en puntuaciones brutas.
- La precisión de las evaluaciones cuantitativas y de los métodos estadísticos.
- La mayor transparencia y comparabilidad de los métodos de valor agregado de evaluación escolar.
- El potencial de una mejor evaluación interna del avance escolar, en especial mediante análisis en el nivel de la escuela (p. ej., analizar las puntuaciones de valor agregado de grupos específicos de estudiantes).
- Los beneficios de una formación amplia y de consultas públicas previas al establecimiento real del sistema basado en modelos de valor agregado.

En este ejemplo, la mayor transparencia y precisión de las estimaciones de valor agregado fueron muy importantes para los maestros y otros interesados. Los niveles de confianza en el sistema aumentaron una vez que los maestros recibieron formación para calcular las estimaciones de valor agregado. Algunos maestros que al principio temían que una nueva medición se empleara para la rendición de cuentas de las escuelas, se adhirieron con entusiasmo a los modelos de valor agregado cuando se dieron cuenta de que era una evaluación mucho más justa que el sistema que ya había estado en uso durante varios años en **Polonia** (Jakubowski, 2007).

Uso de las características socioeconómicas en los modelos de valor agregado

Un argumento básico para elaborar modelos de valor agregado, en lugar de tan sólo emplear puntuaciones brutas de pruebas como mediciones de desempeño escolar, es que las puntuaciones brutas en las pruebas son el resultado acumulado de experiencias de aprendizaje de los alumnos y reciben influencia de muchos factores ajenos al control de las escuelas. Quizá el "factor externo" clave sea la distribución de las características sociales y económicas dentro y entre escuelas relacionadas con el desempeño estudiantil. Muchos estudios muestran que estos factores socioeconómicos influyen en el desempeño y resultados escolares (OCDE, 2007c). Las tablas de desempeño que clasifican a las escuelas según las puntuaciones brutas en las pruebas o la entrada en un nivel educativo más elevado no tienen en cuenta los numerosos factores que afectan a los estudiantes en desventaja y, por tanto, comparan de manera injusta a las escuelas que atienden a estos estudiantes.

En los modelos de valor agregado, una escuela con una población escolar de nivel socioeconómico más bajo que el promedio podría recibir una estimación de valor agregado cercano a cero (es decir, promedio) o superior, aunque la media absoluta del desempeño de sus estudiantes se ubique muy por debajo de la media de todos los estudiantes dentro del sistema escolar. Este punto destaca la importancia de aplicar un análisis de valor agregado. Al hacer inferencias sobre el desempeño escolar es importante tener en cuenta la realidad pues diferentes escuelas enfrentan desafíos muy distintos al educar a sus alumnos. Este poder analítico aumenta con la inclusión de características socioeconómicas contextuales en los modelos de valor agregado. Estos modelos se denominan ***modelos de valor agregado contextualizado*** en este informe. El empleo de características de antecedentes pertinentes genera indicadores de valor agregado escolar más precisos y creíbles.

Dada la necesidad de resultados más creíbles y precisos, los países recopilan y utilizan datos socioeconómicos para los modelos de valor agregado y otras mediciones de desempeño escolar cada vez mas. Sin embargo, se debe observar que las puntuaciones de valor agregado contextualizado de las escuelas pueden ocultar bajos niveles de desempeño estudiantil y por tanto distorsionar los incentivos y la toma de decisiones que tal vez en realidad refuercen desventajas existentes en escuelas con una alta proporción de estudiantes de nivel socioeconómico bajo. Esto tiene un impacto en las propias escuelas y también en el desarrollo de políticas. Los objetivos importantes del uso de los modelos de valor agregado con fines de rendición de cuentas de las escuelas y mejoras escolares incluyen los incentivos creados para elevar el desempeño del alumno y de la escuela, y el uso de datos en el ámbito escolar. Los países que establecieron sistemas basados en modelos de

valor agregado así lo hicieron, al menos en parte, para ofrecer un incentivo más significativo para elevar el desempeño escolar (Bourque, 2005; Ray, 2006). Al publicar los resultados de valor agregado de las escuelas se pueden crear incentivos para administradores escolares, maestros y otros interesados con el fin de elevar el desempeño de las escuelas en esa medida. Sin embargo, el incentivo para elevar el desempeño puede atenuarse en escuelas que tengan puntuaciones sustancialmente mayores de valor agregado contextualizado. Esto puede disminuir las expectativas y reducir los incentivos incluso en escuelas donde la proporción de estudiantes con bajo desempeño absoluto sea preocupantemente alto. Por tanto, la introducción de variables contextuales en el modelo de valor agregado puede tener consecuencias indeseables para los efectos de incentivos en las escuelas. Asimismo, desde la perspectiva de los estudiantes y sus familias, las mediciones de valor agregado pueden despertar menos interés en comparación con mediciones de desempeño absoluto de los estudiantes o del avance del alumno individual.

El uso de características socioeconómicas en modelos de valor agregado contextualizado también pueden tener un impacto negativo en la equidad y eficiencia en la toma de decisiones, si bien esto depende de cómo se emplee la información de valor agregado. Hay beneficios derivados del uso de varias medidas (y modelos) de valor agregado para tomar decisiones más informadas que atiendan distintos propósitos de políticas. Considere la situación de las escuelas con una combinación de una alta proporción de estudiantes con características de nivel socioeconómico más bajo y desempeño académico bajo (medido en puntuaciones de pruebas). Los modelos de valor agregado (sin la inclusión de características contextuales socioeconómicas) pueden mostrar que estas escuelas alcanzan una puntuación de valor agregado relativamente baja.[3] La inclusión de características socioeconómicas en un modelo de valor agregado contextualizado puede mostrar que algunas de estas escuelas tienen puntuaciones altas de valor agregado contextualizado. Para este grupo de escuelas, las puntuaciones brutas de pruebas son bajas y la puntuación de valor agregado de las escuelas también es baja. Sin embargo, la puntuación de valor agregado contextualizado es más alta y puede acercarse mucho al promedio. Para determinar si el uso de las características socioeconómicas contextuales en el modelo de valor agregado facilita el avance de objetivos de políticas declaradas, debe considerarse cada objetivo.

Es útil un análisis de la información de valor agregado del sistema para la toma de decisiones en la asignación de recursos en el sistema educativo. Muchos sistemas educativos ofrecen financiación equitativa para que las escuelas con bajo desempeño reciban fondos adicionales. En este ejemplo, la asignación de fondos sería muy distinta si se emplearan modelos de valor agregado contextualizado y no modelos de valor agregado. Para estas escuelas, sus bajas puntuaciones de valor agregado serían una señal para las autoridades correspondientes de que se requieren recursos adicionales para ayudar a los alumnos en estas escuelas de bajo desempeño. Sin embargo, el análisis de puntuaciones de valor agregado contextualizado (que fueron más altas en estas escuelas) indicaría que estas escuelas no requieren recursos adicionales a pesar del desempeño general tan bajo de los alumnos de estas escuelas. La inclusión de información socioeconómica contextual empeora a los estudiantes de estas escuelas en este escenario. Por tanto, es benéfico y necesario analizar los resultados de modelos que incluyan y excluyan características socioeconómicas.

Los análisis de valor agregado también ayudan a los tomadores de decisiones en los ámbitos del sistema educativo y de la escuela a identificar las escuelas, políticas y programas eficaces. El uso de resultados de valor agregado que no incluyan características socioeconómicas contextuales sería engañoso. Un modelo de valor agregado contextualizado identifica mejor a las escuelas con una mayor proporción de alumnos de antecedentes desfavorables que lograron elevar el desempeño escolar. Este análisis no sería posible con el modelo de valor agregado que no incluyera características socioeconómicas contextuales, y sería aún menos probable con sólo puntuaciones brutas de pruebas. Para un sistema de rendición de cuentas de las escuelas parece por tanto más equitativo emplear puntuaciones de valor agregado contextualizado como indicador principal del desempeño escolar. Por las ventajas y desventajas de estas aplicaciones de modelos de valor agregado y de valor agregado contextualizado, lo óptimo sería emplear diversas mediciones diseñadas para atender distintos propósitos de políticas siempre y cuando se forme a los usuarios para interpretar correctamente las diferencias en los resultados escolares con los diferentes modelos. La información referente a las características socioeconómicas de los estudiantes, puntuaciones brutas de las pruebas y resultados escolares tanto de valor agregado como de valor agregado contextualizado permitirían un análisis más detallado en el cual basar diversas decisiones.

El escenario anterior supone una diferencia sustancial entre las puntuaciones de valor agregado de las escuelas y sus puntuaciones de valor agregado contextualizado. Como veremos en el Capítulo 6, esto no siempre es así. Se argumenta que en sistemas con evaluaciones escolares más frecuentes que se incluyan en los modelos de valor agregado, la importancia cuantitativa y estadística de las características socioeconómicas contextuales se reduce al grado de que tienen un efecto insignificante en los resultados de valor agregado de las escuelas. Este tema se analiza más en el Capítulo 6, pero debe observarse aquí que será importante en la implementación de un sistema basado en modelos de valor agregado para analizar el grado en el que existan estas diferencias entre las escuelas y en el sistema educativo.

Los modelos de valor agregado contextualizado contribuyen a ganarse la confianza de los interesados clave, a quienes les preocupa el tratamiento de las escuelas y maestros que educan a alumnos con desventajas sociales y económicas. La inclusión de estas variables no sólo produce modelos más precisos, sino también envía una señal a esos interesados. Es importante comunicar el mensaje de que un modelo de valor agregado contextualizado mejora el proceso de medición de los alumnos; las dificultades adicionales que enfrentan alumnos indígenas o de antecedentes socioeconómicos más bajos, por ejemplo, tendrán que ser compensadas mediante acciones educativas. Como se ve en la Parte III, los gobiernos emprendieron varias medidas importantes para ganarse la confianza de maestros, funcionarios escolares, padres y otros interesados clave en los sistemas basados en modelos de valor agregado. Estas medidas se centran en aspectos de diseño y uso de los modelos de valor agregado, como la forma de presentar los resultados y de ayudar a los interesados a interpretar bien los resultados escolares. Al conformar una base de conocimientos para emplear los modelos de valor agregado, los interesados otorgan más confianza a los resultados de valor agregado y al sistema que los aprovecha. Esto aligera muchas inquietudes iniciales (Jakubowski, 2008).

RENDICIÓN DE CUENTAS DE LAS ESCUELAS MEDIANTE LOS MODELOS DE VALOR AGREGADO

Durante la década pasada fue más común la adopción de sistemas de rendición de cuentas de las escuelas en varios países (OCDE, 2007a; Kane y Staiger, 2002; Goldstein y Spiegelhalter, 1996; Hanushek y Raymond, 2004; Braun, 2006a; Taylor y Nguyen, 2006). Este desarrollo puede verse como parte de una tendencia internacional más amplia hacia el establecimiento de sistemas que midan el desempeño del sector público en términos de eficacia y eficiencia. El propósito de estos sistemas es facilitar las comparaciones de uso de recursos, resultados y productividad en las instituciones de sectores como salud y educación (OCDE, 2008a).

Las medidas para instituir sistemas de rendición de cuentas a menudo se originan de la preocupación de que existe una heterogeneidad sustancial en el desempeño escolar junto con diferencias significativas en los resultados educativos para subgrupos reconocibles de la población (OCDE, 2007b). El gran aumento del interés en los modelos de valor agregado es una consecuencia de este renovado acento en hacer responsables a escuelas y maestros por su desempeño. Los modelos de valor agregado son una forma de establecer lo que a menudo se denomina ***la rendición de cuentas basada en pruebas***. Si bien la rendición de cuentas de las escuelas es –o debe ser– más amplia que una mera rendición de cuentas usando pruebas, esta última representa con frecuencia un papel predominante. Esto se debe, en parte, a la rentabilidad relativa de aplicar pruebas y, en parte, a la aparente objetividad de sus datos. Sin embargo, es difícil, si no imposible, incorporar todos los aspectos del desempeño escolar en un solo indicador, y algunos aspectos pueden ser inherentemente inconmensurables (Dixit, 2002). Los modelos de valor agregado utilizan puntuaciones de evaluación escolar, de modo que el desempeño en estas evaluaciones sería el interés central de un sistema de rendición de cuentas de las escuelas basado sólo en puntuaciones de valor agregado.

La rendición de cuentas de las escuelas es un componente del sistema de supervisión y las funciones regulatorias del sistema que se efectúa mediante un área o agencia en un ministerio de educación nacional o estatal. Las regulaciones norman, entre otras cosas, la forma de operar de cada componente del sistema, las credenciales necesarias de los diversos profesionales del sistema y los requerimientos para el currículo y evaluación de cada nivel. La supervisión se refiere a los diversos mecanismos mediante los cuales las autoridades inspeccionan el funcionamiento del sistema, así como la forma de reportar los hallazgos internamente, a interesados diversos y al público en general (Caldwell, 2002). Al principio, la rendición de cuentas se centraba en que las escuelas cumplieran con las regulaciones sobre diversos aspectos del funcionamiento escolar, como la cantidad de días de instrucción, tamaño de la clase, credenciales de los maestros, libros de textos empleados, así como varios asuntos relacionados con la administración financiera. En resumen, el interés recaía en entradas, insumos y procesos. Ahora es más común considerar la rendición de cuentas de las escuelas en relación con mediciones de ***resultados***. Afirmar que "las escuelas deben responsabilizarse por su desempeño" debe implicar más que requerir una simple descripción de lo que pasa en las escuelas en un periodo determinado. La rendición de cuentas ahora requeriría que las escuelas ofrezcan un análisis justificativo o explicación de sus resultados. Un aspecto de la rendición de cuentas respecto de la calidad educativa es si los alumnos obtienen progresos satisfactorios en cada año de instrucción. Otro aspecto es si cumplen con los estándares de

las autoridades. Respecto de la meta de calidad, ¿todos los grupos de estudiantes alcanzan las metas de calidad en proporciones más o menos iguales? Quizá sea posible evaluar la meta de eficiencia al preguntar si las escuelas que operan en ambientes más difíciles funcionan de manera tan eficaz como las escuelas con relativamente menos problemas. Esta última pregunta se aborda al relacionar entradas con resultados.

El interés central de los sistemas de rendición de cuentas de las escuelas difiere entre países. Algunos países se interesan mucho más en el desempeño de las escuelas individuales, mientras que en otros sistemas educativos hay más interés en el sistema y relativamente pocas referencias a la rendición de cuentas de las escuelas y en ocasiones relativamente pocos métodos para evaluar el desempeño escolar (OCDE, 2007a, 2008b). Varios países miembros de la OCDE, en años recientes, comenzaron a establecer sistemas de rendición de cuentas de las escuelas. Como ejemplo, en **Noruega**, el concepto de medición de la rendición de cuentas de las escuelas cambió un poco en los últimos años. Con base en recomendaciones de una comisión gubernamental, se estableció un sistema nacional de rendición de cuentas de las escuelas en 2004. El elemento central de este sistema es un sitio de Internet de acceso libre con información detallada de todas las escuelas noruegas. Además de la información administrativa básica, el sitio contiene una gran cantidad de indicadores sobre uso de recursos, ambiente y resultados de aprendizaje. Los indicadores de desempeño escolar son sobre todo promedios o distribuciones escolares brutos de pruebas nacionales o de exámenes administrados centralmente. El objetivo del sistema es mejorar el tipo y nivel de información sobre desempeño escolar para diversos interesados, como el gobierno mismo. No hay sanciones ni recompensas directas por ningún indicador. Se pretende que sea una herramienta central en el proceso de desarrollo escolar, al identificar las buenas prácticas en las escuelas que lo hagan bien, e identificar las escuelas con potencial de mejoras. Las autoridades centrales pueden usar el sistema para supervisar el nivel general de avance del sector educativo, y las locales y dueños de escuelas, el nivel de avance de sus propias escuelas en comparación con las demás. Por último, padres, estudiantes y el público en general ahora cuentan con información extensa y estandarizada sobre aspectos escolares de particular interés para ellos, en lugar de tener que depender en evidencias más anecdóticas y asistemáticas (Haegeland, 2006).

La elaboración y publicación de mediciones de desempeño por sí mismas pueden proporcionar incentivos implícitos e indirectos (monetarios o no) a los directores escolares y maestros (Glenn y de Groof, 2005). Además, los sistemas de rendición de cuentas del sector público también pueden incluir sanciones y recompensas explícitas, tanto para instituciones como para sus empleados. El gobierno, al publicar información sobre desempeño escolar y quizá conectar recompensas al desempeño, puede inducir a los maestros y administradores a responder con un incremento de sus esfuerzos para mejorar el desempeño (Bourque, 2005). Para analizar el tipo y eficacia de los incentivos creados mediante el desarrollo y uso de modelos de valor agregado en sistemas educativos, deben analizarse los incentivos creados para maestros y directores escolares. En los modelos de valor agregado, la escuela es la unidad de rendición de cuentas, y, por tanto, las mejoras de aprendizaje se dirigen a mejoras en la escuela como unidad organizativa. Sin embargo, necesita reconocerse que el impacto en el aprendizaje organizativo de las escuelas se logra sobre todo mediante el impacto en los maestros y en el aula (OCDE, 2005, 2009).

Aunque las escuelas sean la unidad en que se mide el resultado, los incentivos se crean para los maestros y directores escolares individuales, sea de forma colectiva o personal. Estos incentivos tienen dos efectos principales: los efectos incentivos creados para maestros y directores, y, en potencia, la clasificación o selección que ocurre en el mercado laboral en estas profesiones en respuesta a esos efectos (Lazear, 2000; OCDE, 2009).

La estructura de los incentivos influye en las acciones de organizaciones y trabajadores en organizaciones públicas y privadas (Ballou, 2001; Doeringer y Piore, 1985). No hay razón para creer que no sería así también en el sector educativo. Debe esperarse que los maestros y directores escolares respondan a los incentivos positivos y negativos que influyan en la educación que reciben los alumnos (OCDE, 2009). Lavy (2002) ilustra los efectos positivos de un experimento en el sistema de educación en **Israel**, en el que se recompensó a los maestros por aumentar las puntuaciones estudiantiles en pruebas. En una estructura de incentivos diseñada con cuidado, los maestros recibieron diversas recompensas monetarias por mejorar el desempeño estudiantil. Se registraron sustanciales efectos positivos con aumentos del desempeño estudiantil, lo que reflejó la creación de los incentivos. Además, los incentivos se estructuraron de tal manera que se recompensó a los maestros de alumnos de bajo desempeño con más desventajas. Se ofrecieron mayores recompensas a los maestros que lograron incrementos en el desempeño de estudiantes que antes eran de bajo desempeño o se consideraban menos favorecidos. Los resultados positivos fueron claros en estudiantes de bajo desempeño que lograron avances significativos. Esto ilustra las posibilidades para las autoridades educativas en el diseño de estructuras de incentivos para obtener ganancias en el desempeño escolar en áreas donde se consideren más valiosas.

Los incentivos más directos que se crean con los modelos de valor agregado son los que identifican el valor agregado de maestros individuales y ofrecen recompensas proporcionales. Los modelos de valor agregado escolares difieren en su interés, pero aún así ofrecen diversos incentivos a maestros y directores escolares. Estos incentivos emplean resultados tanto monetarios como no monetarios, y llegan a tener efectos diversos. Las puntuaciones de valor agregado escolares se usan para proporcionar incentivos para todos los maestros de una escuela, o pueden fragmentarse las puntuaciones para identificar grupos particulares de maestros (p. ej., maestros de una materia particular). Cabe notar que se crea una capa adicional de complejidad al identificar incentivos de los modelos de valor agregado escolares, que son el interés principal de este informe, pues la unidad de medición (la escuela) difiere del objetivo de los incentivos (maestros). Esto otorga más responsabilidad al liderazgo de las escuelas para asegurar que todos los maestros y miembros del personal trabajen juntos para alcanzar los objetivos de la escuela.

Los incentivos adoptan diversas formas para maestros y directores escolares, y difieren con el nivel de resultados, deseados o no, de un sistema que emplee modelos de valor agregado. Los resultados del desarrollo de sistemas que utilicen modelos de valor agregado, en general, se pueden colocar en las siguientes categorías:

- Resultados monetarios directos: Adoptarían la forma de recompensas y sanciones que alteren la compensación financiera de maestros y directores. Los ejemplos serían bonos financieros a partir de una puntuación de valor agregado escolar positiva o alta (OCDE, 2007a; Figlio y Kenny, 2006).

- Resultados no monetarios: Consisten sobre todo en recompensas como más desarrollo profesional y cambios de responsabilidades laborales, así como la recompensa personal de trabajar en una escuela exitosa. Esto puede generar mayor satisfacción laboral y el prestigio que proviene de una mejor posición en la comunidad profesional. Cabe notar que estos resultados a menudo son evidentes junto con resultados en las otras tres categorías (OCDE, 2005).
- Resultados del lugar de trabajo y escolares: Se pueden otorgar diversas recompensas y sanciones a las escuelas que tengan un gran efecto en maestros y directores. Los ejemplos comunes tienen impacto en la autonomía escolar, al otorgar a las escuelas con desempeño alto más autonomía y colocar a las escuelas de bajo desempeño en periodos de observación y/o designarles intervenciones de inspectores escolares (o sus equivalentes nacionales). Esto crea un incentivo significativo para mejorar resultados tanto por el estigma de estar en observación como el deseo de mayor autonomía por parte de maestros y directores. En casos extremos, estas sanciones llegan también a clausurar las instalaciones y despedir al personal (Ray, 2006; van de Grift, 2007).
- Resultados profesionales: Incluyen resultados tanto monetarios como no monetarios. Se reciben en el transcurso de las carreras de maestros y directores, y provienen de los beneficios de trabajar en una escuela de alto desempeño. Esto depende de la interpretación y uso de la información escolar de valor agregado en el mercado laboral que debe tener un impacto benéfico en la paga futura y las expectativas de ascensos para los miembros del personal asociados a escuelas de alto desempeño (Ladd y Walsh, 2002).

Estos cuatro efectos pueden presentarse juntos u operar de forma independiente. Como los modelos de valor agregado se centran en diversos aspectos de desempeño escolar, los modelos se estructuran para centrarse en resultados particulares según el objetivo del sistema, y la fuerza del incentivo depende de las dimensiones de los resultados, o de las recompensas y sanciones.

Si bien gran parte del interés se dedica a los incentivos directos, las investigaciones revelan que se presenta un efecto igual o mayor en la productividad organizacional mediante los efectos de clasificación y selección en el mercado laboral para maestros y directores (Lazear, 2000). La cooperación de los efectos de la clasificación y selección son ligeramente distintos de los efectos de incentivos directos. Los efectos de los incentivos directos de los sistemas que emplean mediciones de desempeño de valor agregado se centran en el cambio en el trabajo y en la instrucción que ofrecen los maestros actuales, mientras que los efectos de clasificación y selección se centran en el impacto en el mercado laboral de la gente que decide convertirse en maestros y quienes abandonan la profesión. Un análisis de la introducción de un sistema basado en modelos de valor agregado debe incluir los efectos tanto de incentivos directos como de clasificación y selección en el mercado laboral.

El análisis de los efectos de los incentivos se centra en los incentivos para que los maestros y directores escolares incrementen las puntuaciones de valor agregado de estudiantes y escuelas. Se crea un incentivo que busca que los maestros alteren su conducta laboral para aumentar el desempeño escolar. Los efectos de clasificación y selección ocurren porque estos incentivos atraen a las personas de la profesión que creen que pueden aumentar la puntuación de valor agregado de una escuela. Por intuición, esto afectaría la composición de nuevos maestros

al mercado laboral magisterial. De manera correspondiente, afectaría la composición de la retención de maestros, y sería más probable que los maestros menos aptos para contribuir al valor agregado de las escuelas abandonaran la profesión (Lazear, 2000). Estos maestros se sustituirían con los nuevos que se creyesen capaces de contribuir a las puntuaciones de valor agregado de las escuelas. En teoría, las dimensiones de estos efectos dependen en gran medida de las dimensiones de los incentivos. Por ejemplo, si el avance en una profesión depende en gran medida de las puntuaciones de valor agregado de las escuelas y hay beneficios monetarios y no monetarios sustanciales en dicho avance profesional, se magnifican los efectos tanto de incentivos como de clasificación y selección. Sin embargo, la eficacia de estos cambios depende de indicadores y evaluación de desempeño precisos y transparentes, y de cómo se incorporen al sistema más amplio de la evaluación de escuelas y maestros.

Las puntuaciones brutas de pruebas pueden proporcionar incentivos indeseados por la relación imprecisa entre puntuaciones brutas de las pruebas y el desempeño de las escuelas. Las puntuaciones de valor agregado ofrecen una medición más precisa de desempeño escolar que mejoraría los flujos de información en el mercado laboral. Es por tanto posible moldear los incentivos para que su impacto favorezca los resultados de políticas deseados. Un buen ejemplo de esto es estructurar los incentivos para que la mayor parte de su efecto se dirija a estudiantes en desventaja o de bajo desempeño (OCDE, 2009). Por ejemplo, se pueden crear incentivos para que los maestros y directores se trasladen a escuelas de niveles socioeconómicos bajos, donde las mejoras en valor agregado reciban recompensas mayores. De esta manera, el sistema sería capaz de contrarrestar la tendencia en muchos sistemas educativos donde es más probable que los maestros más experimentados trabajen en escuelas con estudiantes de mejor nivel socioeconómico (OCDE, 2005). Pocos sistemas educativos hoy en día vinculan las remuneraciones de maestros y directores directamente a los resultados de los modelos de valor agregado. Sin embargo, debe observarse que el efecto de la clasificación y selección en el mercado laboral es tan importante como los incentivos directos. Este efecto también dura más que los incentivos directos. Como ejemplo, considere al director en una etapa relativamente temprana de su carrera que dirige una escuela en una comunidad de estado socioeconómico relativamente bajo. Ahora, considere un sistema basado en modelos de valor agregado que utilice pruebas en el idioma de instrucción, ciencias y matemáticas en los Grados 3, 5 y 7 en la escuela. Aunque este sistema no se vinculara de modo directo con la remuneración del director, hay un claro incentivo profesional para mejorar el desempeño en estas pruebas en tanto aún queda una porción considerable de su carrera como director. Si el director consigue elevar la puntuación de valor agregado de la escuela, este logro serviría para el mercado laboral. Tras cinco años en esa escuela, es posible que entrase en el mercado laboral y citara las puntuaciones de valor agregado que muestren su capacidad de elevar el desempeño escolar en una escuela de nivel socioeconómico bajo. El director tiene una ventaja relativa sobre otras personas que compitan por un puesto y por tanto esperaría una recompensa proporcional en la medida en que el mercado laboral para directores las ofrezca. Este incentivo aumentaría si se asignaran más recursos para recompensar a los directores y maestros de estos alumnos. Sin embargo, Ladd y Walsh (2002) ilustran que si las mediciones de desempeño de la escuela no se especifican bien y los incentivos no están bien estructurados, sucede lo contario de este patrón: los maestros se irían a las escuelas que atiendan a los alumnos con más ventajas sociales.

Las dimensiones de estos efectos recibe la influencia de la estructura del mercado laboral para maestros y directores escolares, que varía considerablemente por país. Por ejemplo, un sistema educativo con un mercado laboral más flexible y un grado de autonomía escolar relativamente más alto sería capaz de crear mejores incentivos profesionales. Otro factor clave es la medida en la que la información de valor agregado esté disponible y pueda utilizarse tanto por empleadores como empleados para informar sus contrataciones, despidos y movilidad en general entre escuelas dentro del mercado laboral. Sin embargo, las cuatro categorías de incentivos enlistadas antes ilustran el punto de que los incentivos para elevar el desempeño estudiantil y escolar pueden crearse en sistemas educativos que no ofrezcan incentivos monetarios directos o pagos por mérito a los maestros. Los incentivos no monetarios, de lugar de trabajo y profesionales también facilitan el aumento del desempeño escolar. Esto reviste particular importancia por los relativamente pocos países miembros de la OCDE que ofrecen pagos por desempeño a los maestros (OCDE, 2007a).

Incentivos y resultados indeseados

Siempre que se crea una medición de desempeño existe el potencial de resultados negativos o indeseados si los procesos o incluso los resultados se manipulan para crear erróneamente una medición de desempeño positiva. La manipulación puede ser resultado directo de incentivos perversos creados mediante el establecimiento del objetivo de desempeño. Estos incentivos no deseados surgen cuando la medición de desempeño tiene tanto un gran impacto en los actores como se centra en un aspecto escolar que no refleje el propósito y objetivos verdaderos o generales de las escuelas. Por desgracia, esto es común en mediciones de desempeño escolar si la medición de desempeño se define de manera demasiado estrecha, como al centrarse en una materia específica o en un nivel de desempeño específico, o la medición no mide con precisión el desempeño escolar.

Sin duda, la elección de evaluaciones para los modelos de valor agregado crea un incentivo para mejorar el desempeño en dichas evaluaciones. Un incentivo perverso en potencia puede generar un resultado indeseado si los recursos se dedican a aumentar el desempeño en esas evaluaciones específicas a expensas de otras áreas escolares (Nichols y Berliner, 2005). Sin embargo, debe notarse que esto es un resultado indeseado sólo si no se corresponde con la finalidad pretendida. Un mayor acento en las evaluaciones que crean la medición del desempeño escolar puede ser una consecuencia deseada y una característica de diseño del sistema de administración de desempeño. Existen los mismos incentivos si un nivel de desempeño específico es el interés de la medición de desempeño. Por ejemplo, si se calcula el valor agregado para estudiantes que alcancen un nivel de referencia de alfabetización específico, se crea el incentivo para centrarse en un subgrupo particular de estudiantes a expensas de otros estudiantes. Por tanto, es preciso mucho cuidado al usar puntuaciones de valor agregado para identificar escuelas como de desempeño ***bajo*** o ***alto***. A menos que se especifiquen de manera explícita objetivos específicos, como niveles mínimos de alfabetización con las consecuencias detalladas, la medición de desempeño de una escuela debe centrarse en el desempeño de estudiantes de todas las capacidades.

Como ya vimos, las dimensiones del incentivo creado dependen de las acciones provenientes de la medición de desempeño. Cuanto mayor sea el impacto en escuelas y maestros (p. ej., recompensas o sanciones financieras), mayores serán los incentivos creados. Además, el grado

al que se alteren las prácticas de enseñanza y el currículo con la implementación de un sistema dependerá del grado de autonomía que posean la escuela y los maestros. Si bien la mayoría de los sistemas educativos otorga grados significativos de autonomía a las escuelas y maestros respecto de las prácticas de enseñanza que aplican, muchos tienen un currículo prescrito (OCDE, 2007a). Sin embargo, dentro de un currículo prescrito suele quedar margen para permitir a las escuelas y maestros destacar ciertos aspectos y aplicar prácticas dadas, como que una evaluación escolar se centre en mediciones particulares. Un ejemplo muy citado del impacto de las medidas de desempeño escolar es el de "enseñar para el examen" en sistemas con pruebas ambiciosas (Haney y Raczek, 1993; Kohn, 2000).

Otro problema es la posibilidad de estrechar el currículo. Muchos sistemas no incorporan la evaluación escolar en todas las materias. La viabilidad de esta restricción y varias limitaciones de recursos puede impedir una estructura de evaluaciones escolares de este tipo. En cambio, suelen emplearse evaluaciones en sólo unas cuantas materias básicas (véase el Cuadro 4.1). Reducir la cantidad de materias evaluadas puede crear un incentivo para adaptar el currículo escolar y las prácticas de enseñanza para alcanzar mejores mediciones de desempeño en las materias de interés para la medición de desempeño de valor agregado, lo que podría disminuir el acento en toda la gama de materias disponibles para los alumnos. Este efecto de reducir el foco de interés se aplica a todo tipo de mediciones de desempeño, no sólo a las puntuaciones de valor agregado.

La mayoría de los países sólo incluyen dos o tres materias en sus evaluaciones estudiantiles adecuadas para los modelos de valor agregado. Las más comunes son el idioma de instrucción, las matemáticas y las ciencia (véase el Cuadro 4.1). Los directores y maestros tienen por tanto el incentivo de centrarse más en las materias incluidas en la medición de desempeño. Sin embargo, es importante notar que no hay evidencia sistemática de una reducción de las materias enseñadas en las escuelas sujetas a estas mediciones de desempeño (Jacob, 2002). No obstante, en un estudio de escuelas en **EUA**, O'Day (2002) descubrió que la especificación de la prueba usada en las pruebas importantes se convirtió en las especificaciones de los currículos de varias escuelas.

Los incentivos que se centran en mediciones de desempeño definidas de manera más estrecha no deben verse sólo en un contexto negativo. Un interés mayor en resultados particulares de evaluaciones escolares puede tener un efecto positivo, en particular si se considera que las escuelas o sistemas tienen objetivos mal alineados. Esto puede ser particularmente cierto si un interés mayor en las áreas de evaluación tiene un efecto positivo de seguimiento en otras áreas de instrucción y aprendizaje no incluidas en la medición de resultados. Por ejemplo, un sistema que proporcione un incentivo para aumentar el interés en el desempeño estudiantil en medidas particulares de matemáticas puede tener un efecto positivo en el aprendizaje estudiantil en otras áreas. Esto puede ocurrir por dos razones. En primer lugar, la mejora de los aspectos medidos de matemáticas facilitaría el aprendizaje escolar en otras áreas de matemáticas y en otras materias. En segundo lugar, un mayor interés en mejorar el desempeño escolar en matemáticas alentaría a otras áreas de la escuela a aprender de estas experiencias y aumentar la eficacia en toda la escuela. Esto puede tener un efecto de seguimiento de desempeño estudiantil en áreas no medidas.

El impacto en el currículo puede ser el objetivo de una política directa, pero deben evaluarse los impactos deseables e indeseables para evitar las consecuencias indeseables de esta elección. Como estas consecuencias pueden ser positivas y negativas, parece adecuado que las autoridades educativas supervisen estos resultados mediante la elaboración de modelos de valor agregado en su sistema educativo. Esto se añadiría a la información en el sistema educativo que pueda ayudar al desarrollo escolar y del sistema. Se destaca aquí la elección de medición de resultados para ilustrar el punto de que si los modelos de valor agregado tendrán un efecto en las escuelas, es importante la elección de las áreas temáticas que sean evaluadas. Existen problemas semejantes con el proceso de cálculo de las puntuaciones de valor agregado de las escuelas en múltiples evaluaciones. Aunque una puntuación de valor agregado se calcula por evaluaciones en cada materia, si la puntuación de valor agregado de una sola escuela va a aplicarse a un sistema de rendición de cuentas de las escuelas, es necesario elegir entre puntuaciones de valor agregado de diferentes materias. Como ya vimos, se eligen materias específicas si es necesario dar más interés a esa área de aprendizaje. De otro modo, se calcula y emplea un promedio de varias materias. Sin embargo, en estas circunstancias, el promedio puede ocultar diferencias entre materias (Wilson, 2004). También debe notarse que la elección de mediciones de evaluación en los modelos de valor agregado no debe oscurecer la necesidad de utilizar otras mediciones al tomar decisiones sobre mejoras escolares y otros objetivos de políticas. Estas mediciones pueden incluir datos sobre las entradas escolares y diversas mediciones de procesos escolares.

ELECCIÓN DE LA ESCUELA CON INFORMACIÓN DE VALOR AGREGADO

La eficacia de las decisiones de rendición de cuentas de las escuelas depende en gran medida de la precisión y de lo adecuado de la medición de desempeño del que se hace responsables a las escuelas. Por tanto, la información de valor agregado necesita ser precisa y transparente, lo que se mejora con la publicación de los resultados de valor agregado de las escuelas. Esta información también sirve para mejorar la elección de escuela. Sin embargo, debe observarse que en varios países no es posible elegir escuela. Las familias no tienen el derecho de elegir la escuela a la que va su hijo. En la mayoría de estos sistemas, el estudiante tan sólo asiste a la escuela local, independientemente de los deseos de la familia (OCDE, 2006). En otros países es posible elegir escuela con la limitación de las escuelas a las que pueden asistir los estudiantes y los requerimientos de ingreso para escuelas determinadas, lo que representa un obstáculo más para la libre elección de escuela. Además, los países quizá no establezcan requerimientos legales o administrativos para la elección de escuela, pero la cercanía geográfica de las escuelas, y su capacidad de atender una demanda elevada, limitan el grado real de la libertad de elección de escuela.

Mucho se ha escrito sobre la elección de escuela y cómo mejoran los sistemas educativos al permitir a los estudiantes y sus familias elegir la que mejor sirva a sus necesidades (Hoxby, 2003). Mediante este mecanismo, la educación mejora conforme las necesidades de aprendizaje de los estudiantes se satisfacen mejor (Levacic, 2001). Las familias eligen la escuela para sus hijos por varias razones: cercanía, los programas que ofrece, el grupo de compañeros al que se integraría su hijo y la orientación religiosa son sólo algunas razones por las que las familias eligen escuela. Las puntuaciones de valor agregado de las escuelas también se convertirían en un factor importante para familias y estudiantes al elegir la escuela a la que deseen asistir (OCDE, 2006).

Las señales que envían los estudiantes y familias al elegir las escuelas que mejor satisfacen sus necesidades son elementos básicos de los beneficios propuestos de una mayor elección de escuela dentro de los sistemas educativos. Conforme los estudiantes y sus familias acuden a las escuelas que se ajustan mejor a sus necesidades educativas, esto ofrece a las escuelas, administradores y autoridades correspondientes información clara sobre las escuelas que los padres y familias consideran más eficaces (Hoxby, 2003). Esto proporciona información para las decisiones de asignaciones de recursos, los procesos y programas que se ofrecen y ponen en marcha al interior de las escuelas, y también para el aprendizaje en el ámbito del sistema. Un aspecto clave de la provisión de información para la elección de escuela es que los interesados conozcan el desempeño escolar (OCDE, 2006). Aunque esto tiene claras implicaciones para la rendición de cuentas de las escuelas ante dichos interesados, también facilita la participación de los interesados en la mejora del desempeño escolar. Cuando los interesados tienen acceso a información fiable y mediciones precisas de desempeño escolar, están en posibilidades de comprometerse con las escuelas en sus acciones para elevar el desempeño. Para hacerlo, los interesados deben ser capaces de interpretar adecuadamente la información de valor agregado. Esto se analiza en el Capítulo 2 y la Parte III de este informe.

El empleo de la información escolar sobre evaluaciones y desempeño difiere entre países miembros y asociados de la OCDE. En unos cuantos países hay relativamente poca información sobre el desempeño estudiantil en pruebas o evaluaciones nacionales. Más o menos dos tercios de los países miembros de la OCDE hacen pública la información sobre evaluaciones e inspecciones escolares. Poco menos de la mitad de estos países reporta hacerlo para mejorar la toma de decisiones en un sistema de elección de escuela (OCDE, 2007a). Desde 2001, la *Swedish National Agency for Education* publica en **Suecia** para toda la nación información sobre los resultados escolares individuales y otros datos (Antelius, 2006). El propósito es facilitar la identificación de los factores que influyen en los resultados escolares y contribuir como antecedentes para discusiones y análisis de oportunidades, procesos y resultados en las escuelas. Dicha agencia también publica los resultados escolares esperados por escuela.[4] Los resultados escolares esperados se estiman mediante análisis de regresión lineal.[5] El residuo, calculado como la diferencia entre resultados escolares (en términos de puntos de grado promedio) y el resultado esperado de la escuela, sirve después como indicador del desempeño escolar, dada la composición de estudiantes en las escuelas. Sin embargo, éstas no son mediciones de valor agregado, y por tanto no estiman la contribución escolar individual al avance de los estudiantes con el tiempo.

En **Francia**, el ministerio de Educación publica resultados de desempeño escolar que miden el desempeño de los alumnos en escuelas al obtener el *baccalauréat*. Éstas no son mediciones de valor agregado, pero el propósito del ministerio de Educación al publicar los indicadores de desempeño de los *lycées* cada año es dejar disponible la información del desempeño de los servicios de educación pública nacionales y dar a los encargados de las instituciones educativas las herramientas adecuadas que los ayuden a mejorar la eficacia de sus políticas y programas (MNEHER, 2006). La publicación de resultados es sensible, pues no hay una definición única de lo que constituyen "buenos resultados" de un *lycée* individual. Por ejemplo, queda sin despejarse la cuestión referente a los criterios que deben adoptarse al evaluar los resultados de un *lycée*. En este caso, los objetivos de estudiantes y padres pueden diferir. Algunos destacan obtener el *baccalauréat* en una serie dada, y por ende están dispuestos a repetir un año o a

cambiar de institución para hacerlo, mientras otros prefieren completar toda su educación en un *lycée*. Otros más tan sólo desean su *baccalauréat* tan pronto como sea posible. En general, se considera que de poco sirve establecer una lista o clasificación de los *lycées* de mejor desempeño, y que puede determinarse cualquier cantidad de indicadores que correspondan a las diversas expectativas de diferentes personas. Como resultado se dan dos guías para delinear los indicadores de desempeño de los *lycées*:

- Dar puntos de vista complementarios sobre los resultados de los *lycées*.
- Ofrecer una evaluación relativa de la contribución de las instituciones, que tome en cuenta las características de sus estudiantes.

Se supone que los padres, personal educativo nacional, periodistas y gran parte de los actores públicos y privados se interesan en evaluar el desempeño de cada *lycée* individual y la contribución que hace al nivel inicial de los alumnos inscritos allí. Al publicar los indicadores de desempeño de los *lycées* cada año, el ministerio se esfuerza en dar información que responda esta cuestión tan sensible (MNEHER, 2006).

En el **Reino Unido** se emplearon puntuaciones brutas de pruebas para facilitar la elección de escuela antes de la elaboración de su extenso sistema de análisis de valor agregado. En 1992 se introdujeron los Cuadros de Desempeño Escolar con el propósito de dar información a los padres cuando eligieran escuela, y dar a las escuelas un incentivo para elevar sus estándares. Los primeros cuadros mostraron resultados de las pruebas GCSE de jóvenes de 16 años de edad (junto con un indicador para los niveles A de los de 18 años de edad). En 1996 se elaboraron los primeros cuadros para las escuelas primarias con resultados de las nuevas pruebas de la Etapa 2, de chicos de 11 años de edad. Con el tiempo, los cuadros llegaron a incluir más indicadores, en parte como resultado de la mayor cantidad de información disponible en el ámbito nacional. En 2002 se incluyeron las primeras puntuaciones de valor agregado de todas las escuelas secundarias, y siguieron las de las escuelas primarias un año después. Los objetivos de los cuadros son aún ofrecer datos nacionales consistentes y accesibles sobre el desempeño de las escuelas, dar información a los padres y al público de forma más general, y asegurar que las escuelas sean responsables de sus resultados (Ray, 2006). Los cuadros son intensivos en sus recursos para presentarse de forma precisa cada año, y deliberadamente se limitan a una variedad determinada de indicadores clave. Por tanto, no ofrecen resultados o valores agregados de cada materia tomada. Se canaliza a los usuarios a los informes de inspección del *National School Inspectorate* para la imagen más completa de una escuela determinada. También se comunica a los usuarios que las mediciones de valor agregado representan un mejor estimado del desempeño escolar que los resultados brutos que no toman en cuenta el aprovechamiento anterior. Como ya observamos, los nuevos Perfiles Escolares también incluyen las mediciones de valor agregado de los Cuadros de Desempeño Escolar, cuya presentación se analiza en el Capítulo 2.

El uso de los modelos de valor agregado del desempeño escolar permite elegir escuelas con base en mediciones más precisas. Por tanto, debe aumentar la eficacia de un sistema de elección de escuela al grado que el desempeño escolar determine la elección de la escuela más adecuada. Las mejoras en la toma de decisiones se deriva de que los padres estén mejor informados sobre el desempeño de las escuelas. La elección eficaz de escuela se facilitaría

más si la información y puntuaciones de valor agregado provinieran de diferentes grupos de estudiantes (Wilson, 2004). Esto permitiría que los padres y estudiantes de esos grupos eligieran mejor la escuela que satisface sus necesidades educativas. Como ya analizamos, quienes deciden pueden utilizar la información reunida de observaciones de las escuelas que las familias elijan para satisfacer sus necesidades. Si las elecciones de las familias están mejor informadas gracias a los modelos de valor agregado, las decisiones mediante el sistema escolar también estarán mejor informadas. Esto aumenta la eficiencia del sistema de dos maneras: con las familias en posibilidades de enviar a sus hijos a las escuelas que se ajusten mejor a sus necesidades educativas, y con un sistema escolar capaz de aprender de estas elecciones y desarrollar prácticas escolares que produzcan un mejor desempeño. La elección de escuela tendrá un impacto positivo reducido en un sistema educativo que carezca de indicadores significativos de desempeño escolar. Los padres y las familias no pueden tomar decisiones informadas, las escuelas y las autoridades educativas no pueden emprender respuestas que mejoren el desempeño ante un patrón cambiante de demanda basado en mediciones precisas de desempeño escolar, y las escuelas no reciben una recompensa adecuada por su desempeño.

La provisión de información de valor agregado promueve una cultura de toma de decisiones basada en datos que fomenta las mejoras escolares. Esta toma de decisiones permitiría respuestas eficaces a los cambios de demanda de educación. Es benéfico ofrecer más que una sola medición de desempeño para brindar información al elegir escuela. La provisión de datos de valor agregado junto con datos de puntuaciones "brutas" de pruebas proporciona a padres y familias más información con la cual elegir escuela. Al elegir la escuela que mejor se ajuste a sus necesidades, las familias pueden interesarse también en el desempeño general de los alumnos en las escuelas y en las diferencias de las mediciones de valor agregado de desempeño escolar. Los esfuerzos por educar a las familias y al púbico en general para interpretar las mediciones de valor agregado y las diferencias con puntuaciones brutas de aprovechamiento demostrarán ser benéficas para el sistema de elección de escuela. Las iniciativas para informar y educar a los usuarios de los datos de valor agregado se han considerado cruciales en varios países. Se analizan con más detalle en la Parte III de este informe, que trata sobre la implementación de modelos de valor agregado.

CONCLUSIÓN

Tres objetivos clave de políticas educativas (porcesos de mejora escolar, rendición de cuentas de las escuelas y elección de escuela) se presentan aquí por separado, pero a menudo se les considera objetivos complementarios, en especial por los crecientes niveles de autonomía escolar en varios sistemas. Por intuición, los mayores niveles de precisión que se alcanzan con los estimados de valor agregado conforme se mide el desempeño escolar aumentan los impactos de eficiencia de las iniciativas de descentralización en el sistema educativo. Conforme la descentralización traslada las responsabilidades de la toma de decisiones al ámbito escolar, la información de valor agregado permite que esta toma de decisiones se realice de manera informada. Faculta a las escuelas para asignar recursos de forma más eficiente y para modificar la educación que ofrecen de modo que mejoren sus resultados de valor agregado. Pero esta toma de decisiones requiere cierto grado de autonomía escolar que permita a las escuelas ajustar la educación que proporcionan para cumplir mejor con las demandas de estudiantes y padres en un sistema que destaca una mayor elección de escuela.

En algunos sistemas educativos se combinan la descentralización del sistema escolar, el sistema de elección de escuela y los mecanismos de financiación de las escuelas para ofrecer un incentivo que haga que las escuelas compitan por los estudiantes y, por tanto, por mayores recursos presupuestarios. La elaboración de un sistema basado en modelos de valor agregado incrementaría la eficacia de este sistema. La descentralización permite que las escuelas respondan a los cambios de demanda educativa para atraer a más estudiantes (Sandstrom y Bergstrom, 2005). Por los estudiantes adicionales que atraigan, las escuelas también reciben más recursos de la unidad administrativa central, pues la financiación se asigna por alumno. Esto depende de un sistema de elección de escuela que permita a los padres y familias escoger la escuela que mejor se ajuste a sus necesidades. Estas elecciones requieren que esté disponible la información en la que las familias basan sus decisiones. Conforme los modelos de valor agregado ofrezcan mediciones más precisas del desempeño escolar, mejoraría la toma de decisiones y los estudiantes elegirían las escuelas con mayores puntuaciones de valor agregado. Después se recompensaría de forma adecuada a estas escuelas por su mejor desempeño de valor agregado. La mayor eficacia de usar información de valor agregado para promover la elección de escuela mejora así la eficacia de la asignación de recursos en el sistema educativo.

NOTAS

1. Formalmente, en el Reino Unido, se denominan Cuadros de Aprovechamiento y Logros de Escuelas y Colegios (*Achievement and Attainment Tables*). Por razones de espacio en este informe los denominamos "Cuadros de Desempeño Escolar".

2. El valor agregado se modela en el Reino Unido sobre las evaluaciones estudiantiles al final de cada una de las 4 Etapas Clave de escolaridad. La instancia conocida como la *Qualification and Curriculum Authority* mantiene las evaluaciones y el currículo nacional. La Etapa 1 cubre el Año 1 y el Año 2 en las escuelas primarias, y se evalúa a los alumnos al final del Año 2, cuando la mayoría cuenta con 7 años de edad. La Etapa 2 cubre del Año 3 al Año 6, que suelen verse como el final de la educación primaria. La Etapa 3 abarca del Año 7 al Año 9, los primeros tres años de educación secundaria. La Etapa 4 cubre los dos años finales de la educación secundaria, y la mayoría de las evaluaciones se realiza al final del último año (Año 11). La calificación principal es el GCSE (*General Certificate of Secondary Education*).

3. Como se aprecia en el Capítulo 5, algunas de estas escuelas tendrían puntuaciones de valor agregado altas, pero, para este ejemplo, se considera a las escuelas que alcanzaron una puntuación de valor agregado baja.

4. Esto se aplica sólo a las escuelas en el nivel educativo obligatorio.

5. Un modelo de regresión de mínimos cuadrados ordinarios usa el promedio del grado de la escuela como variable dependiente, y el género, antecedentes extranjeros y escolaridad paterna, como variables independientes.

CAPÍTULO 2

Presentación e interpretación de los modelos de valor agregado

Presentación de la información de valor agregado **45**
- Estudio de caso: Reino Unido **46**

Identificación de cambios significativos del desempeño escolar **53**

Creación de estándares y puntos de referencia con información de valor agregado **54**

Presentación en los medios **57**

Como se describió previamente, los modelos de valor agregado son un tipo de modelo estadístico que estima las contribuciones de las escuelas en el avance de los estudiantes sobre objetivos educativos declarados o prescritos (p. ej., logro cognitivo), medidos en al menos dos momentos. Los modelos de valor agregado generan resultados comparativos aunque no ofrecen mediciones absolutas del avance del aprendizaje de los estudiantes, sino mediciones de las contribuciones relativas de las escuelas al aprendizaje del estudiante, cuando el aprendizaje se mide por los cambios de las puntuaciones en los exámenes a lo largo del tiempo. Los resultados de los modelos de valor agregado varían según el modelo adoptado, y la presentación de los resultados puede estructurarse de forma diferente según la unidad o nivel de análisis para ajustarla al propósito y público al que se dirige. Las medidas de valor agregado se pueden estimar y presentar por estudiante individual, materias, grados escolares y escuelas. También es posible calcular y presentar las medidas de valor agregado por áreas regionales o alguna unidad más específica. Sin embargo, cabe observar que las puntuaciones compuestas de valor agregado que ofrecen una medición única para grupos de escuelas pueden originar problemas de interpretación si la intención es analizar aspectos de desempeño escolar y si hay variación del desempeño escolar dentro de las regiones o áreas locales especificadas.

Este capítulo analiza las formas de presentar los modelos de valor agregado para procurar una interpretación eficaz que permita avanzar en los objetivos deseados de políticas. Esto incluye un análisis de las ventajas y los riesgos potenciales de clasificar el desempeño de las escuelas (p. ej., desempeño "***alto***" y "***bajo***"). Se ofrecen varios ejemplos de presentación de la información sobre valor agregado que procuran una interpretación precisa. Estos ejemplos ilustran las ventajas de elaborar un sistema completo que aproveche las puntuaciones de valor agregado de las escuelas, por ejemplo, para crear puntos de comparación y criterios como base para el desarrollo de acciones referentes a los objetivos de las políticas. El capítulo termina con un análisis de la presentación de información de valor agregado en los medios y la necesidad de asegurar que esta cobertura no distorsione el flujo de información ni impida actuar en favor de las escuelas.

Con los modelos de valor agregado se puede clasificar a las escuelas según su desempeño ***alto*** o ***bajo*** (o un punto intermedio). Es obvio que estas clasificaciones no son necesarias, y que la información sobre el valor agregado puede evaluarse y aprovecharse sin colocar a las escuelas en categorías particulares. Puede ser políticamente difícil para los administradores, las autoridades educativas y los interesados clasificar a una escuela como de desempeño bajo o inferior, y es importante que se basen en criterios estadísticos o conceptuales válidos (p. ej., puntuación estadística de valor agregado significativamente distinta de la media) y que los criterios no se elijan de manera arbitraria. Es una decisión importante que podría conllevar un impacto considerable en las escuelas y, según la estructura del sistema escolar, en la asignación de sus recursos y desarrollo. La dificultad para los administradores y los tomadores de decisiones en ocasiones puede venir de presiones contrarias. Por una parte, es necesario establecer una clasificación de manera oportuna de modo que puedan emprenderse las acciones adecuadas para remediar problemas, como atender a estudiantes y a escuelas con desempeño considerado inferior, tan pronto como sea posible. En estos casos, la información sobre valor agregado necesita traducirse en acciones relevantes para que las escuelas con puntuaciones estadísticas de valor agregado significativamente inferiores al promedio durante

dos años, por ejemplo, se clasifiquen como escuelas de desempeño bajo y se emprendan acciones con rapidez para evaluar a la escuela y se inviertan más recursos en la enseñanza de sus estudiantes. Por otra parte, trabajar en una escuela clasificada como de bajo desempeño puede tener efectos negativos en los directores de escuelas, maestros, estudiantes y otros interesados. El grado e impacto de estos efectos negativos depende de la estructura del sistema en que se aplica la clasificación y las acciones que resulten de ella. Por tanto, hay una considerable presión para que se obtengan mediciones precisas y así sea más justa y exacta la clasificación de las escuelas como de ***bajo*** o ***alto*** desempeño.

En el Capítulo 6 se identifican algunos de los problemas estadísticos y metodológicos que deben considerarse en la elaboración de modelos de valor agregado y en la interpretación de sus resultados. Entre estos problemas se encuentran posibles errores de medición y fuentes potenciales de sesgos en sus estimaciones. Es necesario tener en cuenta estos y otros problemas semejantes dada la dificultad de clasificar a las escuelas como de ***bajo*** o ***alto*** desempeño. Estas cuestiones estadísticas que se analizan en la interpretación de dicha clasificación implican que es más fácil saber cuándo ***no*** clasificar a una escuela como de desempeño bajo o inferior que cuándo es adecuado hacerlo mediante los modelos de valor agregado. Esta dificultad necesita equilibrarse con el imperativo de utilizar los resultados de los modelos de valor agregado como base de acciones que puede incluir la clasificación de las escuelas y a continuación las intervenciones educativas y otras medidas requeridas. La estabilidad de resultados de un año al siguiente también se analiza en el Capítulo 7 y la Parte III de este informe. Este análisis incluye la recomendación de emplear un promedio de tres años de resultados de valor agregado como principal indicador para aplicar y presentar resultados.

Es importante que la discusión sobre los problemas estadísticos y metodológicos no se considere como precauciones excesivas con respecto a los modelos de valor agregado, o que los tomadores de decisiones sean demasiado cautelosos al clasificar a las escuelas que requieran acciones específicas, pues esto podría socavar las razones para establecer un sistema basado en los modelos de valor agregado. El uso de los modelos de valor agregado para crear un sistema para la mejora o la rendición de cuentas escolares requiere que se evalúe a las escuelas y se tomen las decisiones consecuentes. Estas decisiones deben alinearse con la meta de mejorar el sistema escolar, como parte de objetivos de políticas generales. ***Si se colocan demasiados obstáculos frente a los administradores o las autoridades educativas antes de clasificar a una escuela, esto puede impedir que se emprendan las acciones necesarias.*** Por ejemplo, piense en un sistema en el que las escuelas clasificadas como de bajo desempeño reciben más evaluaciones y asistencia. Una vez que se clasifica a una escuela como de bajo desempeño, recibe la visita de un inspector escolar, quien realiza una valoración en persona de las razones del bajo desempeño. De acuerdo con los resultados de esta evaluación adicional, se otorgan más asistencia o recursos, se procuran mayores oportunidades para el desarrollo profesional de los docentes y para el desarrollo organizativo de la escuela, y se establece un sistema de supervisión del desempeño futuro. Por ende, los obstáculos o resistencia a la clasificación de desempeño ***bajo*** impedirían asimismo la aplicación de iniciativas para mejorar el desempeño de esta escuela y de sus estudiantes.

PRESENTACIÓN DE LA INFORMACIÓN DE VALOR AGREGADO

La presentación de la información y resultados de valor agregado de las escuelas debe tener en cuenta la cuestión de cómo exponer de la mejor manera posible asuntos estadísticos complicados

a un público lego en estadística. La Parte III de este informe destaca la importancia de la educación y formación de los interesados en el establecimiento de un sistema basado en modelos de valor agregado. También es importante que la presentación de información de valor agregado sea clara y transparente de modo que los interesados aprovechen al máximo los beneficios de establecer un sistema así. Como se observará en los ejemplos posteriores, hay diversos pasos para allanar los problemas de interpretación y facilitar una comprensión eficaz de las clasificaciones de valor agregado y su empleo para la consecución de los objetivos de políticas educativas. A pesar de la complejidad propia de los modelos de valor agregado, las presentaciones simplificadas de estos modelos y la información relacionada son útiles para lograr la claridad requerida.

Los resultados de valor agregado son numéricos y se presentan de diversas formas, como puntuaciones continuas o una puntuación específica superior o inferior a un resultado de desempeño esperado de la escuela. Según la estructura de la base de datos y el tipo de análisis, las estimaciones de valor agregado de la escuela se presentan por características de la materia, grado o estudiante. Las posiciones relativas de varias escuelas pueden servir como punto de partida para los análisis del desarrollo de la escuela. Estos análisis también deben tener en cuenta otras características escolares, como el perfil del personal docente, movilidad estudiantil y problemas locales y comunitarios particulares. Como se ve en los diversos sistemas educativos, hay variadas posibilidades de presentación y uso de la información de valor agregado.

Estudio de caso: Reino Unido

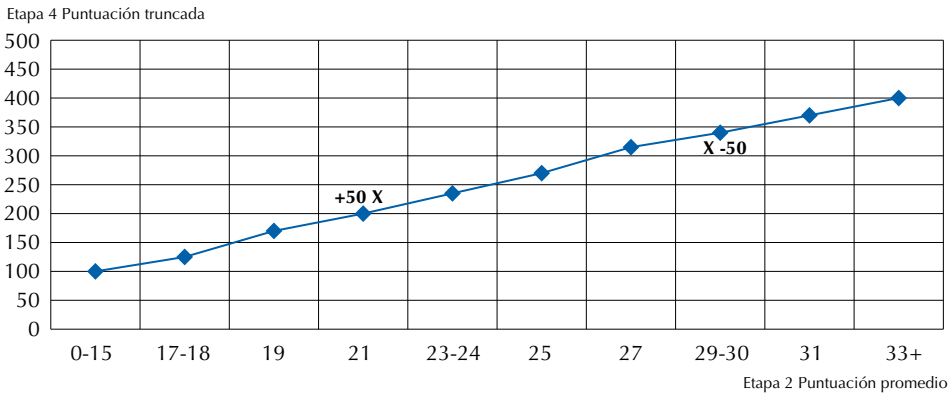
En el **Reino Unido** se han utilizado varios modelos de valor agregado conforme el sistema se ha desarrollado en el tiempo. Asimismo, se han empleado diversos modelos para analizar aspectos particulares del sistema escolar. Un modelo contextualizado de valor agregado es el más complejo utilizado en el **Reino Unido** pues controla la influencia de varias características socioeconómicas en los cambios del rendimiento de los estudiantes. También incluye otras características que influyen en el avance estudiantil ajenas al control de las escuelas, como el mes de nacimiento del alumno (véase el Cuadro 4.2). Esto se analiza junto con el "método de la mediana" de presentación de resultados diseñado y utilizado para ilustrar de manera simplificada el cálculo de la puntuación de valor agregado de una escuela. Más adelante se presenta un ejemplo del método de la mediana. En el **Reino Unido** se emplea para exponer las bases de los modelos de valor agregado y explicar como deben interpretarse de modo que las escuelas los empleen sin dificultad. Los resultados escolares de valor agregado según el método de la mediana se han publicado en forma de gráficas que muestran la mediana de los resultados a partir de un punto de rendimiento anterior. Esto se adaptó para el cálculo de las puntuaciones de las escuelas, que provienen del promedio de cada escuela de las diferencias entre el resultado actual por estudiante, por un lado, y la mediana del resultado nacional por estudiante, por otro, sobre su puntuación de rendimiento anterior.

El método de la mediana se diseñó por su simplicidad y claridad, y porque se integra con facilidad al ciclo de producción de las tablas de desempeño que se han empleado, y aún se emplean, en el **Reino Unido**. Este método también permite que las escuelas calculen sus propias puntuaciones de valor agregado referentes a la información sobre los resultados nacionales esperados. En lugar de emplear un modelo de regresión, el método se basó en las líneas medianas conocidas ya en las escuelas por desarrollos previos en esta área. En este sistema, una escuela considera la

puntuación de rendimiento anterior de cada estudiante y lo compara con la línea de mediana, y la diferencia es la contribución de la puntuación de valor agregado de ese estudiante a la puntuación de valor agregado de la escuela. La Gráfica 2.1 ofrece un ejemplo de cálculo con una línea de mediana. Un estudiante obtuvo 50 puntos más de lo "esperado" en la Etapa 4, según su puntuación de rendimiento anterior en la Etapa 2. Otro estudiante consiguió 50 puntos menos de lo esperado. La suma de las distancias verticales hacia la línea de mediana, dividida entre la cantidad total de estudiantes, es la puntuación de valor agregado de la escuela.

Gráfica 2.1

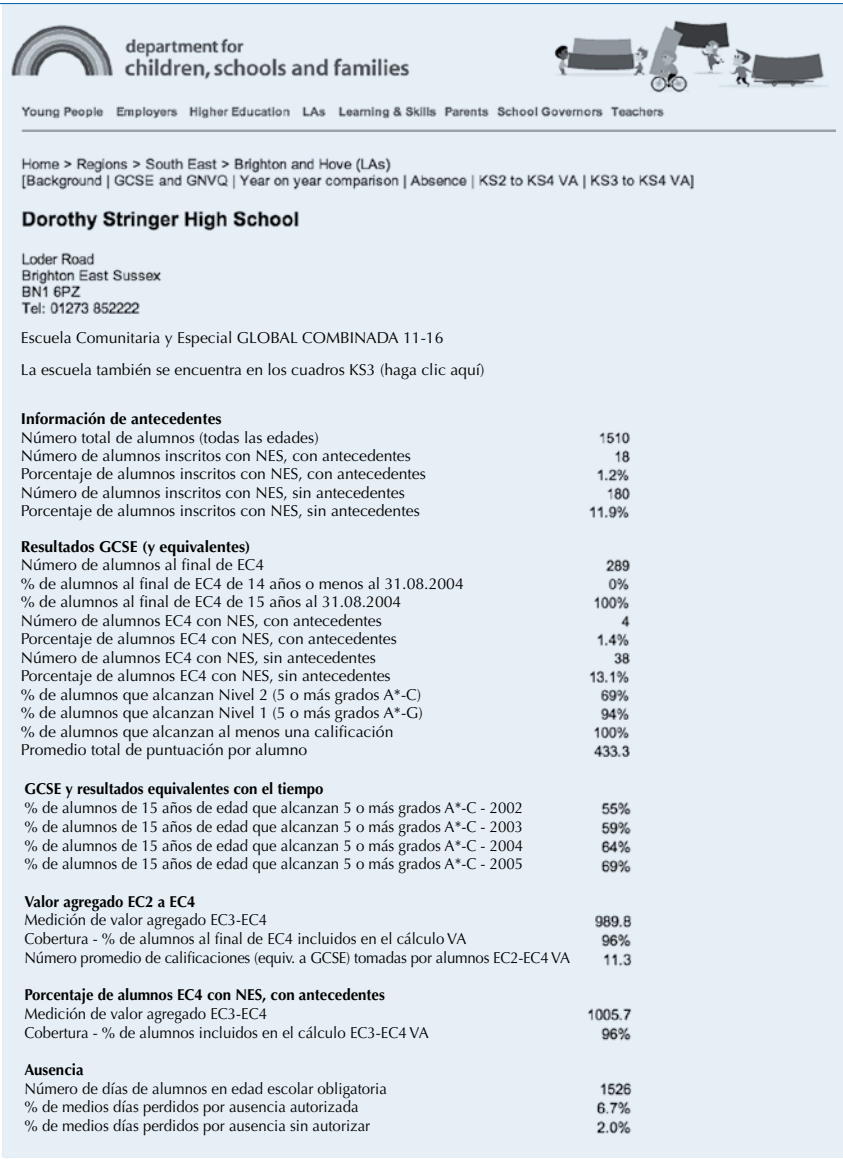
Ejemplo de cálculo de la línea de mediana del valor agregado en el Reino Unido



Así, las escuelas también calculan y verifican con facilidad sus propias puntuaciones de valor agregado con relación a los "resultados esperados" respecto de una línea de mediana nacional. Las escuelas también pueden emplear un sencillo modelo de regresión de mínimos cuadrados ordinarios, que ofrece una fórmula para calcular resultados "esperados", para calcular y verificar resultados. La razón principal para usar el método de la mediana en el **Reino Unido** es su simplicidad de interpretación y comprensión. Un método que ilustra los resultados de un alumno típico empleando un modelo de regresión para que las escuelas calcularan sus propias puntuaciones de valor agregado se consideró poco adecuado para un público lego en estadística. La información de valor agregado se presenta por medio de los Cuadros de Desempeño Escolar, y se desarrolla y publica tanto en Internet como en folletos para cada Autoridad Local (*Local Authority*). Los Cuadros de Desempreño Escolar incluyen una variedad limitada de estadísticas sobre las escuelas. Los datos de valor agregado se presentan junto con otros sobre logros generales y el contexto escolar. En la Gráfica 2.2 se ve cómo se presentarían las puntuaciones de valor agregado de 2005, con base sólo en el método de la mediana y puntuaciones de rendimiento previas para una escuela secundaria (véase *http://www.dfes.gov.uk/performancetables/*). El resultado de valor agregado se incluye junto con resultados brutos y alguna información contextual. Aquí, la puntuación 989.8 de las Etapas Clave 2-4 significa que los estudiantes de esta escuela obtuvieron, en promedio, 10.2 puntos de valor agregado menos que la mediana de estudiantes por cada nivel de rendimiento previo.

Gráfica 2.2

Ejemplo del sitio de Internet de los Cuadros de Desempeño Escolar del valor agregado*



department for children, schools and families

Young People Employers Higher Education LAs Learning & Skills Parents School Governors Teachers

Home > Regions > South East > Brighton and Hove (LAs)
[Background | GCSE and GNVQ | Year on year comparison | Absence | KS2 to KS4 VA | KS3 to KS4 VA]

Dorothy Stringer High School

Loder Road
Brighton East Sussex
BN1 6PZ
Tel: 01273 852222

Escuela Comunitaria y Especial GLOBAL COMBINADA 11-16

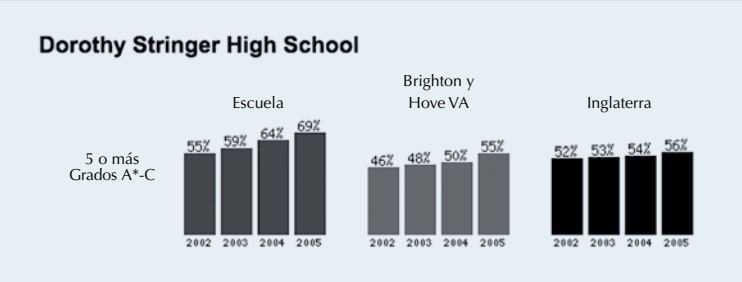
La escuela también se encuentra en los cuadros KS3 (haga clic aquí)

Información de antecedentes	
Número total de alumnos (todas las edades)	1510
Número de alumnos inscritos con NES, con antecedentes	18
Porcentaje de alumnos inscritos con NES, con antecedentes	1.2%
Número de alumnos inscritos con NES, sin antecedentes	180
Porcentaje de alumnos inscritos con NES, sin antecedentes	11.9%
Resultados GCSE (y equivalentes)	
Número de alumnos al final de EC4	289
% de alumnos al final de EC4 de 14 años o menos al 31.08.2004	0%
% de alumnos al final de EC4 de 15 años al 31.08.2004	100%
Número de alumnos EC4 con NES, con antecedentes	4
Porcentaje de alumnos EC4 con NES, con antecedentes	1.4%
Número de alumnos EC4 con NES, sin antecedentes	38
Porcentaje de alumnos EC4 con NES, sin antecedentes	13.1%
% de alumnos que alcanzan Nivel 2 (5 o más grados A*-C)	69%
% de alumnos que alcanzan Nivel 1 (5 o más grados A*-G)	94%
% de alumnos que alcanzan al menos una calificación	100%
Promedio total de puntuación por alumno	433.3
GCSE y resultados equivalentes con el tiempo	
% de alumnos de 15 años de edad que alcanzan 5 o más grados A*-C - 2002	55%
% de alumnos de 15 años de edad que alcanzan 5 o más grados A*-C - 2003	59%
% de alumnos de 15 años de edad que alcanzan 5 o más grados A*-C - 2004	64%
% de alumnos de 15 años de edad que alcanzan 5 o más grados A*-C - 2005	69%
Valor agregado EC2 a EC4	
Medición de valor agregado EC3-EC4	989.8
Cobertura - % de alumnos al final de EC4 incluidos en el cálculo VA	96%
Número promedio de calificaciones (equiv. a GCSE) tomadas por alumnos EC2-EC4 VA	11.3
Porcentaje de alumnos EC4 con NES, con antecedentes	
Medición de valor agregado EC3-EC4	1005.7
Cobertura - % de alumnos incluidos en el cálculo EC3-EC4 VA	96%
Ausencia	
Número de días de alumnos en edad escolar obligatoria	1526
% de medios días perdidos por ausencia autorizada	6.7%
% de medios días perdidos por ausencia sin autorizar	2.0%

Notas:
* Algunas secciones del texto en inglés de la imagen original han sido traducidas para facilitar comprensión de la gráfica.
LAs = Autoridad local (*Local Authority*)
GCSE = Certificado General de Educación Secundaria (*General Certificate of Secundary Education*)
GNVQ = Calificaciones Vocacionales Nacionales Generales de los Niveles 1 y 2 (*General National Vocational Qualifications*)
KS = Etapa clave (*Key Stage*)
VA = Valor agregado (*Value-added*)
NES = Necesidad de Educación Especial (*Special Educational Needs, SEN*)
Fuente: Departamento para Niños, Escuelas y Familias, RU (2008).

Gráfica 2.2 (continuación)

Ejemplo del sitio de Internet de los Cuadros de Desempeño Escolar del valor agregado*



Nota:
* Algunas secciones del texto en inglés de la imagen original han sido traducidas para facilitar comprensión de la gráfica.
VA = Valor agregado (*Value-added*)
Fuente: Departamento para Niños, Escuelas y Familias, RU (2008).

El uso de los Cuadros de Desempeño Escolar se extendió con el avance de RAISEonline, programa de *software* interactivo que permite analizar datos del avance de escuelas y estudiantes. Este programa es un buen ejemplo de como se pueden presentar puntuaciones e información adicional de valor agregado para facilitar su análisis en los ámbitos de la escuela y del sistema educativo. La presentación de la información de valor agregado y un enlace interactivo en el **Reino Unido** fue un paso importante para facilitar el uso de estos datos en el ámbito escolar y para permitir a los interesados emplearlos con el fin de avanzar hacia los objetivos de sus políticas. Los objetivos básicos de la introducción de RAISEonline fueron permitir que las escuelas analizaran los datos de desempeño con mayor profundidad como parte de un proceso de autoevaluación escolar; ofrecer un conjunto común de análisis a las escuelas, autoridades locales, inspectores escolares y Socios de Mejora Escolar (*School Improvement Partners*); y ofrecer un mejor apoyo a la enseñanza y al aprendizaje (Ray, 2006). Al disponer de una cantidad considerable de información para las escuelas primarias y secundarias, junto con los elementos interactivos del *software* permiten a los usuarios profundizar en los datos para realizar un mejor análisis del valor agregado de alumnos y escuelas. Entre las características básicas de RAISEonline se encuentran:

- Inclusión de informes y análisis que cubren los logros y avances de los alumnos en las Etapas 1, 2, 3 y 4, con funciones interactivas que permiten explorar hipótesis sobre el avance estudiantil.
- Aportación de información contextual sobre la escuela, con comparaciones nacionales con otras escuelas.
- Análisis mediante preguntas, lo que permite a las escuelas analizar el desempeño de los alumnos en áreas específicas del currículo.
- Establecimiento de objetivos que apoyen a las escuelas en el proceso de supervisión, desafío y apoyo del avance estudiantil.
- Administración sencilla de datos que permite manejarlos y editarlos en el nivel del alumno y crear campos definidos por escuelas y grupos de enseñanza.

Esta información está disponible en el sitio de Internet RAISEonline, para el cual se proporciona un nombre de usuario y contraseña a los directores de escuela con el fin de asegurar que sólo ellos accedan a la información de su escuela (Ray, 2006). En la Gráfica 2.3 se observa un ejemplo de presentación de información de valor agregado de una escuela secundaria del **Reino Unido**. La información se presenta de manera gráfica y tabulada para facilitar la interpretación de los resultados escolares con comparaciones con su desempeño anterior.

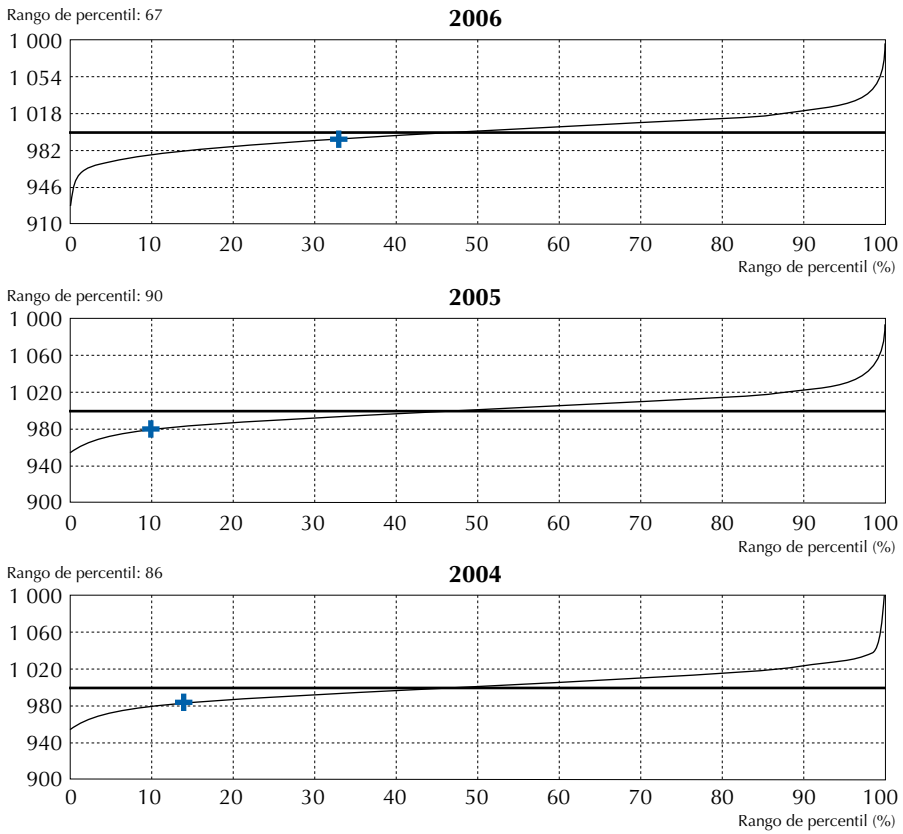
Gráfica 2.3

Ejemplo de información de valor agregado contextualizado proporcionada para una escuela secundaria del Reino Unido

URN de una escuela secundaria: 999999 DfES Núm. 9999999 Mediciones de avance

Valor agregado contextual Etapas Clave 2 a 4: GENERAL

Esta sección ofrece la medición general de valor agregado contextual (VAC) para la escuela en relación con la media nacional de 1 000. La escuela se ubica dentro de la distribución nacional para ilustrar la variedad de puntuaciones VAC alcanzadas por otras escuelas típicas que reciben recursos públicos.



Nota: Algunas secciones del texto en inglés de la imagen original han sido traducidas para facilitar comprensión de la gráfica.
Fuente: Departamento para Niños, Escuelas y Familias, RU (2008).

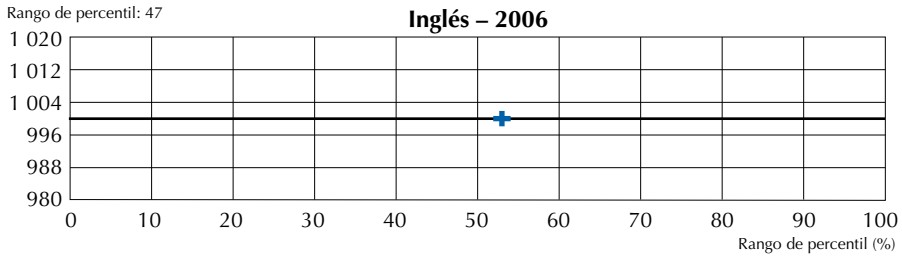
Gráfica 2.3 (continuación)

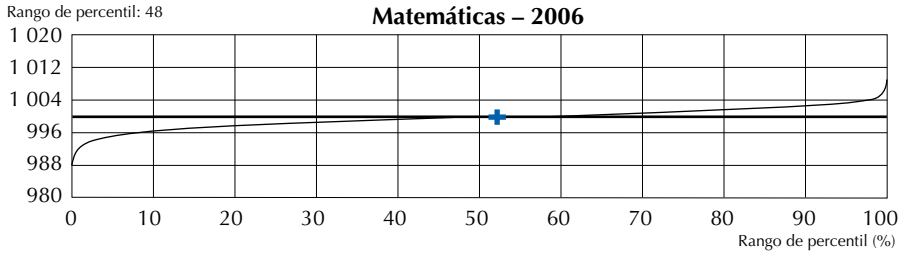
Ejemplo de información de valor agregado contextualizado proporcionada para una escuela secundaria del Reino Unido

URN de una escuela secundaria: 999999 DfES Núm. 9999999 Mediciones de avance

Valor agregado contextual Etapas Clave 2 a 4: POR MATERIA

El análisis en esta sección se centra en el valor agreagdo contextual (VAC) de las materias centrales del Currículo Nacional (Inglés y Matemáticas) en el año en curso. Para todos los análisis VAC basados en materias, el aprovechamiento anterior en los modelos de VAC se basó en una combinación de las tres materias básicas.





Valor agregado contextual Etapas Clave 2 a 4: GENERAL Y MATERIAS

Esta sección presenta las puntuaciones de valor agregado generales y por materia de la escuela en relación con la media nacional de 1 000. Donde un valor de VAC mostró un cambio estadísticamente significativo en comparación con el año anterior, se muestra ↑ o ↓ para indicar la dirección de dicho cambio.

		2004		2005		2006	
Todas las materias	Cohorte por VAC	172		175		175	
	Puntuación escolar de VAC	982.0	↓	980.9		994.5	↑
	Intervalo de confianza de 95% +/-	9.8		9.5		9.4	
	Significación	Sig-		Sig-			
	Rango de percentil	85		90		67	
	Cobertura	98%		98%		98%	
Inglés / Idioma inglés	Cohorte por VAC	166		168		172	
	Puntuación escolar de VAC	995.7	↓	997.3	↑	1 000.3	↑
	Intervalo de confianza de 95% +/-	1.0		1.0		1.0	
	Significación	Sig-		Sig-			
	Rango de percentil	99		91		47	
	Cobertura	94%		94%		96%	
Matemáticas	Cohorte por VAC	164		170		171	
	Puntuación escolar de VAC	998.9		998.8		1 000.2	
	Intervalo de confianza de 95% +/-	1.1		1.1		1.0	
	Significación	Sig-		Sig-			
	Rango de percentil	72		72		48	
	Cobertura	93%		95%		96%	

Nota: Algunas secciones del texto en inglés de la imagen original han sido traducidas para facilitar comprensión de la gráfica.
Fuente: Departamento para Niños, Escuelas y Familias, RU (2008).

Como se aprecia en las gráficas, esta escuela secundaria tiene puntuaciones de valor agregado contextualizadas por debajo del promedio nacional en cada uno de los tres años presentados. Para evitar malas interpretaciones, el promedio nacional se estableció en una puntuación de 1 000. Se observó que no convenía establecer la puntuación promedio en cero, pues implicaba que las escuelas que se desempeñaran por debajo del promedio recibirían una puntuación negativa, y se pensó que las connotaciones de una puntuación negativa tendrían efectos adversos entre los interesados. Además, las puntuaciones negativas complican la interpretación, pues puede interpretarse que muestran una caída del desempeño general de los estudiantes (Ray, 2006). Las puntuaciones de valor agregado de las escuelas son relativas respecto del desempeño de todas las escuelas y, por tanto, una puntuación de valor agregado negativa (inferior al promedio) no implica necesariamente que disminuyera el rendimiento de los estudiantes en general. Para evitar esta malinterpretación, la puntuación de valor agregado promedio se fijó en 1 000. La escuela secundaria presentada como ejemplo tuvo una puntuación de valor agregado contextualizada de 994.5 en 2006 en todas las materias. Esto fue inferior al promedio nacional pero mejor que su puntuación de valor agregado contextualizada de 980.9 en 2005.

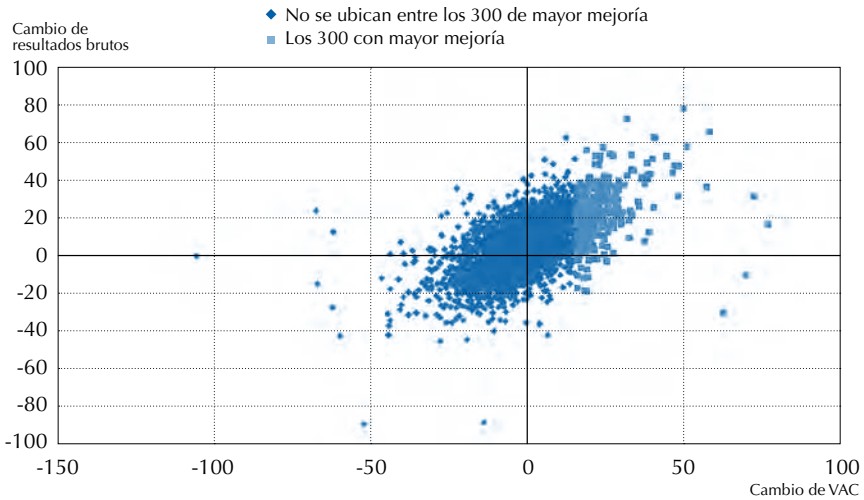
El análisis de la mejora escolar presentado en el Capítulo 1 destacó las oportunidades para que las escuelas valoraran los resultados de valor agregado identificando variaciones de desempeño dentro de las mismas escuelas. Como se ilustra en el cuadro anterior, la puntuación de valor agregado contextualizada de todas las materias en 2006 (994.5), para la escuela ejemplo, fue inferior a la puntuación en Inglés (1 000.3) y Matemáticas (1 000.2), que se situaron justo por encima del promedio nacional. Si bien no debe verse esto como prueba concluyente de un desempeño bajo en las otras materias, sí es una señal de que debe investigarse el desempeño en esas áreas. Un análisis más profundo de información de valor agregado específica podría explicar las causas de estas diferencias, y una evaluación interna proporcionaría buenas ideas para mejorar la escuela en esas áreas.

Como veremos en el Capítulo 3, RAISEonline permite que las escuelas efectúen diversos análisis de su desempeño, incluyendo el análisis del rendimiento de alumnos individuales. Las escuelas comparan los cambios de las puntuaciones de valor agregado contextualizadas con sus puntuaciones brutas. También puede efectuarse esta comparación en el nivel del sistema para analizar la relación entre el avance del valor agregado contextualizado de las escuelas y los resultados brutos de los alumnos. En la Gráfica 2.4 se ofrece un ejemplo con el *software* de RAISEonline en el **Reino Unido**. Estas cifras muestran que es posible identificar las escuelas cuyas puntuaciones de valor agregado contextualizadas subieron sin que mejorasen sus resultados brutos. Estas escuelas quizá fueran más eficaces en 2006, si logran mantener sus estándares con menos recursos. Los diversos tipos de mejora/descenso se categorizarían de acuerdo con los posibles cambios de logros anteriores (arriba/abajo/igual) y valor agregado (arriba/abajo/igual). Bryk *et al*. (1998) analizan esto como "perfiles de productividad de grado", y es una forma más de ver cómo las comparaciones especializadas entre escuelas, así como las comparaciones longitudinales de una escuela en particular, pueden ser criterios sólidos para centrar la atención del personal de una escuela.

Gráfica 2.4

Cambios de valor agregado contextualizado (VAC) comparados con cambios en los resultados brutos de logros

*Etapa 4 Puntuación Promedio (*Average Point Scores*) 2005-06*



Fuente: Ray, A. (2007).

IDENTIFICACIÓN DE CAMBIOS SIGNIFICATIVOS DEL DESEMPEÑO ESCOLAR

Es importante observar que los usuarios pueden identificar si se dieron cambios estadísticamente significativos a lo largo del tiempo. Las puntuaciones de valor agregado significativamente superiores o inferiores al promedio de todas las escuelas ofrecen un mecanismo sólido para clasificar a las escuelas como de desempeño ***alto*** o ***bajo***. En la Gráfica 2.3 se observan cambios de valor agregado negativos significativos entre 2004 y 2005. Esto salta a la vista al apreciar el desempeño en todas las materias y el desempeño en Inglés y Matemáticas. Además, se publicó el intervalo de confianza de 95% para ilustrar la distribución de puntuaciones dentro de este intervalo de confianza. Para 2006, el intervalo de confianza de 95% indica un rango de 9.4 puntos por arriba y por debajo de la puntuación de valor agregado contextualizado de 994.5 (985.1 a 1 003.9). Como el límite superior del intervalo de confianza excede el promedio nacional de 1 000, la puntuación de valor agregado contextualizado de la escuela no es estadísticamente distinta del promedio. En **Polonia**, los modelos de valor agregado generaron polémica por la publicación de los intervalos de confianza de las puntuaciones escolares de valor agregado. Se consideró que había dos claras ventajas en la publicación de las puntuaciones de valor agregado como una estimación de intervalo. En primer lugar, reduciría la facilidad con que las clasificaciones de las escuelas se considerasen como una consecuencia negativa de los modelos de valor agregado. En segundo lugar, contribuiría a que la información de valor agregado se aprovechara no sólo como método de autoevaluación y desarrollo escolar, sino también como método de evaluar las políticas y programas educativos en los ámbitos locales o regionales (Jakubowski, 2007). Por razones semejantes se publicaron intervalos de confianza en **Noruega** (Hægeland, 2006).

Sin duda, los intervalos de confianza requieren una mayor comunicación y formación a los interesados. En la publicación de los Cuadros de Desempeño Escolar en el **Reino Unido**, RAISEonline adjuntó guías para utilizar e interpretar las mediciones de valor agregado. Por ejemplo, el sitio de Internet en 2005 incluyó el mensaje que reproducimos más adelante, diseñado para ayudar a la interpretación y educar a los interesados, incrementando la validez de las puntuaciones de valor agregado en comparación con las puntuaciones brutas de las pruebas. La referencia a la "significación" estadística es necesaria porque las puntuaciones de valor agregado no se acompañan en todos los casos de intervalos de confianza: en Internet en ocasiones ofrece guías sobre el rango de puntuaciones que pueden considerarse como de "promedio amplio" según el tamaño de la escuela.

> *Las mediciones de valor agregado proporcionan la mejor indicación de la eficacia general de las escuelas en estos cuadros. Pero la significación que se otorgue a cualquier medición de valor agregado de una escuela en particular depende, entre otras cosas, de la cantidad de alumnos incluidos en el cálculo del valor agregado. Cuantos menos alumnos, menor confianza se deposita en la medición del valor agregado como indicador de que la eficacia de una escuela se ubique significativamente por encima o por debajo del promedio.*

Declaraciones como esta pretenden dar a conocer a los interesados cómo interpretar las puntuaciones de valor agregado y cómo usarlas para tomar decisiones más informadas (p. ej., con propósitos de mejoras escolares, si las usan directores o maestros; o de contribuir a elegir escuela, si los padres acceden al sitio de Internet). Asimismo, estas declaraciones ofrecen datos claros sobre las limitaciones de los resultados del valor agregado. Esto puede ser útil para los tomadores de decisiones utilizando los resultados escolares, y pueden aligerar algunas preocupaciones de los interesados en la educación sobre la forma de aplicación de estos datos, en particular con fines de rendición de cuentas de la escuela.

CREACIÓN DE ESTÁNDARES Y PUNTOS DE REFERENCIA CON INFORMACIÓN DE VALOR AGREGADO

El viraje de la preocupación pública y gubernamental desde un mero control de los recursos y contenidos de la educación hacia la atención a los resultados provocó, en muchos países, el establecimiento de estándares de calidad del trabajo de las instituciones educativas. Las estrategias de los países para fijar estándares van de la definición de amplias metas y áreas de competencia educativas a la formulación de expectativas de desempeño concisas en áreas temáticas bien definidas. Algunos países han ido más allá, al establecer estándares educativos como meros puntos de referencia e introducir metas de desempeño que los estudiantes de cierta edad o grados escolares deben alcanzar. Es en este contexto que las mediciones de valor agregado representan un papel de particular importancia. La aplicación de modelos de valor agregado para detonar acciones específicas requiere medir el desempeño de las escuelas ya sea entre sí o respecto de un estándar determinado. En cuanto al crecimiento estudiantil, un estándar se puede definir directamente en términos de crecimiento promedio que exceda un umbral predefinido. Otra estrategia es fijar objetivos de crecimiento para cada estudiante con base en su situación actual, su posición en relación con el estándar de logros vigente y, quizá, con datos históricos sobre la distribución de logros de estudiantes en situación semejante en años anteriores. Por ejemplo, un indicador de desempeño escolar se basaría en una comparación de los avances reales y esperados de un alumno que fomentara que el personal de la escuela

atendiera las necesidades de todos los estudiantes (hay variantes de estos esquemas en McCall *et al.* 2004, y Doran y Azumi, 2004). Asimismo, es posible valorar el crecimiento en distintas regiones de la escala. Hill *et al.* (2005) describen una metodología para elaborar "tablas de valores" que capturen las ideas de los tomadores de decisiones sobre los avances estudiantiles y sirvan después para fijar estándares de desempeño. Pueden aplicarse varias estrategias para crecer, pero a menudo se alejan de los modelos de valor agregado debido a la naturaleza de las proyecciones de crecimiento.

Una vez determinados los estándares para cada criterio, se diseña una matriz de decisión para guiar acciones específicas. Por ejemplo, consideremos que los umbrales del desempeño (in) satisfactorio y ejemplar se establecen respecto de cada uno de los tres criterios (situación actual, cambio en el tiempo o alguna combinación de ambos), y que el análisis se efectúa sólo para toda la escuela. La combinación de los estándares de desempeño produce nueve categorías con una puntuación de valor agregado que coloca a la escuela en una de estas categorías. La matriz de decisión especifica el tratamiento que detonan las puntuaciones de cada categoría. Por ejemplo, se recompensaría a las escuelas que alcanzaran el nivel ejemplar en los tres criterios durante dos años consecutivos. Por otra parte, las escuelas que no alcanzaran el nivel satisfactorio en dos o más criterios en un año dado quedarían sujetas a una revisión externa.

Los ejemplos anteriores ilustran cómo los resultados de valor agregado ofrecen una base para actuar. En los sistemas de desarrollo escolar y mejora escolar, hay beneficios al especificar estas acciones y las clasificaciones de desempeño escolar que las ponen en marcha. Las puntuaciones de corte predeterminadas servirían como puntos detonadores de acciones como la autoevaluación de una escuela o la valoración de un inspector escolar, como sucede en los **Países Bajos** (Van de Grift, 2007). Diseñar un esquema así requiere un análisis de la distribución de las puntuaciones de valor agregado en cada sistema educativo. El ejemplo en el **Reino Unido** ilustra cómo categorizar los resultados de valor agregado. Se consideraron cinco categorías posibles en las que se identificaría a las escuelas que mejoran si:

i. Presentan uno de los mayores crecimientos en las puntuaciones de valor agregado (p. ej., 100 o 10%).

ii. Realizan un cambio estadísticamente significativo (en un nivel de confianza de 95%).

iii. Se desplazan entre distintas partes de la distribución (p. ej., de "baja" [cuartil más bajo] a "promedio" [intermedio]).

iv. Se desplazan entre distintas partes de la distribución definida en términos de desviaciones estándar del promedio, o de "significativamente bajo" a "significativamente por arriba".

v. Mejoran algún umbral predefinido.

Estas categorías son particularmente útiles al clasificar escuelas, activar respuestas a políticas y programas, y evaluar el desempeño general de la escuela y el sistema. Se sometió a análisis la cantidad de escuelas que realizaron un cambio estadísticamente significativo en su desempeño entre 2005 y 2006 (como la opción *ii*). El Cuadro 2.1 presenta una comparación entre las puntuaciones de valor agregado del nivel de escuela de 2005 y el modelo de valor agregado contextualizado de 2006 empleado en el **Reino Unido**. Este cuadro utiliza información de RAISEonline para ver si la puntuación de valor agregado contextualizado de una escuela aumentó o disminuyó significativamente entre 2005 y 2006. En este sitio de Internet también

se presenta de manera gráfica la información de la significación estadística en torno a las puntuaciones de valor agregado de las escuelas. Las gráficas muestran puntuaciones de valor agregado contextualizado de años consecutivos con intervalos de confianza alrededor de dichas puntuaciones de modo que no se interpreten en exceso los cambios pequeños. El modelo compara resultados de alumnos de 16 años de edad con logros anteriores a los 11 años, y toma en cuenta diversos datos contextuales. Emplea un modelo multinivel que reduce las puntuaciones en escuelas más pequeñas y tiene la ventaja de disminuir la inestabilidad del modelo. La primera columna es la puntuación de todo el modelo de valor agregado con base en la puntuación promedio de todas las materias. Las demás columnas son modelos de resultados en inglés y matemáticas (con el mismo conjunto de variables contextuales de entrada). Esto ilustra la proporción de escuelas que registran cambios significativos en un año determinado y brinda información para la planificación de políticas y programas diseñados para responder a dichos cambios. La comparación permite tomar decisiones mejor informadas respecto de la asignación de recursos y brinda una panorámica de los usos de las puntuaciones de valor agregado para clasificar el desempeño escolar. Al establecer categorías de desempeño, conviene, durante la fase piloto de implementación, analizar la cantidad de escuelas que se categorizarían en cada clasificación de desempeño.

Cuadro 2.1

Cantidad de escuelas por año con cambios anuales consecutivos significativos en puntuaciones de valor agregado contextualizado entre 2005 y 2006

	Valor agregado en todas las materias	Valor agregado en inglés	Valor agregado en matemáticas
Aumento significativo en comparación con 2005	318	696	452
Descenso significativo en comparación con 2005	430	481	422
Sin cambio significativo desde 2005	2 337	1 908	2 211
Datos faltantes	27	27	27
Cantidad total de escuelas	3 112	3 112	3 112

Fuente: Elaboración propia con base en datos de Raise online, RU.

El cuadro muestra que en tres cuartas partes de las escuelas no hubo cambios significativos de valor agregado en todas las materias entre 2005 y 2006, pero en Inglés esto sólo se observa en 60% de las escuelas. Más escuelas registraron una mejoría estadísticamente significativa (22%) que un descenso estadísticamente significativo en su puntuación de valor agregado contextualizada (15%). Los cambios más extensos anuales también fueron más visibles en inglés que en matemáticas. Esto es congruente con los hallazgos en el **Reino Unido** sobre puntuaciones brutas de logros y con los resultados de valor agregado en **Eslovenia** y **Polonia** que revelaron más estabilidad en las puntuaciones de valor agregado de matemáticas y ciencias de las escuelas que en iengua y humanidades.

No cabe duda de que esta interpretación de resultados de valor agregado ofrece una base tangible para emprender, por ejemplo, acciones de mejora escolar. Esto es importante, pues los sistemas de gestión del desempeño aplicados en algunos sistemas educativos, en particular los basados en puntuaciones brutas de pruebas, proporcionan mediciones menos precisas y por ende son menos aptos para distinguir entre diferencias estadísticamente significativas de desempeño escolar (Ladd y Walsh, 2002). Cabe observar también que este análisis se enfoca a cambios anuales. Como se

analiza en las Partes II y III, este informe destaca los beneficios de ofrecer promedios móviles de tres años de puntuaciones de valor agregado de las escuelas para controlar de manera adecuada la inestabilidad aleatoria de los estimados de valor agregado de las escuelas.

PRESENTACIÓN EN LOS MEDIOS

Por el impacto de la cobertura mediática en el desarrollo y recepción de programas y políticas educativos, es importante determinar el tipo de cobertura de medios para la información de valor agregado y su manejo para la implementación eficaz de sistemas de modelos de valor agregado. En sistemas donde las familias puedan elegir a qué escuela enviar a sus hijos, la provisión de resultados de modelos de valor agregado contribuye a una elección eficaz de escuela. Publicar los resultados también afecta a los maestros y directores escolares, y a menudo forma parte de un sistema de rendición de cuentas escolar. Esto puede ser en particular visible si la publicación de resultados adopta la forma de clasificaciones escolares en que se base un sistema escolar de recompensas y sanciones, y también si las clasificaciones atraen mucha atención de los medios.

En algunos países, muchos padres se enteran de los resultados de valor agregado en los medios. En el **Reino Unido**, los medios dedican considerable atención al desempeño escolar y a la publicación de resultados escolares. Asimismo se aplican esfuerzos considerables para mejorar la interpretación de los resultados de valor agregado. Como ejemplo, la Gráfica 2.5 es un extracto de *The Guardian* (19 de enero de 2006) que, junto con los demás periódicos de formato grande, publicó las cifras escolares de cada Autoridad Local en orden alfabético (si bien cabe notar que se titula "Tablas de Ligas").

Gráfica 2.5

Extracto del periódico *The Guardian* (19 de enero de 2006) que presenta datos de valor agregado y de otros tipos


League tables

Barking and Dagenham

School/college	No of GCSE students	% achieving A*-C at GCSE	Average GCSE point score	Value added KS2-KS4	Number of A-level students	Average A-level point score
All Saints RC	182	88	515.8	1036	64	236.9
Barking Abbey	272	53	349.7	985.5	125	233.3
Dagenham Park	202	41	290.6	973.4	19	109.5
Eastbrook	261	42	318.4	971.1	47	171.3
Eastbury	252	39	315.4	973.5	50	193.6
Robert Clack	254	68	428.3	987.6	68	225.7
Sydney Russell	248	45	306.1	961.9	34	214.7
The Warren	251	35	291.2	938.1	57	207.4
Barking College					67	130.1

	TABLAS DE LIGAS					
Escuela/Colegio	Núm. de estudiantes con GCSE	% que alcanza A*-C en GCSE	Puntuación GCSE promedio	Valor agregado EC2-EC4	Número de estudiantes de Nivel A	Puntuación promedio de Nivel A
Barking and Dagenham						
All Saints RC	182	88	515.8	1 036	64	236.9
Barking Abbey	272	53	349.7	985.5	125	233.3
Dagenham Park	202	41	290.6	973.4	19	109.5
Eastbrook	261	42	318.4	971.1	47	171.3
Eastbury	252	39	315.4	973.5	50	193.6
Robert Clack	254	68	428.3	987.6	68	225.7
Sydney Russell	248	45	306.1	961.9	34	214.7
The Warren	251	35	291.2	938.1	57	207.4
Barking College					67	130.1

Nota: Se ha traducido el texto del extracto original para facilitar su comprensión.
Fuente: Derechos reservados *Guardian News & Media Ltd 2006*. Based on Ray, A. (2006).

Esta clase de artículos periodísticos también permite explicar las cifras con base en la información publicada en el sitio de Internet de los Cuadros de Desempeño Escolar. En enero 19 de 2006, *The Times* publicó una "tabla de liga" en que se clasificó a las escuelas (Gráfica 2.6) y mostró a las escuelas con el mejor valor agregado de las Etapas 2-4 (muchas de las cuales fueron pequeñas escuelas independientes que, como ya analizamos, presentan una variación más grande de puntuaciones de valor agregado). Esto puede considerarse un avance significativo respecto de la publicación de tablas de ligas con base en puntuaciones brutas de pruebas, e ilustra el avance que puede efectuarse en la presentación de mediciones de desempeño escolar mediante modelos de valor agregado.

Gráfica 2.6

Extracto de *The Times* (19 de enero de 2006) que presenta una "tabla de liga" de valor agregado

LAS DE MAYOR VALOR AGREGADO

	Alumnos	Valor agregado	% alumnos 5+A*-C
Islamiyah School, Blackburn	29	1088.2	83
Parsons Mead School, Ashtead	22	1081.5	91
Selly Park Tech College for Girls, B'ham	130	1077.6	84
Tayyibah Girls' School, London	17	1076.1	100
Casterton School, Carnforth	44	1075.8	100
Gloucestershire Islamic Secondary for Girls	20	1075.5	59
Pattison College, Coventry	15	1075.3	93
Wellington College, Crowthorne	132	1073.9	89
Bryanston School, Blandford Forum	128	1073.3	92
King's School, Bruton	55	1073.1	82
Feversham College, Bradford	36	1072.8	75
St Teresa's School, Dorking	56	1071.7	86
Ibstock Place School, London	55	1071.7	96
Queen Margaret's School, York	61	1071.5	88
Jamia Al-Hudaa Residential, Nottingham	18	1070.5	45
Taunton School, Taunton	95	1070.0	91
St James's School, Malvern	30	1069.9	73
Manor House School, Leatherhead	35	1069.2	97
St Edmund's School, Canterbury	69	1069.0	92
Wychwood School, Oxford	22	1066.5	74
Kassim Darwish Gmr for Boys, Manchester	21	1066.5	95
St Mary's School Ascot, Ascot	54	1066.3	100
Royal School Hampstead, London	16	1065.8	93
Stonar School, Melksham	42	1065.3	89
Tonbridge School, Tonbridge	142	1065.0	93
Bowbrook House School, Pershore	14	1063.9	67
Red House School, Stockton-on-Tees	44	1063.7	93
St Paul's School, London	160	1063.1	99
Leicester Islamic Academy, Leicester	46	1063.1	93
Abu Bakr Girls' School, Walsall	32	1063.0	44
St Antony's Leweston School, Sherborne	40	1062.9	87
Al-Mahad-Al-Islam School, Sheffield	13	1062.9	23
Cranleigh School, Cranleigh	124	1062.6	98
Denstone College, Uttoxeter	75	1062.3	92
Manchester Islamic High School for Girls	51	1062.3	88
Guru Nanak Sikh VA School, Hayes	61	1062.2	95
Bedford School, Bedford	133	1061.9	92
Queenswood School, Hatfield	65	1061.8	89
The Towers Convent School, Steyning	33	1061.3	100
Tormead School, Guildford	85	1061.2	99
St Mary's School, Shaftesbury	54	1061.1	91
Rye St Antony School, Oxford	41	1061.1	98
St Nicholas' School, Fleet	29	1061.1	100
Dean Close School, Cheltenham	89	1061.0	84
Culcheth Hall School, Altrincham	19	1060.8	95
Eastbourne College, Eastbourne	111	1060.4	94
Wimbledon High School, London	89	1060.2	100
Babington House School, Chislehurst	17	1060.2	93
Pipers Corner School, High Wycombe	62	1059.9	97
Wycombe Abbey School, High Wycombe	86	1059.5	100

Nota: Se ha traducido el texto del extracto original para facilitar su comprensión.
Fuente: Ray, A. (2006).

Como se observa, puede haber esfuerzos fructíferos para reducir la atención exclusiva a los resultados brutos de las pruebas. Además, pueden destacarse las cuestiones estadísticas en el análisis de la publicación y presentación gráfica de los resultados de valor agregado. Todo análisis de los intervalos de confianza en torno a las puntuaciones de valor agregado de las escuelas debe abarcar el de las implicaciones para la formación de "tablas de ligas" por parte de los medios. Se pensó que la publicación de los intervalos de confianza reduciría el grado de malinterpretación de las tablas de ligas. Sin embargo, se vio que, una vez que se produce un punto estimado o que se presenta información de valor agregado de las escuelas, siempre existe la posibilidad/tendencia de crear tablas de ligas. Al publicar información de valor agregado se puede hacer más énfasis en aspectos particulares. Por ejemplo, el desarrollo de modelos de valor agregado contextualizados en el **Reino Unido** se ha visto reflejado en la cobertura mediática. El sitio de Internet de la BBC británica permite a los usuarios ver las tablas de ligas escolares más recientes de escuelas del **Reino Unido**. El Recuadro 2.1 presenta una descripción extensa sobre cómo se deben interpretar las puntuaciones de valor agregado contextualizado:

Recuadro 2.1 Descripción del valor agregado contextualizado en los medios del Reino Unido

Los resultados incorporan una compleja puntuación de valor agregado contextual (VAC) de la Etapa 2 a la Etapa 4 diseñada para mostrar el avance de los alumnos.

Esto se logra al comparar sus logros con los de otros alumnos en el país que registraron logros iguales o semejantes en sus resultados de pruebas a los 10 u 11 años de edad en 2002.

La VAC incluye nueve factores que se sabe afectan el logro de los alumnos pero son ajenos al control escolar:

- *Sexo*
- *Necesidades educativas especiales*
- *Origen étnico*
- *Candidato a alimentos escolares gratuitos*
- *Lengua materna*
- *Movilidad*
- *Edad*
- *Bajo cuidado*
- *IDACI (medición posterior de privaciones)*

Lo que la VAC hace es predecir cuál debe ser el logro del niño con base en el logro real de otros niños con logros previos semejantes y antecedentes también semejantes.

La idea es que la forma de la cual se desempeñaron en realidad –mejor o peor que otros– se debe a la influencia de la escuela.

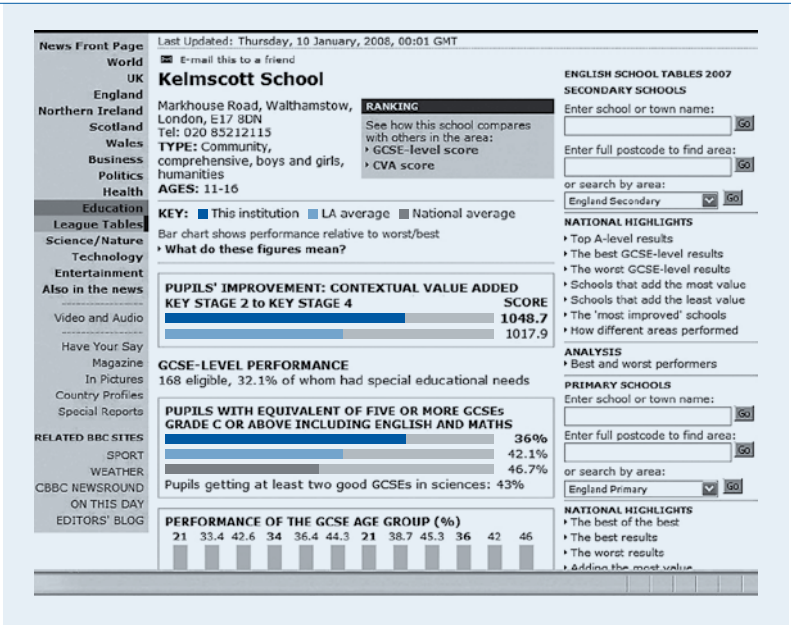
Se promedian las puntuaciones individuales de los alumnos para dar una puntuación a la escuela en su conjunto, a la cual se aplica otro cálculo para generar al final una cantidad alrededor de 100.

Fuente: *BBC News* (2007).

Esta información provino originalmente del Ministerio de Educación inglés que subrayaba la importancia de las puntuaciones de valor agregado contextualizado en la medición de desempeño escolar y destacaba los riesgos de depender sólo de puntuaciones brutas de pruebas. El Ministerio explicó que la introducción de puntuaciones de valor agregado contextualizado permitiría más equidad y justicia en la publicación de resultados de desempeño escolar. Esto fue benéfico para la escuela presentada en la Gráfica 2.7 que logró un promedio superior en la puntuación de valor agregado contextualizado. Esto reviste una importancia particular: esta escuela no alcanzó un lugar tan ***alto*** en la autoridad local en otras mediciones, como desempeño estudiantil en su Certificado de Graduación de Secundaria, lo que indica que esta escuela atiende a una proporción creciente de estudiantes de condiciones socioeconómicas más bajas, lo que disminuye los resultados generales predichos de la escuela. La atención a las puntuaciones de valor agregado contextualizado presentaron una imagen más favorable de esta escuela que la que habría resultado si se atendiese sólo a las puntuaciones brutas de pruebas, o, en este caso, a los grados del Certificado de Graduación de Secundaria.

Gráfica 2.7

Resultados de valor agregado de una escuela del Reino Unido disponibles en el sitio de Internet de la BBC, 2008



Fuente: *BBC News* (2008).

El grado de atención mediática e interés en las clasificaciones escolares en el **Reino Unido** no existe en otros países con información disponible sobre desempeño escolar. Para algunos países, la publicación de resultados es común y no despierta mucho interés de los medios. En cambio, en la **Comunidad Flamenca de Bélgica**, no es común la publicación de estos resultados.

Se ha visto un gran interés mediático en los resultados escolares y fueron los medios los que tomaron la iniciativa de publicar informes parciales de inspecciones, disponibles sólo sobre pedido. En parte, como reacción a ello y a la atención de los medios, se emprendió una iniciativa en 2007 para publicar los informes de inspecciones escolares en un sitio de Internet (*http://www.ond.vlaanderen.be/doorlichtingsverslagen/*). Es difícil determinar por qué la publicación de resultados escolares en un país no atrae la atención mediática que tiene en otros países. Se supondría que son importantes los factores institucionales y culturales, como es la historia de las mediciones de desempeño con base en resultados. Los objetivos del sistema y los métodos con que se introducen también pueden ser un factor importante. Los resultados de valor agregado que se emplean en un sistema de rendición de cuentas con posibles repercusiones fuertes para los directores de escuela, maestros y familias pueden causar una reacción más intensa que un sistema basado en mejoras escolares internas. Esto destaca los beneficios de una comunicación clara sobre cómo se usarán los resultados de valor agregado escolares y cómo se obtienen. Estos asuntos se analizan más en la Parte III de este informe.

CAPÍTULO 3

Aplicaciones de los modelos de valor agregado para procesos de mejora escolar

Escuelas como organizaciones de aprendizaje **64**
- Análisis de los perfiles de valor agregado de las escuelas **66**
- Aplicación de modelos de valor agregado para ayudar a las iniciativas de mejora escolar **68**
- Aplicación de modelos de valor agregado para proyecciones de desempeño **71**

Uso dirigido de los modelos de valor agregado y equidad **74**

Aplicación de modelos de valor agregado para mejorar el sistema de evaluación escolar **77**

Apéndice:
Establecimiento de un sistema basado en modelos de valor agregado: Pasos básicos en la fase de implementación **83**

En el desarrollo de un sistema basado en modelos de valor agregado, el objetivo debe ser un impacto positivo en el ámbito escolar con el fin de aumentar el desempeño de las escuelas y del sistema educativo en su conjunto. El impacto en el ámbito escolar dependerá de la aplicación pretendida con la información de valor agregado y del marco en que se desarrollen los modelos de valor agregado. Este capítulo aborda la presentación e interpretación de los modelos de valor agregado del Capítulo 2. Aquí nos centramos en ilustrar cómo analizar la información de valor agregado dentro de las escuelas o en los ámbitos regional o de un sistema orientado a las mejoras escolares. Como con gran parte de este informe, un tema recurrente es el desarrollo de toma de decisiones basadas en datos dentro de las escuelas que operan como organizaciones de aprendizaje y dirigen mejoras en el sistema. Primero se analizan estos temas y después se dan ejemplos de la aplicación de modelos de valor agregado. Este análisis recurre a ejemplos del **Reino Unido** y de Tennessee, en **EUA**, considerados los que mejor ilustran la forma de aprovechar los modelos de valor agregado para promover mejoras escolares.

ESCUELAS COMO ORGANIZACIONES DE APRENDIZAJE

La información de modelos de valor agregado sirve para diversos fines de mejora escolar, pero sólo si la utilizan los actores que influyen en los procesos y/o resultados. En un sistema educativo, entre los actores más importantes destacan los maestros y los directores. Por tanto, es imperativo asegurar que tengan la capacidad de interpretar de manera eficaz la información de valor agregado y de actuar en consecuencia. Como la escuela es la unidad de acción, el interés en las mediciones de rendición de cuentas y desempeño se sitúa, al menos al principio, en el ámbito escolar. Por intuición, es probable que las iniciativas escolares ofrezcan grandes beneficios a las escuelas mejor capacitadas para utilizar la información con la cual elaborar y establecer mediciones de rendición de cuentas y desempeño (Caldwell y Spinks, 1998). El maestro en el aula, más que la escuela como unidad organizativa, es el que ejerce el mayor efecto en el aprendizaje estudiantil (OCDE, 2009). Por ello es esencial que los efectos de las mediciones de rendición de cuentas y desempeño puedan transferirse a los maestros y sus actividades en el aula. Esto requiere que la información se transmita con eficacia a los maestros y directores, y que esta información siga fluyendo en las escuelas para una mejora escolar continua (Senge, 2000). Para que esto suceda, debe reconocerse que las escuelas son sistemas organizativos complejos que emplean información para mejorar. Hay complicaciones para obtener, divulgar y utilizar información, así como varias barreras que impiden el uso eficiente de información para la mejora escolar (O'Day, 2002). Si la información de valor agregado se usa en un sistema con fuertes mediciones de rendición de cuentas, hay una mayor probabilidad de que la información que fluya esté distorsionada. La presencia de sanciones fuertes para escuelas y maestros crea el incentivo de distorsionar la información como forma de autoprotección frente a resultados deficientes (Lazear, 2000). Este comportamiento se extiende después para distorsionar la intervención que resulte de ese proceso. Por ejemplo, si se coloca a una escuela en un periodo de prueba restrictivo que requiera más información sobre los procesos escolares y desempeño estudiantil, se puede dificultar por la distorsión y restricción de información por parte de maestros y directores. Esto puede restringir de manera grave un programa de desarrollo escolar y dificultar el aprendizaje en el ámbito del sistema.

Por estos problemas potenciales es importante observar que los modelos de valor agregado resuelven muchas distorsiones asociadas a otras mediciones de desempeño escolar. Los

indicadores de desempeño que no miden con precisión el progreso estudiantil a menudo generan consecuencias indeseables, como el hecho que las escuelas seleccionen sólo a estudiantes de alto desempeño para continuar a los siguientes años y obliguen a los alumnos menos capaces a abandonarlas (Meyer, 1997). Esta selección ocurre porque, con estas mediciones, el desempeño escolar se correlaciona directamente con las habilidades innatas de los alumnos, siendo entonces muy importante para las escuelas quiénes son los alumnos que se presentan a la prueba. Sin embargo, con los modelos de valor agregado, el interés en el progreso estudiantil retira muchos de estos incentivos. El desempeño escolar se juzga con mediciones precisas de avance de desempeño estudiantil, de modo que se niega el incentivo de sólo retener a los alumnos de mejor desempeño (Wilson, 2004).

La divulgación de información de los modelos de valor agregado debe desarrollarse de manera que tome en cuenta la complejidad de la información y la estructura del flujo de dicha información dentro de las escuelas. En este sentido, es importante ver a las escuelas como unidades organizativas que operan dentro de sistemas más grandes de los que obtienen recursos y reciben restricciones. O'Day (2002, p. 294) sostiene que "los sistemas de rendición de cuentas fomentarán una mejoría en la medida en que generen y centren su atención en la información pertinente para enseñar y aprender, motiven a los individuos y escuelas a usar esa información y emprendan esfuerzos para mejorar la práctica, construir la base de conocimientos necesaria para interpretar y aplicar la nueva información para mejorar dicha práctica, y asignen recursos para todo esto". Los tipos de escuela, como en otras organizaciones, mejor equipados para transmitir las mediciones de rendición de cuentas y desempeño escolares a los maestros de aula individuales son las que tienen mayores niveles de colaboración y confianza entre colegas y, así, flujos de información más eficaces. Es más probable encontrar este tipo de cultura escolar en las escuelas que ya se interesen más en la responsabilidad colectiva por el aprendizaje escolar de forma proporcional a dicha colaboración y confianza. Por tanto, se obtienen mayores beneficios de las mediciones de rendición de cuentas y desempeño en escuelas con estas características organizativas. Por desgracia, las escuelas de bajo desempeño son a menudo las que tienen niveles deficientes de colaboración y confianza entre colegas y un sentido débil de responsabilidad colectiva por el aprendizaje escolar. Por tanto, existe el riesgo de que sea menos probable que se alcancen los objetivos de los modelos de valor agregado en las escuelas que suelen necesitar más iniciativas de mejoras escolares bien dirigidas y eficaces. Muchas escuelas operan con un alto grado de autonomía para los maestros referente a sus prácticas de enseñanza (OCDE, 2004, 2008b). Este grado de autonomía opera en contextos organizativos de alta colaboración entre colegas, pero también actúa como barrera para el flujo de información e incrementa la complejidad en la implantación de cambios en el ambiente escolar. Esto quizá explique por qué algunas escuelas responden bien a las intervenciones basadas en resultados de modelos de valor agregado y otras generan resultados deficientes por periodos prolongados, a pesar de recibir intervenciones que beneficiaron a otras escuelas (O'Day, 2002). Vencer estas barreras organizativas negativas es esencial para divulgar e interpretar con eficacia los modelos de valor agregado y después diseñar e implementar las iniciativas de mejora escolar pertinentes basadas en dicha información.

Los esfuerzos por mejorar los aspectos organizativos de las escuelas son un interés creciente en varios sistemas educativos en los países miembros de la OCDE (OCDE, 2005, 2008a, 2008b).

Muchos esfuerzos estan dirigidos a fomentar el desarrollo de la colaboración eficiente entre colegas y aumentar el interés del desarrollo escolar general. Estas acciones facilitarían el uso eficaz de la información derivada de los modelos de valor agregado, aparte de los beneficios fundamentales de las mejoras en la colaboración y confianza entre colegas y de la creación de un sentido de responsabilidad colectiva por el aprendizaje estudiantil. Además de la formación específica para interpretar la información de valor agregado, pueden desarrollarse programas para facilitar la colaboración entre colegas y mejorar las políticas organizativas que faciliten la comunicación eficaz entre maestros, directores y personal escolar. Necesitarían reconocer la complejidad de los modelos de valor agregado mismos, interpretar los resultados de la información obtenida y después idear la forma de aplicarla en el contexto organizativo de las escuelas para alcanzar los objetivos de desempeño y rendición de cuentas.

En algunos sistemas educativos, se coloca a las escuelas en un periodo de prueba o tienen una mayor colaboración con inspectores escolares o asesores pedagógicos u otros organismos externos como resultado de una medición de desempeño baja (OCDE, 2007a). Estos sistemas se consideran más intervencionistas por sus acciones para mejorar los resultados de las escuelas que, por ejemplo, un sistema que se interese más en la rendición de cuentas administrativa. Estas intervenciones se beneficiarían de un mayor énfasis en los factores organizativos que no sólo deben beneficiar los esfuerzos de mejoras escolares, sino también facilitar la divulgación y uso de información recopilada de los modelos de valor agregado. Las intervenciones capaces de reunir información de las escuelas e investigar las causas de un desempeño alto o bajo pueden tener un impacto mayor en el desempeño escolar y, en el largo plazo, en el desempeño del sistema.

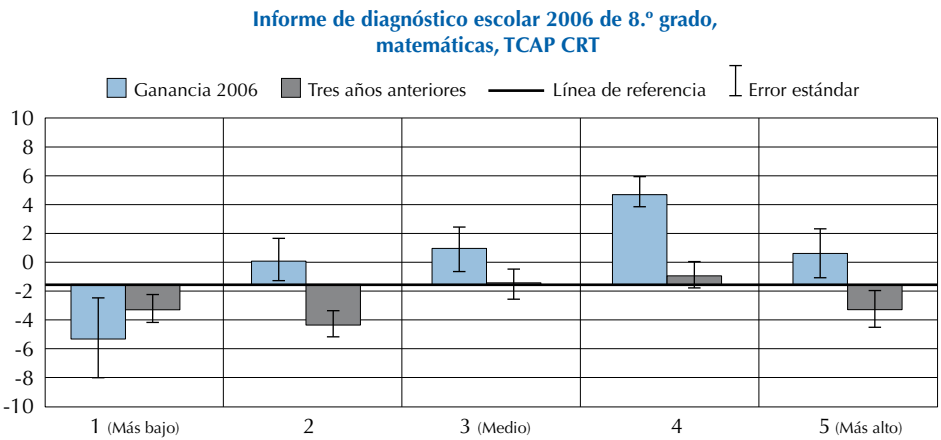
Análisis de los perfiles de valor agregado de las escuelas

El análisis de la información de valor agregado con fines de mejora escolar se beneficia de los análisis de los datos del alumno y de su desagregación por características de los alumnos. Esto permite que las escuelas individuales elaboren o analicen su "perfil de valor agregado". Por ejemplo, supongamos que todos los alumnos de 8.º grado en un área local o unidad administrativa se categorizan en quintiles elaborados con sus registros de desempeño previos. Se puede ajustar un modelo de valor agregado a los datos asociados a los estudiantes de un quintil particular. Al aplicar este análisis a cada quintil se genera un perfil de valor agregado de cinco componentes por escuela. Como ejemplo se presenta un perfil de este tipo de una sola escuela en la Gráfica 3.1. Es un ejemplo del sistema basado en modelos de valor agregado empleado en el estado de Tennessee (TVAAS), **EUA**, tema de la siguiente sección de este capítulo.

Primero observamos, tomando como referencia el aprovechamiento anterior, que los alumnos de esta escuela son relativamente más competentes que la población escolar del distrito. Además, por cada quintil excepto el más bajo, el valor agregado estimado de la escuela es positivo, y excede el estimado basado en los tres años anteriores. Para el director de la escuela éste es un buen perfil, si bien el resultado débil del quintil más bajo es preocupante y necesita más investigación. Tras hablar con los maestros y estudiar los registros escolares, el director cae en la cuenta de que este grupo incluye a estudiantes con discapacidades que tienen Planes de Educación Individualizada *(Individualised Education Plans*) y a alumnos con un registro de aprovechamiento bajo en matemáticas que no están claramente discapacitados.

Gráfica 3.1

Ejemplo del perfil de matemáticas de una escuela en el TVAAS



			Subgrupos por aprovechamiento previo				
			1 (Más bajo)	2	3 (Medio)	4	5 (Más alto)
Matemáticas	Línea de referencia		0.0	0.0	0.0	0.0	0.0
	2006	Ganancia	**-3.7**	**1.7**	**2.5**	**6.4**	**2.1**
		Error estándar	2.8	1.4	1.5	1.1	1.7
		Núm. de alumnos	31	39	47	66	34
	Tres años anteriores	Ganancia	**-1.7**	**-2.7**	**-0.0**	**0.7**	**-1.7**
		Error estándar	1.0	0.9	1. 0	0.9	1.3
		Núm. de alumnos	111	110	110	115	112

Notas:
Se ha traducido el texto del ejemplo original para facilitar su comprensión.
TVAAS = *Tennessee Value-added Assessment System*
TCAP = Programa de Evaluación Global de Tennessee (*Tennessee Comprehensive Assessment Program*)
CRT = Prueba con criterio referenciado (*Criterion Referenced Test*)
Fuente: Reel, M. (2006).

Las ganancias (relativamente) decepcionantes que aparecen en los modelos se ubican en este último subgrupo. El director también descubre que estos estudiantes están catalogados, por lo que la mayoría de los maestros no ha asumido el reto de enseñar a una amplia variedad de capacidades estudiantiles. Con esta información, el director puede reunirse con el supervisor de la materia de matemáticas y con los maestros responsables para discutir posibles estrategias que mejoren la tasa de progreso de los alumnos de bajo aprovechamiento.

Las medidas que se adoptan para mejorar un aspecto específico del desempeño escolar tienen efectos positivos de flujo hacia otras áreas de las escuelas con ganancias de eficiencia inadvertidas (Mante y O'Brien, 2002). Por ejemplo, suponga que la información del análisis de valor agregado muestra que una gran porción del bajo desempeño de una escuela se deriva de las dificultades de enseñar a alumnos con una lengua materna distinta al idioma de instrucción.

Análisis complementarios de las puntuaciones de valor agregado de los alumnos individuales revelan que el desempeño de estos estudiantes se ubica en general muy por debajo del nivel de la mayoría de los alumnos en distintas escuelas en los mismos grados escolares. Tras conversar con el inspector escolar o el administrador responsable, se obtiene información de valor agregado de otras escuelas para identificar a las escuelas con prácticas exitosas para estudiantes con un idioma materno distinto del de instrucción. Entonces pueden organizarse redes de aprendizaje con maestros y directores de las escuelas pertinentes que permitan al personal escolar aprender de las experiencias de los demás y compartir las mejores prácticas en estas y otras áreas. Los maestros y directores también agradecerían saber que su escuela no es la única con dificultades, lo que reduce la sensación de fracaso que pueden motivar las mediciones de puntuaciones brutas, y refuerza que se requiere y es posible el aprendizaje constante tanto dentro de las escuelas como en todo el sistema escolar. La red educativa se beneficiaría de los datos de valor agregado para destacar las diferencias en las tasas de progreso entre grupos de estudiantes tanto dentro de las escuelas como entre ellas. En el ámbito del sistema, un análisis de los resultados de la escuela junto con los resultados de otras escuelas puede revelar un patrón particular de grupos étnicos, por ejemplo, y que el desempeño de estos grupos viene siendo un tema prioritario sostenido en el tiempo. Puede decidirse entonces que es necesario dedicar más recursos a la educación de estos alumnos y dirigirlos a las escuelas que requieren más formación y recursos para enseñar a estos estudiantes. De hecho, las diferencias de desempeño entre distintos grupos étnicos y la población general son claros en varios países. Por esta razón, varios países incluyen variables de "país de origen" en sus modelos en lugar de una variable simple de "inmigrante" o una que identifique si el idioma materno de los alumnos es el mismo que el idioma de instrucción.

Aplicación de modelos de valor agregado para ayudar a las iniciativas de mejora escolar

En **EUA**, la historia de los modelos de valor agregado difiere por estados. Se usaron por algún tiempo con fines de rendición de cuentas y mejoras escolares en los estados de Carolina del Sur y Florida. Por otra parte, el estado de Tennessee, las ciudades de Dallas, Texas, y Milwaukee, Wisconsin, así como varios consorcios de reformas escolares, emplearon abundantemente los modelos de valor agregado para mejoras escolares sin un vínculo directo con la rendición de cuentas de las escuelas (Braun, 2006a). Como Tennessee tiene un programa bien establecido de desempeño escolar y un sistema de informes muy refinado, la siguiente sección se centra en cómo las escuelas distritales de Tennessee emplean los modelos de valor agregado con el fin de ilustrar el potencial de los análisis de valor agregado.[1]

Estudio de caso: Tennessee, EUA

Tennessee fue el primer estado en adoptar formalmente los análisis de valor agregado como parte de una iniciativa de desarrollo escolar. Fascinado por la obra de William Sanders, entonces profesor en la Universidad de Tennessee, el estado aprobó una legislación en 1993 que requería que las escuelas y distritos recopilaran y transmitieran datos estudiantiles al profesor Sanders. Esto permitió a Sanders efectuar los cálculos propios de su modelo de valor agregado, denominado Sistema de Evaluación de Valor Agregado Educativo de Tennessee (*Tennessee Value-added Assessment System,* TVAAS).[2] La legislación prohibía de manera explícita el uso de los resultados de los modelos de valor agregado para rendición de cuentas de escuelas o maestros. En cambio, se usarían exclusivamente para el desarrollo escolar y, además, se dejó

a cada distrito la decisión de utilizar el TVAAS. Los datos de entrada del TVAAS se basan en el desempeño estudiantil en el Programa de Evaluación Global de Tennessee (*Tennessee Comprehensive Assessment Program,* TCAP), que consiste en una batería de pruebas de aprovechamiento de opción múltiple. Estas pruebas, administradas en la primavera, ofrecen información referente tanto a normas como a criterios. Por cada combinación de materia-grado se generan informes de los niveles de distrito, de escuela y de estudiante individual.

Desde el principio se reconoció que si se quería que el TVAAS tuviera el impacto deseado, los educadores en todo el estado tendrían que someterse a un proceso de inducción y formación. Los análisis estadísticos son complejos, y con razón quienes no son estadísticos los consideran una proverbial "caja negra". Los educadores primero deben convencerse de que los resultados que genere el sistema son tanto pertinentes como justos. Después deben recibir una formación para interpretar de manera adecuada los resultados y, para esto, se diseñaron informes especialmente para facilitar el proceso y alentar su empleo eficaz. Por último, tendría que haber apoyo del Departamento de Educación del estado para que los directores se aseguraran de que este esfuerzo no era pasajero y no se les dejaría solos sin ningún apoyo adicional una vez completada la fase inicial.

La introducción del TVAAS recibió mucho apoyo de los directores escolares en sus 15 años de aplicación. La formación para los educadores es constante, pues hay una corriente continua de nuevos ingresados en el sistema educativo del estado. Las imágenes y el texto que las acompaña en esta sección pretenden dar al lector una muestra del sistema.[3] La introducción en todo el estado comprendió un proceso de tres fases que implicó a miles de supervisores de currículo, directores escolares, directores regionales y personal del departamento estatal. La Fase I fue informativa, diseñada para ofrecer una introducción general al TVAAS y a la estructura de los informes que genera el sistema. La Fase II fue la implementación inicial, que incluyó una revisión del TVAAS y un análisis e interpretación guiados de datos locales. Esta fase también abordó estrategias para informar a los padres y a la comunidad más amplia sobre el TVAAS. La Fase III constituyó la etapa avanzada, en la que el TVAAS y otras fuentes de información se integraron a un proceso de toma de decisiones basado en datos que afectó directamente las acciones del personal escolar y las asignaciones de recursos. La siguiente meta fue facilitar el desarrollo de una cultura de mejora escolar continua que se basó en un sólido cimiento empírico con información, en parte, de los resultados del TVAAS. Desde luego, la meta final era mejorar el nivel de aprovechamiento escolar, para lo cual se supervisarían y evaluarían varios indicadores. Para ilustrar cómo funciona el sistema, a continuación se presentan varios extractos de la biblioteca del informe del TVAAS, junto con comentarios explicativos. El TVAAS desarrolla un modelo de valor agregado longitudinal de varias materias que recoge datos de cuatro materias (lectura/artes del lenguaje, matemáticas, ciencias y ciencias sociales) de los Grados 3 al 8. Los análisis se efectúan por cada distrito escolar, y los informes escolares proporcionan resultados del año en curso, los dos años anteriores y el promedio de los tres años.

La Gráfica 3.2 contiene una parte del informe del TVAAS de matemáticas de una escuela media (secundaria) de los Grados 5 a 8. En la parte superior, los efectos escolares estimados se expresan en unidades de curva equivalente normal (*Normal Curve Equivalent,* NCE) y se acompañan de estimaciones de sus errores estándar. Para facilitar la lectura, el color de cada celda varía si el efecto escolar estimado es mayor que el crecimiento estándar (cero) –azul (B)–, no más

de una unidad de error estándar por debajo de cero –azul claro (LB)– o más de una unidad de error estándar inferior a cero –negro (BL)–. En este último caso, las celdas se etiquetan G* si el efecto escolar estimado es más de dos errores estándar inferior a cero. Por ejemplo, en el Grado 6 (2006), el efecto escolar estimado es de 3.6 unidades de escala con un error estimado de 0.8. Por tanto, en el informe tiene color azul. Es importante recordar que esta estimación es un estimador empírico-bayesiano, por lo que el estimador directo de la ganancia media de la escuela se "contrajo" hacia el promedio del distrito, y la cantidad de la contracción depende de la precisión relativa de ese estimador (véase el Capítulo 6 para un mayor análisis de la contracción de los estimadores de valor agregado).

La estructura matricial facilita las comparaciones entre grados dentro de años (horizontal) y de grados dentro de años (vertical). Para esta escuela, parece haber una mejora general con el tiempo en cada grado, y los resultados más sólidos están en el Grado 6. En el extremo derecho de la sección, los resultados de la escuela (promediados por grados) se comparan tanto con los del crecimiento estándar como con los del estado.

Gráfica 3.2

Ejemplo de un Informe Escolar TVAAS, 2006

Matemáticas, TCAP CRT

Ganancia estimada de la media escolar NCE

Grado	5	6	7	8	Ganancia media NCE de grados en relación con	
Crecimiento estándar	0.0	0.0	0.0	0.0		
Prom. estatal de 3 años consecutivos	2.4	1.7	1.6	1.5	Crecimiento estándar	Estado
Ganancia media NCE 2004		3.2 B	-2.3 G*	-2.7 G*	**-0.6**	**-2.2**
Error estd.		0.8	0.8	0.8	0.5	0.5
Ganancia media NCE 2005	1.0 B	6.2 B	-2.4 G*	2.1 B	**1.7**	**-0.1**
Error estd.	1.2	0.8	0.7	0.8	0.4	0.4
Ganancia media NCE 2006	-0.1 LB	3.6 B	1.0 B	2.2 B	**1.7**	**-0.1**
Error estd.	1.2	0.8	0.8	0.7	0.4	0.4
Ganancia NCE prom. 3 años		4.3 B	-1.2 G*	0.5 B	**0.9**	**-0.6**
Error estd.		0.5	0.4	0.4	0.2	0.2

Puntuaciones estimadas de la media escolar NCE

Grado	5	6	7	8
Año base estatal (1998)	50	50	50	50
Prom. de 3 años estatal consecutivos	54.8	54.1	53.3	53.5
Media 2003		48.4	52.8	51.4
Media 2004		53.2	46.1	50.1
Media 2005	49.3	49.3	50.6	48.2
Media 2006	49.6	56.3	50.4	52.9

Notas:
Se ha traducido el texto del ejemplo original para facilitar su comprensión.
TCAP = Programa de Evaluación Global de Tennessee (*Tennessee Comprenhensive Assessment Program*)
CRT = Prueba con criterio referenciado (*Criterion Referenced Test*)
NCE = Curva equivalente normal (*Normal Curve Equivalent*)

B: Media estimada de ganancia NCE igual o mayor que el crecimiento estándar
LB: Media estimada de ganancia NCE inferior al crecimiento estándar, pero menos que un error estándar
BL: Media estimada de ganancia NCE inferior al crecimiento estándar por al menos uno, pero menos de dos errores estándar
G*: Media estimada de ganancia NCE inferior al crecimiento estándar por al menos dos errores estándar

Fuente: Reel, M. (2006).

La sección inferior de la Gráfica 3.2 traduce los resultados escolares a puntuaciones medias de NCE. Esto permite al lector considerar el registro escolar desde dos perspectivas. En el Grado 6 (2006), la media NCE es 56.3, correspondiente a un desempeño promedio marginalmente mayor que el promedio estatal de tres años, de 54.1. En otros grados en 2006, la media NCE de la escuela es marginalmente inferior a los promedios correspondientes de tres años del estado. (Observe que las comparaciones con las escuelas fuera del distrito siempre se dan en términos de niveles de aprovechamiento, nunca en términos de estimaciones de valor agregado.)

Aplicación de modelos de valor agregado para proyecciones de desempeño

Al combinar las trayectorias estudiantiles observadas con un perfil de valor agregado estimado de la escuela es posible predecir (proyectar) el desempeño de un alumno. El valor de este ejercicio es que permite a las escuelas y administradores determinar, con la tasa de crecimiento esperado de un grupo particular de estudiantes, qué proporción de estudiantes cumplirá con el estándar de aprovechamiento deseado en uno o más años. Esto facilita la planificación y la asignación de recursos, y destaca las áreas de categorías de desempeño bajo y alto de estudiantes y escuelas. Además, cuando un resultado predicho se queda corto, se da a la escuela una clara indicación de que debe proponerse un objetivo de valor agregado con el fin de alcanzar el nivel deseado de éxito (Doran e Izumi, 2004; McCall, Kingsbury y Olson, 2004; Hill *et al.,* 2005; Wright, Sanders y Rivers, 2006).

El establecimiento de objetivos en las escuelas es una parte importante del proceso de mejoras escolares en el **Reino Unido**. Los objetivos se fijan en relación con los resultados de pruebas y no con el valor agregado (hacerlo significaría que los objetivos no se calcularían de forma tan sencilla como las puntuaciones de valor agregado en relación con el promedio nacional, y por tanto sería estadísticamente imposible mejorar para todas las escuelas), pero la estrategia de valor agregado, al tomar en cuenta el aprovechamiento previo del alumno, apuntala el establecimiento de objetivos de desempeño. Se ha de tener cuidado de alentar el establecimiento de objetivos para estudiantes, escuelas y autoridades locales que no sean simples extrapolaciones del desempeño previo. Hay varias formas de hacer esto, pero la estrategia general ha sido dar información sobre la clase de resultados que se esperarían en el futuro si una escuela, por ejemplo, mejorase su valor agregado al nivel de escuelas con similares características (en términos de aprovechamiento previo promedio) que en ese momento tengan un valor agregado más alto.

Hay implicaciones de políticas respecto de la inclusión de variables más ***contextuales*** en los modelos de valor agregado para fijar objetivos. Se corre el riesgo de que se formen bajas expectativas para estudiantes que en ese momento hagan menos avances en promedio (p. ej., estudiantes de antecedentes socioeconómicos más bajos). Por otra parte, las escuelas con aprovechamiento previo alto pero pocos estudiantes de antecedentes socioeconómicos bajos fijarían objetivos más ambiciosos si estos datos contextuales se incluyeran en los modelos. En **Noruega**, las variables contextuales no se incluyen en los modelos de valor agregado publicados, pues se considera que no representan bien la intención del programa e incluso operan para afianzar las desigualdades existentes.

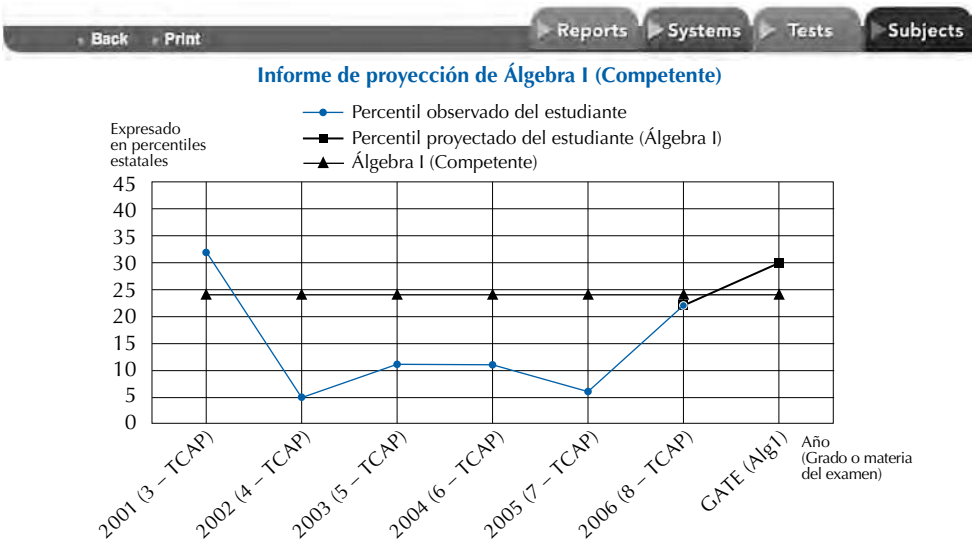
Una estrategia muy distinta es la consideración de patrones de movimiento de una categoría a la siguiente en el curso de uno o más años lectivos. A grandes rasgos, el interés recae en cuán

exitosa es una escuela para ayudar a los estudiantes a desplazarse de categorías más bajas a las más altas. La mejor manera de presentar los datos pertinentes es en forma matricial, en donde las filas representen las categorías en el año inicial, y las columnas, las del año final. No es necesario que la cantidad de categorías en los dos años sea la misma. Pueden proponerse distintos índices que resuman las trayectorias de los estudiantes de una escuela, y la diferencia entre el valor del índice de la escuela y el valor del índice del conjunto de escuelas es una medición del valor agregado de la escuela (para más detalles, véase Betebenner, 2008; Braun, Qu y Trapani, 2008).

La Gráfica 3.3 presenta un Informe de Proyección basado en el TVAAS. Describe la trayectoria de un alumno de bajo desempeño a lo largo del 7.º grado. Al combinar el registro del alumno con el valor agregado de la escuela para estudiantes en el quintil más bajo se obtienen proyecciones del desempeño de este estudiante en el 8.º grado y en Álgebra I en el siguiente año. Sin duda, se espera que el alumno mejore su clasificación relativa en el 8º grado al Percentil 22, si bien aún está por debajo del umbral de competencia (Percentil 24).

Gráfica 3.3

Ejemplo de un Informe de Proyección TVAAS



Materia: Álgebra I							
	TCAP CRT (Matemáticas)						Álgebra I (Competente)
Año (Grado o materia del examen)	2001(3)	2002(4)	2003(5)	2004(6)	2005(7)	2006(8)	Proyección (Probabilidad)
NCE estatal/Puntuación	38	12	26	24	20	36	
Percentil	32	5	11	11	6	22	30 (63.2%)

Notas:
Se ha traducido el texto del ejemplo original para facilitar su comprensión.
TCAP = Programa de Evaluación Global de Tennessee (*Tennessee Comprehensive Assessment Program*)
CRT = Prueba con criterio referenciado (*Criterion Referenced Test*)
NCE = Curva equivalente normal (*Normal Curve Equivalent*)

Fuente: Reel, M. (2006).

Sin embargo, se espera que para el siguiente año alcancen el Percentil 30, que lo colocaría en la categoría de competencia. La probabilidad estimada de que alcancen la competencia para el 9º grado es de 63.2%.

Estas proyecciones, agregadas por grupos de estudiantes, representarán un papel importante en los informes que el estado someta al gobierno federal como parte del régimen de rendición de cuentas escolar. Sin embargo, también son muy útiles internamente para establecer mejoras escolares. Si se proyecta que una gran cantidad de estudiantes caerá por debajo del estándar de competencia, la escuela tiene una advertencia temprana de que debe abordar con firmeza los factores pedagógicos o de otra clase, que retardan el avance estudiantil. Incluso en el caso presente, las proyecciones no son garantías. La escuela debe esforzarse mucho para mantener su valor agregado positivo y supervisar el desempeño futuro de los alumnos con el fin de verificar que las proyecciones son precisas, en un sentido probabilístico. Sin duda, una empresa de esta clase requeriría formación y apoyo para los directores, así como la infraestructura necesaria para estos análisis.

La habilidad para usar modelos de valor agregado como advertencia temprana es importante considerando datos alternativos. Los datos de resultados escolares, como las tasas de retención y de avance a un nivel educativo más alto, muestran retrasos temporales propios entre la identificación de un desempeño en declive y la identificación de ese problema mediante los datos. Los modelos de valor agregado ofrecen datos más ágiles, pues las tasas de retención y el avance son menos sensibles a los cambios de desempeño escolar. Los datos de valor agregado basados en las evaluaciones estudiantiles en múltiples años proporcionarían información más oportuna que permitiría una identificación expedita de problemas potenciales. Así, se facilitarían las acciones para abordar dichos problemas.

En la Gráfica 3.4 se ofrece otra presentación de un tema semejante, que ilustra cómo se pueden comunicar las proyecciones para que las empleen las escuelas individuales. La gráfica presenta resultados de estudiantes de 8.º grado en una escuela que tomaban su primer curso de álgebra.

Gráfica 3.4

Ejemplo del Informe Escolar del TVAAS (Álgebra)

TVAAS

Home Search Contact Us Help Admin User Settings Logout

Back Print

Reports Systems Schools Tests Subjects

Prueba	Año	N	Puntuación estudiantil media	Puntuación media del percentil	Media del percentil predicho	Puntuación del percentil predicho	Efecto escolar	Percentil del efecto escolar	Promedio escolar respecto del estatal
Álgebra I	2004	49	579.6	82	584.8	85	-4.7	35	Sin datos
	2005	43	595.6	90	570.4	78	22.3	96	Superior
	2000	58	583.0	85	571.8	79	10.2	81	Superior
	Promedio de 3 años	150	585.5	86	575.6	80	9.3	80	Superior

Para ver con más detalle los informes, haga clic en las cifras o palabra subrayadas.

Nota: Se ha traducido el texto del ejemplo original para facilitar su comprensión.
Fuente: Reel, M. (2006).

Se muestran las puntuaciones medias del año en curso, los dos años anteriores, así como el promedio de tres años. La "puntuación media predicha" (Columna 6) implica un cálculo basado en las trayectorias de las puntuaciones de alumnos durante el séptimo grado y su ganancia esperada si se inscribieran en una escuela distrital normal. Así, en 2006, la puntuación media alcanzada de 583.0 excedió la puntuación media esperada de 571.8 en 11.2 puntos de la escala de puntuación. El estimador correspondiente (empírico-bayesiano) del valor agregado de la escuela es 10.2. Esto coloca a la escuela en el Percentil 81 del valor agregado entre las escuelas del distrito. Es importante observar que la puntuación media predicha en 2006 es 13 puntos más baja que en 2004, cuando el valor agregado de la escuela fue de -4.7, para colocarla en el Percentil 36 entre las escuelas del distrito de ese año. Esto indica que los cambios en la composición estudiantil en la escuela bajaron la puntuación predicha, y de nuevo revela la importancia de usar modelos de valor agregado en lugar de centrarse en las puntuaciones brutas de pruebas.

USO DIRIGIDO DE LOS MODELOS DE VALOR AGREGADO Y EQUIDAD

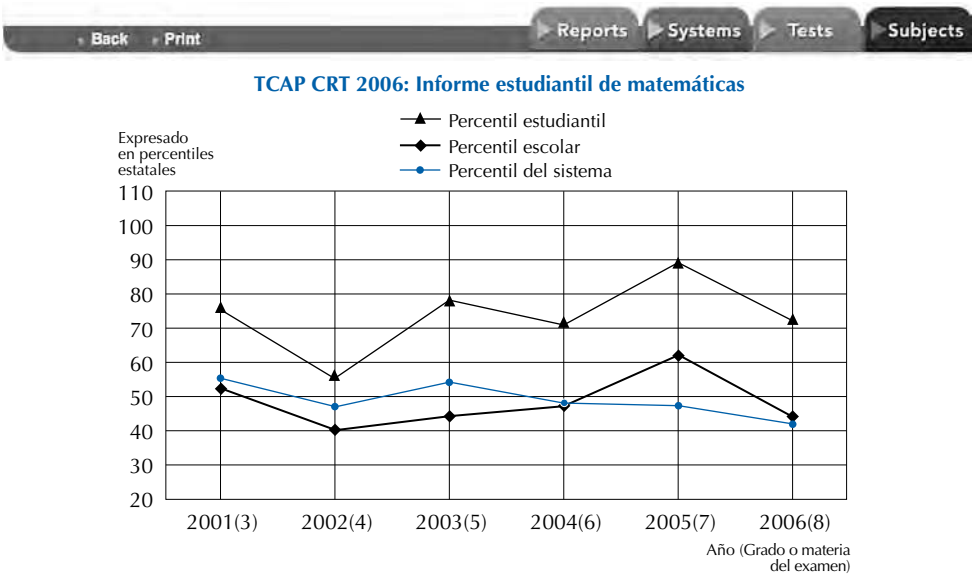
Los análisis sobre el empleo de los modelos de valor agregado han resaltado las ventajas de centrarse en grupos particulares de escuelas o estudiantes, o incluso en políticas y programas particulares. Para las autoridades educativas en muchos países de la OCDE es una prioridad establecer políticas dirigidas a grupos particulares de escuelas y estudiantes de desempeño bajo o alto (OCDE, 2007c). Las puntuaciones de valor agregado sirven para identificar a las escuelas específicas en las cuales desarrollar programas adecuados, y para supervisar su impacto. Son varias las ventajas de usar mediciones de valor agregado en lugar de puntuaciones brutas de aprovechamiento. Con las puntuaciones brutas de aprovechamiento es posible identificar a estudiantes de bajo desempeño y las escuelas en donde están. No obstante, esta información no puede usarse para analizar el progreso escolar. Por ejemplo, estos estudiantes, que quizá provengan de ambientes socioeconómicos bajos, ¿siempre fueron de bajo desempeño? ¿Qué escuelas (y tal vez también programas) tienen el mayor valor agregado para estos estudiantes, y qué puede aprenderse de los éxitos? Son preguntas básicas sobre equidad en los sistemas educativos. Abordan directamente la cuestión de que los estudiantes de bajo desempeño tal vez estén atascados en el fondo de la distribución o sean capaces de alcanzar altos niveles de competencia.

El análisis de datos de valor agregado permite a los maestros, directores y autoridades correspondientes sumergirse en los datos de estudiantes de bajo desempeño para comprender mejor sus trayectorias de aprendizaje. Por ejemplo, en el **Reino Unido**, los tomadores de decisiones analizan los datos de estudiantes en niveles específicos de desempeño. Se distingue entre estudiantes que de verdad mejoran con el tiempo, estudiantes estancados en niveles de bajo desempeño y estudiantes que en realidad reducen su medición de valor agregado de desempeño (Ray, 2006). Son distinciones importantes porque no sólo ofrecen información considerable sobre el aprendizaje y la educación escolar de estos estudiantes, sino porque también guían el desarrollo de las políticas y programas adecuados que beneficien más a esos estudiantes. Tras poner en marcha los programas, los análisis adicionales de los resultados de valor agregado con los datos de las siguientes evaluaciones facilitan la supervisión de la eficacia de dichos programas. De nuevo, esto en gran medida no es posible con análisis de datos brutos de aprovechamiento. La base de datos del TVAAS

que apoya el análisis de valor agregado también posibilita el registro del desempeño de un estudiante individual. La Gráfica 3.5 presenta una trayectoria de seis años de un estudiante particular acompañada de las trayectorias medias de la escuela y del distrito (sistema). De acuerdo con los indicadores del nivel de desempeño, este estudiante excedió el umbral de posición "Avanzada" desde el 5.º grado. Al mismo tiempo, el declive sustancial de la clasificación relativa del Grado 7 al 8 es causa de inquietud. La revisión de las trayectorias de otros estudiantes de posición semejante revela patrones entre materias y dentro de las escuelas que quizá revelen problemas más sistemáticos y se puedan ofrecer respuestas a los problemas que enfrentan algunos estudiantes.

Gráfica 3.5

Ejemplo de trayectorias de desempeño comparativas de TVAAS



Materia: Matemáticas						
	TCAP CRT (Matemáticas)					
Año (Grado o materia de examen)	**2001(3)**	**2002(4)**	**2003(5)**	**2004(6)**	**2005(7)**	**2006(8)**
NCE estatal/Puntuación	66	55	67	63	79	68
Percentil	75	55	78	71	89	72
Nivel de desempeño			AD	AD	AD	AD

Notas:
Se ha omitido el nombre del estudiante y se ha traducido el texto del ejemplo original para facilitar su comprensión.
TCAP = Programa de Evaluación Global de Tennessee (*Tennessee Comprehensive Assessment Program*)
CRT = Prueba con criterio referenciado (*Criterion Referenced Test*)
NCE = Curva equivalente normal (*Normal Curve Equivalent*)

Niveles de desempeño
NP – No competente
P – Competente
AD – Avanzado

Fuente: Reel, M. (2006).

En ocasiones, el interés recae en los efectos escolares estimados para un subconjunto de la población de escuelas que contribuyeron al análisis completo de valor agregado. Por ejemplo, suponga que se desea comparar el desempeño aparente de dos grupos de escuelas, cada una con un programa educativo distinto. La estrategia más sencilla sería extraer los efectos escolares estimados provenientes del análisis completo. Sin embargo, si los dos grupos constituyen una fracción relativamente pequeña de la población mayor de escuelas, quizá se efectuaría un nuevo análisis de valor agregado para sólo los dos grupos de escuelas. La cuestión es si es necesario este análisis auxiliar. Haegeland *et al.* (2005) efectuaron un estudio para comparar los resultados de estas dos estrategias con datos noruegos e informaron que las diferencias en las comparaciones fueron insignificantes. Aunque no es más que un hallazgo, puede suponerse que sucede así en general. Sin embargo, no hay que olvidar que las varianzas asociadas a los efectos escolares estimados son muy heterogéneas, y se debe tener cuidado al elaborar estadísticas de examen adecuadas para las comparaciones deseadas.

En el **Reino Unido** se han dirigido diversos programas a grupos de escuelas (p. ej., el programa de Escuelas Especializadas o la Beca de Incentivos al Liderazgo [Ray, 2006]). Las puntuaciones de valor agregado sirven como información para supervisar las iniciativas de políticas de esta clase. Además de proporcionar información sobre valor agregado general, las puntuaciones de valor agregado de las escuelas muestran cuánta variación hay entre escuelas dentro de estas políticas. Si bien el valor agregado no se emplea para la financiación directa de las escuelas, sí ha servido como método para seleccionar escuelas determinadas. Por ejemplo, algunas escuelas se designan como de "Alto desempeño" y reciben la responsabilidad adicional de ayudar a las escuelas locales más débiles o de participar en otros proyectos.[4] Así, las escuelas identificadas reciben más fondos para ayudar a las escuelas vecinas y centrarse en actividades especiales, como aprendizaje vocacional o estudiantes con necesidades educativas especiales. Los criterios que deben cumplirse se basan en mediciones de valor agregado en diferentes Etapas de los tres años anteriores.

Las políticas dirigidas quizá requieran modelos de valor agregado más complejos desarrollados con variables correspondientes al grupo o subgrupo pertinente de las escuelas o estudiantes en cuestión. En el Capítulo 2, el análisis ilustró que es más fácil presentar y comunicar los modelos más sencillos a los interesados. Por tanto, hay que equilibrar entre el deseo de presentar un modelo más fácilmente comunicable y elaborar un modelo estadísticamente más robusto pero también más complejo. Como estos análisis suelen dirigirse más a un público interno que a uno externo, se aligera un poco el problema de comunicar modelos más complejos. Los análisis de programas particulares para los tomadores de decisiones y administradores no tienen los mismos requerimientos de divulgación y presentación que los resultados de valor agregado para políticas de elección de escuela que ayudan a padres y familias a elegir escuela. La mayor complejidad de los análisis estadísticos se analiza con más facilidad en la forma adecuada para este tipo de análisis (p. ej., un informe o resumen) que los análisis presentados en cuadros grandes para el público en general.

También se aplican análisis más profundos para investigar subgrupos de escuelas específicos. Esto se hace para conocer mejor esas escuelas y también para asegurarse de que no haya problemas de especificación deficiente con el modelo más sencillo. Por ejemplo, si se clasifica a las escuelas como de bajo desempeño, puede aplicarse el modelo más complejo para asegurar que los resultados de estas escuelas no sean resultado de emplear el modelo más sencillo: Esta "doble verificación" también contribuye a comunicar la precisión de los procedimientos a los

interesados. Los análisis que utilizan modelos de valor agregado más complejos pueden ser importantes para las escuelas que exhiben mayor inestabilidad en los resultados del modelo de valor agregado más sencillo. De nuevo, esto tendría la ventaja de saber más sobre estas escuelas y de verificar si la falta de estabilidad puede controlarse con otra especificación.

El análisis de modelos más complejos que incluyan variables más contextuales es beneficioso para análisis específicos de los factores contextuales que afectan al progreso del desempeño estudiantil. Éstos pueden no ser estrictamente modelos de valor agregado, sino modelos de regresión que estudian las variables escolares y contextuales en el primer nivel de un modelo de valor agregado. Sin duda, hay una cantidad sustancial de oportunidades para efectuar estimaciones más complejas de valor agregado y otras multivariadas para analizar aspectos particulares en el sistema educativo.

APLICACIÓN DE MODELOS DE VALOR AGREGADO PARA MEJORAR EL SISTEMA DE EVALUACIÓN ESCOLAR

El éxito de iniciativas para objetivos de políticas de rendición de cuentas escolar, elección de escuela o mejoras escolares reside en las evaluaciones eficaces de desempeño escolar. Un mensaje central de este informe es que los modelos de valor agregado ofrecen mediciones más precisas de desempeño escolar que las basadas en puntuaciones brutas. No obstante, este informe sostiene que estas mediciones deben complementar los métodos existentes de evaluación escolar. Combinar información de valor agregado con la complementaria sobre entradas y procesos escolares facilita la toma de decisiones eficaz basada en datos en todo el sistema escolar. Esta toma de decisiones puede extenderse a discusiones entre maestros y directores sobre problemas escolares, como el ambiente escolar y las políticas y programas de las escuelas, que puedan generar información importante sobre las acciones particulares que tal vez se requieran para abordar los problemas que quizá hayan producido una puntuación baja de valor agregado.

Las iniciativas de mejoras escolares requieren una evaluación de la situación presente para identificar las áreas que necesitan mejorarse y las que ofrezcan ejemplos de las mejores prácticas. En varios países, un sistema de evaluación escolar no se considera por ende tan sólo otra forma de rendición de cuentas escolar. En cambio, se ve como otro mecanismo para desarrollar e impulsar iniciativas de mejoras escolares. En **Portugal**, el uso de mediciones de desempeño forma parte de un sistema más amplio de evaluación escolar. El programa, Evaluación Integrada de Escuelas, pretende contribuir al mantenimiento de calidad educativa al identificar los puntos fuertes y débiles del funcionamiento de las escuelas y del sistema escolar en general. Los objetivos principales de la Evaluación son valorar tanto el aprendizaje como la calidad de la experiencia escolar de los alumnos, identificar puntos fuertes y débiles del desempeño escolar, inducir procesos de autoevaluación en las escuelas, recopilar información y caracterizar el desempeño del sistema educativo, y regular el funcionamiento del sistema educativo (IGE, 2001). Las medidas para alcanzar estos objetivos se beneficiarían en gran medida de los modelos de valor agregado.

En el **Reino Unido**, los datos de valor agregado para los Cuadros de Desempeño Escolar incluyen una variedad limitada de estadísticas sobre las escuelas: los datos de valor agregado se presentan junto con hechos sobre aprovechamiento general e información contextual escolares. Para las mejoras e inspección escolares se emplea una variedad más amplia de mediciones de valor agregado, gráficas y otros datos mediante el *software* RAISEonline que se ilustró en el Capítulo 2.

En el mismo año que se introdujeron los Cuadros de Desempeño Escolar se reformó la inspección escolar con la creación de la Inspección Escolar Nacional. Este organismo supervisa a todas las escuelas y Autoridades Locales en el Reino Unido, y sus inspectores tienen acceso a los datos de aprovechamiento escolar, en forma de Informes de Desempeño y Evaluación (*Performance And Assessment* [PANDA] Reports).[5] Por tanto, los datos de estos informes representan un papel importante en el sistema de rendición de cuentas escolar, pues forman parte de la base de evidencias con que los inspectores juzgan el desempeño escolar. Los informes de la Inspección se publican y en ellos se clasifica a las escuelas como Sobresaliente, Buena, Satisfactoria o Inadecuada; las escuelas en esta última categoría pueden colocarse dentro de "mediciones especiales" o recibir una Notificación de Mejora.[6] Tanto las puntuaciones escolares de valor agregado como de otro tipo de análisis de valor agregado se han empleado en otros ámbitos: en la publicación de información para padres y escuelas, en la selección de escuelas con fines particulares y como parte de la estrategia de establecer objetivos. El principal objetivo de RAISEonline es ofrecer a todas las escuelas un producto de *software* gratuito que les permita analizar sus propios datos y compararlos con los patrones nacionales y los resultados y valor agregado que alcanzan escuelas de alto desempeño. Las escuelas usan el RAISEonline como parte del proceso de autoevaluación y de establecimiento de objetivos que efectúan con ayuda de los Socios de Mejoras Escolares. Los datos también están disponibles para que los inspectores escolares juzguen el grado al que la escuela mejora o tiene la capacidad de mejorar. Un aspecto importante es que las estadísticas no están disponibles para el público en general.

La Inspección holandesa, **Países Bajos**, está revisando sus operaciones para evaluar e incrementar el desempeño escolar. Aunque siempre se ha considerado crucial un programa extenso de evaluación escolar, se cree que hay beneficios al centrarse en áreas específicas para guiar las evaluaciones escolares y la asignación de recursos con el fin de evaluar y elevar el desempeño escolar. Esto generó interés en los indicadores de resultados escolares y también en los indicadores de organización y procesos escolares. Se identificaron cinco indicadores clave de resultados:

- La media de resultados escolares al final del periodo de tres años es más de la mitad de una desviación estándar por debajo del nivel que debe esperarse de la población estudiantil de la escuela.
- Más de 10% de estudiantes son considerados de bajo desempeño en lectura y aritmética.
- Más de 5% de estudiantes repitió un año en la escuela.
- Se transfiere a más de 2% de los estudiantes a educación primaria especial o a centros expertos designados.
- Ocurren incidentes de violencia física al menos una vez al mes en la escuela.

Un estudio reciente descubrió que 24% de las escuelas primarias tendría al menos uno de estos indicadores de resultados y por tanto requeriría más recursos o inspecciones. En esta etapa, una falta de datos impide que se efectúen análisis de valor agregado en todas las escuelas de los **Países Bajos**, pero se considera que ayudarían en gran medida a la Inspección en la identificación de las escuelas dada la mayor precisión de las mediciones de valor agregado y las ventajas inherentes de medir si se registran mejoras en estos estudiantes de bajo desempeño. Complementar estos indicadores de resultados escolares es un foco de atención en ocho procesos de enseñanza-aprendizaje: que se detallan más en el Recuadro 3.1.

Recuadro 3.1 Interés en procesos de enseñanza y aprendizaje específicos de inspecciones escolares en los Países Bajos

En un esfuerzo por concentrar el sistema de evaluaciones escolares que alimentan las iniciativas de mejoras escolares, la Inspección holandesa identificó ocho estándares de organización y procesos escolares que miden o influyen en los procesos de enseñanza y aprendizaje. Los ocho estándares son:

	Estándar	Indicador
1	El currículo cubre los objetivos de aprovechamiento y se ofrece a todos los estudiantes en su totalidad.	Los métodos y materiales para cubrir los objetivos de aprovechamiento de las materias de idioma holandés y aritmética/matemáticas.
		El currículo de la materia de idioma holandés y aritmética/matemáticas se ofrece en su totalidad a todos los alumnos hasta incluir el nivel del año 8.
2	El tiempo de enseñanza se invierte de manera eficiente.	Se evita la pérdida innecesaria de tiempo de enseñanza.
3	Los maestros dan explicaciones claras, organizan la lección con eficiencia y mantienen el interés de los alumnos en sus actividades.	Los maestros explican las cosas con claridad.
		Los maestros organizan las lecciones con eficiencia.
		Los maestros mantienen el interés de los alumnos en sus actividades.
4	Los maestros adaptan el currículo, tiempo de enseñanza, instrucción y tiempo dedicado al aprendizaje de la materiá en función de las diferencias entre los estudiantes.	Los maestros adaptan el currículo en función de las diferencias entre los alumnos.
		Los maestros adaptan el tiempo de enseñanza y aprendizaje en función de las diferencias entre los alumnos.
		Los maestros adaptan la instrucción en función de las diferencias entre los alumnos.
		Los maestros adaptan el tiempo dedicado a que los estudiantes aprendan la materia en función de las diferencias entre los alumnos.
5	El ambiente escolar se caracteriza por la seguridad y el respeto entre la gente.	Los maestros se aseguran de que los estudiantes se traten entre sí con respeto.
		La escuela salvaguarda la seguridad social del alumnado y el personal.
6	Los maestros supervisan sistemáticamente el avance de sus alumnos.	La escuela emplea un sistema cohesivo de instrumentos y procedimientos para supervisar el desempeño educativo y el desarrollo de sus alumnos.
		Los maestros supervisan sistemáticamente el avance de sus alumnos.
7	Los maestros proporcionan suficiente cuidado y asistencia a los alumnos en riesgo de rezagarse.	Para estudiantes con necesidades especiales, los maestros sistemáticamente establecen las medidas pertinentes.
		La escuela proporciona un cuidado sistemático.
		La escuela evalúa los efectos del cuidado que proporciona.
8	La administración de la escuela supervisa la calidad de la educación.	Cada año, la escuela evalúa sistemáticamente la calidad de los resultados.
		Cada año, la escuela sistemáticamente evalúa la calidad de la organización del proceso de enseñanza-aprendizaje.
		La escuela trabaja sistemáticamente en mejorar las actividades.
		La escuela salvaguarda la calidad de la organización del proceso de enseñanza-aprendizaje.

La evaluación de estos estándares se complementaría con información de valor agregado para ofrecer una evaluación escolar más completa y permitir el análisis de las relaciones entre estos estándares y las puntuaciones de valor agregado tanto dentro como entre las escuelas. Estos análisis permitirían saber al interior de las escuelas cómo mejorar estas áreas para elevar el desempeño de estudiantes y de la escuela. Asimismo, facilitaría las mejoras dentro de la Inspección holandesa, pues elaborarían sus evaluaciones de desempeño sobre estas áreas con las características más asociadas a las mejores puntuaciones de valor agregado.

Por sí mismos, ninguno de estos asuntos o estándares identificados por la Inspección holandesa ofrece un indicador singular de desempeño escolar. En cambio, como sucede en otros países, se combinan numerosos indicadores para proporcionar un perfil escolar que sirva para evaluar a las escuelas y elaborar iniciativas de mejoras escolares. En este ambiente, las puntuaciones de valor agregado proporcionan un papel útil como "ancla cuantitativa" para el desarrollo y análisis del perfil de la escuela. De esta manera, el uso de modelos de valor agregado permite construir un marco evaluativo más exacto. Desde luego, la estrategia particular adoptada dependerá del propósito de la evaluación escolar, así como de la amplitud y naturaleza de las mediciones para construir el perfil de la escuela. Incorporar mediciones de valor agregado a un perfil escolar más amplio ofrece una panorámica más completa del desempeño escolar y, en potencia, del desempeño de diferentes aspectos dentro de cada escuela. Esto tiene efectos indirectos para la calidad de las iniciativas de mejoras escolares. Las inspecciones verifican las conclusiones provenientes del análisis de las puntuaciones de valor agregado e incrementan la cantidad de información sobre la intervención adecuada. Esto es particularmente importante porque los resultados de los modelos de valor agregado sólo son indicadores, y la información necesita complementarse con información más detallada sobre los procesos escolares y de enseñanza para determinar la acción o intervención adecuada.

La información de valor agregado también sirve para incrementar la eficiencia del sistema de evaluación escolar y de instituciones como las inspecciones escolares, a menudo en el centro de dichos sistemas. Las ganancias de eficiencia provienen tanto de la mejor dirección de inspecciones escolares individuales como de una mejor asignación de recursos que se centre en las escuelas en las que más se necesite aplicar instrumentos evaluativos.

El análisis de la información de valor agregado identifica áreas clave en las cuales centrar una evaluación escolar para incrementar la eficiencia general y permitir una evaluación más profunda de áreas clave del desempeño escolar. Antes de inspeccionar una escuela, los encargados de hacerlo cuentan con información sobre el valor agregado de la escuela por áreas temáticas, niveles de años y estudiante. El análisis permite a quien efectúe la evaluación centrarse en los aspectos clave. Un elemento importante de la mayor eficiencia en el **Reino Unido** es la naturaleza exhaustiva de RAISEonline. Este *software* interactivo permite que las escuelas y los inspectores analicen la información de valor agregado para, por ejemplo, identificar las puntuaciones de valor agregado de los estudiantes en materias particulares y en niveles de años específicos para comprender mejor dónde tiene éxito la escuela y dónde necesita mejorar.

Los modelos de valor agregado no incluyen mediciones de entradas financieras y por tanto no pueden dar una forma de análisis de rentabilidad. El análisis de impacto diferencial de varias entradas en la educación escolar por ende no puede obtenerse mediante los modelos de valor agregado. Sin embargo, al ofrecer una medición de resultados más precisa, es posible efectuar un análisis más extenso del impacto de varias asignaciones de recursos. Además, la información de evaluaciones por niveles por año y materia puede ser en particular pertinente si los resultados de valor agregado se centran en materias particulares o muestran que es en materias particulares donde el desempeño estudiantil es bajo o alto en comparación con otras materias. Por ejemplo, si los resultados de valor agregado muestran que en el idioma de instrucción y ciencia los estudiantes se desempeñan en un nivel mayor que el de matemáticas,

esto indicaría que se requiere más información respecto de la materia. En algunos países es muy común aplicar evaluaciones por materia en lugar de evaluaciones de todas las materias, y esto puede ser en particular útil en una situación como la descrita (OCDE, 2007a).

Se pueden asignar más recursos a las escuelas, o áreas dentro de ellas, que muestren resultados deficientes de valor agregado. Para una inspección escolar, un sistema de inspecciones aleatorias se complementaría con inspecciones determinadas por las puntuaciones de valor agregado de una escuela. El componente aleatorio asegura que cualquier escuela aún puede ser objeto de una evaluación en cualquier momento, mientras que el componente determinado por las puntuaciones de valor agregado de una escuela dirige las inspecciones a las escuelas que no avanzan al nivel deseado. Las ganancias de eficiencia se incrementarían si una puntuación de valor agregado particular, como una que categorizara a una escuela como de bajo desempeño, supusiera automáticamente una inspección escolar.

La evaluación de los procesos escolares es subjetiva por naturaleza, y complementa la información de valor agregado. La calidad de las evaluaciones subjetivas de las prácticas escolares y de enseñanza se evalúa y después se mejora con los modelos de valor agregado. Una medición precisa del desempeño escolar permite el subsiguiente desarrollo de evaluaciones subjetivas de "lo que funciona", que es la base de las evaluaciones escolares. Como ya ilustramos, numerosas prácticas organizativas y de enseñanza se evalúan en evaluaciones escolares y por las inspecciones escolares. Estas prácticas a menudo se evalúan respecto de lo que se considera una "buena práctica". Se supondría que la definición de una buena práctica evoluciona con el tiempo, conforme avanza la comprensión de la enseñanza y la escolaridad eficaces. Por tanto, es importante evaluar y desarrollar continuamente lo que en verdad se considera una buena práctica. Parece pertinente incorporar alguna forma de medición de resultados a estas decisiones. Como las estimaciones de valor agregado son mediciones más precisas del desempeño escolar, los resultados pueden brindar información al desarrollo organizativo tanto de las inspecciones escolares como de la aplicación de las evaluaciones escolares. Lo que hoy en día se considera una "buena práctica" en las escuelas puede analizarse después de sus puntuaciones de valor agregado para determinar la validez de dichos juicios.

Apéndice

ESTABLECIMIENTO DE UN SISTEMA BASADO EN MODELOS DE VALOR AGREGADO: PASOS BÁSICOS EN LA FASE DE IMPLEMENTACIÓN

Este informe identifica varios objetivos para el desarrollo de un sistema basado en modelos de valor agregado e ilustra el uso potencial en varias aplicaciones y programas. La siguiente sección destaca los pasos principales que necesitan emprenderse en la fase de ejecución. Estos aspectos se analizan con más detalle en la Parte III de este informe, pero se presentan aquí para subrayar la importancia de vincular los objetivos y el uso de la información de valor agregado a la necesidad de una implementación fructífera. También destaca la manera en que se abordan muchos de los aspectos más técnicos en la fase de implementación. Los pasos analizados a continuación no comprenden una lista exhaustiva, y los detalles de cada actividad se analizan de modo más completo en el resto de este informe. Se presenta aquí el proceso requerido en la implementación de un sistema basado en modelos de valor agregado para hacerlos más accesibles a las autoridades educativas y a los administradores.

Fase 1: Establecimiento de objetivos de políticas y mediciones de desempeño escolar

- Identificar de manera explícita los objetivos de políticas para el establecimiento de un sistema basado en modelos de valor agregado. Esto implica una especificación de los usuarios de la información de valor agregado y de cómo interpretar las puntuaciones de valor agregado de las escuelas para alcanzar los objetivos. Esto debe abarcar:
  - Si las puntuaciones de valor agregado de las escuelas se clasificarán en categorías de desempeño. Si las puntuaciones de valor agregado se usarán para clasificar a las escuelas como de desempeño alto o bajo, es necesario determinar cómo se va a establecer esta clasificación, es decir, cómo se relacionan con los criterios estadísticos predeterminados y/o conceptuales válidos. Es necesario identificar los objetivos de hacer estas clasificaciones, incluso las acciones que se emprenderán una vez que la escuela se clasifique en una categoría particular. Después es necesario determinar cómo comunicar esa clasificación a la escuela y si se comunicará al público.
  - Si la información de valor agregado se va a usar internamente como herramienta para desarrollar iniciativas de mejoras escolares, esto influirá en otras decisiones (como de elección de datos y modelo), y conviene planear, asignar recursos y diseñar pronto el programa piloto para evaluar dichos objetivos.
  - Si se va a publicar la información de valor agregado, la forma de hacerlo debe desarrollarse más en la etapa piloto, pero deben establecerse los parámetros de la publicación de modo que se revisen durante el proceso piloto.
  - Cómo se va a usar la información de valor agregado en las estructuras evaluativas y mecanismos existentes mediante los cuales ya se evalúa a la escuela (p. ej., inspecciones escolares o instituciones equivalentes).

- Al determinar la medición de valor agregado en que se base el desempeño escolar, la categorización de la medición de desempeño y si se usará una variable continua, categórica o dicótoma en los modelos de valor agregado. Esto debe vincularse a las acciones que se deriven de las puntuaciones de valor agregado de las escuelas y los incentivos creados al interior de las escuelas.
- Revisar la estructura existente de las evaluaciones estudiantiles para determinar si se necesitan más evaluaciones o si necesita alterarse la estructura existente para ajustarse a los objetivos de los modelos de valor agregado.
- Establecer un marco para identificar con claridad las evaluaciones escolares particulares con que se va a medir el desempeño escolar. El marco debe permitir:
  - La identificación de las evaluaciones escolares adecuadas para los modelos de valor agregado dentro de la estructura existente de las evaluaciones estudiantiles.
  - La identificación de las materias y los niveles de grado/año en que deben tener lugar las evaluaciones.
  - La identificación del interés de las evaluaciones estudiantiles (p. ej., estándares mínimos de alfabetización o medición de desempeño continuo de todos los estándares).
  - La consideración de la forma cómo pueden afectar al desempeño escolar y a los incentivos dentro de las escuelas estas decisiones sobre la elección de evaluaciones. Por ejemplo, ¿es demasiado específico el enfoque en los conocimientos básicos de aritmética para medir el desempeño de escuelas enteras, y evaluaciones más amplias distribuirían con más equidad los incentivos para aumentar el desempeño dentro de la educación escolar?
  - La revisión y tal vez mayor desarrollo de los instrumentos de evaluación para asegurar que se empleen para los modelos de valor agregado. Es de particular importancia que la escala de las evaluaciones permita una interpretación de desempeño significativa y cambios temporales en las mediciones de desempeño con los datos longitudinales.

Fase 2: Presentación y uso de la información de valor agregado

- Dados los objetivos de políticas y la estructura de las evaluaciones estudiantiles que apoyan el sistema basado en modelos de valor agregado. Decidir sobre el método más adecuado para presentar la información de valor agregado. Esto debería de tomar en cuenta a los actores comprometidos con las políticas y mediante la retroalimentación de las escuelas piloto sobre la presentación y uso de resultados.
- Si se van a publicar los resultados de valor agregado de las escuelas, determinar qué medición (o mediciones) de valor agregado en particular se usará y cómo se presentará (p. ej., como tema único o con otra información).
- Elaborar guías para interpretar la información de valor agregado, y deben incluir la categorización de puntuaciones de escuelas con vínculos entre dichas clasificaciones y políticas y programas relacionados. Por ejemplo, esto puede incluir identificar qué puntuaciones escolares se clasificarían como de desempeño bajo o alto, y las acciones

que partan de dichas clasificaciones. Si se van a emprender acciones específicas en función de los resultados de valor agregado, deben identificarse estos "puntos críticos" y comunicarlos a los interesados.

- Para fines de rendición de cuentas de las escuelas, determinar la medición (o mediciones) específica a usar. Con el propósito de rendición de cuentas escolar, es conveniente usar una sola medición de desempeño, y debe efectuarse un análisis de las implicaciones de estas elecciones. Por ejemplo, una medición que se centre sólo en los niveles mínimos de alfabetización concentrará la atención de las escuelas, de manera tanto positiva como negativa, en materias específicas y en estudiantes de capacidades específicas de desempeño. Un interés en materias específicas proporciona incentivos semejantes. Por otra parte, una medición de valor agregado que promedie las puntuaciones de valor agregado de todas las materias oculta las discrepancias de desempeño entre las materias.
- Dados los beneficios de usar un promedio de tres años de los resultados de valor agregado de las escuelas, delinear estrategias para el uso de los datos provisionales. Esto se centraría en las acciones que se derivan de los resultados de valor agregado, en cómo se apoyan esas acciones en los datos provisionales y en cómo se publican los resultados provisionales (si esa es la intención). El uso de datos provisionales debe asegurar que se aborden los problemas de desempeño escolar o estudiantil de modo que se aligere el impacto de los retrasos inherentes de un promedio móvil de tres años.

Fase 3: Calidad de los datos

- Revisar los sistemas de datos en las escuelas y de la infraestructura más amplia para recopilar y divulgar los datos con el fin de evaluar las capacidades existentes acordes con los requerimientos de un sistema bien instalado de modelos de valor agregado. Esta revisión puede incluir una evaluación de las capacidades para el empleo de información de valor agregado en el ámbito escolar y por parte de otras instituciones (p. ej., inspecciones o actores como estatales de educación).
- Tras la revisión de los sistemas de información y de la estructura de las evaluaciones escolares presentes, quizá sea necesario establecer una base de datos completa que cumpla con los requerimientos de los modelos de valor agregado. Necesitan determinarse los requerimientos de datos de un sistema de valor agregado, y diseñarse (de ser necesario) el sistema de recopilación de datos e información proporcional. Este sistema después puede evaluarse de nuevo durante el programa piloto.
- Determinar la muestra de estudiantes a incluir en los modelos de valor agregado. Esto se centra en gran medida en identificar a las escuelas y estudiantes que necesitan identificarse y, de ser necesario, excluirse de la muestra principal. Por ejemplo, en varios sistemas se excluye de la muestra principal a las escuelas y estudiantes con necesidades especiales (aunque aún puede aprenderse mucho al calcular su valor agregado). Otro problema es asegurar que haya un sistema de registro y seguimiento de estudiantes capaz de identificar la movilidad de éstos entre escuelas entre los periodos de evaluación previo y en curso. En sistemas educativos con mecanismos de seguimiento explícito de estudiantes (p. ej., entre registros educativos académicos y vocacionales), también se necesita registrar si los estudiantes se desplazan entre estos registros educativos, pues esto puede afectar al

cálculo del valor agregado y a menudo se relaciona con problemas de carencias. Estos temas deben revisarse y desarrollarse principalmente durante la etapa piloto del proceso de implementación. Esta revisión incluiría análisis de valor agregado del desempeño de subgrupos específicos de la población con el fin de evaluar si deben incluirse con la muestra principal al estimar las puntuaciones de valor agregado de las escuelas.

- Efectuar un análisis del uso de características socioeconómicas contextuales específicas en los modelos de valor agregado. Esto dependerá de los objetivos generales del sistema y del modelo empleado, lo que también recibirá la influencia de la cantidad y frecuencia de evaluaciones estudiantiles y de la distribución de desempeño general de las escuelas.
- Establecer si los requerimientos de datos y el sistema de información soportarán sólo los modelos de valor agregado o también las instituciones (sobre todo escuelas) que usarán la información para poner en marcha las políticas y programas determinados. Tal vez se necesiten una base de datos y un sistema de información más completos para admitir a los usuarios adicionales y el desarrollo del programa en el futuro.

Fase 4: Elección de un modelo adecuado de valor agregado

- El programa piloto se emplea para evaluar la validez de distintos modelos de valor agregado. Necesitarán estimarse varios modelos de valor agregado a partir de los datos obtenidos en la fase piloto (con datos de las evaluaciones estudiantiles previas cuando sea posible). La fase piloto sirve después para evaluar las ventajas y desventajas de distintos modelos de valor agregado y, por ende, brindar información para elegir el modelo más adecuado.
- Al elegir un modelo, es importante identificar cómo guiarán la elección del modelo los objetivos de políticas y el uso propuesto de las puntuaciones de valor agregado de las escuelas. Necesitan expresarse ciertos objetivos de políticas antes de analizar los diversos modelos. Esto incluye identificar la forma de la variable dependiente, cómo se usará la información de valor agregado y si se generarán categorías de desempeño escolar.
- Es necesario identificar los criterios estadísticos y metodológicos con que se analizarán los distintos modelos de valor agregado. El análisis con los datos piloto durante la fase de implementación se concentraría en:
  - La varianza de cada modelo de valor agregado. Esto debe analizarse para evaluar lo adecuado de los modelos particulares. Quizá se prefieran modelos específicos si pueden identificar una mayor cantidad de escuelas estadísticamente diferentes del promedio o algunos criterios predeterminados.
  - El uso de datos socioeconómicos contextuales y las funciones de distintos componentes de datos en un análisis de valor agregado. Deben efectuarse análisis para evaluar el impacto de la inclusión y exclusión de características específicas en las puntuaciones de valor agregado de las escuelas y la estimación de valor agregado.
  - El sesgo potencial en el modelo que necesita analizarse (así como su posibilidad de reducción) durante la fase piloto de implementación. Se analiza la importancia de los datos perdidos, y han demostrado su conveniencia las comparaciones con datos y análisis existentes.
  - Las suposiciones sobre datos perdidos. Esto se evalúa respecto de los resultados que resulten en la recopilación de datos piloto. Después deben desarrollarse procedimientos para reducir la frecuencia de los datos perdidos.

  - Los estimados de valor agregado de escuelas pequeñas. Esto se pone a prueba y se recomienda tanto el análisis como la presentación de los resultados escolares.
  - La estabilidad de las puntuaciones de valor agregado de las escuelas y cómo se ven afectadas por la clasificación del desempeño escolar y la elección del modelo específico de valor agregado. En estos análisis es importante considerar no sólo el nivel general de estabilidad, sino los cambios en puntuaciones escolares individuales. Después se efectúan análisis de las causas de tal inestabilidad y para identificar si algunas escuelas en particular son más susceptibles de inestabilidad en sus resultados escolares.
- Es importante analizar el impacto de diferentes modelos con los objetivos de políticas prescritos y el uso planeado de los datos. Es decir, es importante analizar el impacto de la elección de modelo en distintas escuelas por el uso planeado de estas puntuaciones. Este análisis no sólo debe centrarse en el modelo general (p. ej., adecuación), sino también en el impacto para escuelas individuales. Esto conformaría la base de la recomendación de un modelo preferido de valor agregado en un informe piloto.

Fase 5: Estrategias de comunicación y compromiso con el interesado

- Idear una estrategia de comunicación y compromiso con actores clave que lo incluya en el desarrollo del sistema basado en modelos de valor agregado. Se elabora una estrategia de comunicación que exprese con claridad los objetivos y lógica del sistema, los modelos de valor agregado que se emplean, y el uso e interpretación de los resultados de valor agregado de las escuelas.
- Alinear el énfasis de la estrategia de comunicación con los objetivos de políticas. Deben describirse con claridad las mediciones con que se juzgará el desempeño escolar, y mencionarse las consecuencias para los diversos niveles de desempeño escolar.
- Para el análisis en el ámbito escolar, contar con la infraestructura adecuada, y elaborar paquetes de guías e información para directores y maestros referentes a la manera de interpretar la información de valor agregado y de emplearse con fines de mejoras escolares. Preparar información semejante para los padres y los medios.

Fase 6: Formación

- Delinear programas de formación que se dirijan a los usuarios específicos. La formación para directores y maestros se concentraría en la manera en que se derivan los resultados de valor agregado y en que se emplean al interior de las escuelas con fines de mejoras escolares. Esto puede incluir formación en análisis estadístico y en el uso del sistema de información requerido. La retroalimentación de los interesados durante el programa piloto debe facilitar un mayor refinamiento de los programas de formación y destacar áreas de importancia para maestros y directores.
- Dirigir la formación para padres y familias a la interpretación de las puntuaciones de valor agregado presentada al público en general para facilitar su comprensión. La publicación de resultados escolares puede inducir una forma de responsabilidad par parte de los padres. Pueden hacerse disponibles paquetes de formación e información que describan cómo se calculan los resultados y qué significan en términos de desempeño escolar y la educación que reciben los alumnos. Esta formación también puede hacerse disponible a los medios y expertos en educación.

Fase 7: Programa piloto

- Estructurar el programa piloto de modo que permita a las autoridades educativas evaluar y desarrollar todos los aspectos del sistema basado en modelos de valor agregado y las políticas y programas correspondientes en torno al uso de la información de valor agregado. Esto incluye:
  - Los asuntos de operación e implementación, desde la implementación de evaluaciones de alumnos hasta la recopilación, análisis y divulgación de datos y otra información de valor agregado. Si bien un programa piloto a menudo se efectúa con una muestra de escuelas, algunos sistemas educativos tendrán acceso a datos de evaluación estudiantil exhaustivos. De ser posible, conviene iniciar el programa con el conjunto de datos completo para evaluar la infraestructura requerida, en particular si se diseñó durante la fase piloto. Estimar el valor agregado del conjunto de datos completo también facilitaría el análisis de la elección de modelo.
  - Aunque no es el interés principal de este informe, el programa piloto debe servir para el análisis de lo adecuado de las evaluaciones escolares empleadas.
  - Las estimaciones sobre los datos piloto ofrecen el análisis requerido para elegir la especificación más adecuada del modelo de valor agregado al evaluar diferentes modelos respecto de los criterios predeterminados.
  - Si se decide que las puntuaciones de valor agregado se convertirán en categorías específicas de desempeño, se evalúa la aplicabilidad del esquema de clasificación. Si se van a elegir categorías específicas (p. ej., escuelas de bajo desempeño) con base en criterios específicos, la cantidad de escuelas que entren en cada categoría de desempeño se estima con los modelos de valor agregado en consideración.
  - El desarrollo de las estrategias de comunicación y compromiso de actores clave que participen en una revisión de estrategias existentes dentro de las escuelas. Debe incluirse la información de directores, maestros y otros interesados en estas revisiones para trabajar sobre la eficacia de diversas iniciativas y para un mejor desarrollo de las estrategias de comunicación y compromiso. Las opiniones de estos grupos también ayudarían a los tomadores de decisiones a determinar los temas clave por incluirse en un sistema de control de calidad de la implementación.

- Al realizar el programa piloto, alinear las decisiones sobre el tamaño y características de la muestra de escuelas con los objetivos del sistema general de modelos de valor agregado. Esto requiere identificar áreas prioritarias, como escuelas en comunidades desfavorecidas, y generar el marco muestral adecuado.

- Emplear el programa piloto para evaluar las acciones vinculadas a los resultados de los modelos de valor agregado. Las acciones incluyen la clasificación de escuelas en categorías de desempeño, la asignación de recompensas y sanciones, el desarrollo de iniciativas específicas y la aplicación de evaluaciones adicionales. Identificar cómo se ejecutarán dichas acciones, con base en las puntuaciones específicas de valor agregado establecidas (de ser así) y las acciones correspondientes delineadas.

- Incluir en el programa piloto un informe o una serie de recomendaciones basadas en los hallazgos y experiencia de efectuar el programa piloto. Esto destacaría los problemas que necesitan abordarse antes de la implementación real. Este informe también puede incluir los resultados del análisis del modelo de valor agregado más adecuado y una evaluación del impacto en los actores clave. Asimismo debe informar sobre las áreas principales que deben ser el centro de atención de un sistema de control de calidad empleado durante la implementación real del sistema basado en modelos de valor agregado.

Fase 8: Desarrollo continuo

- Establecer un sistema de control de calidad con recursos adecuados que se centre en los datos recopilados, las capacidades del sistema de información utilizado, los modelos de valor agregado empleados, las políticas y programas que se supone promueven, y el impacto en los actores clave.
- Analizar con este sistema de control de calidad no sólo los resultados generales de los modelos de valor agregado, sino también los resultados de las escuelas individuales para asegurar que el modelo aún avanza hacia los objetivos de políticas deseados. Este sistema subrayaría puntuaciones específicas de la escuela (p. ej., las de menos estabilidad en el curso de los años) y analizaría diversos problemas de muestreo y datos. Asimismo revelaría problemas de evaluación que ameriten atención.
- Efectuar análisis para el desarrollo continuo del modelo (o modelos) de valor agregado en uso. Esto pretendería mejorar el "ajuste" de la especificación y adecuarse a cualquier cambio de datos u objetivos de políticas. Si se hacen cambios al modelo subyacente de valor agregado, debe analizarse el impacto en las escuelas.

NOTAS

1. Es importante resaltar que Tennessee también emplea modelos de valor agregado para obtener estimaciones del valor agregado de los maestros, pero un análisis de esa aplicación trasciende el alcance de este informe.

2. Ese sistema ahora se conoce como Sistema de Evaluación de Valor Agregado Educativo (*Education Value Added Assessment System,* EVAAS), y es aplicado por el profesor Sanders y sus colegas bajo los auspicios de una empresa privada. En Braun (2006b) se presenta una descripción abreviada del EVAAS.

3. Las gráficas y las descripciones se tomaron con permiso de la presentación de la doctora Mary Reel, Directora de Valoración & Evaluación, del Departamento de Educación de Tennessee, durante el *ETS National Forum on State Assessment and Student Achievement,* de la *Educational Testing Service,* celebrado del 13 al 15 de septiembre de 2006, Princeton, New Jersey, EUA.

4. Hay más información en la sección sobre Escuelas Especializadas en Alto Desempeño: *http://www.standards.dfes.gov.uk/specialistschools/.*

5. Informe del Indicador de Preinspección de Contexto y Escolar (*Pre-inspection Context and School Indicator [PICSI] Report*).

6. Los informes de inspección están en *http://www.ofsted.gov.uk/reports/.*

Parte II

El diseño de modelos de valor agregado

Como ya se mencionó, en este informe, el término ***modelos de valor agregado*** se refiere a una clase de modelos estadísticos que estiman las contribuciones relativas de las escuelas al progreso escolar respecto de los objetivos educativos declarados o prescritos (p. ej., logro cognitivo) medido en al menos dos momentos. En la medida en que dicho avance es un resultado deseable de la escolaridad, los modelos de valor agregado pueden ofrecer por tanto una fuente valiosa de información. De hecho, como aclara la Parte I, el resultado de los modelos de valor agregado es útil de muchas maneras para autoridades educativas y para funcionarios escolares. Hoy en día se aplican muchos modelos de valor agregado, cada uno con sus propias ventajas y desventajas. La Parte II de este informe identifica los aspectos básicos del diseño de modelos de valor agregado y después presenta las descripciones de algunos de los más comunes. Más adelante se analizan diversos aspectos estadísticos y metodológicos para ayudar a las autoridades educativas y a los administradores en el diseño de modelos de valor agregado y en la elección del modelo más adecuado para el desarrollo escolar y para supervisar el avance hacia objetivos especificados en su sistema educativo.

Este informe distingue entre los ***modelos de valor agregado*** y los ***modelos de logro contextualizado***. Los primeros siempre emplean al menos una medición previa de logro académico relevante como base para tomar en cuenta diferencias en los alumnos inscritos entre las escuelas. Por otra parte, los modelos de logro contextualizado no incorporan mediciones de logro anteriores. La Parte II presenta algunos resultados empíricos relacionados con las ventajas de incorporar datos de exámenes anteriores a las estimaciones de la eficacia escolar. No existe un acuerdo universal sobre el conjunto de modelos estadísticos que se puedan denominar "de valor agregado". Por ejemplo, si se dispone de dos puntuaciones de pruebas por estudiante (p. ej., puntuaciones de matemáticas en grados consecutivos), y las puntuaciones se expresan en una escala común, es posible calcular la diferencia (es decir, la puntuación de ganancia individual). La ganancia promedio en los alumnos inscritos se considdera como medida del valor agregado de la escuela. Además, la diferencia entre las puntuaciones de ganancia promedio entre escuelas, o la diferencia entre la puntuación de ganancia promedio de una escuela y la media de todas las puntuaciones de ganancia

promedio de todas las escuelas, puede tratarse como medición de la eficacia relativa de la escuela. Estos modelos tienen propiedades estadísticas problemáticas porque los ajustes para estudiar la variación entre escuelas debida a las condiciones de entrada de los estudiantes son débiles. Por esta razón ya no se tratarán en este volumen. Sin embargo, el lector debe ser consciente de que los modelos de puntuaciones de ganancia han sido discutidos en la literatura correspondiente.

¿Cuáles son las bases del análisis de valor agregado? Para comenzar, los datos de las puntuaciones de las pruebas de una gran cantidad de escuelas se recopilan y organizan de acuerdo con los requerimientos del modelo empleado. Al menos, la base de datos debe contener, por cada estudiante: la identificación de su escuela; puntuaciones estandarizadas en al menos dos ocasiones consecutivas; información sobre variables demográficas y de contexto del estudiante.* Una vez que el modelo se aplica a los datos, el resultado es un conjunto de números, uno por escuela. Estos números representan una función semejante a la de los residuos de una regresión común. Es decir, representan la parte del resultado de la escuela (la puntuación promedio de los estudiantes) que no puede atribuirse a las diversas variables explicativas incluidas en el modelo. Como los residuos, estos números promedian cero. El número adjudicado a una escuela determinada se interpreta de manera provisional como medida del desempeño relativo de la escuela; es decir, se toma como la estimación de la diferencia entre la contribución de la escuela al aprendizaje de sus estudiantes y la contribución promedio al aprendizaje de los estudiantes de todas las escuelas de donde provienen los datos. Por tanto, estos números son estimaciones del valor agregado de la escuela. Supongamos, por ejemplo, que el análisis se centra en el desempeño escolar en una prueba determinada. Por construcción, el residuo o estimación de valor agregado para el promedio de las escuelas es cero. En consecuencia, una estimación del valor agregado positiva significa que la escuela correspondiente parece haber realizado una contribución superior al promedio, y una estimación negativa significa que la escuela correspondiente parece haber hecho una contribución inferior al promedio. En este último caso, aún es posible, incluso probable, que los estudiantes de dicha escuela hayan obtenido puntuaciones de ganancia positivas durante el periodo en estudio.

En el ejemplo anterior es importante reconocer que la estimación del valor agregado de una escuela depende de las escuelas que se incluyan en el estudio, pues las estimaciones del valor agregado se definen de manera relativa. Es decir, el modelo pretende registrar las diferencias de resultados entre las escuelas relacionadas con las diferencias en las características de los estudiantes entre las escuelas. El modelo ajustado, y su éxito al explicar la varianza de los resultados, estará determinado por los datos de las escuelas que se empleen. El uso de otro conjunto de escuelas llevará a un ajuste del modelo distinto. La diferencia entre el resultado de una escuela y lo que se prevería a partir del modelo ajustado (es decir, el resultado promedio) se denomina valor agregado de la escuela, pues esa es la parte del resultado que no se explica por las características medidas de los estudiantes. Como se señala en el párrafo anterior, las estimaciones de valor agregado así definidas son simples residuos de un

* Si bien la mayoría de los modelos de valor agregado emplean datos que no provienen de pruebas, algunos no lo hacen. El ejemplo más conocido es el modelo del Sistema de Evaluación de VA Educativo (EVAAS).

modelo de regresión, y por tanto, se dice que se definen de manera relativa. La noción de un indicador de desempeño escolar definido respecto de una recopilación determinada de escuelas contrasta con los indicadores basados en ganancias de puntuaciones, que suelen definirse de manera absoluta. Esto no es una desventaja, pero debe tenerse presente al interpretar los resultados de valor agregado. En muchas aplicaciones, el interés se dirige a las escuelas cuyas contribuciones estimadas son sustancialmente distintas del promedio (es decir, muy positivas o muy negativas). Con este fin, la mayoría de los modelos de valor agregado también generan un error estándar estimado del valor agregado estimado de la escuela. La proporción entre el estimador del valor agregado y su error estándar sirve para determinar si la estimación para la escuela es estadística y significativamente distinta del promedio. Desde luego, desde el punto de vista político, la significación estadística debe considerarse junto con su pertinencia práctica.

Las estimaciones de valor agregado de las escuelas se calculan por separado para cada grado o año, y, de ser así, son especialmente útiles para fines diagnósticos. Sin embargo, para efectos de síntesis, se calcula un indicador compuesto de valor agregado de una escuela promediando las estimaciones de valor agregado de los distintos grados de la escuela. Aunque es una medida conveniente, se recomienda que las escuelas con grados muy distintos no se comparen entre sí utilizando estos estadísticos de síntesis, pues las propiedades estadísticas de los estimadores de valor agregado pueden variar de un grado a otro. Si bien las estimaciones de valor agregado suelen denominarse "efectos (estimados) de la escuela", debe tenerse en mente que incluso en las mejores circunstancias estos efectos de la escuela estimados sólo se pueden aproximarse a las "verdaderas" contribuciones de la escuela sobre las puntuaciones de ganancia de los alumnos en las pruebas. El término "efecto" se toma de la literatura estadística, y por lo general no implica una atribución causal. Es igualmente importante señalar que los análisis estadísticos por sí solos no descubren las razones de las diferencias (aparentes) del desempeño escolar. Estas explicaciones requieren visitas *in situ* y la recopilación de información cualitativa más amplia sobre las actividades de enseñanza y de aprendizaje en la escuela. Por último, las escuelas tienen muchas otras metas además de mejorar las puntuaciones en las pruebas. Así, las evaluaciones escolares deben tener en cuenta una amplia variedad de indicadores que abarque, aunque sin restringirse, a medidas de valor agregado basadas en pruebas.

Como indicamos al principio, los modelos de valor agregado pretenden estimar las contribuciones de la escuela al aprendizaje de los alumnos individuales. La palabra "contribución" denota el papel que juegan las escuelas en la generación del resultado de interés (por ejemplo, el aumento de las puntuaciones de las pruebas como medida del avance de los alumnos en el aprendizaje), tomando debidamente en cuenta la influencia de otros factores relacionados con este resultado. Así, la intención es dotar a las estimaciones de los modelos de valor agregado de una interpretación causal. Es decir, suele interpretarse que la diferencia entre las contribuciones estimadas de dos escuelas refleja diferencias en su eficacia para promover el aprendizaje de los alumnos. Es comprensible que los tomadores de decisiones deseen efectuar estas inferencias causales basadas en un análisis estadístico. Si de verdad se aislase la contribución de una escuela, se tendría una base sólida para emprender acciones de varios tipos. Sin embargo, por la clase de datos de que suele disponerse y la

realidad de la limitada asignación de estudiantes en las escuelas, las inferencias causales pueden ser problemáticas. Por lo común, las inferencias causales provienen de grandes experimentos aleatorios, como los que suelen realizarse en agricultura o medicina. En la versión más sencilla, hay dos grupos: un grupo control y uno experimental. Las unidades individuales se asignan al azar a uno de los dos grupos. Las unidades del primer grupo reciben un tratamiento estándar (o placebo), y las unidades del segundo grupo, el tratamiento de interés. La diferencia entre los resultados promedio de ambos grupos es una medida de la eficacia relativa del tratamiento de interés en comparación con el estándar. El uso de la aleatorización junto con el tamaño grande de las muestras reduce la probabilidad de que una diferencia sustancial en los resultados se deba a alguna combinación de fluctuaciones azarosas y la acción de factores no observados.

Los modelos de valor agregado son un esfuerzo por capturar las virtudes de un experimento aleatorio cuando no se ha efectuado ninguno. En ambientes educativos, los estudiantes casi nunca se asignan al azar a las escuelas, sino que la geografía y el costo son los dos mayores determinantes. Así, los datos escolares se consideran producto de un estudio observacional y no de un experimento estadístico. Por esta razón, las comparaciones directas de escuelas en términos de puntuaciones de ganancia promedio pueden ser engañosas. Como veremos más adelante, la mayoría de los modelos de valor agregado adoptan una estrategia más elaborada para informar de las puntuaciones de ganancia ajustadas por las diferencias entre diversas características de los alumnos. Estos ajustes pretenden considerar las diferencias en las poblaciones de alumnos entre escuelas que quizá puedan estar relacionadas con dichas ganancias. La idea es aislar la contribución relativa de la escuela por sí misma (su personal, políticas y recursos) sobre el aprendizaje de los estudiantes.

El empleo adecuado de los modelos de valor agregado descansa en la comprensión de la distinción entre descripción estadística e inferencia causal (Rubin, Stuart y Zanutto, 2004). Suponga, por ejemplo, que la ganancia promedio de los estudiantes durante un año en la Escuela Alfa es de 8 puntos, y la ganancia promedio de los alumnos de la Escuela Beta es de 12 puntos. Esa es una descripción. Sin embargo, como resultado de la aplicación de un modelo particular de valor agregado, obtenemos "efectos de la escuela" estimados, mismos que podemos tratar como indicadores del desempeño escolar relativo. Por ejemplo, suponga que el efecto asociado a la Escuela Alfa es 2, y el efecto asociado a la Escuela Beta es 5 (observe que el efecto estimado de la escuela por lo general será numéricamente distinto de la simple ganancia media de la escuela). La interpretación deseada de estos efectos es que si los estudiantes de la Escuela Alfa se hubiesen inscrito en la Escuela Beta, su ganancia promedio habría sido de $5 - 2 = 3$ puntos mayor. Es decir, los resultados del análisis de valor agregado reciben una interpretación causal.

Sin embargo, la transición de la descripción a la inferencia estadística está llena de dificultades, pues los alumnos de la Escuela Alfa no se inscribieron en la Escuela Beta. Además, no se asignó a los estudiantes inscritos en las escuelas Alfa y Beta de manera aleatoria a estas escuelas, sino que se inscribieron mediante diversas elecciones individuales. Así, aquí no se cumplen las condiciones de un experimento aleatorio. Interpretar las diferencias de los efectos escolares estimados como diferencias de eficacia escolar requiere suponer que la aplicación del modelo tuvo en cuenta todas las diferencias pertinentes entre los alumnos de

ambas escuelas. Por desgracia, muy pocas veces observamos o controlamos los factores que determinan la elección de una escuela. Si hay factores sin observar que son determinantes tanto para la elección de escuela como para el aprovechamiento, la interpretación causal directa será difícil, pues no se ha probado de manera adecuada otras hipótesis contrarias y posibles. De hecho, es el papel de las posibles explicaciones alternativas lo que distingue la inferencia causal de la descripción simple.

En realidad, se distinguen al menos dos tipos de inferencia causal en estos entornos (Raudenbush y Willms, 1995; Raudenbush, 2004). El primero, llamado efecto Tipo A, se relaciona mucho con lo recién descrito y sse relaciona con la situación en que a los padres les interesa elegir la escuela en la que a sus hijos les vaya mejor. Pueden obtener una respuesta aproximada al ver a niños semejantes a su hijo en cada escuela y después determinar qué grupo obtuvo mejores resultados. La diferencia en el desempeño sería el efecto Tipo A en este caso. Si bien la superioridad observada en el desempeño quizá se deba en parte a diferencias inadvertidas entre los dos grupos, no existe una razón para no preferir la escuela en apariencia más eficaz. Sin embargo, el efecto Tipo A no es un instrumento adecuado para evaluar el desempeño ni la rendición de cuentas de la escuela. La razón es que la diferencia media de desempeño entre escuelas puede deberse a una combinación de diferencias entre los contextos en los que las escuelas operan y las diferencias entre las prácticas escolares. Raudenbush y Willms (1995) definen "contexto escolar" como los factores sobre los cuales los educadores tienen poco control, como la composición demográfica de la escuela y el ambiente de la comunidad en donde funciona la escuela. Definen "práctica escolar" como el conjunto de las estrategias instructivas, las estructuras organizativas y actividades de liderazgo de la escuela, que, en principio, están bajo control del personal escolar. Aunque los padres manifiesten relativa indiferencia ante las contribuciones relativas de ambos componentes, Raudenbush y Willms (1995) sostienen que los administradores y las autoridades educativas deben interesarse más en las contribuciones de la práctica escolar, pues por lo general están bajo control del personal de la escuela. Así, a los administradores y las autoridades correspondientes les agradaría diferenciar las contribuciones del contexto escolar y de la práctica escolar sobre las ganancias de los alumnos y aislar la diferencia de desempeño debida a las diferencias en las prácticas escolares. Esto constituiría el efecto Tipo B.

Aparte de cierta ambigüedad respecto de lo que debe clasificarse como práctica escolar, Raudenbush y Willms (1995) observan que es esencialmente imposible obtener estimaciones no sesgadas de los efectos Tipo B a partir de los datos de un sistema escolar estándar. Incluso los efectos Tipo A son perfectamente estimables sólo en circunstancias ideales, muy poco probables de ver en la práctica (para más discusión sobre los problemas de obtener estimaciones no sesgadas de las contribuciones escolares sobre aprendizaje de los estudiantes véase McCaffrey *et al.,* 2003; Braun, 2005a; van de Grift, 2007). Si bien estos problemas pueden ser desalentadores, es necesario hacer notar que cualquier indicador empírico de desempeño escolar es falible, pues está sujeto a variabilidad y sesgos. En realidad, el análisis de valor agregado se ha estudiado con más rigor que otras estrategias, como las visitas de inspección y otras similares. En consecuencia, cuando se aplica e interpreta bien, un análisis de valor agregado genera un indicador del nivel escolar que, junto con otros indicadores, otorga un retrato informativo del funcionamiento escolar. De hecho, como las

estimaciones de valor agregado tienen una base empírica distinta de la mayoría de los demás indicadores, representan una contribución particularmente valiosa para la agenda de revisión del desempeño escolar. El análisis de valor agregado sirve como la primera etapa de un proceso en el que, por ejemplo, se examinen las relaciones entre las estimaciones de valor agregado y varias características escolares, con el fin de evaluar patrones útiles o novedosos. Es importante señalar que la utilidad de las estimaciones de valor agregado es sustancialmente mayor que la de las medidas de desempeño escolar basadas en la comparación de puntuaciones brutas de pruebas empleadas en algunos países miembros de la OCDE (OCDE, 2007a), o incluso que los resultados de los modelos de logro contextualizados. Este informe subraya la mayor credibilidad de las estimaciones de valor agregado. No obstante, es crucial analizar las advertencias y suposiciones aplicables al emplear modelos de valor agregado para avanzar hacia los objetivos de las políticas educativas.

CAPÍTULO 4

Consideraciones de diseño

Datos de la evaluación de alumnos **99**
- Concepto de validez **100**
- Error de medición **102**
- Escalas de puntuaciones **103**
- Resultados de evaluación en una escala ordinal **104**
- Estructura de evaluaciones de los alumnos en países participantes **106**

Filosofía del ajuste y características contextuales **110**
- Importancia de las características contextuales **111**
- ¿Qué características contextuales socioeconómicas? **114**
- Datos del nivel de escuela **119**

El diseño de un aparato, sea un modelo estadístico o una casa, está definido por el uso que se le pretende dar, los recursos disponibles y las limitaciones relevantes. A esta mezcla debe añadirse la experiencia del diseñador con aparatos semejantes o relacionados. En el contexto de los modelos de valor agregado, hay varios factores de diseño básicos, como calidad de los datos, integridad y cobertura de los datos, filosofía del ajuste estadístico, complejidad técnica, transparencia y costo. A continuación abordaremos cada uno.

1. ***Evaluación de los estudiantes y calidad de datos de las pruebas.*** Como los modelos de valor agregado trabajan con datos generados por evaluaciones de los estudiantes, se debe considerar en primer lugar la naturaleza y calidad de dichos datos. En particular, ¿reflejan adecuadamente los datos lo que los alumnos saben y pueden hacer respecto de las metas curriculares establecidas? Esta es la esencia de la validez de las puntuaciones de las pruebas, y debe abordarse de diversas maneras. Las cuatro preguntas más importantes son: ¿ofrecen las pruebas evidencia respecto de todas (o al menos de las más importantes) metas curriculares?, ¿todos los alumnos se someten a examen en condiciones comparables?, ¿las puntuaciones son lo bastante precisas para apoyar las inferencias pretendidas?, y ¿las puntuaciones de las pruebas están protegidas contra influencias inadecuadas y/o corrupción? Si las respuestas son afirmativas, se puede considerar emplear los modelos de valor agregado.
2. ***Integridad y cobertura de los datos.*** Deben evaluarse con cuidado los procedimientos con que los datos brutos de las pruebas se transforman en archivos de datos útiles, así como su carácter exhaustivo. Por lo general se necesitan registros de los alumnos de dos o más años para los modelos de valor agregado, y no es poco frecuente que falten algunas puntuaciones en los archivos de datos longitudinales debido a registros inexactos, ausencias de alumnos y emigración o inmigración. En general, cuanto mayor sea la proporción de datos perdidos, más débil será la credibilidad de los resultados. Además, algunos modelos de valor agregado emplean datos de pruebas de múltiples materias y/o datos auxiliares derivados de características del alumnado (p. ej., género, raza/origen étnico, estado socioeconómico). De nuevo, deben evaluarse la integridad y el alcance de estos datos.
3. ***Filosofía del ajuste.*** Los modelos de valor agregado difieren en el grado en el que se realizan ajustes de las características de los alumnos. Para algunas clases de modelos, estos ajustes son la base principal para tratar las estimaciones como indicadores de las contribuciones causales de las escuelas. Al hacer ajustes, debe tenerse cuidado con la elección de características, pues las que se miden con errores también introducen sesgos. Esto puede ocurrir al realizar ajustes con características que puedan haber recibido parcialmente la influencia de políticas escolares pudiendo introducir sesgos indeseados en las estimaciones de desempeño escolar. Algunas de estas características son las actitudes del alumnado hacia la escuela o la cantidad promedio de tarea semanal. En otras clases de modelos, cada estudiante se emplea como su propio "control" y, por tanto, los modelos no incorporan ajustes explícitos. En cambio, aprovechan la covariación de los datos de pruebas recopilados en múltiples materias y muchos años, o incorporan al estudiante como "efecto fijo". Más adelante abundaremos en estas variantes.
4. ***Complejidad técnica.*** Los modelos de valor agregado ahora varían desde modelos de regresión simple a modelos en extremo complejos que requieren bases de datos amplias y procedimientos computacionales de última generación. En general, puede argumentarse

que los modelos más complejos ofrecen mejores estimaciones de desempeño escolar sin influencias de factores distractores, si bien aún hay polémica al respecto. La desventaja es que, por lo común, mientras mayor sea el nivel de complejidad, mayores serán los requerimientos de personal y de tiempo para establecer y validar el sistema. Los modelos más complejos suelen requerir datos más completos (años y materias), de modo que la disponibilidad de los datos limita la complejidad de los modelos que pueden ser considerados. Además, las mayores dificultades para comunicar los resultados y el uso de los modelos más complejos puede reducir la transparencia del sistema y aumentar los problemas cuando se quiere obtener el apoyo de los interesados.

5. ***Transparencia.*** Si bien la noción de "valor agregado" es atractiva por intuición, su introducción en ambientes escolares puede generar controversia, en particular si los motivos para introducirla se ven con sospecha entre algunos interesados. Si es relativamente sencillo explicar el funcionamiento del modelo en lenguaje no técnico, muchas de esas sospechas se disipan. Por otra parte, si el modelo de valor agregado se presenta como una "caja negra" a cuyo funcionamiento interno sólo puede acceder un grupo elitista de tecnócratas, será más difícil obtener la aceptación general. Los modelos más sencillos suelen ser más transparentes y, en consecuencia, más favorecidos por razones políticas, aunque sean menos deseables técnicamente.

6. ***Costo.*** La proporción más grande del costo se va en la recopilación de datos y en la elaboración de una base de datos útil. La primera por lo general se asigna al presupuesto de instrucción, pues las puntuaciones de pruebas tienen fines académicos. No obstante, la elaboración y mantenimiento de una base de datos adecuada puede ser considerable, por representar el costo de introducir un sistema nuevo de indicadores de desempeño escolar, pudiendo necesitar llegar (y proporcionar formación) a varios interesados. Los costos reales de funcionamiento del modelo, con los análisis secundarios y producción de informes, son relativamente modestos, en especial después de uno o dos años. Sin embargo, las consideraciones y magnitudes de costos varían sustancialmente por país. Los aspectos relacionados con los costos e implementación de sistemas basados en modelos de valor agregado se analizan en la Parte III de este informe, que se dedica a los asuntos de la implementación.

Los dos primeros problemas son los ladrillos esenciales para elaborar un sistema basado en modelos de valor agregado. Se analizan a continuación para la identificación de aspectos básicos que deben enfrentar los administradores y las autoridades educativas al elaborar una base de datos eficaz para la modelización del valor agregado. Los problemas tercero y cuarto se analizan después, donde se abordan consideraciones estadísticas y metodológicas. Sin embargo, por la importancia de estos asuntos, también se analizan en otras áreas de este informe, en particular en los Capítulos 5 y 6, donde se introducen varios tipos de modelos de valor agregado. Los problemas quinto y sexto recién citados se tratan en este informe como asuntos de presentación e implementación.

DATOS DE LA EVALUACIÓN DE ALUMNOS

Este informe no se detiene demasiado en el desarrollo de instrumentos de evaluación que se usan con los modelos de valor agregado. El interés de este informe se centra en el desarrollo y

empleo de modelos de valor agregado. La literatura sobre evaluación educativa y las decisiones básicas necesarias para el desarrollo de instrumentos de evaluación es muy extensa. Esta literatura describe los diversos métodos utilizados para evaluar el razonamiento general y las competencias específicas por materia. Este informe no evalúa esta literatura; sin embargo, la siguiente discusión sí aborda algunas decisiones referentes al marco evaluativo que influye en el desarrollo de modelos de valor agregado, así como la forma en que las escuelas, administradores y autoridades correspondientes usan los resultados. También se analizan los marcos de evaluación de alumnos de los países participantes para ilustrar los diversos modos de abordar estas cuestiones. Parece claro que en la mayoría de los sistemas educativos no se desarrolló un marco de evaluación de alumnos con el objetivo explícito de generar datos para modelos de valor agregado. En cambio, los modelos de valor agregado se desarrollaron para aprovechar los datos generados por las evaluaciones escolares existentes. El análisis del diseño del marco evaluativo debe ofrecer información a las autoridades educativas y los administradores en su esfuerzo por elaborar evaluaciones que mejoren la utilidad de un sistema basado en modelos de valor agregado.

En algunos países, el desarrollo e implementación de un currículo nacional se acompañó de la elaboración de un marco evaluativo y de su correspondiente conjunto de evaluaciones. Los resultados de estas evaluaciones servirían como entradas para distintos tipos de modelos de valor agregado. También es posible aplicar modelos de valor agregado a los datos provenientes de pruebas estandarizadas que se administran en múltiples jurisdicciones con currículos distintos. Sin embargo, así se complica el desarrollo de estas pruebas y la interpretación de los resultados de los modelos de valor agregado. En el diseño de la prueba estandarizada puede haber problemas de sesgos cuando la evaluación se alinea de forma más estrecha con un currículo que con otro. También hay dificultades al estimar las contribuciones de la escuela al avance del alumnado si se utilizan datos de una evaluación que no se relacione estrechamente con el currículo al que se supone se ajustan las escuelas o al que dedican sus recursos. Puede ser problemático interpretar los resultados de los modelos de valor agregado en este contexto. En muchos países con un sistema federal, el currículo se diseña en el ámbito subnacional y por ende puede diferir sustancialmente por regiones. Por tanto, para evitar estas dificultades puede ser prudente aplicar los modelos de valor agregado separadamente dentro de cada jurisdicción subnacional. Asimismo pueden obtenerse ventajas políticas e institucionales al emplear los modelos de valor agregado para supervisar y dar información del desarrollo del sistema en el mismo ámbito administrativo en el que residen las principales responsabilidades de toma de decisiones. Como es natural, estas consideraciones varían en función del país respecto de la naturaleza del sistema nacional, así como de la estructura jerárquica de la toma de decisiones educativas en dichos países.

Concepto de validez

Las puntuaciones de las pruebas son la materia prima de un análisis de valor agregado y, sin duda, las propiedades de estas puntuaciones serán críticas en la calidad de los efectos escolares estimados resultantes. Muchos análisis dependen de la suposición de que las puntuaciones son "lo bastante buenas", sin especificar lo que implica esto ni efectuar investigaciones empíricas sobre la manera de determinar estas puntuaciones. Quizá la suposición de un carácter adecuado se basa en que, en la mayoría de los casos, las puntuaciones de pruebas se emplean

sobre todo para tomar decisiones sobre los alumnos, y sólo después para estudios de eficacia escolar. No obstante, es sin duda adecuado revisar las características deseables de los datos de las puntuaciones de las pruebas en el contexto de un análisis de valor agregado. Como señala la discusión inicial de este capítulo, debe establecerse la validez y la confiabilidad de la prueba de evaluación del aprovechamiento académico. Los dos riesgos principales para la validez son las deficiencias en la representación del constructo y los altos niveles de varianza irrelevante del constructo (Messick, 1989).

Respecto del primer riesgo, la preocupación principal está relacionada con las pruebas mal diseñadas o que abordan sólo algunas metas de aprendizaje, o que tienen un énfasis temático inadecuado. Por lo general, esto sucede por una falta de experiencia entre quienes elaboran las pruebas y/o limitaciones financieras que restringen los tipos de preguntas incluidas en la prueba. Por ejemplo, muchas pruebas estandarizadas constan sólo de preguntas de opción múltiple para reducir el costo de calificarlas. En consecuencia, quizá no se examinen bien otras metas de aprendizaje más importantes con este formato. Una preocupación relacionada es el grado en el que la secuencia de pruebas afecta a la instrucción. Es decir, si las pruebas se alinean con el currículo cambiante, será probable que haya un "cambio de constructo" conforme los alumnos avancen a grados de escolaridad más altos. Esto es muy adecuado para hacer inferencias sobre la competencia de los alumnos en cada grado, pero puede generar sesgos en las estimaciones de valor agregado si las escalas de puntuación de diferentes años se vinculan de forma vertical. Véase Martineau (2006) para abundar en esto.

En cuanto al segundo riesgo, la preocupación gira en torno a las desviaciones significativas de la administración estandarizada, preguntas mal elaboradas o ambiguas, y problemas como baja confiabilidad. Por ejemplo, las preguntas que requieren que el alumno proporcione respuestas escritas deben ser calificadas por maestros contribuyendo a la baja confiabilidad porque los procedimientos de puntuación no están bien implementados o supervisados. Por fortuna, esta clase de problemas técnicos se resuelven con formación y práctica. La aplicación eficaz debe asegurar a los directores de las escuelas que el desempeño de los estudiantes en las pruebas es una medida razonable de su situación académica. De no ser así, las escuelas cuyo desempeño en apariencia no alcance el estándar cargarán la culpa a las pruebas y las inferencias incorrectas realizadas a partir de los análisis llevarán a decisiones deficientes en diversos niveles. Otra dificultad potencial es que los resultados de las pruebas de algunas escuelas se manipulen en un esfuerzo por lograr una mejor puntuación de esa escuela en el valor agregado. Esto representa un caso particularmente pernicioso de varianza de la irrelevancia del constructo. Estos problemas se aligeran un poco con la estructura del marco de evaluaciones de los alumnos y su función en los programas de rendición de cuentas y mejoras escolares. En la Parte I se aborda la creación de incentivos que tal vez generen estos resultados no deseables.

Otra consideración al investigar la calidad de las pruebas se relaciona con los instrumentos de evaluación, si y cómo se preparan cada año los instrumentos de evaluación. En caso de que se emplee la misma forma (o en esencia la misma) cada año, es probable que su integridad se vea comprometida con el tiempo y el desempeño en las pruebas aumente pero no se acompañe de un mejor aprendizaje (Koretz, 2005). Esta "inflación de puntuaciones de las pruebas" socava la credibilidad de los análisis de valor agregado, en particular si su magnitud varía

por escuelas. Si se elaboran diferentes formas cada año, la nueva forma debe equipararse con la anterior para conservar la comparabilidad de la escala (Kolen y Brennan, 2004). Un error de equiparación sustancial, que incorpore tanto varianza de medida como sesgos, también compromete las estimaciones de valor agregado. Por último, los análisis de valor agregado suelen emplear escalas de puntuación de pruebas con vínculos verticales entre grados (Harris *et al.,* 2004). Estrategias diferentes con vinculaciones distintas generan escalas de puntuaciones con propiedades también diferentes que, a su vez, llegan a tener un efecto sustancial en las estimaciones de valor agregado (Patz, 2007).

De manera más general, la validez de las pruebas abarca tanto la validez de constructo como la validez consecuente (Messick, 1989). Esta última se refiere a la adecuación de las inferencias y acciones que se basen en las puntuaciones. Que esas puntuaciones tengan consecuencias no es problema; en cambio, la cuestión es si su empleo se justifica dados su contexto y propósito. Así, las puntuaciones de las pruebas son válidas para un uso pero no para otro. La validez no es una cuestión de "todo o nada": es cuestión de grado. Sin embargo, si hay preocupaciones graves relacionadas con la validez de constructo y consecuente, quizá no sea recomendable proceder con un análisis de valor agregado, al menos hasta que se aborden de manera razonable dichas preocupaciones.

Error de medición

Otra característica de las puntuaciones de las pruebas es la confiabilidad, que es una medición de la replicabilidad del proceso de medición. La confiabilidad es una cantidad sin dimensión (es decir, no se expresa en unidades de medida) que adopta valores entre 0 y 1. La confiabilidad alta (es decir, valores cercanos a 1) significa que los alumnos obtendrían calificaciones semejantes si se sometieran a otra prueba de estructura y formato paralelos a la que sí respondieron. Por otra parte, si hay mucho "ruido" en el proceso de examen, se reduce la confiabilidad. Algunas características de la prueba que determinan la confiabilidad son aspectos de diseño (como extensión, formatos de las preguntas, etc.) y la calidad de la forma de calificar las respuestas. La confiabilidad baja es un riesgo para la validez porque significa que los resultados del análisis de valor agregado pueden ser materialmente distintos si se aplicase de nuevo la prueba.

La confiabilidad es un indicador que resume un aspecto de la calidad de la prueba. Un término muy relacionado es el error de medición, que se expresa en unidades de puntuación de escala y sirve para cuantificar la incertidumbre asociada a las puntuaciones de pruebas observadas. A grandes rasgos, una confiabilidad alta corresponde a un error de medición bajo. Sin embargo, hay ventajas en la representación de la replicabilidad de puntuaciones de pruebas en términos de error de medición. En muchas pruebas es posible calcular el error de medición asociado a cada punto de la escala de reporte. Por lo común, el error de medición es menor cerca del centro de la escala, donde suele encontrarse la mayoría de las puntuaciones de los alumnos, y es mayor en los extremos de la escala. Este fenómeno es un resultado directo de la forma de diseñar y elaborar las pruebas. Los problemas se pueden agravarse cuando se mide el progreso en el desempeño de los alumno a lo largo del tiempo, pues se puede inducir un mayor error de medición al igualar diferentes evaluaciones de los estudiantes (Doran y Jiang, 2006). La asunción estándar de los modelos de regresión es que cada valor observado del criterio

proviene de una distribución con la misma varianza. Así, el hecho de que el error de medición no sea uniforme en la escala de medida (llamado heterocedasticidad) causa problemas cuando las puntuaciones de las pruebas son usadas como criterio. Si no se tiene en cuenta la heterocedasticidad, las estimaciones pueden resultar sesgadas. En este momento poco se sabe de la relación entre el grado de desviación de un error de medición uniforme y el sesgo resultante. Para abundar al respecto, véase McCaffrey *et al.* (2003: 103).

El error de medición también puede causar problemas cuando las puntuaciones de las pruebas se emplean como variables de control en un modelo de regresión. La asunción usual es que las variables de control están medidas sin errores. Se sabe bien que cuando las puntuaciones de las pruebas se emplean como variables de control, el error de medición causa un sesgo descendente en las estimaciones de los coeficientes de regresión correspondientes. Al tomar datos de dos estados de **EUA**, Ladd y Walsh (2002) investigaron la magnitud de este sesgo. Los modelos fueron de regresión lineal estándar que incorporaron puntuaciones de las pruebas de años anteriores pero no características de los alumnos. Estos modelos se utilizaron en Carolina del Norte y del Sur con fines de evaluación de escuelas. Observaron que los efectos estimados en las escuelas que atendían a estudiantes de capacidades menores (basándose en su desempeño del año anterior) descendieron sustancialmente, y que los efectos estimados para las escuelas que atendían a alumnos de capacidades mayores ascendieron también de manera sustancial. Es decir, los resultados del análisis de valor agregado de escuelas no favorecieron a las escuelas que atendían a alumnos más débiles y favorecieron a las que atendían a alumnos más capaces. Más aún, muestran cómo este sesgo se reduce considerablemente si se emplean las puntuaciones de las pruebas de años anteriores como variables instrumentales. Sin ellas, deben emplearse otras características de los alumnos relevantes cuando estén disponibles. Se abunda en esto en el Capítulo 6.

Las propiedades de la distribución de las puntuaciones de las pruebas son también relevantes en la implementación e interpretación de un análisis de valor agregado. La asunción estándar es que las puntuaciones se distribuyen de acuerdo con la distribución gaussiana (normal), al menos condicional de las demás variables (características de los alumnos) del modelo. Las desviaciones leves de esta suposición no deben causar alarma. Sin embargo, los efectos sustanciales de "piso" o "techo" pueden ser problemáticos. Por ejemplo, si la prueba de un grado particular es relativamente fácil para muchos estudiantes inscritos en un subconjunto de escuelas, la distribución de las ganancias de sus puntuaciones tendrá una desviación pronunciada hacia el lado de la distribución más bajo. Los estimadores de valor agregado para esas escuelas tendrán sesgos descendentes en comparación de los que habrían obtenido si la prueba hubiese sido lo bastante difícil para esos estudiantes.

Escalas de puntuaciones

Si bien la elaboración de evaluaciones de estudiantes y las pruebas no son el tema central de este informe, el asunto de las escalas de puntuaciones de las pruebas se considera demasiado importante para no mencionarlo. Es común que las puntuaciones "brutas" de las pruebas se transformen en otra escala para informar y para análisis secundarios. Estas transformaciones hacen que parezca que las puntuaciones de las pruebas son comparables de un año al siguiente. Sin embargo, la verdadera comparabilidad depende de una aplicación cuidadosa

de las especificaciones de la prueba y, de ser necesario, de ajustes de puntuaciones mediante un proceso especial llamado equiparación (de pruebas). Las desviaciones graves de la comparabilidad año con año quizá no sean especialmente problemáticas para los alumnos si sólo se comparan con otros de la misma cohorte. Sin embargo, son problemáticas para el análisis de valor agregado porque significa que la distribución de las puntuaciones de ganancias varía al paso de los años (Harris *et al.,* 2004). Si los efectos de la escuela se obtienen a partir del análisis de datos de múltiples cohortes, esta variación introduce varianza de la irrelevancia del constructo.

En algunos ambientes, las pruebas de fin de año se administran en cada grado y las puntuaciones brutas de las pruebas de distintos grados se "vinculan verticalmente" para generar una escala única para todos los grados. Hay varios procedimientos para efectuar la vinculación vertical, y cada una produce una escala de grados con diferentes propiedades que generan distintas estimaciones de los efectos de la escuela (Patz, 2007). Aunque no se requiere construir una escala común de grados para aplicar muchos modelos de valor agregado, las puntuaciones de pruebas vinculadas verticalmente a menudo se emplean como archivo de entrada para un análisis de valor agregado. En estas situaciones, los usuarios deben tener en mente las características de la escala vertical y cómo puede afectar a las estimaciones del modelo de valor agregado. Deben pensarlo dos veces antes de tratar la escala como escala de intervalo (es decir, una para la cual las diferencias de escala tengan el mismo significado a lo largo de toda la escala). Aunque es tentador hacerlo, pocas veces se justifica, y se recomienda una vía más conservadora.

Resultados de evaluación en una escala ordinal

Hasta aquí hemos supuesto que las puntuaciones de las pruebas se reportan en una escala con suficientes valores para tratarla como si de verdad fuese continua. Sin embargo, en algunos ambientes, se informa de las puntuaciones finales en una escala gruesa que comprende apenas dos categorías ordenadas. Por ejemplo, las autoridades pueden establecer dos estándares que denoten "logro competente" y "avanzado". Cada estándar se representa con una puntuación, o punto de corte, en la escala de reporte original. Después se clasifica a los alumnos en una de tres categorías ("menos que competente", "competente" y "avanzado") según donde se sitúe su puntuación. Si bien los modelos de valor agregado convencionales no deben aplicarse en esos casos, es posible, de todos modos, efectuar un análisis de valor agregado. Si sólo hay dos categorías, se puede emplear una regresión logística o regresión probit en lugar de los modelos usuales basados en la distribución normal. Si hay más de dos categorías, pueden usarse modelos polítomicos de regresión logística o modelos probit ordenados. Para ahondar en este tipo de modelos, véase Fielding, Yang y Goldstein (2003).

Los problemas de validez y confiabilidad también son importantes en los datos ordinales. Si las categorías se determinan mediante alguna variante de un procedimiento de punto de corte, debe evaluarse la validez del procedimiento (Hambleton y Pitoniak, 2006). Si las categorías corresponden a etapas de una escala de desarrollo, debe evaluarse el apoyo teórico y empírico para la escala. En ambos casos, la confiabilidad se relaciona con la probabilidad de que un alumno sea asignado a la categoría apropiada. Colocarlo en la categoría equivocada es un tipo de error de medición que induce sesgo en la estimación. Mientras mayor sea el error

de medición (y más baja la confiabilidad), menos creíble serán las estimaciones del valor agregado de la escuela.

En la mayoría de los países participantes, las razones para implementar un sistema de valor agregado basado en ciertas evaluaciones son desplazar la atención de los directores de escuelas, maestros y alumnos a mejorar el desempeño en esas medidas y el aprendizaje de los alumnos en las disciplinas académicas correspondientes. La elección de materias y grados, así como la naturaleza de las evaluaciones, deben hacerse con cuidado, pues es probable que afecten a las acciones de los interesados. En particular, las deficiencias de las evaluaciones pueden llevar a obtener puntuaciones de los alumnos más altas que no se están asociadas a las mejoras deseadas en el aprendizaje de los estudiantes. Este sería el caso de una falta de validez consecuente. Las decisiones referentes a la forma de emplear el desempeño de los alumnos para la evaluación de escuelas puede alterar los incentivos y, por tanto, el comportamiento de directores y maestros (Burgess *et al.,* 2005). Por lo común, las puntuaciones de los alumnos se transforman o resumen en indicadores de desempeño que dan información al proceso de toma de decisiones. Una distinción básica es la que se establece entre los indicadores de desempeño discretos y continuos. Si se evalúa a una escuela con base en un indicador discreto, hay un incentivo natural para concentrar los recursos en la mejora de dicho indicador. Por ejemplo, un análisis de valor agregado que se centre en la proporción de niños que alcanzan o exceden un nivel particular de lectura alienta a las escuelas a centrar su atención en los estudiantes que se ubican por debajo del nivel de alfabetización pero que es probable que alcancen ese nivel con el apoyo adecuado. Por otra parte, en este ejemplo hay pocos incentivos para que la escuela mejore las puntuaciones de los alumnos que ya superan ese nivel, o centrarse en los estudiantes muy por debajo de dicho nivel. En cambio, es más probable que un análisis de valor agregado que se centre en un indicador continuo aliente una asignación de recursos más uniforme, aunque es posible que los alumnos que parezcan mejor ubicados para obtener mayores ganancias reciban más atención. Por ejemplo, quizá sea más fácil mejorar el desempeño de alumnos de mayor aprovechamiento que el de los de menor aprovechamiento. Esto no sólo genera distorsiones al interior de las escuelas, sino también dificulta las comparaciones entre escuelas. Es decir, las escuelas con mayores proporciones de alumnos con antecedentes favorables (como sea que se midan) pueden recibir puntuaciones de valor agregado más altas porque sus alumnos por lo general lograrían ganancias más grandes. De ser éste el caso y si los maestros de escuelas con puntuaciones de valor agregado más altas acordasen beneficios especiales, habría un claro incentivo para que los maestros se dirigieran a las escuelas con mayores proporciones de alumnos con antecedentes favorables.

Sin embargo, es posible introducir una fuerza que contrarreste esto con una ponderación diferencial de las puntuaciones de ganancias. Por ejemplo, se puede acordar una mayor ponderación a las mejoras en el extremo inferior de la escala en comparación con el extremo alto. Como es más probable hallar alumnos de estado socioeconómico bajo en el extremo bajo de la escala, este esquema de ponderaciones ofrecería incentivos adicionales para que los directores de escuelas y maestros se centren en elevar el desempeño de esos estudiantes e incluso induzca a los maestros más eficientes a trabajar en esas escuelas. Estas cuestiones se abordan en la Parte I, que ilustra estos sistemas y las implicaciones de varias estructuras de incentivos.

Estructura de evaluaciones de los alumnos en países participantes

Algunas decisiones referentes al diseño y uso de modelos de valor agregado dependen de la naturaleza de los datos de evaluación disponibles. A continuación se analizan los datos de evaluación recopilados en cada país para ilustrar las diferencias entre países, así como las estrategias que mejoran los datos y así aumentan la utilidad de las políticas de los análisis de valor agregado. En algunos países, la elección de evaluaciones para los análisis de valor agregado se determina en esencia por la estructura del sistema educativo. Por ejemplo, si el sistema escolar se organiza en sectores de primaria y secundaria, y las escuelas pertenecen a uno u otro, por lo general los análisis de valor agregado sólo se basan en evaluaciones administradas en un periodo proporcional al tiempo que los alumnos pasarían normalmente en una escuela primaria o secundaria. Desde la perspectiva de los análisis de valor agregado, es problemático si una evaluación tiene lugar a la mitad de la educación primaria de los alumnos y la segunda mitad de su educación secundaria. El Cuadro 4.1 detalla las evaluaciones de los alumnos que pueden usarse en análisis de valor agregado en los países participantes, e ilustra las diferencias entre países en las materias cubiertas. Cabe observar que en algunos países la falta de comparabilidad de evaluaciones es una barrera para el desarrollo de un análisis de valor agregado.

Hay variación considerable en las edades y niveles de grado/año en que se recopilan los datos de evaluación de los estudiantes. Al considerar los datos de evaluación de los alumnos que servirán para los análisis de valor agregado, la edad a la que se evalúa a los alumnos moldea la medición de resultados mediante la cual es posible apreciar los efectos de las escuelas en el progreso de los estudiantes. Las evaluaciones en algunos países se centran en la educación primaria, y en otros, en la secundaria y en la educación media. Países como **Bélgica (Comunidad Flamenca)** y **República Checa** concentran sus evaluaciones en los primeros grados, que facilita el uso de modelos de valor agregado en el desarrollo del sector de educación primaria. Por otra parte, la estructura de los marcos de evaluación estudiantil en países como **Eslovenia**, **Noruega**, **Polonia**, **Portugal** y **Suecia** facilita, en su mayoría, el desarrollo de modelos de valor agregado centrados en el sector de educación secundaria. En **Dinamarca** hay evaluaciones de matemáticas y lectura en educación primaria y secundaria, y evaluaciones adicionales en ciencias e inglés sólo en educación secundaria. Las materias incluidas en el marco de evaluación estudiantil reflejarán las prioridades del sistema nacional y tendrán un impacto en el uso e interpretación de los modelos de valor agregado. Si sólo se evalúa matemáticas en años determinados, sólo se medirá el valor agregado de matemáticas. Si se desea crear un indicador de valor agregado de base más amplia, sin duda se requieren evaluaciones de alumnos de una variedad más extensa de materias. En general, se evalúa a los estudiantes en más materias en la educación secundaria, en particular en la educación media, donde los resultados de los exámenes de todas las materias (es decir, exámenes nacionales) se emplean para modelos de valor agregado (según el tipo de modelo de valor agregado que se use). En niveles más bajos, las evaluaciones se concentran sólo en unas cuantas áreas. En la mayoría de los países, éstas son matemáticas, ciencias y el idioma nacional o el de instrucción (con hincapié en lectura y/o redacción de dicho idioma).

Cuadro 4.1

Exámenes de alumnos en países participantes que en potencia se usarían para modelos de valor agregado

	Año-Nivel	Materias
Bélgica (Fl.)	Año 1-6	Matemáticas, idioma de instrucción.
	Año 1-6	Matemáticas, lectura, ortografía.
	Año 6 (año final de ISCED 1)	Matemáticas, lectura, naturales (subdominio de estudios ambientales), francés, sociedad.
	Año 8	Áreas intracurriculares ("aprender a aprender", "recopilación y procesamiento de información"), biología, francés, sociedad.
Dinamarca	Año 2, 4, 6, 7, 8,	Lectura, matemáticas, inglés, ciencias.
	Año 9 y 10	Todas las materias obligatorias (evaluadas por maestros).
	Educación media	Lectura, matemáticas, inglés, ciencias.
Eslovenia	Año 6	Lengua materna, matemáticas, primer idioma extranjero.
	Año 9	Lengua materna, matemáticas, una materia escolar obligatoria (decidida por el ministerio).
	Educación media (Año 13)	Vocacional: Lengua materna, matemáticas o primer idioma extranjero, dos materias escolares y específicas del currículo.
	Educación media (Año 13)	General: Lengua materna, matemáticas, primer idioma extranjero y dos de 30 materias opcionales.
España	4 (primaria), 8 (secundaria)	Matemáticas, idioma de instrucción: ciencias sociales con educación cívica, ciencias, tecnologías de la información y comunicación, otra.[1]
Francia	Examen nacional (bachillerato al término de la educación media)	Cubre 15 materias por estudiante.
Noruega	Año 5, 8	Exámenes nacionales en matemáticas, lectura, inglés (lectura).
	Año 10	Exámenes externos (matemáticas, noruego o inglés). Todas las materias obligatorias (evaluadas por maestros).
	Año 11,12,13	Exámenes y evaluaciones de maestros en varias materias.
Polonia	Año 6 (examen de salida de primaria)	Prueba de competencia de materias.
	Año 9 (examen de salida de secundaria)	Humanidades, matemáticas, ciencias.
	Año 12 (examen de salida de educación media)	Examen final de preparatoria (polaco es obligatorio y después hay evaluaciones en varias materias adicionales).
Portugal	Año 4, 9	Matemáticas, portugués.
	Año 12	Todas las materias requeridas para certificación y entrada a la educación superior.
Reino Unido	Etapa 1: Año 2	Lectura, redacción, matemáticas.
	Etapa 2: Año 6	Lectura, redacción, matemáticas, ciencias.
	Etapa 3: Año 9	Inglés, matemáticas, ciencias.
	Etapa 4: Año 11	Una amplia variedad de materias, que en su mayoría cuentan para los mejores 8 resultados del alumno.
Rep. Checa[2]	13 (Estado Maturita)	Idioma checo, idioma extranjero y uno de matemáticas, ciencias sociales, ciencia o tecnología.
	Año 5, 9	Idioma checo, matemáticas, idioma extranjero, habilidades de aprendizaje.
Suecia	Año 9, grados finales	Evaluación de 16 materias.
	Año 5, prueba estandarizada	Inglés, matemáticas, sueco.
	Año 9, prueba estandarizada	Inglés, matemáticas, sueco.
	Educación media, grados finales	Promedio de puntos de grado, todas las materias por cada alumno (30-35 materias).
	Educación media, prueba estandarizada	Inglés, matemáticas, sueco.

1. Matemáticas e idioma de instrucción se evalúan al año. Otras materias se evalúan con menor frecuencia.
2. Recopilación de datos en etapa piloto en 2008. El proyecto que recopila datos en el Año 13 se transformará en Examen final estatal de Maturita en 2010; años 5 y 9 no continuarán.
Fuente: Información proporcionada por expertos consultados.

La frecuencia de las evaluaciones varía considerablemente por país. Debe observarse que el sistema de evaluaciones en algunos países no permite a la fecha análisis de valor agregado como se definen en este informe. Nuestra definición destaca que se requiere una evaluación previa para medir el valor agregado. Además, las evaluaciones tienen que ser comparables de modo que apoyen las inferencias deseadas referentes a la relación de distintos factores con el progreso de los alumnos. En **Dinamarca** y el **Reino Unido** se elaboraron marcos de evaluación de alumnos que abarcan los sectores educativos de escuelas primarias y secundarias. En el **Reino Unido** se identificaron etapas básicas en el progreso de alumnos a lo largo de su paso por la escuela, con evaluaciones en los Años 2, 6, 9 y 11. La **Comunidad Flamenca de Bélgica** es el único país participante que tiene datos anuales de evaluación de estudiantes, si bien sólo del nivel de primaria. Las pruebas anuales resuelven parcialmente algunos de los problemas estadísticos y metodológicos con los modelos de valor agregado analizados más adelante en este informe, y deben aumentar la utilidad de los resultados.

La frecuencia de las evaluaciones tiene un impacto en la elección del modelo de valor agregado, así como para decidir si se debe o no incluir características de antecedentes de los alumnos. Estas decisiones a su vez afectan la interpretación de los resultados del modelo. Las decisiones referentes a la frecuencia de las evaluaciones dependerán de la naturaleza del currículo y de las prioridades respecto de la supervisión del progreso de los alumnos en diversos momentos de la carrera escolar. Para países que se preparan a elaborar un marco de evaluaciones estudiantiles y a emplear modelos de valor agregado, hay ventajas en registrar los avances mediante evaluaciones estudiantiles más frecuentes.

Como veremos en el Capítulo 6, incrementar la cantidad de mediciones de aprovechamiento anteriores mejora en gran medida la precisión y credibilidad de los análisis de valor agregado. Por tanto, es tentador alentar evaluaciones de alumnos más frecuentes. Sin embargo, existe la preocupación de que las evaluaciones adicionales añadan una carga indebida a las escuelas y reduzcan la cantidad de tiempo efectivo de enseñanza. Es decir, las pruebas no sólo ocupan tiempo del día escolar, sino también imponen requerimientos organizativos de actividades previas y posteriores a la evaluación. Las autoridades educativas pueden ponderar los beneficios de incrementar la frecuencia de evaluación respecto de estas cargas y los costos financieros. Además, las pruebas pueden añadir presión a los alumnos que también puede tener consecuencias negativas. Esto se refleja en el Cuadro 4.1, que muestra que en la mayoría de los sistemas educativos se evalúa hoy en día a los alumnos sólo en unos cuantos niveles anuales, y en materias y áreas de aprendizaje seleccionadas.

Como vimos en la Parte I, el uso de los resultados de las pruebas con propósitos ambiciosos crea incentivos para influir en el desempeño de los alumnos de estas evaluaciones de manera no adecuada. La práctica de "enseñar para la prueba" es una de estas consecuencias indeseables, pero hay varios casos documentados en que varios indicadores escolares y pruebas ambiciosas pueden manipularse (y se han manipulado) de manera que genere resultados no deseados (Nichols y Berliner, 2005). Surgen otros problemas si la puntuación de valor agregado de una escuela se manipula de forma más directa. Considere un escenario en que se emplean dos evaluaciones para estimar el valor agregado de una escuela. Suponga que la primera evaluación se realiza en el Año 3 y la segunda en el Año 6. Sin duda, el valor agregado de una escuela aumenta si hay una diferencia positiva más grande entre las evaluaciones.

Por tanto, hay un incentivo para elevar las puntuaciones de los alumnos en el Año 6 y para bajar las puntuaciones (de esos mismos estudiantes) en el Año 3. Esto se lograría aconsejando a los alumnos que no se tomen tan en serio el examen del Año 3, como lo harían de otro modo, o incluso al alentarlos a que de forma deliberada no lo respondan bien. Algunas acciones más radicales serían estructurar el currículo de modo que los alumnos no se preparen bien para la evaluación del Año 3. Aún así, pueden elaborarse estrategias para reducir la probabilidad de estas actividades no deseadas. Por ejemplo, el efecto perverso de los incentivos se contrarresta al imponer objetivos de desempeño para la evaluación del Año 3. De manera más general, las escuelas deben tener un incentivo para elevar el desempeño de estudiantes en todas las evaluaciones, y alinear así sus intereses con los de los alumnos. Considere el marco de evaluación anual en la **Comunidad Flamenca de Bélgica**, donde cada evaluación (salvo la del Año 1) tiene una doble función. Así, la evaluación del Año 3 es una medición final de desempeño en el análisis de valor agregado entre el Año 2 y el Año 3 (o Año 1 y Año 3), y también una medición de desempeño anterior en el análisis de valor agregado entre el Año 3 y el Año 4, o algún otro año subsecuente. Esta doble función mitiga el incentivo de reducir el desempeño en la evaluación del Año 3. Ocurriría una excepción si las autoridades educativas hicieran más hincapié en la medición de valor agregado en un año específico.

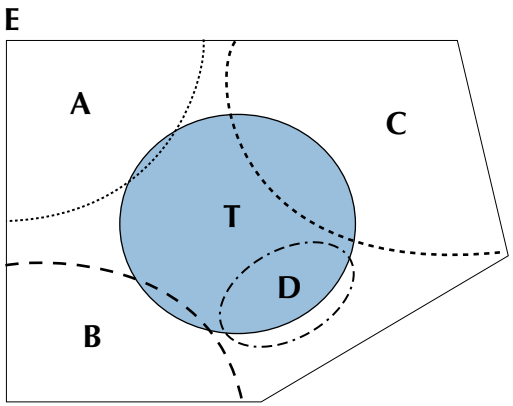
También se puede promover que las escuelas eleven el desempeño de los estudiantes en la evaluación inicial al hacer que dicha evaluación forme parte de procedimientos administrativos generales o de programas o políticas educativas. Por ejemplo, el desempeño de los alumnos en la evaluación inicial o previa se vincularía a un sistema de inspecciones escolares y procedimientos de evaluación escolar. Asimismo, las medidas de la evaluación pueden formar parte de un marco más amplio de mediciones escolares que faciliten una elección escolar eficiente. Como vimos en la Parte I, publicar estas mediciones a menudo crea incentivos positivos para elevar el desempeño del alumnado. Aparte de las consideraciones para alinear incentivos, deben establecerse procedimientos adecuados para asegurar que todas las evaluaciones sean justas y no contengan errores. Debe estandarizarse la aplicación de los exámenes, y el marcado de los documentos del examen debe ser confiable y quedar protegido contra la manipulación en todas las etapas del proceso. Esto generará mayor confianza en los resultados de la evaluación y en el análisis de valor agregado siguiente. También debe notarse que algunos países utilizan evaluaciones estandarizadas elaboradas externamente, y otros dependen de exámenes escolares. Unos cuantos sistemas, como en el **Reino Unido**, en ocasiones emplean ambas clases de evaluación, si bien todas las calificaciones en la Etapa 4 se evalúan de manera externa. Sin embargo, en las Etapas 2 y 3, los datos se recopilan de evaluaciones externas y de los maestros. Se emplean datos de evaluaciones externas porque se supone que son más creíbles y comparables, así como que poseen propiedades psicométricas superiores. En la Etapa 1, los exámenes no se puntuaron externamente, y ha habido cierta preocupación sobre la robustez de los datos (véase Tymms y Dean, 2004). Desde 2005, todos los resultados de la Etapa 1 (de alumnos de siete años de edad) se basan en evaluaciones de los maestros. Si bien esto puede presentar el riesgo de sesgos (en contraste con una evaluación estandarizada), existe la posibilidad de que los datos sean más válidos, pues los maestros recurren a una mayor variedad de evidencias durante un periodo dado que un solo examen aplicado en una ocasión. Si se emplean evaluaciones de maestros, deben someterse a supervisión externa para asegurar su comparabilidad y validez.

FILOSOFÍA DEL AJUSTE Y CARACTERÍSTICAS CONTEXTUALES

Con el fin de obtener efectos estimados de la escuela, la mayoría de los modelos de valor agregado efectúa un ajuste de regresión en las puntuaciones de las pruebas de los alumnos. El ajuste pretende "nivelar el campo de juego", es decir, retirar de las comparaciones entre escuelas los efectos de confundido por las diferencias sistemáticas en las poblaciones de alumnos en el ingreso. Al hacerlo, se espera que el análisis de valor agregado sea más fructífero al "aislar" las contribuciones de escuelas individuales sobre el avance académico de sus alumnos que cuando se compara a las escuelas basándose sólo en el aprovechamiento de los alumnos. Si bien esta estrategia es prudente y muy común, es importante observar que el ajuste estadístico debe efectuarse con cuidado y sin olvidar las posibles consecuencias negativas. Con esto en mente, los siguientes párrafos presentan una explicación simplificada del ajuste estadístico, al ilustrar los puntos fuertes y débiles del procedimiento.

Gráfica 4.1

Ilustración gráfica del proceso de ajuste estadístico



Suponga que la meta es estimar el desempeño relativo de una escuela. Este es el objetivo o parámetro de interés. El círculo (con la letra "T" en la Gráfica 4.1) representa el valor verdadero del parámetro. El estimador obtenido de una comparación sin ajustar que se representa con la figura de cuatro lados (letra "E"). En este caso, el estimador es demasiado grande. Es decir, usamos las áreas de las figuras para señalar sus magnitudes. "E" puede ser más grande que "T" porque los alumnos de la escuela son más aventajados que los de la escuela promedio. Al reconocer que las escuelas no se asignan de forma aleatoria a los alumnos (o *viceversa*), recurrimos al ajuste estadístico de las características de los estudiantes medidas para crear un campo de juego más nivelado. Se supone que cada ajuste modifica a "E" para acercarlo a "T". En la Gráfica 4.1, el efecto del ajuste se representa con una figura contenida en "E" que puede o no superponerse con "T".

El primer ajuste ("A") reduce el área de "E". El nuevo estimador, "E-A", se acerca más a "T" que a "E". Observe que "A" se superpone un poco con "T", lo que indica que parte del ajuste aisló una pequeña porción de la diferencia verdadera. Sin embargo, el nuevo estimador aún es demasiado grande. Otros ajustes para las siguientes dos características ("B" y "C") ofrecen un estimador "E-A-B-C", más cercano a "T". Sin embargo, en el caso de "C" hay una superposición considerable con "T", lo que significa que se ajustó demasiado. Por último, el ajuste "D" aisló una buena porción de "T" pero relativamente poco de la parte de "E" fuera de "T". Esto significa que hubo un sobre-ajuste. El estimado resultante, "E-A-B-C-D", puede estar más cerca de "T", pero ser menor que "T" en lugar de mayor. Un ajuste más, con efectos semejantes a los de "D", puede arrojar un estimador más deficiente que los anteriores. La lección aquí es que el ajuste estadístico debe efectuarse con todo cuidado.

En la mayoría de los modelos de valor agregado, aislar la contribución de las escuelas requiere estimar la relación entre las puntuaciones de los alumnos y las diversas variables socioeconómicas y contextuales. Aunque hay problemas de medición que necesitan abordarse al aislar los múltiples impactos en el desempeño estudiantil, es útil para los tomadores de decisiones analizar tanto el grado de la relación entre desempeño estudiantil y las características contextuales específicas como, en algunos casos, analizar los resultados de valor agregado de grupos particulares de alumnos. El análisis de estos datos ofrece información para la elaboración de políticas en diversas áreas, como una financiación escolar equitativa.

Importancia de las características contextuales

El programa PISA de la OCDE no produce mediciones de valor agregado y se alinea más con lo que se clasifica como modelos de aprovechamiento contextualizado en este informe. Los hallazgos más recientes de PISA confirman evidencias previas de que el nivel socioeconómico de los alumnos es uno de los mayores predictores de desempeño escolar con estos modelos (OCDE, 2007a). Estos hallazgos son congruentes con la literatura, que documenta el vínculo estadístico entre las variables de antecedentes individuales y familiares, por una parte, y la educación de los jóvenes, por otra (OCDE, 2007d; Haveman y Wolfe, 1995). Además, esta relación se extiende al incluir características del vecindario o comunidad y compañeros (Ginther, Haveman y Wolfe, 2000; Brooks-Gunn *et al.,* 1993; Corcoran *et al.,* 1992; Mayer, 1996). Estos análisis estiman la fortaleza de la relación entre diversos factores y una sola medida o resultado de desempeño. Entre estos factores se encuentran características de antecedentes individuales y diversas características socioeconómicas contextuales, así como características escolares. Como vimos en la Introducción de este informe, la característica clave que distingue a los modelos de valor agregado es la inclusión de una medida de aprovechamiento anterior comparable con la cual aislar de manera más precisa la contribución de la escuela al progreso de los alumnos. Cuando se incluyen mediciones de rendimiento previo en el modelo de regresión, a menudo se reduce mucho la contribución de las características contextuales que expliquen las diferencias en los resultados de los alumnos. Ballou, Sanders y Wright (2004) señalan que cuando se dispone de un conjunto rico de medidas de aprovechamiento previas y actuales, el ajuste de características demográficas de los alumnos tiene un impacto mínimo en los efectos escolares estimados. Además, a pesar de favorecer por lo general la inclusión del nivel socioeconómico como variable de antecedente del alumno, McCaffrey *et al.* (2003, 2004) concluyen que controlar los factores socioeconómicos y demográficos de los estudiantes sin mediciones de rendimiento previo

no basta para anular los efectos de las características de antecedentes en todos los sistemas escolares, en especial los sistemas que atienden a estudiantes heterogéneos. Así, tomadores de decisiones deben tener cuidado al interpretar las mediciones de desempeño escolar a partir de modelos de aprovechamiento contextualizado.

En el diseño de modelos de valor agregado, las autoridades educativas y los administradores deben considerar con cuidado el empleo de características socioeconómicas contextuales. Quienes conocen mejor los modelos de aprovechamiento contextualizado, es bien conocida la importancia de las características socioeconómicas contextuales como predictoras del aprovechamiento estudiantil. En consecuencia, el análisis de la sección precedente respecto del papel atenuado de estas características en los modelos de valor agregado quizá sea un tanto sorpresivo. Los análisis de datos noruegos y portugueses muestran que el empleo de características contextuales es mucho más importante en modelos de aprovechamiento contextualizado que en modelos de valor agregado. Hægeland y Kirkebøen (2008) ofrecen una ilustración empírica de cómo los estimadores de desempeño escolar se ven afectados por la elección de variables contextuales socioeconómicas en modelos tanto de aprovechamiento contextualizado como de valor agregado. Los autores observan que ajustar el rendimiento previo y el nivel socioeconómico de los alumnos no son estrategias mutuamente excluyentes al estimar el desempeño escolar. También queda claro que el papel de los factores contextuales difiere por país y tipo de modelo utilizado. Sin embargo, los hallazgos del estudio noruego referentes a la influencia de las características del nivel socioeconómico en los estimadores de valor agregado también aparecieron en el estudio longitudinal portugués. El análisis de los datos noruegos aclara el uso de las variables contextuales en modelos de valor agregado e ilustra las diferencias en este punto con los modelos de aprovechamiento contextual. El estudio comparó los resultados de cuatro modelos, al incorporar una cantidad creciente de datos socioeconómicos como variables de control. La comparación de los resultados mostró que añadir características socioeconómicas aumentó la cantidad de varianza explicada de las puntuaciones de los alumnos y redujo la dispersión de la distribución de indicadores de desempeño escolar en los modelos de aprovechamiento contextualizado. Esto es congruente con la literatura, que muestra que las características socioeconómicas se correlacionan con el desempeño estudiantil y no se distribuyen de manera uniforme entre escuelas. Sin embargo, sus resultados indican que, en sus modelos de valor agregado, los efectos de incluir variables de nivel socioeconómico adicionales están limitados debido a la presencia de medidas de rendimiento previo. Muestran que un modelo de valor agregado simple que contenga sólo información demográfica básica (género y año de nacimiento), además de medidas de rendimiento previo, tiene un poder explicativo mucho mayor que el modelo más completo de aprovechamiento contextualizado. La inclusión de características socioeconómicas adicionales a este modelo de valor agregado sólo tuvo un efecto menor en el poder explicativo del modelo y en los estimadores de desempeño escolar. Por otra parte, incorporar mediciones de desempeño anterior adicionales tuvo un impacto mayor en el poder predictivo del modelo.

A pesar de estos hallazgos, la inclusión de características socioeconómicas en un modelo de valor agregado puede tener consecuencias para determinadas escuelas. Respecto de los datos noruegos, el mayor impacto para una escuela individual con la inclusión de todo el vector de características contextuales en el modelo de valor agregado correspondió a la mitad de

una desviación estándar de la distribución del desempeño escolar estimado. Este resultado subraya la importancia –al desarrollar un sistema basado en modelos de valor agregado– de efectuar análisis de sensibilidad no sólo en los parámetros del modelo general, sino también en los estimadores escolares individuales. Los cambios sustanciales en los estimadores de valor agregado deben estimular más investigación, pues pueden indicar problemas con los datos. De forma ideal, estos tipos de análisis deben efectuarse durante la etapa piloto del proceso de implementación.

Aunque es sugerente el análisis de los datos noruegos, no podemos extraer una conclusión general de este ejercicio. Las consecuencias de incluir (más) variables contextuales socioeconómicas en un modelo de aprovechamiento contextualizado pueden variar por niveles, años y países. Si las características socioeconómicas sólo se relacionan con el nivel inicial de desempeño y no con la tasa de crecimiento, no habría ningún beneficio en la inclusión de estas características en los modelos de valor agregado. Por otra parte, habría algún beneficio si estas características se correlacionaran con el crecimiento de desempeño de los alumnos. En algunos países miembros de la OCDE, la inclusión del "año de nacimiento" en el modelo de valor agregado captura el efecto de "repetición" o retención de grado, fenómeno correlacionado negativamente con el nivel socioeconómico (OCDE, 2007c). También es posible que la inclusión del "año de nacimiento" capture el efecto de la edad de entrada diferencial en el sistema educativo. Al emplear un modelo de aprovechamiento contextualizado (modelo de componentes de varianza) con datos de PISA 2000, Ferrão (2007a) muestra que la "repetición" explica 45% de la variabilidad del desempeño de alumnos portugueses en matemáticas (medido por PISA). Desde el punto de vista educativo, la inclusión de la variable "año de nacimiento" como covariante en el modelo de valor agregado puede generar polémica y debe abordarse de forma adecuada en cada país.

Un análisis de datos portugueses (representativos de la región Cova da Beira) reveló hallazgos semejantes a los del análisis noruego respecto del efecto de incluir varias características socioeconómicas en modelos de valor agregado (Ferrão, 2008). Este análisis usó datos recopilados al principio y final del año académico 2005-06 de alumnos inscritos en los grados 1.º, 3.º, 5.º, 7.º y 8.º. La variable de respuesta fue la puntuación en matemáticas en una prueba estandarizada equiparada[1] con el rendimiento previo en matemáticas (Ferrão *et al.,* 2006). Entre las características socioeconómicas analizadas se encuentran las que miden la escolaridad paterna y la elegibilidad estudiantil para alimentos y libros escolares gratuitos. La elegibilidad para alimentos escolares gratuitos es una medición común en estimaciones similares que incluyan características contextuales socioeconómicas (véase Goldstein *et al.,* 2008; Braun, 2005a; Ballou, Sanders y Wright, 2004; McCaffrey *et al.,* 2004; Sammons *et al.,* 1994; Thomas y Mortimore, 1996). El asunto de interés fue la sensibilidad de los estimadores escolares de valor agregado ante diversas operacionalizaciones de variables individuales del constructo de estado socioeconómico. Los resultados revelaron correlaciones cercanas a 0.90, lo que sugiere que el empleo de aproximaciones alternativas simples puede generar resultados comparables (Ferrão, 2007a). Sin embargo, es importante observar que las clasificaciones de algunas escuelas experimentan grandes cambios con el tiempo. Aunque estos hallazgos son un tanto alentadores, debe investigarse más con otras características comunes, con atención al empleo de covariantes múltiples.

Al considerar el empleo de características socioeconómicas, también deben tomarse en cuenta la frecuencia y amplitud de las evaluaciones de los estudiantes. Si se evalúa con frecuencia a los alumnos en varias materias, y la cantidad de puntuaciones de las pruebas es correspondientemente grande, la contribución de variables de antecedentes en los modelos de valor agregado se reduce en gran medida. Sin embargo, si hay evaluaciones menos frecuentes y una brecha mayor entre las evaluaciones de los alumnos, la contribución potencial de las variables de antecedentes es mayor. Por ejemplo, si un estudiante que se evaluó en el Año 3 no se evalúa de nuevo hasta el Año 6, las variables contextuales como el nivel socioeconómico pueden correlacionarse de manera estrecha con la tasa de crecimiento del alumno durante este periodo de tres años. Aparte de las consideraciones técnicas, puede ser recomendable incluir características socioeconómicas en un modelo de valor agregado con el fin de ganar la confianza de los interesados. Una estrategia sería presentar los resultados de diferentes modelos que incluyan algunas, todas o ninguna característica socioeconómica y de antecedentes disponible. La importancia de esta estrategia dependerá del propósito del uso los estimadores escolares de valor agregado. Las preocupaciones de los interesados pueden ser mayores si se aplica un rígido sistema de rendición de cuentas escolar o docente que si los estimadores de valor agregado se emplearan sólo con propósitos de mejoras escolares.

¿Qué características contextuales socioeconómicas?

Conviene recordar que los efectos estimados de la escuela que generan los modelos de valor agregado representan las contribuciones combinadas de las acciones y políticas de la escuela con los efectos del ambiente escolar que surgen por las interacciones entre estudiantes y su impacto en el clima escolar, la actitud hacia lo académico y otras variables escolares. En la medida en que los ajustes de características individuales y escolares no capturen por completo estos efectos de la relación entre iguales, las mediciones del desempeño escolar estimado no son estimadores imparciales de las contribuciones escolares al aprendizaje de los alumnos. Observe también que la interpretación de las medidas del desempeño escolar estimado depende de las variables que se empleen para el ajuste. Cada conjunto de variables establece de manera implícita el "nivel del campo de juego" donde se comparan las escuelas. Es decir, cuando declaramos que las medidas de desempeño escolar estimado nos dan la clasificación relativa del desempeño de las escuelas, siendo el resto de las condiciones iguales, es el ajuste del modelo lo que determina lo que define "todo lo demás". Debe tenerse en mente que el propósito principal de incluir variables explicativas en el modelo es reducir sesgos en las mediciones del desempeño estimado de la escuela. Para cumplir con esta meta, esas variables deben relacionarse con el resultado y distribuirse de manera diferencial entre las escuelas. Cuanto más fuerte sea la relación y mayor la variación entre escuelas, el ajuste tendrá un mayor efecto deseado. En todo caso, la adición de estas variables por lo general incrementará la precisión de la predicción.

Las características de los estudiantes que suelen emplearse en el proceso de ajuste abarcan variables como género, raza/origen étnico y nivel de escolaridad paterna. Estas características por lo general se asocian al aprovechamiento académico (OCDE, 2007b; Lissitz *et al.,* 2006). Si estas características se distribuyen de forma desigual entre las escuelas, de no tomarse en cuenta generarán estimadores sesgados del valor agregado de las escuelas. Es decir, en ausencia de ajuste, las escuelas en que se inscriben alumnos con más características "favorables", en promedio, tendrán ventaja en comparación con las escuelas en que se inscriban estudiantes

con características menos "favorables", en promedio. Un análisis de datos existentes y datos recopilados durante el programa piloto debe mostrar qué características contextuales son las adecuadas para incluirse en los modelos de valor agregado. Al hacerlo, debe reconocerse que la inclusión de (múltiples) mediciones de rendimiento previo por lo general debilitará la relación entre puntuaciones de las pruebas actuales y características socioeconómicas. Al mismo tiempo, la inclusión de ciertas características en el modelo quizá sea valiosa para la aceptación pública y tenga un impacto en las puntuaciones de valor agregado de escuelas individuales.

El éxito del proceso de ajuste depende de lo adecuado del modelo así como del alcance y calidad de las variables usadas en el ajuste. Respecto de lo primero, el ajuste por lo general se efectúa al aplicar un modelo de regresión lineal. Si la relación no es lineal, el modelo no es específico y los estimadores de valor agregado estarán sesgados. El problema en ocasiones se aligera al introducir interacciones entre los predictores. Por ejemplo, tal vez para ciertos grupos inmigrantes haya una brecha de género en desempeño que tenga una magnitud distinta e incluso dirección distinta de la observada en el grupo mayoritario. El modelo de regresión lineal estándar carecería de especificidad y los estimadores de valor agregado resultantes estarían sesgados. El sesgo quizá sea en particularmente problemático si los miembros del grupo minoritario se concentran en ciertas escuelas, lo que en varios sistemas puede ser probable.

Respecto de la segunda consideración, las limitaciones en la recopilación de datos suele ocurrir que sólo un pequeño conjunto de características de los alumnos estén disponibles para el análisis. Si hay características sin medir que se relacionen de manera independiente con el resultado, el modelo de ajuste no estará especificado y, de nuevo, los estimadores resultantes estarán sesgados en cierto grado. Además, la calidad de los datos siempre es una preocupación, pues la calidad deficiente genera aumentos en la varianza y en el sesgo de los efectos estimados de la escuela. Surgen imprecisiones cuando los datos se obtienen de informes de los mismos alumnos, en especial de los más jóvenes. Los datos de informes paternos son problemáticos si los cuestionarios son ambiguos o si los padres no conocen bien el idioma. Incluso los datos administrativos de archivos escolares pueden tener errores graves.

Una ventaja de los modelos de valor agregado es que permiten una evaluación cuantitativa de la magnitud de la desventaja asociada a características particulares (p. ej., origen étnico, ingreso, nivel de escolaridad familiar) en relación con el crecimiento de los estudiantes, no sólo en relación con el aprovechamiento de los alumnos en un momento dado. Los patrones que surgen con el tiempo en estas relaciones son importantes para la elaboración de políticas. Por ejemplo, ¿hay formas particulares de desventaja, se mantienen en el transcurso de la educación de los alumnos, y el impacto de dicha desventaja se extiende o declina con el tiempo? Además, el empleo cuidadoso de los resultados de modelos de valor agregado posibilita identificar a las escuelas más exitosas en la mejora del desempeño de alumnos en desventaja. Esto genera la diseminación de las "mejores prácticas" entre las escuelas, si hay canales que faciliten la transferencia de dicha información.

El análisis realizado por Hægeland y Kirkebøen (2008) demostró, entre otras cosas, que según estándares internacionales **Noruega** tiene un conjunto extenso de datos contextuales disponibles de alumnos para los análisis. Sin duda, el nivel de disponibilidad de datos difiere por país y, por lo general, es la disponibilidad de los datos lo que limita las características contextuales que se incluyen en varios modelos. Por otra parte, la disponibilidad de mediciones del rendimiento

académico previo puede aligerar la necesidad de un conjunto extenso de variables contextuales. En la mayoría de los países se recopila algún tipo de información demográfica de los alumnos y la incluyen en sus modelos de valor agregado. En el Cuadro 4.2 se detalla la variedad de datos contextuales recopilados y disponibles para los modelos de valor agregado en los países participantes. La edad, género y una variable que indica el estado de inmigrante y/u origen étnico del alumno son las principales características incluidas en los países.

Los resultados de varios países ilustran la importancia de incluir una medida de la edad de los alumnos (Ray, 2006; Hægeland *et al.,* 2005). Incluso cuando se excluye a alumnos de edad madura o que repiten un año o grado, la edad de los estudiantes en un año o grado determinado puede variar hasta un año en algunos sistemas. Se ha visto que la edad tiene una relación estadísticamente significativa con el progreso de los alumnos y, por tanto, con la estimación del valor agregado de la escuela. El registro de edad varía por país y, en parte, refleja las diferencias en los métodos de recopilación de datos. En algunos países, los datos de inscripción escolar especifica la fecha de nacimiento de los alumnos, y en otros, la falta de estos datos significa que hay otras fuentes de datos administrativos o que los datos (edad o rango de edad exactos) se obtienen directamente de los alumnos.

El género de los alumnos es una característica presente en la mayoría de los análisis de valor agregado en los países participantes. Esta característica no suele influir en las puntuaciones de valor agregado de las escuelas, pues la distribución de alumnos femeninos y masculinos por lo general es uniforme (con la obvia excepción de escuelas de un solo género). Sin embargo, el género puede ser importante para análisis más detallados de información de valor agregado que fomente iniciativas de mejora escolar. Las diferencias en el desempeño de estudiantes masculinos y femeninos han recibido en los últimos años más atención conforme las alumnas alcanzan niveles más elevados de desempeño y aprovechamiento que los varones en varias áreas y diversas medidas de aprovechamiento. No obstante, la magnitud y quizá la dirección del efecto esperado según una variable de género tal vez difiera según la medición. En algunos países, las comparaciones de desempeño muestran que los alumnos varones rinden más en áreas como matemáticas y ciencias, y las alumnas, en lectura y redacción (OCDE, 2007a; 2007b). Estas disparidades de género quizá no tengan impacto en las estimaciones de valor agregado. Sin embargo, puede ser útil separar por género los análisis de valor agregado en materias específicas, pues los resultados indicarían la necesidad de políticas y programas específicos que abordasen dichas disparidades.

El estado de inmigrante y/u origen étnico se identifica de manera diferente por país y refleja diferencias en la mezcla étnica, el interés de las políticas y los datos disponibles. En algunos países se incluye en los modelos una sola variable que refleja el estado migratorio. En otros se incluyen grupos étnicos o la región específica de donde proviene el alumno inmigrante, pues algunos grupos están en relativa desventaja en comparación con el grupo mayoritario. Los resultados de un análisis de valor agregado por grupos específicos de estudiantes pueden indicar la necesidad de fragmentar más la población estudiantil. Por ejemplo, el análisis de una sola variable que identifique la situación migratoria puede generar una distribución bimodal o una distribución de puntuaciones que incluya distintos grupos. Esto indicaría que grupos étnicos o inmigrantes particulares avanzan con distinto ritmo y que también difieren las contribuciones de la escuela a esos progresos. Hay ciertas evidencias de que estos patrones persisten e incluso crecen con el

tiempo (Borjas, 1995, 2001). Otros análisis pueden indicar qué grupos deben identificarse por separado. En estas situaciones, ni siquiera una medición simple del estado migratorio capturará por completo la desventaja que enfrentan distintos grupos inmigrantes, y por ende no será tan útil para iniciativas de políticas. En algunos casos, las variables de interacción pueden ser útiles, en particular si hay heterogeneidad económica sustancial con grupos étnicos particulares. Para reflejar dichos cambios se requiere flexibilidad tanto en la recopilación de datos como en la tecnología de la información con que se recopilen los datos. Los administradores y los tomadores de decisiones requieren esta flexibilidad para especificar mejor los modelos de valor agregado y generar resultados más útiles, así como se requieren recopilaciones de datos *ad hoc* para objetivos específicos de políticas, como programas dirigidos a regiones o grupos específicos de estudiantes. En algunos países, una preocupación se refiere a las barreras del idioma en el progreso de los alumnos, en particular cuando el idioma de instrucción difiere del hablado en casa o del materno de los alumnos. Se considera que estas barreras revisten particular importancia (desde las perspectivas tanto educativa como política) cuando estos estudiantes manifiestan un desempeño deficiente en varias áreas de estudio.

En el Cuadro 4.2 se organizan variables contextuales por categorías. Esta categorización se efectuó con propósitos ilustrativos y no necesariamente aplica a un país específico. Como ayuda para sus modelos, la mayoría de los países recopila mediciones de dificultades de aprendizaje estudiantil, nivel de escolaridad familiar, nivel de recursos económicos y beneficios de seguridad social. Esto último también puede considerarse una medición de recursos económicos. Algunos países también recopilan características relacionadas con la estructura familiar del alumnado que haya mostrado afectar resultados, como estado civil de los padres, si el alumno vive fuera del ámbito familiar, y una medición del tamaño de la familia (Amato y Keith, 1991). Es importante observar que algunas características son fijas y no cambian durante la estancia del alumno en la escuela, pero otras características pueden cambiar con el tiempo. La recopilación de datos y los sistemas de almacenamiento deben ser lo bastante flexibles para incluir ambas clases de características.

Las características socioeconómicas recopiladas por países se concentran en el nivel de escolaridad paterna y en el de ingreso familiar. Las características que denotan si los estudiantes y/o sus familias reciben beneficios de seguridad social, como apoyo educativo o para el sostenimiento del hogar, también se incluyen en algunos países. Éstas pueden ser indicaciones del nivel de recursos económicos disponibles para los alumnos y sus familias. En la **Comunidad Flamenca de Bélgica** se recopilan diversos datos para formar un registro de estudiantes "En riesgo". **Noruega** también incluye mediciones sobre el nivel del ingreso familiar y la incidencia de desempleo de los padres durante los 10 años anteriores a la evaluación.

En la mayoría de los países se recopilan características que identifican a los alumnos con dificultades en el aprendizaje. La tipología de necesidades de aprendizaje varía por país, y por lo general se alinea con recopilaciones de datos existentes en el sistema educativo. Aunque no se consideran indicadores de necesidades especiales de aprendizaje, varios países incluyen datos que identifican si el alumno repitió un grado en la escuela. Esto es en particular importante si el estudiante repite el grado en el que se administra la evaluación, o un grado entre la evaluación en curso y la anterior. Los estimadores de la contribución de una escuela al avance estudiantil entre ambas evaluaciones pueden tener sesgos por las diferencias en la cantidad de años de instrucción.

Cuadro 4.2

Datos contextuales recopilados en países participantes que en potencia se usarían en modelos de valor agregado

	Información demográfica	Estado migratorio	Dificultades de aprendizaje del alumno	Estructura familiar	Escolaridad familiar	Recursos económicos	Beneficios de seguridad social
Bélgica (Fl.)	Edad, género, país de nacimiento del alumno y ambos padres, edad al inmigrar	Idioma hablado con la madre en casa, antecedentes migratorios	Dificultades de aprendizaje identificadas, historia de repetir grados	Alumno criado fuera de su hogar (p. ej., padres adoptivos, instituciones) como factor del estado de alumno en riesgo (BAR)	Calificación educativa materna		Beca de estudios, ingreso por sustitución de vivienda, hogar dependiente de beneficios de seguridad social como factor del estado de alumno en riesgo (BAR)
Eslovenia	Edad, género		Necesidades educativas especiales				
España	Edad, género	País de nacimiento del alumno y sus padres, edad de inmigración, idioma hablado en casa	Alumnos con necesidades especiales de aprendizaje, historia de repetición de grados	Cuestionario sobre estructura familiar	Niveles educativos de los padres	Niveles de ocupación de los padres, posesiones culturales y de otro tipo en casa	Becas del alumno
Francia	Edad, género, lugar de nacimiento	Nacionalidad, lugar de nacimiento	Clase de los alumnos, opciones de materia			Ocupación de los padres (dividida entre 4 categorías ocupacionales), tamaño de la familia	Ayuda financiera recibida
Noruega	Edad, género, graduación en menos años de lo esperado	Nacido fuera de Noruega, país/región de origen, edad de inmigración		Estado civil de los padres, edad de los padres al nacimiento del primer hijo, número de hermanos y medios hermanos, orden de nacimiento	Mayor nivel de escolaridad completo de los padres	Ingreso familiar, patrimonio familiar (con base en el patrimonio familiar gravable)	Incidentes de desempleo de los padres en los 10 años anteriores
Polonia	Edad, género		Dislexia				
Portugal	Edad, género	Idioma hablado en casa	Calificaciones del alumno, repetición de grados, necesidades educativas especiales	Número de hermanos	Escolaridad de los padres (clasificación ISCED)	Ocupación de los padres, computadora en casa, Internet en casa	Derecho del alumno a apoyos (depende del ingreso familiar)
Reino Unido	Edad, género, grupo étnico	Inglés como primer idioma (alumno)	Alumno registrado con necesidades especiales de aprendizaje			Carencias de ingreso del vecindario (medidas con datos de código postal)	Alumno con derecho a Alimentos Escolares Gratuitos (dependiente del ingreso familiar)
Rep. Checa	Edad, género, lugar de nacimiento		Estudiantes con necesidades especiales de aprendizaje		Último nivel de escolaridad terminado de los padres	Categorías de ocupación de los padres	
Suecia	Edad, género, lugar de nacimiento, grupo étnico	Antecedentes migratorios de estudiantes y padres, año de inmigración			Mayor nivel educativo completo de los padres	Ingreso familiar	Beneficios sociales de la familia

Nota: A diferencia del Cuadro 4.1 con 11 países, Dinamarca no se incluge aqui.
Fuente: Información proporcionada por expertos consultados.

Datos del nivel de escuela

Hasta aquí, el análisis se ha centrado en los ajustes de las características del alumno. También es posible ajustar las características de la escuela o contextuales.[2] Estas características pueden ser agregaciones de variables de los alumnos individuales (p. ej., medias de puntuaciones de exámenes) o de las que sólo se definen en el ámbito escolar (p. ej., composición racial/étnica de la población escolar, nivel socioeconómico de la comunidad). Si bien estas variables se incorporan con facilidad en un modelo, persiste el peligro de sobre-ajustar. Es decir, si la variable contextual se asocia al verdadero desempeño escolar, ajustar por esa variable sesga los estimadores de los efectos de la escuela. Así, se recomienda cautela al decidir efectuar estos ajustes.

En algunos países, el tipo de escuela se incorpora como covariante, aunque esto puede no extenderse a una distinción entre escuelas del gobierno y ajenas al gobierno, pues en ocasiones no se incluye a estas últimas en los análisis de valor agregado. Puede disponerse de más información referente al nivel de los recursos escolares y, en cierto grado, a los procesos escolares. Incorporar covariantes del nivel de la escuela puede ser en particular útil para quienes se interesen en el desarrollo escolar. Los análisis que se centran en ciertos tipos de escuelas o en grupos particulares de alumnos (p. ej., alumnos con necesidades especiales de aprendizaje) son más útiles cuando se emplean variables tanto contextuales como escolares para ajustar los resultados de los alumnos. Un ejemplo es la evaluación de programas, cuando éstos se desarrollan en algunas escuelas pero no en otras. En algunos ambientes, también pueden incorporarse datos del aula para análisis más detallados de valor agregado de los maestros. Como ejemplo, en la **Comunidad Flamenca de Bélgica** se recopila información sobre el uso de libros de texto determinados, género y experiencia del maestro, si hay computadora en la clase, uso de computadoras e internet en las lecciones, y tiempo de enseñanza dedicado a la materia. Estos análisis se aplican con facilidad en análisis de valor agregado más específicos. Los análisis que computan estimadores de valor agregado sobre prácticas escolares para determinar si tienen en cuenta una cantidad sustancial de la varianza en los estimadores de valor agregado son análisis secundarios eficaces, y ofrecen otra opción para los tomadores de decisiones.

Deben darse los pasos adecuados para asegurar la integridad de todos los datos, sin importar si forman parte de una recopilación más amplia de datos administrativos o si se reunieron junto con otros datos para algún uso particular en el análisis de valor agregado. Ray (2006) señala que algunas covariantes escolares están sujetas a manipulación por parte de las autoridades escolares. Con algunos modelos, el impacto de un cambio de la covariante en el valor agregado de la escuela puede anticiparse, y, por tanto, hay un incentivo para cambiar el valor en la dirección deseada. Por ejemplo, en los modelos de valor agregado contextualizado utilizados en el **Reino Unido**, cuanto mayor sea la proporción de alumnos de una escuela sin clasificar respecto de su origen étnico, mayor es su valor agregado, con el resto de condiciones constantes. Así, a la escuela le convendría no inquirir o no informar sobre el origen étnico de sus alumnos. De manera muy sensata, Ray sostiene que los modelos seleccionados deben diseñarse para reducir estos incentivos perversos. Lo ideal sería que estos datos se recopilasen fuera del marco de la evaluación de los alumnos y se reunieran en un sistema que no implicara a la administración escolar, con el fin de reducir la probabilidad de corrupción de los datos.

NOTAS

1. Equiparación mediante preguntas comunes.

2. No son posibles estos ajustes con modelos que incorporen efectos escolares fijos.

CAPÍTULO 5

Modelos ilustrativos de valor agregado

Modelos de valor agregado: Breve reseña histórica 122
Modelos de valor agregado de regresión lineal 124
Modelos de componentes de varianza o de efecto aleatorio 125
Modelos de efectos fijos de valor agregado 126
El modelo Dallas, TX 127
Modelos de respuesta de efecto aleatorio multivariado 129
Análisis de la curva de crecimiento 131
Conclusión 133

Este capítulo comienza con una breve reseña histórica sobre los antecedentes de los modelos de valor agregado, y posteriormente presenta varios modelos de valor agregado para ofrecer algunos ejemplos útiles en sistemas educativos. El objetivo de este capítulo no es presentar una lista o revisión completa de los diversos tipos de modelos de valor agregado, pues esto trasciende el alcance y propósito de este informe. En cambio, los tipos de modelos presentados ilustran algunas de sus diferencias e ilustran cómo se manejan problemas específicos con distintos procedimientos de los modelos. Las características del diseño analizadas en el Capítulo 4 afectan a estos modelos en grados diversos, y cada modelo tiene tanto ventajas como desventajas respecto del conjunto completo de temas. Se presentan cinco categorías generales de modelos de valor agregado: modelos de regresión lineal, de componentes de varianza, de efectos fijos, de respuesta multivariada con efecto aleatorio y alguna discusión sobre los análisis de curva de crecimiento. Los modelos de valor agregado sirven para estimar efectos escolares anuales o acumulados, pero en varios modelos presentados como ejemplos aquí, el efecto de la escuela se mide como efecto anual en lugar de acumulado.

El análisis de estos tipos de modelos también debe brindar información para elegir el modelo más adecuado, en vista de los problemas metodológicos analizados en el Capítulo 6. También debe destacarse que este informe no es partidario de un modelo respecto de otro. En cambio, señala cómo algunos modelos son más adecuados debido a los distintos objetivos de políticas y a las limitaciones con que deben efectuarse los análisis. No obstante, durante la elaboración de un sistema de análisis de valor agregado, es imperativo examinar varios modelos para evaluar su relativa adecuación respecto de diversos criterios.

MODELOS DE VALOR AGREGADO: BREVE RESEÑA HISTÓRICA

Los antecedentes de los modelos actuales de valor agregado se remontan a la década de 1960 y pueden agrupase en tres grandes líneas: la investigación sobre la efectividad de las escuelas, las políticas enfocadas en la rendición de cuentas y el desarrollo de técnicas estadísticas (Martínez-Arias, Gaviria y Castro, 2009). En este contexto, la era moderna de la investigación sobre "efectos escolares" comenzó, al menos en **Estados Unidos de América** (**EUA**), con el *Informe Coleman*, que estudió las relaciones de escuelas y familias con el aprovechamiento académico estudiantil (Coleman, 1996). Esto complementó varios estudios en Europa que abordaron problemas de inequidad en términos de análisis intergeneracionales que comparaban resultados a lo largo de generaciones (Carlsson, 1958; Glass, 1954). Los estudios posteriores sobre eficacia escolar también efectuaron comparaciones cuantitativas entre escuelas. En la fase inicial se identificó a las mejores escuelas al comparar el promedio de las puntuaciones en los exámenes de los alumnos. A menudo, el siguiente paso de los investigadores fue seleccionar una pequeña cantidad de estas escuelas realizando análisis más detallados con el fin de identificar los elementos de su práctica que explicaban su éxito. El propósito final fue divulgar los hallazgos con el fin de efectuar mejoras escolares más ampliamente (Madaus, Airasian y Kellaghan, 1980).

Pronto se reconoció que las clasificaciones de las escuelas basadas en la puntuación "bruta" de las pruebas de los alumnos mantenían una estrecha correlación con el nivel socioeconómico de éstos (Coleman *et al.*, 1996; McCall, Kingsbury y Olson, 2004). Bethell (2005), por ejemplo, analizó algunas polémicas que surgieron por las tablas que comparaban puntuaciones brutas de pruebas en el **Reino Unido**. Se emplearon análisis multivariados transversales para abordar y resolver estos problemas. En la versión más sencilla de estos análisis, las puntuaciones

promedio de las pruebas de las escuelas se incluyeron en un análisis de regresión junto con un número de características demográficas relevantes (agregadas) de los alumnos. La idea fue clasificar a las escuelas con base en los residuos de sus regresiones. Estos residuos a menudo se calificaron como "efectos escolares". Las escuelas con residuos positivos mayores se consideraron ejemplares y dignas de más estudios. Las escuelas con residuos negativos significativos se consideraron problemáticas y también requirieron más estudios, pero de diferente índole. Se propusieron otras estrategias de ajuste y se compararon las diferencias resultantes en las clasificaciones de escuelas (Dyer, Linn y Patton, 1969; Burstein, 1980).

Más adelante se popularizaron modelos transversales más complejos y se emplearon con métodos que toman en cuenta la estructura jerárquica de los sistemas escolares, con los estudiantes anidados en clases, las clases en escuelas y éstas en distritos o áreas locales (Aitkin y Longford, 1986; Goldstein, 1987; Willms y Raudenbush, 1989). Los estimadores provenientes de estos modelos se hicieron mas complejos y ahora son comunes en análisis educativos en los países miembros de la OCDE, por ejemplo. Estas estimaciones transversales se denominan en este informe como ***modelos de aprovechamiento contextualizado***. Estos modelos multivariados ofrecen una medida del desempeño escolar, pero se consideró que estos análisis no contenían el marco analítico requerido para clasificarse como modelos de valor agregado. Los modelos de aprovechamiento contextualizado estiman la magnitud de los factores contribuyentes al desempeño o aprovechamiento estudiantil en un momento dado. Un ejemplo habitual es un modelo de regresión que incluya como predictores un vector de antecedentes socioeconómicos o características contextuales de los alumnos y una variable que identifique la escuela a que asiste cada alumno respecto de alguna medida de aprovechamiento. El ajuste de la puntuación bruta con la inclusión de características contextuales proporciona medidas que reflejan mejor la contribución de las escuelas al aprendizaje de los estudiantes que el empleo de las puntuaciones "brutas" de exámenes para medir el desempeño escolar. Los resultados de estos modelos transversales apoyan los análisis teóricos del papel de la familia en determinar los resultados socioeconómicos de las personas, y a menudo revelan que el principal factor explicativo del nivel de aprovechamiento de los alumnos son los antecedentes socioeconómicos paternos (OCDE, 2007b; Haveman y Wolfe, 1995; Becker, 1964). La información acerca del papel de los antecedentes socioeconómicos de los alumnos en el aprovechamiento educativo, si bien interesante e importante, no suele generar suficiente información que permita a las autoridades educativas formular políticas, tomar decisiones sobre la responsabilidad de la escuela y guiar procesos de mejora escolar. No obstante, estos modelos de aprovechamiento contextualizado representan una clara mejoría sobre el empleo de resultados sin ajustar y puntuaciones brutas de aprovechamiento para evaluar el desempeño escolar.

Un avance significativo fue la elaboración de modelos de valor agregado que empleaban múltiples mediciones de desempeño de los alumnos para estimar el impacto (o agregado de valor) de las escuelas individuales en las medidas del desempeño de esos alumnos. Una evaluación importante de los modelos de valor agregado fue la de Fitz-Gibbon (1997), a quien le fue solicitado por el gobierno del **Reino Unido** asesorar sobre la elaboración de un sistema basado en modelos de valor agregado. Fitz-Gibbon concluyó que un modelo así podía ser la base de un sistema nacional de valor agregado estadísticamente válido y comprensible para la mayoría de las personas. Independientemente de la polémica desatada en **EUA** por la

legislación federal aprobada en 2002 *No Child Left Behind* (NCLB por sus siglas en inglés), su enfoque en la rendición de cuentas de las escuelas y las exigencias a las autoridades educativas estatales sobre medición y progreso anual de alumnos representó un impulso importante al desarrollo de modelos estadísticos de medición, incluyendo los de valor agregado en ese país. De esta forma, actualmente se ha llegado a concebir a los modelos de valor agregado como aquellos que emplean datos que registran las trayectorias de las puntuaciones de exámenes de alumnos individuales en una o más materias en uno o más años (Mortimer *et al.,* 1988; Goldstein *et al.,* 1993; SCAA, 1994; Sanders, Saxton y Horn, 1997; Webster y Mendro, 1997; Rowan, Correnti y Miller, 2002; Ponisciak y Bryk, 2005; Choi y Seltzer, 2005; McCaffrey *et al.,* 2004; McCaffrey *et al.,* 2003; McCaffrey *et al.,* 2005, Martínez-Arias, Gaviria y Castro, 2009). Mediante varias clases de ajustes, los datos de avance de los alumnos se transformaron en indicadores del valor agregado escolar.

MODELOS DE VALOR AGREGADO DE REGRESIÓN LINEAL

Este primer conjunto de modelos emplea la regresión lineal simple para ajustar las puntuaciones de resultados de las pruebas por alguna combinación de puntuaciones de rendimiento previo y características del alumno o del contexto. Una forma del modelo es:

(1) $$y_{ij(2)} = a_0 + a_1 y_{ij(1)} + b_1 X_{1ij} + \ldots + b_p X_{pi} + \varepsilon_{ij}$$

donde:

i subíndice para estudiantes dentro de las escuelas j,

$y_{ij(2)}$ = puntuación de la prueba final,

$y_{ij(1)}$ = puntuación de la prueba anterior,

$\{X\}$ denota un conjunto de características del alumno y familiares,

a_0, a_1, b_1, … b_p representan un conjunto de coeficientes de regresión y

ε_{ij} representa las desviaciones individuales y distribuidas normalmente con una varianza común para todos los estudiantes.

Se representa el valor predicho para el estudiante i en la escuela j por $\hat{y}_{ij(2)}$, basado en el ajuste de la Ecuación (1) a todo el conjunto de datos. Después, el valor estimado para la escuela j se toma como el promedio sobre sus estudiantes de los residuos ajustados: $ave_i\{y_{ij(2)} - \hat{y}_{ij(2)}\}$.

Así, si los estudiantes de la escuela j alcanzan puntuaciones en sus pruebas finales mayores que el promedio (en comparación con estudiantes de otras escuelas con valores predichos semejantes), el residuo correspondiente tiende a ser positivos, lo que genera un valor agregado estimado positivo para la escuela. Hay muchas variantes del modelo básico. En particular, si se dispone de puntuaciones de pruebas anuales de años anteriores o de otras materias, pueden acomodarse sin dificultad. Hay más ejemplos en Ladd y Walsh (2002), y Jakubowski (2007). Para que este método genere estimados consistentes se requiere que las covariantes incluidas no estén correlacionadas con el término de error, lo cual puede incluir un efecto de la escuela además de errores idiosincrásicos. Aún más, no toma en cuenta la estructura del término de error, característico de algunos de los modelos que se ilustran más adelante.

MODELOS DE COMPONENTES DE VARIANZA O DE EFECTO ALEATORIO

Otro tipo de modelos implican dos ecuaciones de regresión: una regresión en el nivel del alumno, como en la Ecuación (1) anterior, y una regresión de la escuela que modela la variación en los interceptos escolares ajustados obtenidos con la regresión en el nivel del alumno. Una ventaja técnica de los llamados modelos jerárquicos (o multinivel) es que tienen en cuenta el agrupamiento de estudiantes dentro de las escuelas, lo que genera estimadores más precisos de la incertidumbre que acompaña a los estimadores de valor agregado.

Una formulación habitual de estos modelos es:

(2) $$y_{ij(2)} = a_{0j} + a_1 y_{ij\,(1)} + b_1 X_{1ij} + \ldots + b_p X_{pij} + \varepsilon_{ij}$$

$$a_{0j} = A + \delta_{0j}$$

donde:

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

$$\delta_{0j} \sim N(0, \tau^2).$$

Se supone que cada residuo en ambas ecuaciones es independiente de los demás residuos. La lógica de la segunda ecuación es que se considera que los interceptos escolares ajustados $\{a_{0j}\}$ se distribuyen de forma aleatoria en torno a una gran media (A), y las desviaciones de esa media se toman como estimadores del valor agregado de la escuela. El interés recae en las escuelas con desviaciones grandes (positivas o negativas). Esta clase de modelos se emplean en los modelos de "valor agregado contextual" que se aplican en el **Reino Unido**, si bien los estimadores de valor agregado escolar reales se obtienen mediante más análisis y cálculos. Más adelante se analiza el modelo utilizado en el **Reino Unido**.

Este tipo de modelos suele denominarse de "efectos aleatorios" porque los parámetros con que se pretenden capturar las contribuciones escolares al desempeño de los alumnos se tratan como variables aleatorias. En consecuencia, el efecto estimado para una escuela en particular recibe la influencia de los datos de todas las demás escuelas, así como los datos de la propia escuela. Los estimadores resultantes en ocasiones se llaman estimadores "de contracción", porque suelen representarse como promedio ponderado del estimador ordinario de cuadrados mínimos para la escuela y un estimador relacionado con los datos para todas las escuelas. La combinación específica depende tanto del modelo como de los datos disponibles. Los estimadores contraídos están sesgados, pero por lo general tienen un error cuadrático medio menor que los estimadores de mínimos cuadrados ordinarios.

Con los modelos multinivel, la varianza residual se particiona en dos niveles: el de alumnos (Nivel 1) y el escolar (Nivel 2). Estos son los "efectos aleatorios" del modelo. Dentro de un sistema educativo, es posible tener otros niveles. Por ejemplo, *dentro* de las escuelas, los estudiantes se agrupan en clases, pero si no hay datos nacionales sobre grupos de enseñanza, no puede modelarse este nivel. Los residuos del Nivel 1 muestran variación en los resultados de los alumnos en relación con sus escuelas. Los residuos del Nivel 2 muestran los resultados de las escuelas en relación con los resultados nacionales esperados, dadas las covariantes incluidas. Estos residuos de Nivel 2 son las puntuaciones de valor agregado escolares.

Un modelo muy relacionado es el modelo de componentes de varianza (véase Raudenbush y Willms, 1995, p. 321) con un conjunto diferente de covariantes de nivel uno y/o dos, según el tipo de efecto escolar (*tipo A* o *tipo B*) que el analista pretenda estimar. El modelo es como sigue:

(3) $y_{ij} = \mu + \beta_W\left(x_{ij} - \bar{x}_j\right) + \beta_b\bar{x}_j + u0_j + \varepsilon_{ij}$

donde y_{ij} es el resultado de la puntuación de la prueba del estudiante *i* en la escuela *j*; x_{ij} es el rendimiento previo del estudiante, $\bar{x}j$ es la media de la muestra en el rendimiento previo de la escuela *j*; u_{0j} es el componente aleatorio de la escuela, también llamado efecto aleatorio o valor agregado de la escuela *j*, que se supone distribuido normalmente con una media de cero y varianza σ^2_{u0}; y ε_{ij} es el componente aleatorio del alumno individual que se supone distribuido idéntica, independiente y normalmente con una media de cero y una varianza σ^2_{ε}. Los parámetros fijos μ, $ß_w$ y $ß_b$ representan, respectivamente, la media de las puntuaciones en las pruebas, el coeficiente de regresión dentro de la escuela que relaciona el rendimiento previo del estudiante con la puntuación resultante de la prueba, y la pendiente escolar intermedia.

Antelius (2006, p. 4) ilustra cómo calcular, con un modelo de componentes de varianza, el valor agregado de las escuelas de educación media en **Suecia**. Se supuso que los grados obtenidos al completar la educación obligatoria reflejaban el conocimiento previo de los estudiantes y los antecedentes educativos, mientras que las calificaciones de la escuela de educación media mostraban el nivel de conocimientos que alcanzaron los estudiantes en las materias básicas (matemáticas, ciencias naturales, sueco, inglés, ciencias sociales, actividades artísticas, educación física y estudios de higiene y religiosos). Las mediciones de cada escuela se presentan por un periodo de tres años para determinar si este valor cambia con el tiempo o no (Antelius, 2006).

En **Portugal** se consideró el análisis de tres modelos de componentes de varianza para la región de Cova da Beira, que implicó una muestra representativa de estudiantes de niveles educativos de primaria, secundaria y media (Vicente, 2007). En cada modelo se incluyó un conjunto diferente de variables predictoras: un modelo nulo, un modelo Tradicional de Valor Agregado (TVA) que abarcó el nivel socioeconómico y el rendimiento previo de los estudiantes, y además un modelo que incluyó otras variables del nivel del alumno, como género, si el estudiante calificó como de necesidades especiales, si asistieron al jardín de infancia, tipo de clase en la educación primaria y repetición de grados (TVA+). La correlación entre los estimadores de valor agregado generados por los modelos Nulo y TVA varió de 0.61 a 0.94 según el grado. En cambio, con excepción de las puntuaciones del 3er grado, los valores de la correlación entre los estimadores TVA y TVA+ fueron iguales o más grandes que 0.96. Ferrão y Goldstein (2008) también evaluaron el impacto del error de medición en esos estimadores.

MODELOS DE EFECTOS FIJOS DE VALOR AGREGADO

Una estrategia distinta emplea los llamados modelos de efectos fijos. Como su nombre indica, estos modelos representan las contribuciones de las escuelas como parámetros fijos, a diferencia de los modelos de efectos aleatorios, donde se supone que las contribuciones escolares son variables aleatorias con una distribución común. En los modelos de efectos aleatorios, las correlaciones entre covariantes y efectos aleatorios introducen sesgos en los

estimadores de los efectos de la escuela. No existe ese problema con los modelos de efectos fijos, y esto, se sostiene, es su principal ventaja. Por otra parte, los efectos estimados de la escuela pueden variar considerablemente de un año al otro, pues no hay "contracción". A continuación se presenta una versión sencilla de este modelo:

(4) $y_{ij(2)} = a_0 + a_1 y_{ij(1)} + \sum_k b_{kij} X_{kij} + \theta_j + \varepsilon_{ij}$

donde:

θ_j = efecto de la escuela j.

Hægeland y Kirkebøen (2008) analizan con un modelo de efectos fijos el valor agregado escolar en **Noruega**. Proporcionaron una ilustración empírica de la forma cómo los estimadores de desempeño escolar se veían afectados por la elección de las variables socioeconómicas contextuales que se incluyeran en modelos de aprovechamiento contextual o de valor agregado. Los autores observan que ajustar el rendimiento previo de los alumnos y ajustar el nivel socioeconómico no son estrategias mutuamente excluyentes para estimar el desempeño escolar. También queda claro que el papel de los factores contextuales puede diferir por países y tipo de modelo empleado.

EL MODELO DALLAS, TX

Un modelo muy conocido que combina las características de diferentes clases de modelos es el de dos etapas empleado en Dallas, Texas, presentado en Webster y Mendro (1997; véase también Webster, 2005). El papel de la primera etapa fue ajustar las variables de las puntuaciones de las pruebas de los alumnos (puntuaciones actuales y anteriores) que aparecieran en la segunda etapa. El ajuste se efectuó con varias características de los alumnos relevantes. En la segunda etapa, la puntuación actual ajustada se regresó sobre las puntuaciones previas ajustadas con un modelo jerárquico lineal que tomó en cuenta el agrupamiento de estudiantes dentro de las escuelas. Además, este modelo se adaptó con facilidad a la inclusión de covariantes escolares que mejorasen las características estadísticas de los estimadores resultantes del valor agregado de las escuelas. Específicamente, sea:

(5) $y_{ij} = b_0 + b_1 X_{1ij} + \ldots + b_p X_{pij} + \varepsilon_{ij}$, donde:

i representa a los estudiantes dentro de las escuelas j,

y representa un resultado de puntuaciones actual o anterior,

$\{X\}$ representa a un conjunto de características estudiantiles, como origen étnico/competencia en el idioma, género, nivel de pobreza del estudiante, interacciones de primer y segundo orden entre estas características, y varios indicadores del nivel socioeconómico de su vecindario,

$\{b\}$ representa un conjunto de coeficientes de regresión y

ε_{ij} representa desviaciones individuales distribuidas normalmente con una varianza común para todos los estudiantes.

Así, se estiman los coeficientes de la Ecuación (5) por cada posible elección de y. Por lo general se emplean mínimos cuadrados ordinarios. Sin embargo, el interés no recae en los coeficientes estimados, sino en los residuos de la regresión. Por cada regresión ajustada, se estandarizan los residuos. Suponga que usamos ~ para representar un residuo estandarizado.

La Etapa 2 emplea un modelo de dos niveles. El Nivel 1 adopta la forma siguiente:

(6) $\tilde{Z}_{ij} = c_{0j} + c_{1j}\,\tilde{P}_{ij}^{1} + c_{2j}\,\tilde{P}_{ij}^{2} + \delta_{ij}$

y el nivel adopta la forma:

(7) $c_{0j} = G_{00} + \sum_{k=1}^{m} G_{0k} W_{kj} + u_{0j}$

$$c_{1j} = G_{10} + \sum_{k=1}^{m} G_{1k} W_{kj}$$

$$c_{2j} = G_{20} + \sum_{k=1}^{m} G_{2k} W_{kj}.$$

En el Nivel 1:

i es el subíndice para los estudiantes dentro de las escuelas j,

$\tilde{Z}_{ij}$ representa la puntuación de prueba actual ajustada del estudiante,

$\tilde{P}_{ij}^{1}$ y $\tilde{P}_{ij}^{2}$ son las puntuaciones de pruebas de rendimiento previo ajustadas del estudiante,

$\{c\}$ representa un conjunto de coeficientes de regresión y

δ_{ij} representa desviaciones individuales distribuidas normalmente con una varianza común para todos los estudiantes.

Observe que el término "ajuste" se refiere a los resultados obtenidos al efectuar el análisis de la Etapa 1. En principio, pueden emplearse más de dos mediciones anteriores de rendimiento previo.

En el Nivel 2:

$\{W\}$ representa un conjunto de características escolares m, incluso varios indicadores de la composición demográfica de la escuela, múltiples indicadores del nivel socioeconómico de la comunidad escolar, movilidad escolar y del tamaño de la escuela,

$\{G\}$ representa una matriz de coeficientes de regresión y

u_{0j} denota una desviación de la escuela específica de su intercepto en la ecuación de Nivel 1 de la regresión lineal general que relaciona los interceptos escolares con las características escolares.

El modelo de la Etapa 2, semejante a un modelo de efectos aleatorios, se ajusta con *software* multinivel. El efecto escolar estimado es de nuevo un estimador de confiabilidad ajustada de u_{0j}. Esto en ocasiones se llama estimador empírico-bayesiano porque es igual al estimador de u_{0j} obtenido de una regresión de cuadrados mínimos para sólo esa escuela contraído hacia el plano de regresión estimado, y su cantidad de concentración es inversamente proporcional a la precisión relativa de dicho estimador (en Braun, 2006b, hay una introducción a la metodología empírico-bayesiana). El índice de desempeño general para una escuela en particular se elabora como promedio ponderado de los efectos escolares estimados de diversos cursos y grados. En Dallas, las ponderaciones se determinaron por adelantado por un grupo planeado de interesados, el *Accountability Task Force*.

En el **Reino Unido** se empleó una versión simplificada de un modelo multinivel para facilitar la interpretación eficaz a los interesados. Un ejemplo de estas acciones es la decisión de no incluir ninguna variable explicativa sobre el componente aleatorio del modelo. Esta decisión simplifica el modelo pero introduce la suposición de uniformidad en el valor agregado entre estudiantes dentro de las escuelas, de modo que ese desempeño se ilustra con una sola puntuación de valor agregado. Una estrategia más compleja es suponer variación dentro de las escuelas produciéndose diversas mediciones por escuela. Una característica significativa de los modelos multinivel es la aplicación de la "contracción", donde las puntuaciones de valor agregado para las escuelas pequeñas tiende a acercarse más a la media nacional, lo que disminuye la probabilidad de que se registren para esas escuelas puntuaciones extremas de valor agregado. El modelo se mantiene relativamente sencillo: en teoría, puede tener más niveles de análisis y más variables explicativas en las partes tanto "fija" como "aleatorias" del modelo.

MODELOS DE RESPUESTA DE EFECTO ALEATORIO MULTIVARIADO

El modelo Sistema de Evaluación de Valor Agregado Educativo (*Education Value-Added Assessment System*, EVAAS) es un ejemplo de modelo de efectos multivariados, longitudinales y mixtos; es decir, se recopilan datos de estudiantes en múltiples materias en varios grados. Si bien el modelo EVAAS aún se actualiza un poco con el tiempo, todavía no se cuenta con versiones publicadas, y una aplicación reciente adopta la forma siguiente:

Sea:

i subíndice de estudiantes,

j subíndice de transiciones y

n_i la escuela a la que asiste el estudiante i.

Entonces, el modelo bivariado tiene la forma:

(8) $$(y_{ij}, z_{ij}) = (\mu_j, \gamma_j) + \sum_{k \le j} (\theta_{n_i k}, \varphi_{n_i k}) + (\varepsilon_{ij}, \delta_{ij}); \ (j = 1, 2, 3)$$

donde:

y_{ij} representa la puntuación de lectura del estudiante,

z_{ij} representa la puntuación de matemáticas del estudiante,

μ_j representa la puntuación de lectura promedio respecto de toda la población,

γ_j representa la puntuación de matemáticas promedio respecto de toda la población,

$\theta_{n_i k}$ representa un efecto de la escuela en lectura,

$\varphi_{n_i k}$ representa un efecto de la escuela en matemáticas y

ε_{ij} y δ_{ij} son los términos de error aleatorio en lectura y matemáticas, respectivamente.

Se supone que los parámetros $\{\mu\}$ y $\{\gamma\}$ son fijos, mientras que se supone que los parámetros $\{\theta\}$ y $\{\varphi\}$ son aleatorios y conjuntamente independientes. Sean $\underline{\varepsilon_i} = (\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i3})$ y $\underline{\delta_i} = (\delta_{i1}, \delta_{i2}, \delta_{i3})$, entonces se supone que $(\underline{\varepsilon_i}, \underline{\delta_i})$ sigue una distribución normal multivariada con vector de media cero y matriz inestructurada de covarianza definida positiva. Condicional sobre los demás parámetros del modelo, se supone que $(\underline{\varepsilon_i}, \underline{\delta_i})$ son independientes entre los estudiantes.

La suposición de normalidad conjunta de los términos de error es crítica en los modelos multinivel de este tipo para corregir la asignación confundida o no aleatoria.

El modelo de capas en ocasiones se denomina *modelo de persistencia* porque los efectos de la escuela en una transición se trasladan a las transiciones siguientes. Por lo general, la matriz de varianza-covarianza para los componentes de error de alumno se deja sin estructurar. Se supone que es común para todos los estudiantes dentro de la cohorte, pero puede variar entre cohortes. En consecuencia, la cantidad de parámetros es grande, y se requieren abundantes datos para una estimación precisa.

Debe quedar claro que tanto los requerimientos de la base de datos como las exigencias de computación son de capital importancia. El modelo EVAAS se implementa con *software* propietario, y el modelo recién descrito ha servido para analizar datos de más de cien distritos escolares durante más de una década. Hace poco se modificó, pero no hay descripciones disponibles aún. Una versión más compleja del modelo EVAAS se emplea para estimar efectos del docente. Los modelos para la escuela y del profesor pueden aplicarse al mismo tiempo, y así se hace, pero hay pocos análisis en la bibliografía referentes sobre cómo usar de manera conjunta los dos conjuntos de efectos estimados.

El atractivo principal del modelo EVAAS es que, dado que está centrado en el progreso de los alumnos en varias evaluaciones, no da ventajas obvias a las escuelas con estudiantes que entran con puntuaciones de pruebas comparativamente altas. Otro atractivo es que no es necesario descartar los registros de los alumnos con datos perdidos. Se espera que falten datos. Estudios recientes apoyan la solidez de los estimadores obtenidos de EVAAS como punto de partida para hacer suposiciones sobre la naturaleza de los datos perdidos (Lockwood y McCaffrey, 2007). Una distinción obvia entre los modelos Dallas y EVAAS es que el segundo no incluye covariantes de los alumnos ni de las escuelas. Como el modelo Dallas emplea datos de sólo dos ocasiones de medida, debe depender de ajustes de covarianza para hacer más equitativas las comparaciones entre escuelas. Además, la consideración de imperativos políticos y la aceptabilidad para los interesados ofrecen un impulso adicional para incorporar las características de los alumnos al modelo de Etapa 1. Por otra parte, Sanders *et al.* (1997) sostienen que, con datos longitudinales multivariados, cada estudiante actúa como su propio "control", y esto elimina la necesidad de incorporar estos datos al modelo (Sanders *et al.,* 1997; Ballou, Sanders y Wright, 2004). Si bien sin duda es cierto que las puntuaciones de ganancia simple mantienen una correlación más débil con las características de los alumnos que las puntuaciones actuales, la afirmación de Sanders no es una certeza matemática y requiere más investigaciones.

Con este fin, Sander y Wright (2004) mostraron que las covariantes de los alumnos pueden incluirse en el modelo EVAAS para maestros sin introducir sesgos en la estimación de los efectos de los profesores (denotados como EVAAS-C). Aplicaron ambos modelos a los datos de un distrito escolar y hallaron que los efectos estimados de los profesores de los dos modelos eran muy semejantes. En otras palabras, los estimadores EVAAS fueron sólidos para la inclusión de covariantes de los alumnos. Es una pregunta pendiente si estos hallazgos se pueden generalizar a otros ambientes y a la estimación de los efectos de los estudiantes.

Para algunos estudiosos, el hecho de que el EVAAS no emplee covariantes de alumnos es una ventaja porque no se sugiere que haya diferentes expectativas para estudiantes con distintos

antecedentes personales. Por otra parte, puede haber situaciones en que las consideraciones no estadísticas, por ejemplo, lleven a la adopción de EVAAS-C y no EVAAS. Debe tenerse presente que ajustar las covariantes de los alumnos en modelos menos exhaustivos que EVAAS puede sesgar las estimaciones de desempeño de la escuela (p. ej., los niveles más altos de escolaridad paterna se correlacionan con escuelas que tienen maestros más cualificados), y ajustar la covariante puede generar un estimador más bajo de desempeño escolar.

Goldstein (1987) ofrece otro ejemplo de un modelo de respuesta multivariada que permite la clasificación cruzada de estudiantes tanto por sus escuelas primarias y como secundarias. Los resultados del modelo de clasificación cruzada sugieren que el valor agregado de la escuela secundaria recibe la influencia de la escuela primaria en particular a la que asistió el estudiante. Otro ejemplo se encuentra en la obra de Ponisciak y Bryk (2005). A partir del trabajo previo del Consorcio de Investigaciones de Escuelas de Chicago, se desarrolló un modelo de clasificación cruzada de tres factores, que denominaron HCM3. El modelo aprovecha los registros longitudinales de los estudiantes en una sola materia. Se efectuaron análisis separados para cada materia. Se clasificó de manera cruzada a los estudiantes por la clase y escuela a la que asistieron en cada grado. Como señalan los autores, su "modelo es una combinación de dos modelos más sencillos: un modelo de dos niveles para el crecimiento del aprovechamiento de los estudiantes con el tiempo y un modelo de dos niveles para el valor que cada escuela y aula agrega al aprendizaje del alumno con el tiempo" (Ponisciak y Bryk, 2005, p. 44).

Aunque la versión final del modelo es muy compleja, la idea básica es muy sencilla. Se supone que cada estudiante tiene una trayectoria de crecimiento lineal latente. La pendiente de esa trayectoria en un año y grado determinados se desvía, positiva o negativamente, por los efectos combinados del aula y la escuela en ese año. Se supone que la desviación es permanente; es decir, persiste hasta la siguiente evaluación y más allá. Observe que en este modelo se supone que las puntuaciones de escalas de las pruebas se tratan como si fueran una escala de intervalos, suposición que, en el mejor de los casos, es una aproximación.

ANÁLISIS DE LA CURVA DE CRECIMIENTO

Debe darse asimismo alguna consideración a los análisis de la curva de crecimiento que utilizan datos longitudinales con más de dos ocasiones de medida de desempeño del alumno para estimar la contribución de las escuelas al crecimiento estudiantil en ese desempeño. Una curva de crecimiento (de desempeño) se describe con una curva de crecimiento de una medida de desempeño (u otro resultado) a lo largo del tiempo. Al estimar las curvas de crecimiento, el modelo suaviza las mediciones observadas para estimar las trayectorias continuas que se cree subyacen a las observaciones. Los modelos de curvas de crecimiento suponen que hay una curva de crecimiento latente que da origen a las puntuaciones en las ocasiones de medición (es por esto que a veces se les llama "modelos de curva de crecimiento latentes"). En un análisis de curva de crecimiento individual, se estima una curva de crecimiento para cada materia para representar el desarrollo a lo largo del tiempo. Con curvas de crecimiento lineales, se estiman dos parámetros de crecimiento, a saber, un parámetro de crecimiento de nivel inicial (intercepto o estado) y un parámetro de tasa de crecimiento (crecimiento o pendiente). Ambos parámetros varían entre individuos, lo que significa que se estima una curva de crecimiento por individuo con un nivel inicial específico y una tasa específica de cambio.

Hay un "modelo de crecimiento básico" para una cohorte que entra en un grado o año particular:

(9) $E[y_{it}] = c_{0i} + c_{1i}t$

Aquí:

i es el subíndice para estudiantes y t es el subíndice para grados,

E representa el operador esperanza,

y representa la puntuación de la prueba y

c_0 y c_1 representan el nivel inicial y la pendiente de crecimiento.

Se supone que el par (c_0, c_1) está distribuido aleatoriamente entre los estudiantes de la cohorte. La Ecuación (10) representa la trayectoria de crecimiento latente para el estudiante i en ausencia de efectos de clase y de escuela. Ahora, v_t denota la desviación de la pendiente por la clase y la escuela la que está inscrito el estudiante en el grado t.

Así:

(10) $E[y_{it}] = c_{0i} + tc_{1i} + \sum_{k=1}^{t} v_k$

El último término del lado derecho, el sumatorio, representa la contribución acumulada de los efectos de clase y de escuela en los grados t. Se supone que $\{v\}$ (los efectos escolares) es aleatorio entre aulas dentro de las escuelas e independiente de los efectos del alumno.

Se introduce mayor complejidad al tener en cuenta la realidad de los sistemas escolares. Por ejemplo, puede haber cambios esenciales en el sistema y afectar a todos los estudiantes que entraron en el sistema en un año dado y están inscritos en un grado particular. Se supone que dichos cambios desplazan la media de la cohorte de ese grado/año. Además, se introduce un efecto aleatorio por cada escuela al considerar los efectos de selección debido a que no se asigna a los estudiantes de manera aleatoria a las escuelas. El modelo también puede extenderse para reflejar los cambios de los efectos de clase y de escuela en el tiempo. Para mayores detalles, consulte Ponisciak y Bryk (2005). La referencia citada contiene un extenso análisis de datos del sistema de Escuelas Públicas de Chicago, así como una comparación de los resultados de HCM3 con los de modelos más sencillos. Choi y Seltzer (2005) propusieron un modelo muy relacionado, que utiliza modelos de regresión de variable latente. Véase también la revisión de Choi, Goldschmidt y Yamashiro (2005).

Como los modelos de curva de crecimiento son un tipo de modelo multinivel (medidas anidadas en los estudiantes), es sencillo incluir un nivel adicional, como el escolar (estudiantes en las escuelas), con el fin de estimar los residuos escolares. Estos residuos escolares reflejan la contribución relativa de una escuela al nivel y tasa de crecimiento de sus alumnos con el tiempo y, así, sirven como puntuaciones de valor agregado de las escuelas. Los modelos de crecimiento son intuitivamente atractivos y pueden aplicarse en sistemas con una mayor cantidad de observaciones de desempeño estudiantil (los modelos de curva de crecimiento no se ajustan a situaciones en las que sólo se dispone de dos mediciones de desempeño de los alumnos). Los modelos dependen en gran medida de la calidad del conjunto de datos longitudinales, y deben considerarse aspectos como movilidad estudiantil y repetición de grado (estos temas se analizan con más detalle en el Capítulo 6).

CONCLUSIÓN

Este capítulo presentó algunos ejemplos básicos de modelos de valor agregado y analizó sus propiedades estadísticas, que ilustran las ventajas y desventajas de su empleo en circunstancias específicas. Cada modelo tiene distintos requerimientos de datos, y por tanto cada uno implica distintos costos asociados a su implementación. Asimismo, diferentes modelos se ajustan a objetivos de políticas y analíticos particulares, por lo que es imposible declarar, *a priori,* que hay un modelo "verdadero" o "mejor" para todos los sistemas educativos. En cambio, es necesario analizar cómo aprovechar cada modelo para cumplir con los objetivos requeridos y los criterios estadísticos deseados durante la etapa de implementación del sistema basado en modelos de valor agregado.

El Capítulo 6 analiza otros criterios destinados a comprender las características estadísticas operativas de diversos modelos de valor agregado de modo que las autoridades educativas y los administradores seleccionen de manera informada un modelo adecuado.

CAPÍTULO 6

Elección de modelo: Aspectos estadísticos y metodológicos

Criterio estadístico: Varianza y estabilidad intertemporal **137**

Criterio estadístico: Sesgo **140**

Criterio estadístico: Error cuadrático de la media **142**
- Datos perdidos **142**

Elección de modelo en análisis de valor agregado **144**

Conclusión **147**

El objetivo de este capítulo es ayudar a los administradores y a las autoridades educativas en su toma de decisiones referentes a la elección del modelo de valor agregado adecuado para su sistema educativo. La decisión de emplear modelos de valor agregado y, de ser así, un modelo en particular, implica muchos factores, tanto técnicos como no técnicos. Algunos aspectos básicos de diseño se abordan en los Capítulos 4 y 5. El tema central de este capítulo son las consideraciones estadísticas y metodológicas cuya importancia radica en que, al explicarlas, se revelan los puntos fuertes y débiles de los diversos modelos en varios contextos. Ni siquiera al juzgarse con criterios puramente técnicos hay ocasiones, si hubiera, en que sobresalga un solo "mejor modelo" para aplicarse en todas las situaciones. Aunque los análisis técnicos pocas veces son definitivos, sí contribuyen a tomar una decisión informada. Además, si se implanta un modelo de valor agregado, la apreciación de sus puntos fuertes y débiles reduce el riesgo de interpretaciones inadecuadas y del mal uso de las estimaciones de las puntuaciones escolares de valor agregado.

Hay tres aspectos estadísticos principales a considerar. El primero es la varianza de los estimadores, incluso su inestabilidad intertemporal, que puede ser un problema particularmente importante debido a la dificultad de desentrañar los cambios verdaderos de desempeño escolar entre varias fuentes de ruido. El segundo aspecto son los sesgos y la solidez como punto de partida de las asunciones en los que descansan los modelos. Por último, está la cuestión del grado de semejanza entre los estimadores de valor agregado que producen los distintos modelos. La Parte III de este informe incluye un análisis de la forma de aplicar en la práctica estos criterios en la elección el modelo más adecuado en la etapa piloto del proceso de implantación. El material de este informe debe permitir a las autoridades educativas emplear la estimación apropiada y ganar la confianza de los interesados en la estimación de valor agregado.

Antes de proceder con la tarea principal del capítulo, vale la pena recordar la razón por la que lidiamos con este conjunto de complejos asuntos. Desde el punto de vista de las políticas, es en extremo importante la capacidad de identificar escuelas inusualmente tanto eficaces como ineficaces. Estos indicadores basados en datos sirven junto con otros indicadores para varios propósitos, como evaluación, mejoras o provisión de información al público. Por intuición, es posible imaginar que los datos longitudinales (agregados) de las puntuaciones de las pruebas sirven para emitir juicios creíbles sobre la calidad escolar. Sin embargo, es muy difícil construir un sistema de evaluación adecuado.

La aplicación de un modelo de valor agregado a un conjunto particular de datos pretende generar estimadores de las contribuciones de las escuelas al avance de los estudiantes. El objetivo es aislar la contribución de la escuela en sí (su personal, políticas y recursos) sobre el aprendizaje estudiantil. En otras palabras, el uso de estos modelos pretende emular (en la mayor medida posible) las condiciones de un experimento aleatorio. Esto no es fácil, y los criterios estadísticos por analizar sirven como base para decidir cuán adecuado es un modelo en particular para alcanzar esta meta en un ambiente específico. El modelo preferido variará entre los sistemas educativos por sus diferentes objetivos, muestras y datos contextuales, así como por la naturaleza de las evaluaciones de los alumnos. Desde un punto de vista práctico, no debe elegirse un modelo sin una extensa prueba piloto, análisis y consultas con varios interesados. Estas consideraciones se analizan más en la Parte III.

CRITERIO ESTADÍSTICO: VARIANZA Y ESTABILIDAD INTERTEMPORAL

Por lo común, la aplicación de un modelo de valor agregado produce un conjunto de estimaciones de los efectos de la escuela, junto con estimadores de las varianzas de dichos estimadores. La varianza (estimada) del efecto de una escuela es una medición de la incertidumbre que acompaña a ese estimador. En general, la cantidad de varianza se determina en gran medida por el modelo de valor agregado particular en uso y por la cantidad de datos disponibles, en especial, la de observaciones que se obtienen de la escuela. Los estimadores de varianza son importantes no sólo porque proporcionan un contrapeso a la inclinación natural a interpretar en exceso pequeñas diferencias entre efectos escolares. También sirven para elaborar intervalos de confianza en torno a los efectos escolares estimados.

Es obvio que se preferiría que las varianzas fuesen lo más pequeñas posible, pues así se generan intervalos de confianza pequeños. Cuando los intervalos de confianza son pequeños en comparación con la dispersión entre las medidas estimadas de desempeño de la escuela, se pueden identificar con facilidad las escuelas "extremas". Es decir, las escuelas con efectos verdaderos sustancialmente mayores (o menores) que el promedio, por lo general se asocian a estimadores relativamente precisos y se juzgan estadística y significativamente diferentes del promedio. De acuerdo con esto, se invierten esfuerzos sustanciales en reducir el nivel de las varianzas de los estimadores de desempeño escolares. Esto suele implicar la obtención de más datos relevantes (p. ej., secuencias más extensas de puntuaciones de pruebas o datos procedentes de pruebas en múltiples materias), así como la selección de un modelo que permita un uso más eficiente de los datos a mano.

Un elemento clave para elegir un modelo de valor agregado adecuado es la estabilidad de los resultados a lo largo del tiempo. Si las puntuaciones de valor agregado de las escuelas fluctúan sustancialmente y, aún mas, de manera aparentemente aleatoria, es difícil confiar en que se obtienen estimadores precisos de la contribución de una escuela al crecimiento del desempeño de los estudiantes. Una reducción de la confianza puede tener graves repercusiones para diversos interesados en el sistema educativo, en particular para quienes soporten un sistema punitivo de rendición de cuentas escolar. Los resultados de estabilidad de las escuelas deben por tanto analizarse en el desarrollo de los modelos de valor agregado y en la supervisión normal del sistema. Sin embargo, como algunos cambios en las puntuaciones de valor agregado de las escuelas se esperan y desean con el tiempo, hay dificultades al determinar si la inestabilidad se debe a cambios reales del desempeño escolar o sólo son fluctuaciones al azar.

Las correlaciones año con año de los estimadores de valor agregado de las escuelas dependen del tamaño de la escuela, del tipo de modelo, de la cantidad de variables contextuales, de la cantidad de años entre resultados de rendimiento previo y de la cobertura de la comparación (todas las escuelas del país o algún subgrupo). Cuando se calculan de forma anual los efectos de las escuelas es común hallar que muchos fluctúan con mucha amplitud. Kane y Staiger (2002) observaron este fenómeno en Carolina del Norte. Algunas escuelas mostrarán cambios inusuales basándose en los datos empleados en el modelo de valor agregado, pero para ciertas escuelas es difícil afirmar si un ascenso o descenso de valor agregado parece "genuino". Pueden emplearse datos más detallados de valor agregado (p. ej., de modelos para materias o subgrupos dentro de una escuela) para establecer si los cambios son verosímiles.

Como ejemplo, se emprendió un análisis de datos del **Reino Unido** de la estabilidad del valor agregado escolar y puntuaciones de valor agregado contextualizado en comparación con la estabilidad de los resultados brutos en las escuelas (Ray, 2007). El Cuadro 6.1 muestra el cambio promedio absoluto en cada una de estas medidas y la desviación estándar de estos cambios. Todas estas estadísticas se presentan en las mismas unidades: puntos de la Etapa 4. Los resultados brutos se incrementaron entre 2005 y 2006, mientras que las puntuaciones de valor agregado y valor agregado contextualizado cambiaron poco en promedio porque son medidas relativas. Es importante observar que las desviaciones estándar de estos cambios son de dimensiones semejantes. Los resultados aquí revelan que, si bien las puntuaciones de valor agregado y de valor agregado contextualizado son más variables que las brutas en términos relativos (p. ej., medidas según las correlaciones entre 2005 y 2006), la estabilidad no es necesariamente menor para el valor agregado en términos absolutos. De hecho, la estabilidad en este caso es un poco mayor para las puntuaciones de valor agregado y de valor agregado contextualizado que para los resultados brutos, y la estimación de valor agregado es la que produce la medición más estable.

Cuadro 6.1

Cambios absolutos en valor agregado contextualizado (VAC), valor agregado (VA) y resultados brutos (APS): Resumen de estadísticas, Etapa 4, 2005-06 (RU)

	Media del cambio	Desviación estándar de los cambios	Cambio del percentil 25	Mediana del cambio	Cambio del percentil 75
Cambio de APS brutas	5.4	14.9	-4.1	4.9	14.2
Cambio de VA	-0.1	12.3	-7.9	-0.4	7.3
Cambio de VAC	-0.3	13.4	-8.1	-0.4	7.5

Fuente: Ray, A. (2007).

Tres factores que afectan la estabilidad de las puntuaciones de valor agregado a lo largo del tiempo, distintos de la variación del desempeño real de las escuelas, son: los cambios en los instrumentos de evaluación, los cambios de los datos que se acompañan (por lo general, contextuales) y la mayor volatilidad de los resultados de escuelas pequeñas. Las características de las puntuaciones de las pruebas varían de un año al siguiente por un control insuficiente de desarrollo de la prueba, problemas en la equiparación de las formas de las pruebas, o incluso cambios planeados. De modo semejante, puede haber cambios en la cantidad, significado y calidad de las variables para el ajuste. Un remedio común recomendado en este informe es emplear medias móviles de tres años para las puntuaciones de valor agregado de las escuelas. Esto tiende a suavizar las fluctuaciones aleatorias y debe proporcionar mediciones más estables. El costo de este procedimiento es que puede dificultar la identificación de cambios verdaderos en la eficacia de las escuelas. Las medias móviles de tres años son aplicables a los resultados de cualquier modelo de valor agregado. En particular, recuerde que los modelos llamados de efectos aleatorios manifiestan una característica importante, a saber, que los estimadores del valor agregado de las escuelas se "contraen" hacia el promedio general de cero, y la cantidad de contracción mantiene una relación inversa con la cantidad relativa de información disponible de la escuela. Así, los estimadores de escuelas pequeñas tienden a experimentar una buena cantidad de concentración, que contribuye a la estabilidad pero, de

nuevo, dificulta más identificar a las escuelas significativamente diferentes del promedio. En cierto sentido, se trata de una versión de la conocida compensación entre los errores de Tipo I y Tipo II. Sin embargo, cabe notar que difieren las opiniones sobre lo adecuado de emplear residuos concentrados en el contexto de un sistema para proporcionar puntuaciones de valor agregado de las escuelas (Kreft y De Leeuw, 1998, p. 52).

Los cambios en las pruebas pueden incrementar o reducir la cantidad de alumnos aprueba o que obtiene mejores calificaciones. Esto puede crear inestabilidad en los indicadores escolares si los modelos dependen de equiparaciones verticales para producir puntuaciones de crecimiento o estadísticas "de progresión".[1] Incluso con puntuaciones de valor agregado que sólo comparen escuelas entre sí y produzcan estimadores centrados en torno al promedio, habría un problema de inestabilidad si los cambios en las pruebas favorecieran a algunas escuelas más que a otras. Por ejemplo, si las tasas de aprobados aumentan en una materia vocacional que forma parte de la medida de resultados de valor agregado y esta materia la estudian sobre todo estudiantes en escuelas determinadas, estas escuelas podrían terminar con puntuaciones de valor agregado más altas que en el año anterior.

Un problema relacionado es la robustez de los resultados de valor agregado con distintos datos. Por ejemplo, suponga que hay dos pruebas de la misma materia, cada una administrada en varios años. Si se aplica el mismo modelo de valor agregado a cada conjunto de datos, ¿cuán semejantes son los resultados? Sass y Harris (2007) efectuaron un estudio para averiguarlo con datos de Florida al estimar los efectos de los profesores y obtuvieron resultados cualitativamente diferentes. Este resultado no sorprende, pues las pruebas se elaboraron con distintos marcos y características psicométricas. No obstante, este hallazgo sirve como recordatorio de que la naturaleza y calidad de los datos de las pruebas puede y debe tener un efecto material en el resultado de los análisis. Hay más elaboración al respecto en Fielding *et al.* (2003) y Lockwood *et al.* (2007).

Cuando el modelo de valor agregado incluye datos contextuales, las discontinuidades también generan inestabilidad. Por ejemplo, en el **Reino Unido**, una Autoridad Local particular que cambie su política sobre el derecho a Alimentos Escolares Gratuitos puede afectar las puntuaciones de valor agregado contextualizado en sus escuelas durante ese año. Al comparar la estabilidad de las puntuaciones de valor agregado contextualizado con puntuaciones brutas, Thomas *et al.* (2007) ilustraron que las correlaciones basadas en puntuaciones brutas son considerablemente más altas. Se observó que las puntuaciones de valor agregado fueron menos estables que los resultados brutos porque estos últimos suelen estar sujetos a factores que las puntuaciones de valor agregado descartaron. Por ejemplo, los resultados de una escuela pueden ser relativamente bajos a lo largo del tiempo porque por lo general tienen estudiantes de ingreso con rendimiento previo bajo y altos niveles de privación; si las puntuaciones de valor agregado miden la variación residual de los resultados tras tomar en cuenta estos factores, hay una mayor posibilidad de inestabilidad en las puntuaciones. Sin embargo, debe observarse que, a pesar de esta inestabilidad, es probable que los resultados de valor agregado sean una medición más equitativa de la eficacia de esta escuela.

Los estimadores para escuelas pequeñas estarán sujetos a una mayor variabilidad de muestreo. Si se realiza un gráfico con las diferencias de un año con siguiente de los efectos de la escuela con el tamaño de la muestra de escuelas se observa un patrón característico con una mayor

dispersión asociada a tamaños de muestra más pequeños y una dispersión insignificante asociada a muestras más grandes. De modo más general, como las estimaciones de los efectos de la escuela son desviaciones de un promedio nacional, el resultado de una escuela también depende de las puntuaciones de ganancias en las pruebas (ajustadas) en otras escuelas. Éstas también varían con los años. En la mayoría de los sistemas educativos, las escuelas más pequeñas son más comunes en el sector de educación primaria que en el de secundaria. De igual manera, es más probable que los estimadores de valor agregado de las escuelas primarias manifiesten mayor inestabilidad relativa, lo que dificulta aislar a las que persistentemente muestran "bajo desempeño". Ray (2007) investigó la cantidad de escuelas primarias que podrían catalogarse como de bajo desempeño basándose en los datos acumulados en tres años. De las 16 200 primarias examinadas, relativamente pocas (424) tuvieron un estimador de valor agregado mayor de una desviación estándar inferior al promedio durante tres años consecutivos. Esto no se calculó con las puntuaciones de valor agregado contextualizado, sino que se basó en el método de la mediana (por tanto, sin contracción). Con el fin de aumentar la cantidad de escuela que calificaban como de bajo desempeño teniendo un valor agregado "bajo" en cada uno de los tres años, la definición de "bajo" debía hacerse menos restrictiva (p. ej., 0.75 desviaciones estándar por debajo del promedio durante los tres años). Sin duda, se puede establecer un criterio basado en los promedios de tres años con el fin de suavizar un poco la inestabilidad. Otras opciones serían excluir a las escuelas inferiores a un tamaño determinado junto con advertencias generales para el usuario acerca de la precisión de evaluar cambios anuales en las puntuaciones de valor agregado. Suavizar a lo largo de los años y/o excluir a las escuelas pequeñas implica un equilibrio entre tener efectos escolares estimados menos afectados por la variación aleatoria y descubrir cambios reales en un periodo posterior. En las argumentaciones dentro del grupo de expertos constituido para la elaboración de este informe, se consideró en general que las escuelas con cohortes anuales de menos de 20 a 30 alumnos eran más proclives a producir resultados menos estables. Sin embargo, se reconoció que el tamaño de la escuela varía notablemente por país y que necesitan incluirse consideraciones prácticas en toda decisión referente a retirar a las escuelas de la muestra o análisis. Se necesitan más investigaciones sobre la estabilidad de los resultados de valor agregado de las escuelas para guiar los juicios sobre su inclusión en la muestra.

CRITERIO ESTADÍSTICO: SESGO

La utilidad de un modelo de valor agregado también depende de la cantidad de sesgo que produce en los estimadores. El sesgo es una medida esencial de imprecisión. Un estimador está sesgado si su valor promedio en muchas replicaciones de un estudio no tiende hacia su valor "real". Por lo general, el sesgo no se reduce sólo con añadir más datos del tipo que ya se incluyó en el modelo. En este informe, el sesgo es fundamentalmente distinto de la varianza porque esta última suele reducirse al incrementar la cantidad de datos disponibles para el análisis.

Asimismo, el sesgo es más difícil cuantificar y aminorar que la varianza porque, en cierto sentido, se ubica "fuera" del modelo. Por ejemplo, suponga que es común en algunos distritos que los estudiantes asistan a clases privadas de preparación para los exámenes. Si estas clases están bien diseñadas, los estudiantes avanzarán académicamente y es de suponerse que esto se refleje en su desempeño en la prueba. Sin embargo, si las puntuaciones de la prueba se emplean

para un análisis de valor agregado, las escuelas a donde asisten estos estudiantes parecerán más exitosas de lo que son en realidad, lo que generará una panorámica distorsionada o "sesgada" de su desempeño relativo. En este ejemplo, el sesgo entra en la estimación de efectos escolares a causa de una variable omitida (asistir a clases de preparación), lo que crea una correlación entre las variables escolares y el término de error. Si bien el cálculo de una varianza se basa en la suposición de que el modelo es correcto, el sesgo suele surgir cuando no se satisfacen los supuestos subyacentes del modelo. Los supuestos se pueden relacionar con la naturaleza de los datos (como las omisiones de variables pertinentes), la estructura del modelo, o ambas cosas (Reardon y Raudenbush, 2009). Así, aunque los estimadores de varianza de los efectos escolares se generan como algo esperado en la mayoría de los modelos de valor agregado, nunca se producen los estimadores de sesgos. Las aproximaciones a los sesgos en ocasiones se calculan de forma analítica. Más a menudo, se obtienen mediante simulaciones en las que se exploran de manera sistemática los puntos de partida de los supuestos.

Los efectos estimados de la escuela estarán sesgados en la medida en que haya un ajuste sistemático deficiente o excesivo (véase el Capítulo 4). Los datos de alumnos disponibles para el análisis muy pocas veces incluyen los antecedentes estudiantiles relacionados con el aprovechamiento académico. Por ejemplo, el nivel de escolaridad paterna suele considerarse un indicador del nivel socioeconómico general. Sin embargo, un modelo especificado por completo del nivel socioeconómico por lo general incluiría también la ocupación (u ocupaciones) de los padres, el ingreso familiar e incluso las transferencias inter generacionales. Queda claro que el nivel de escolaridad paterna por sí solo no hace justicia al concepto de nivel socioeconómico. Por tanto, es probable que un modelo que incorpore sólo el nivel de escolaridad paterna genere un ajuste deficiente. Es decir, los efectos estimados de las escuelas con poblaciones de nivel socioeconómico más alto están sesgadas hacia arriba, y los efectos estimados de las escuelas con poblaciones de nivel socioeconómico más bajos, hacia abajo.

Por desgracia, hay muchas formas de que los sesgos interfieran con los estimadores de desempeño escolar. Considere, por ejemplo, la situación en que la movilidad estudiantil varía entre las escuelas. En escuelas con poblaciones estudiantiles muy móviles, tal vez destinen sustanciales recursos escolares a estudiantes temporales sólo para que se hayan ido antes de presentarse a la prueba o para que no hayan pasado el tiempo suficiente en la escuela para que se les cuente. Esta dificultad incluye también el efecto de los cambios de la composición de la clase en los estudiantes no temporales. Así, cierta cantidad de los esfuerzos escolares no se refleja en los datos del el modelo y puede generar un estimador más bajo del desempeño de la escuela. Si las tasas de movilidad son mayores en escuelas que atienden a poblaciones más desfavorecidas y con menos recursos en general, los estimadores de esas escuelas estarán sesgados hacia abajo. Estos y otros escenarios semejantes sugieren que debe tenerse mucho cuidado al comparar escuelas con patrones de movilidad muy distintos.

El error de medida es otra fuente potencial de sesgos. Es bien sabido que los supuestos de la teoría clásica de la regresión suponen que las variables explicativas en el modelo se miden sin error. En el caso presente, tanto las puntuaciones de pruebas anteriores como las variables contextuales pueden contener cantidades sustanciales de ruido, con la consecuencia de que los estimadores de los coeficientes de regresión usados para el ajuste estén sesgados hacia cero. Ladd y Walsh (2002) muestran que el uso de una sola puntuación de rendimiento previo

genera estimadores de valor agregado con características operativas deficientes. Sugieren puntuaciones de doble retraso (es decir, puntuaciones de dos años antes) como instrumento para la medida del rendimiento previo. Sin embargo, no hay consenso sobre que la puntuación de doble retraso cumpla con los requerimientos de una variable instrumental.

CRITERIO ESTADÍSTICO: ERROR CUADRÁTICO DE LA MEDIA

En la práctica, los supuestos nunca se satisfacen por completo y ningún modelo es perfectamente adecuado. Así, siempre puede haber sesgos. El asunto es la dirección del sesgo y su magnitud (tanto de manera absoluta como en relación con la magnitud de la varianza). El sesgo suele generar mayor preocupación que la varianza, sobre todo porque es un peligro más sutil para la utilidad de los estimadores producidos por un modelo de valor agregado. Tradicionalmente, los expertos en estadística juzgan un estimador con base en una medida de error total, llamada error cuadrático de la media (EMC). Una expresión del EMC es:

$$\text{EMC} = \text{Varianza} + (\text{Sesgo})^2$$

Así, algunos modelos aceptan una pequeña cantidad de sesgo con el fin de reducir la varianza lo bastante para generar un EMC menor. Esta es la estrategia de los modelos de valor agregado que modelan las contribuciones escolares como efectos aleatorios. Generan estimaciones de los efectos escolares concentrados hacia el promedio (al introducir sesgo), pero las varianzas de los estimadores se reducen sustancialmente en comparación con los no basados en datos compartidos entre escuelas. Los primeros suelen tener un EMC menor que los segundos. Otra estrategia para trabajar con las preocupaciones de los ajustes es emplear modelos en los que tanto estudiantes como escuelas se traten como efectos fijos. Esto elimina el problema de errores correlacionados y similares. Sin embargo, cuando la cantidad de estudiantes y escuelas es grande, hay problemas de cálculo que pueden generar mayor incertidumbre con los estimadores escolares de valor agregado que necesitan abordarse a causa de la gran cantidad de parámetros a estimar. Los estimadores de efectos fijos son consistentes pero muy variables porque no hay "préstamo de información" entre escuelas, como sucede con los modelos de efectos aleatorios. Lockwood y McCaffrey (2007) investigaron las propiedades estadísticas de los modelos de efectos aleatorios. Demostraron que, con suficientes datos sobre el rendimiento previo, el sesgo que introduce la correlación entre errores específicos del nivel de alumnos y efectos escolares (aleatorios) es lo bastante pequeño como para ignorarlo. Estos modelos generan estimadores que se concentran en la media, lo que induce un poco de sesgo pero también reduce la varianza. Estos modelos suelen ser los preferidos debido al EMC más bajo resultante. Sin embargo, nunca se debe olvidar el equilibrio presente al usar modelos de efectos aleatorios, pues el préstamo de información produce estimadores menos variables (es decir, más precisos) con el costo de un sesgo.

Datos perdidos

Hasta aquí hemos considerado tres criterios estadísticos bajo el supuesto de que la base de datos en el análisis está completa. Sin embargo, en la práctica muy pocas veces se disfruta de esta circunstancia positiva, en parte porque los modelos de valor agregado son muy exigentes respecto de los datos. Requieren registros del desempeño en las pruebas de los alumnos en una o más materias durante dos o más años. Muchos requieren también características estudiantiles y otros datos contextuales. En la mayoría de los contextos, algunos registros de alumnos

estarán incompletos. Desde luego, más preocupante es la situación en que los estudiantes inscritos están por completo ausentes de la base de datos. Por tanto, es esencial efectuar varias evaluaciones de la calidad de los datos antes de proceder al análisis. Estos asuntos se tratan con más profundidad en la Parte III.

Una cantidad sustancial de datos perdidos, en especial datos de puntuaciones de pruebas, es causa de inquietud respecto de consideraciones de varianza y sesgo, en especial este último. Ahora, sin duda hay casos en que los datos de puntuaciones en las pruebas faltan por razones legítimas, como que el estudiante abandone la escuela, el área/región, o que tome otra forma de evaluación (en especial en un sistema con registros educativos explícitos). Por otra parte, el estudiante quizá faltó a clases el día de la prueba, sin oportunidad de realizarla después. La cuestión, así, apunta a preguntarse si las características de los estudiantes con estos datos perdidos son consistentes con los supuestos del modelo, pregunta que abordamos aquí.

Para empezar, considere primero la situación en que el modelo de valor agregado requiere puntuaciones de pruebas de dos ocasiones sucesivas, así como la medida de otras características de los alumnos. Si todos los registros de los estudiantes contienen la puntuación anterior pero faltan algunas de la actual, debe hacerse algo para remediar la situación. Una posibilidad es tan sólo borrar los registros con datos perdidos y efectuar el análisis del conjunto de registros completos. Por desgracia, es probable que esto produzca estimadores sesgados a menos que los datos perdidos falten al azar. El supuesto de que los datos perdidos falten por completo por azar significa que la distribución de puntuaciones perdidas es la misma que la distribución de puntuaciones observadas (McCaffrey *et al.,* 2003, p. 82). Es poco probable sostener esta suposición en los sistemas escolares. Por ejemplo, esto no se sostiene si es más probable que falten puntuaciones de pruebas de los estudiantes con características desfavorables (es decir, asociadas a ganancias menores) y todo lo demás permanezca igual. En ese caso, las escuelas con mayores proporciones de estos estudiantes y, por lo común, mayores proporciones de registros perdidos, tendrán ventaja en el análisis. Esto es una forma de sesgo.

Los modelos más complejos (p. ej., EVAAS) son capaces de integrar registros completos e incompletos. Los registros incompletos no introducen sesgos si los datos faltan por azar. La suposición de que los datos falten por azar es una suposición más débil que la de que falten *por completo* al azar. Esto significa que, si se incluyen en el modelo las características y puntuaciones de los estudiantes en las pruebas, y se supone que la distribución de las puntuaciones perdidas es la misma que la distribución de las puntuaciones observadas, p. ej., dentro de un grupo de estudiantes con las mismas características y puntuaciones en la prueba en el modelo, las puntuaciones perdidas no son sistemáticamente distintas de las puntuaciones no perdidas. En otras palabras, el proceso que genera el patrón de valores perdidos y los resultados de puntuaciones de pruebas hace que sean independientes entre sí (Rubin, 1976; Little y Rubin, 1987).

Incluso el supuesto más débil de perdidos por azar puede fallar de muchas maneras. Lo hace, por ejemplo, si es más probable que, en un conjunto fijo de características de alumnos, los estudiantes más débiles (es decir, aquellos cuyas trayectorias de puntuaciones en las pruebas son más frágiles) estén ausentes el día de la prueba. Quizá estén ausentes porque así lo quisieron o quizá porque se les alentó a que lo hicieran. Desde luego, es poco probable satisfacer por

completo el supuesto de perdidos por azar. La pregunta entonces es cuán robustas son las estimaciones de los efectos de la escuela si se parte del supuesto de perdidos por azar. Un estudio reciente (McCaffrey *et al.,* 2004) sugiere que, en ciertas condiciones para algunos modelos, hay un grado justo de robustez. En oras palabras, el sesgo en los estimadores que introducen los datos perdidos es relativamente pequeño.

Debe interpretarse con cautela esta buena noticia. En primer lugar, la robustez se debe en parte a la extensión de los datos que emplean estos modelos. Es decir, el efecto de partida del supuesto de perdidos por azar se mitiga por las contribuciones de la abundante información que hay en el modelo. En segundo lugar, los datos perdidos provocan una varianza mayor en los estimadores en comparación con la que se obtendría con datos completos. Así, cantidades sustanciales de datos perdidos reducen la utilidad de los estimadores si, por ejemplo, la meta principal fuera identificar escuelas significativamente distintas del promedio. Si es más probable que las escuelas de verdad menos eficaces tengan bases de datos incompletas, con los modelos de efectos aleatorios sus estimadores de valor agregado experimentarán mayor contracción y será más difícil distinguirlas estadísticamente del promedio.

ELECCIÓN DE MODELO EN ANÁLISIS DE VALOR AGREGADO

Al establecerse un modelo de valor agregado es aconsejable comparar, en lo posible, las características de los estimadores escolares de valor agregado según distintas especificaciones del modelo. Desde un punto de vista práctico, lo más importante es la medida en que los diversos modelos de valor agregado suelen generar resultados semejantes, es decir, si la elección de modelo importa empíricamente. Jakubowski (2007) efectuó un estudio comparativo, con datos de **Eslovenia** y **Polonia**, para comparar distintos modelos de valor agregado respecto de la estabilidad de los resultados. Estos modelos se emplean a menudo en la investigación de valor agregado y algunos se han implementado de forma operativa. No se describen aquí porque se han tratado en la literatura de modelos multinivel (lineal jerárquico o combinado) y de métodos de valor agregado para la evaluación de escuelas (véase Goldstein, 1997; Raudenbush y Bryk, 2002; Snijders y Bosker, 1999).

En ambos países, los datos incluyeron puntuaciones de alumnos individuales de exámenes aplicados al término de la educación primaria y al término de la secundaria. Sin embargo, difirieron la edad de los alumnos y las materias de los exámenes. Es importante observar que los dos países tienen diferencias importantes respecto al tamaño de población, organización escolar y muchas características sociales y económicas. El primer modelo fue el de regresión lineal simple, con cuyos residuos de regresión se calculó el valor agregado de la escuela. El segundo fue un modelo de regresión lineal de efectos fijos. El tercero fue un modelo de efectos aleatorios, en el que se supuso que los efectos escolares tenían una distribución independiente y normal. El cuarto fue un modelo de pendiente aleatoria (o coeficiente aleatorio) en el que se supuso que no sólo los interceptos (efectos escolares), sino también las pendientes de ingresos, tenían una distribución aleatoria y se les permitió variar entre escuelas.

El hallazgo clave fue que las correlaciones entre distintos conjuntos de estimadores de valor agregado fueron muy altas (Jakubowski, 2007). Por tanto, desde un punto de vista práctico, se juzgó que los modelos más sencillos eran preferibles a los más complicados en condiciones en las que la sencillez y accesibilidad fuesen más importantes para los tomadores de decisiones

que su carácter teórico óptimo. El modelo de pendiente aleatoria también ofreció estimadores muy semejantes a los de los modelos más sencillos. El hecho de permitir la variación de las pendientes de ingresos no produjo resultados significativamente distintos por sí solo. Esto no significa que la elección de modelo sea un asunto irrelevante, ni que los modelos más sencillos siempre sean preferibles y que siempre producirán resultados semejantes. En cambio, ilustra que los distintos estimadores de valor agregado pueden no producir resultados sustancialmente distintos y que estas diferencias deben someterse a prueba y analizarse. Comparar los resultados de diferentes modelos de valor agregado respecto de algún conjunto de criterios y objetivos predeterminados debe permitir la identificación de un modelo adecuado. Sin embargo, al revisar estas comparaciones, las correlaciones generales pueden no ser tan importantes como la consistencia de las puntuaciones de valor agregado de las escuelas en ambos extremos de la distribución. Al comparar modelos, debe reconocerse que hay costos y beneficios asociados a los distintos modelos y que, si bien los más complejos pueden generar propiedades estadísticas superiores, como cierta robustez respecto de a los datos perdidos y el sesgo de selección, también pueden ser más costosos en términos de transparencia y, sobre todo de requerimientos de datos en algunos países con recopilaciones de datos centralizados deficientes.

Hay otros estudios importantes. Gray *et al.* (1995) calcularon las puntuaciones de valor agregado de un grupo de escuelas secundarias entre 1990 y 1991, y entre 1991 y 1992, y descubrieron fuertes correlaciones de entre 0.94 y 0.96. Los autores consideran que sus hallazgos, junto con investigaciones previas, sugieren "que hay una buena cantidad de estabilidad en la eficacia de las escuelas de un año a otro" (p. 97). En su estudio más reciente de 63 escuelas secundarias en Lancashire, Thomas, Peng y Gray (2007) hallaron correlaciones en valor agregado contextualizado para años consecutivos de 0.80 a 0.89. Ponisciak y Bryk (2005) también efectuaron análisis comparativos y descubrieron correlaciones modestas entre métodos. En **EUA**, Tekwe *et al.* (2004) realizaron un estudio para comparar los efectos escolares estimados con cuatro modelos y datos de los Grados 3, 4 y 5 de un distrito escolar de Florida con 22 escuelas primarias. Los modelos variaron de sencillos a complejos. Las correlaciones entre los estimadores por modelo por lo general excedieron de 0.90, excepto los que implicaron un modelo multinivel complejo, en donde excedieron de 0.70. Los autores concluyeron que no parece haber ninguna ventaja sustancial de usar modelos más complejos en lugar de un modelo simple de cambio de puntuaciones. En respuesta al análisis de Tekwe *et al.* (2004), Wright (2004) efectuó una simulación con un diseño factorial para distintos parámetros: cantidad de estudiantes, patrones de ganancias y el grado al que los valores perdidos pudieron sesgar las puntuaciones de valor agregado de las escuelas. Comparó un modelo simple de puntuaciones de ganancia con dos modelos longitudinales más complejos. Con un criterio de EMC, concluyó que los modelos más complejos son preferibles en vista de su EMC más bajo en las celdas del diseño en las que es más probable que representen datos de la vida real. También es posible que el tamaño habitual de las estimaciones de los errores estándar que acompañan a las mediciones de las estimaciones de desempeño escolar sea diferente entre los modelos. Por tanto, puede preferirse un método debido a la mayor cantidad de escuelas que pueda distinguirse con precisión del promedio. Sin embargo, la pregunta de que la estabilidad sea "razonable" depende de manera crítica del uso que se dé a las puntuaciones de valor agregado y de las definiciones que se apliquen a nociones como "bajo desempeño". Los resultados recién descritos son consistentes con el trabajo empírico sobre el modelo EVAAS.

La semejanza de las puntuaciones de valor agregado de las escuelas con diferentes modelos ilustra que las elecciones de las autoridades educativas y los administradores no son simples elecciones entre modelos buenos y malos. En general, la mayoría de los modelos producirá resultados semejantes si los datos son los mismos, los datos de pruebas son confiables y sobre todo si se incorporan múltiples medidas de rendimiento previo al proceso de estimación. No obstante, parece que los modelos más complejos, por las limitaciones de los datos disponibles, ofrecen mayor precisión y también parecen menos sensibles a los supuestos de partida. Los modelos son complejos en diversas formas. Un modelo puede introducir complejidad al incluir múltiples puntuaciones de evaluación en múltiples materias, como en el modelo EVAAS. Otro modelo puede tomar en cuenta diversos factores adicionales que afecten las puntuaciones de desempeño (Ponisciak y Bryk, 2005). El mayor nivel de complejidad en cualquiera de estos modelos (o cualquier modelo complejo) sólo es benéfico si capta patrones o fuentes de ruido significativos en los datos. La desventaja reside en el mayor nivel de complejidad y la necesidad de más datos para que se estimen bien los parámetros del modelo. Este equilibrio necesita analizarse en la etapa piloto de la implantación de un sistema basado en modelos de valor agregado, junto con una evaluación de la medida en que se requieren datos adicionales para modelos más complejos.

En la recomendación al gobierno del **Reino Unido** respecto de la implantación de modelos de valor agregado, Fitz-Gibbon (1997, p. 38) encontró que "los indicadores de valor agregado producidos por el procedimiento simple de comparar el desempeño de los estudiantes directamente con el desempeño de estudiantes semejantes, sin importar la escuela a la que asisten, y después sumar las puntuaciones de valor agregado (puntuaciones residuales) produjo indicadores que correlacionaron fuertemente con los indicadores de modelos más complejos que podrían recomendarse los métodos simples". Dadas las ventajas de comunicar a los interesados con modelos más sencillos, este hallazgo se presta a la adopción de estimaciones de valor agregado más sencillas. Estas pueden complementarse con modelos más complejos tanto para análisis internos como para supervisar los resultados de un modelo más sencillo.

Otro aspecto por analizar son las diferencias de modelos de distintas estructuras de puntuaciones de evaluaciones de alumnos. Fielding, Yang y Goldstein (2003) compararon estimadores de valor agregado basados en un modelo multinivel para puntuaciones de escala y un modelo multinivel para puntuaciones ordinales. Los modelos se aplicaron a una base de datos grande del examen para el Certificado General de Educación de Nivel Avanzado (*General Certificate of Education Advenced Level*) en el **Reino Unido**. Para ambas clases de modelos, las covariantes fueron: rendimiento previo anterior del alumno, género, edad, escuela, tipo de financiación y políticas de admisión, y junta de calificadores de exámenes. Se reveló que los coeficientes de correlación y las correlaciones entre las clasificaciones entre los estimadores residuales de la institución y los estimadores de valor agregado de cada par de modelos fueron superiores a 0.96. Sin embargo, si es cierto que los estimadores de valor agregado de una escuela individual pueden diferir sustancialmente entre los modelos, es importante la elección del modelo de valor agregado más adecuado. Por tanto, al comparar el impacto de los distintos modelos, deben identificarse las escuelas individuales en las que haya diferencias significativas. Además, debe destacarse que la consistencia de los hallazgos no implica necesariamente que no existan sesgos o errores de medición.

CONCLUSIÓN

La contribución estimada de una escuela al aprendizaje de sus alumnos varía en función del modelo de valor agregado específico que se emplee. Las diferencias de especificaciones se derivan de diversos factores, como la amplitud de los datos de pruebas (es decir, la cantidad de años y la de materias), el tratamiento de los datos perdidos y las clases de ajustes efectuados. Con estas diferencias, cada modelo de valor agregado conlleva ventajas y desventajas que deben considerarse a la luz del contexto en que se utilizan y de la naturaleza de los datos disponibles. En general, los modelos más complejos tienen mayores demandas de datos, son más difíciles de implementar y evaluar, y plantean mayores desafíos al comunicar su lógica a diversos interesados, incluso al público en general. Surge entonces una pregunta natural: "¿vale la pena usar modelos más complejos?". Mayor complejidad lleva asociada mayores costos, en particular si es preciso recopilar datos adicionales para los modelos más complejos (que a menudo es el caso). Las ventajas de esta mayor complejidad, como una varianza reducida, necesitan ponderarse respecto de los costos. Entre los tomadores de decisiones hay una preferencia comprensible por los modelos de valor agregado más sencillos, pues son más fáciles (y baratos) de desarrollar y se prestan mejor a una comunicación eficaz con los interesados. Sin embargo, si los modelos más sencillos generan una menor especificación, los estimadores de desempeño escolar estarán sesgados y los costos serán mayores a la larga. Estos costos y beneficios diferirán entre los sistemas educativos, y pueden analizarse durante la fase piloto del proceso de implementación para aclarar los equilibrios necesarios.

Dadas las características particulares de cada sistema educativo, los objetivos de los modelos de valor agregado y el tipo de evaluación de estudiantes en que se basan, no es posible identificar un modelo único de valor agregado que se ajuste a todos los sistemas educativos. En cambio, deben analizarse diversos modelos para ver si son adecuados para cada sistema. El análisis de los temas en este capítulo que deben estudiarse para tomar decisiones informadas sobre la elección de modelos abarcó lo siguiente:

- Debe analizarse la varianza en cada modelo de valor agregado para evaluar la adecuación de modelos particulares. Las estimaciones de los errores estándar que acompañan a las estimaciones de los efectos escolares difieren entre los modelos. Puede preferirse un método porque errores estándar menores implican que puede distinguirse con precisión a una mayor cantidad de escuelas del promedio o clasificarlas si alcanzan algún objetivo predeterminado. Pueden efectuarse análisis que comparen los modelos de valor agregado con este criterio en la etapa de implementación. Por ejemplo, pueden ponerse a prueba los datos piloto para identificar el modelo es más adecuado al reducir la varianza y generar así resultados más interpretables.
- En tanto que todos los modelos de valor agregado implican alguna clase de ajuste a la secuencia de puntuaciones brutas de pruebas que acompañan a cada estudiante, el uso de datos socioeconómicos contextuales juega un papel distinto según qué componentes de datos que se incluyan en un análisis de valor agregado. Si bien la necesidad de ajuste fluye de manera natural de la lógica de los modelos de valor agregado, debe hacerse con cuidado, o se generarán estimadores muy engañosos. Deben efectuarse análisis para evaluar el impacto de la inclusión de características socioeconómicas en las puntuaciones de valor agregado de las escuelas y en aspectos del modelo de valor agregado en general (p. ej., la fuerza predictiva del modelo y los errores estándar asociados a los estimados escolares).

- Es necesario analizar el sesgo potencial en el modelo y poner a prueba el potencial para reducirlo durante la fase piloto de implementación. Si bien no es fácil analizar la medida del sesgo en las estimaciones, pueden hacerse aproximaciones y simulaciones para evaluar el sesgo potencial. Puede explorarse el potencial de datos perdidos, y la inclusión o exclusión de variables específicas en el modelo pueden poner de manifiesto problemas específicos. Las comparaciones con puntuaciones brutas reales de pruebas ilustran más el sesgo potencial de las estimaciones.
- Los supuestos sobre los datos perdidos realizados en la especificación de los modelos de valor agregado se comparan con el patrón de datos perdidos de la muestra, y se calculan los estimadores de los efectos de los datos perdidos. También se aplican procedimientos para reducir la frecuencia de datos perdidos en las evaluaciones de los alumnos y otras recopilaciones de datos (p. ej., al fomentar [o desalentar] la participación de alumnos con desempeño alto [o bajo]).
- Una muestra pequeña es un problema por los niveles mayores de incertidumbre que suelen rodear al valor agregado de escuelas con muestras pequeñas y por la estabilidad reducida de las puntuaciones de valor agregado de estas escuelas. Se ponen a prueba los estimadores de valor agregado de escuelas pequeñas y se hacen recomendaciones para tanto el análisis como la presentación de los resultados escolares. En general, los países participantes que consideraron cohortes con menos de 20 a 30 alumnos produjeron estimadores de valor agregado escolar que dificultaron la interpretación de los resultados.
- La estabilidad de las puntuaciones de valor agregado de las escuelas y cómo afecta a la clasificación de escuelas según su desempeño, así como la elección de modelos de valor agregado es también importante. Análisis como los presentados en este informe se realizan para determinar el grado de estabilidad de las puntuaciones escolares y si puede reducirse. En estos análisis es importante considerar no sólo el nivel general de estabilidad (o falta de estabilidad), sino los cambios en las puntuaciones escolares individuales. En ese caso se efectúa un análisis de las causas de dicha inestabilidad y se identifica si algunas escuelas en particular son más susceptibles a la inestabilidad en sus resultados escolares.

Dada la necesidad de modelos de valor agregado sencillos que se comuniquen con eficacia a los interesados, el análisis recién delineado debe comparar los resultados entre modelos relativamente más sencillos y más complejos, y evaluar las diferencias. Si hay pocas diferencias significativas entre estos modelos, quizá convenga usar modelos de valor agregado más sencillos para presentar resultados al público y a algunos interesados. Esto facilitaría la comunicación eficaz y el uso de información de valor agregado para avanzar hacia propósitos de políticas específicas. La presentación de los resultados de modelos más sencillos necesitaría entonces apoyarse en extensos análisis internos continuos que comparasen esos resultados con los obtenidos de modelos de valor agregado más complejos. Los análisis comparativos garantizarían que los modelos más sencillos producen estimaciones precisas y no afectan de manera injusta a escuelas o grupos escolares específicos. Como el modelo se desarrolla con el tiempo, sería necesario efectuar estos análisis de forma continua. Esto es de particular importancia en casos en los que la disponibilidad y requerimientos de datos cambiaran con el tiempo.

Si se decide emplear dos tipos de modelos, se requiere un conjunto de acciones que aligere cualquier discrepancia en los resultados entre el modelo más sencillo y el más complejo.

Como presento en este capítulo, estas discrepancias no necesariamente son habituales en una gran cantidad de escuelas. Además, durante la fase de implementación, la elección del modelo específico que se emplea y presenta a los interesados debe basarse en un análisis que ilustre que esas discrepancias se redujeron. Pero es importante que exista un conjunto predeterminado de criterios para evaluar la validez de los distintos resultados, sobre todo si los resultados de valor agregado se van a usar con fines de rendición de cuentas. Estos criterios deben identificar el origen de la diferencia de los resultados de una escuela y después permitir la identificación de una medida más precisa del desempeño de una escuela. Si la información de valor agregado se usa con fines de mejoras escolares, estos procedimientos ofrecen información más valiosa. En algunos casos, pueden incorporarse al sistema de mejoras escolares. Una discrepancia en los resultados de una escuela puede resultar en una recopilación extensa de datos que ayude a identificar el origen de la discrepancia. Aparte de las acciones para escuelas individuales, el análisis de las discrepancias de los resultados entre modelos de valor agregado más sencillos y más complejos debe entonces impulsar el desarrollo continuo del sistema basado en modelos de valor agregado. Esto debe ayudar a reducir la cantidad y el tamaño de las discrepancias entre modelos sencillos y complejos con el tiempo. Puede ser prudente iniciar los análisis de valor agregado mediante modelos más sencillos, y reservar los más complejos para la investigación e introducirlos quizás en una etapa posterior, cuando se hayan resuelto de manera satisfactoria todos los problemas técnicos.

Nota

1. Un ejemplo en el **Reino Unido** es una estadística simple que se considera en la actualidad (aunque aún no está en uso): la cantidad de alumnos en una escuela que avanza dos niveles del Currículo Nacional o más dentro de una Etapa.

Parte III

Implementación de un sistema basado en modelos de valor agregado

Además de la naturaleza de las bases estadísticas y metodológicas de los modelos de valor agregado, el impacto en las políticas, prácticas y resultados pueden ser insignificantes o incluso negativos si no se efectúa una implementación eficaz. Esta convicción quedó clara en varios países que participaron en la elaboración de esta publicación y dio pie a análisis más detallados de los métodos para establecer un sistema basado en modelos de valor agregado. La Parte III de este informe se elabora sobre el análisis presentado en las Partes I y II para ofrecer una guía con la cual implementar un sistema basado en modelos de valor agregado en los sistemas educativos. Esta guía no es una lista definitiva, ni cada aspecto será aplicable a todos los sistemas educativos. Se presenta, sin embargo, el conocimiento obtenido en varios sistemas educativos y del grupo de expertos con experiencia en establecer sistemas basados en modelos de valor agregado en varios sistemas educativos.

Es necesario abordar varios retos con el fin de establecer con eficacia un sistema basado en modelos de valor agregado. Aquí algunos problemas ya analizados previamente se presentan y se abordan los siguientes temas de implementación: establecer objetivos de políticas y medidas de desempeño escolares, elegir un modelo adecuado de valor agregado, elaborar una base de datos eficaz, llevar a cabo un programa piloto adecuado, supervisar los resultados de los análisis de valor agregado, diseñar una estrategia de comunicación y compromiso con los actores clave y los programas de formación correspondientes, y presentar y usar información basada en resultados de valor agregado. Para completar la Parte III, se presenta una breve lista de los temas principales que deberán abordarse en la implementación de un sistema basado en modelos de valor agregado. Como en otros procesos de reformas educativas, las mejores prácticas y la evidencia internacional sobre temas específicos tendrán que adaptarse al sistema educativo en un país, con los factores limitantes y las oportunidades que esto represente. En este contexto, una etapa de ***investigación y análisis*** tendrá que formar parte del proceso de diseño, planeación y implementación de un sistema de evaluación basado en modelos de valor agregado.

CAPÍTULO 7

Establecer objetivos de políticas y elegir el modelo adecuado de valor agregado

Determinación de la variable con la cual medir el valor agregado **155**
- Medidas categóricas y continuas **156**

Identificación del modelo adecuado de valor agregado que aborde mejor los objetivos de políticas **157**

Elaboración de una base de datos eficaz **160**

Importancia del programa piloto para los modelos de valor agregado **165**

El análisis de valor agregado sirve para impulsar varios objetivos de políticas y programas. La implementación de un sistema basado en modelos de valor agregado para promover objetivos específicos de políticas requiere tomar varias decisiones clave y recorrer varias etapas. Esto se deriva desde los tres principales objetivos de políticas analizados en la Parte I de este informe: procesos de mejora escolar, mecanismos de rendición de cuentas y elección de escuela.

Los esfuerzos por realizar ***mejoras escolares*** reciben una gran ayuda procedente de la información de valor agregado, en particular en los sistemas que permiten que las escuelas usen los resultados de valor agregado para desarrollar y supervisar iniciativas de mejora escolar. Las características básicas que afectan las acciones de implementación se centran en el uso de la información de valor agregado para apoyar e impulsar sistemas de toma de decisiones basados en datos que faculten a las escuelas y a otros responsables de decidir en el análisis de la variación del desempeño escolar y de los alumnos. Esto proporciona información para asignar mejor los recursos e identificar las áreas con mejores prácticas y las que necesitan atención.

Mecanismos de rendición de cuentas escolares utilizan las puntuaciones de valor agregado de las escuelas para responsabilizarlas por su desempeño. La rendición de cuentas puede adoptar numerosas formas: aquellas vinculadas a la financiación escolar, a intervenciones específicas para escuelas de bajo desempeño o a consecuencias en la remuneración de administradores, directores y maestros. También pueden establecerse sistemas de rendición de cuentas más implícitos que centran la atención en los resultados escolares sin vínculos explícitos a los recursos, autonomía o remuneración. El primer paso en la implementación de un sistema de rendición de cuentas escolar basado en modelos de valor agregado es considerar cuáles son los acuerdos de la rendición de cuentas actual y la forma en que los cambios pueden afectar a los interesados. Un componente básico del compromiso fructífero de los interesados es ofrecer claridad en los objetivos y operaciones de un sistema basado en modelos de valor agregado. Respecto a los acuerdos sobre la rendición de cuentas escolar, surgen cuestiones cruciales sobre recompensas y sanciones, y el nivel al que se aplicarán. Este informe se ha centrado sólo en las mediciones de valor agregado de la escuela, pero los modelos de valor agregado pueden servir también para fomentar la rendición de cuentas de maestros individuales (Braun, 2005b; McCaffrey *et al.,* 2004; McCaffrey *et al.,* 2003), y es importante expresar de manera explícita esta distinción por el impacto potencial en interesados clave y en el desarrollo de estimaciones de valor agregado específicas.

La ***elección de escuela*** puede ayudar al desarrollo de los sistemas educativos al permitir a los padres y familias elegir la escuela que mejor se ajuste a sus necesidades. Así, las escuelas se motivan en sus desarrollos educativos para satisfacer las necesidades de padres y familias. Los beneficios de un sistema que facilite la elección de escuela residen en la suposición de que padres y familias cuentan con la información requerida para distinguir entre escuelas. Las medidas de valor agregado tienen un valor incalculable porque proporcionan medidas mucho mejores de desempeño escolar en comparación, por ejemplo, con puntuaciones brutas en los exámenes. Estas medidas perfeccionadas deben permitir adoptar mejores decisiones y por tanto mejorar la correspondencia entre escuelas y las necesidades de padres y familias. A su vez, esto debe dar a las escuelas mejor información ofreciendo acciones educativas que atraigan a los estudiantes y sus familias. Si el fomento de la elección de escuela es un objetivo básico de la implementación de un sistema basado en modelos de valor agregado,

sería también beneficioso revisar el grado real en que los padres y familias están en disposición de elegir entre escuelas. En algunos países hay requerimientos legislativos y regulatorios que restringen la elección de escuela, y en otros hay limitaciones institucionales, geográficas y de recursos para las decisiones de las familias (OCDE, 2006). En estas circunstancias, la información adicional puede tener un beneficio reducido en la elección de escuela. Una revisión de estas circunstancias debe ofrecer un contexto importante para decidir respecto del uso de la información de valor agregado.

Una cuestión crucial en la puesta en marcha de un sistema basado en modelos de valor agregado es si se publicarán los resultados de valor agregado de las escuelas, y en qué forma. Sin duda, se requiere la publicación de resultados para ampliar la elección de escuela en un sistema educativo. La Parte I de este informe ofrece numerosos ejemplos para presentar los resultados escolares conforme a propósitos particulares. Es conveniente detallar la presentación de resultados en las primeras etapas del proceso de implementación. Esto puede contribuir al desarrollo de modelos específicos de valor agregado y a emplear los resultados de valor agregado para categorizar el desempeño escolar. La decisión de cómo presentar los resultados de valor agregado de las escuelas debe ponerse a prueba y después desarrollarse en la etapa piloto del proceso de implementación, y es crucial para obtener el compromiso eficaz de los actores clave en el proceso (NASBE, 2005).

Si bien la declaración de objetivos puede considerarse un requisito para la elaboración de toda política o programa en un sistema bien reglamentado, expresar de forma explícita estos objetivos influye decisiones como identificar el modelo adecuado de valor agregado, la forma de publicación de las puntuaciones de valor agregado de las escuelas y una estrategia de comunicación que gane la confianza de los actores. Si la información de valor agregado se va a usar para evaluar el desempeño escolar y moldear las iniciativas de mejoras escolares, es importante considerar cómo se incorporará dicha información al sistema existente de evaluación escolar para aumentar su eficacia. En la mayoría de los países miembros de la OCDE, el sistema actual de evaluación escolar utiliza inspecciones escolares (o una institución semejante) y/o autoevaluaciones escolares (OCDE, 2007a). Como se presenta en la Parte I, hay varios métodos para incrementar tanto la eficiencia como la eficacia de las evaluaciones escolares. Por ejemplo, puede elaborarse un sistema mediante el cual los resultados de valor agregado desencadenen evaluaciones escolares específicas. Esto aumenta la eficiencia mediante la identificación de escuelas de bajo desempeño o grupos de alumnos en riesgo, y aumenta el flujo de información cuando se establecen mecanismos para permitir que las escuelas de alto desempeño compartan sus mejores prácticas.

DETERMINACIÓN DE LA VARIABLE CON LA CUAL MEDIR EL VALOR AGREGADO

Tras definir de manera explícita los objetivos del desarrollo de los modelos de valor agregado es necesario especificar la medida (o medidas) con las cuales se calibrará el desempeño de las escuelas. Esto requiere identificar los instrumentos adecuados de evaluación de alumnos y la variable (o variables) dependiente que se usará en los modelos de valor agregado. La construcción de esta variable debe relacionarse directamente con los objetivos de desarrollo del sistema basado en modelos de valor agregado. Por ejemplo, si el objetivo es que los alumnos alcancen niveles mínimos de alfabetización y de conocimientos de aritmética,

entonces pueden identificarse los instrumentos de evaluación y la variable adecuada con los cuales medir el desempeño de valor agregado de las escuelas que mida así el desempeño estudiantil por encima de estos niveles.

Un modelo de valor agregado puede centrarse en diversos aspectos del desempeño de las escuelas. Las decisiones sobre el desempeño pueden afectar al tipo de modelo elegido y asimismo a las acciones de políticas y programas que partan del uso del modelo de valor agregado. Las decisiones concernientes a las áreas temáticas y a los niveles de grados o años en los que se usarán las evaluaciones de alumnos en los modelos de valor agregado revisten particular importancia en tanto que delinean los aspectos de una escuela en los que se mide el desempeño. Por tanto, estas decisiones definen lo que importa en una escuela cuando se estiman las puntuaciones de valor agregado escolar para promover la rendición de cuentas escolar, la elección de escuela o las mejoras escolares. Si se evalúa a los alumnos sólo en matemáticas o el idioma de instrucción, por ejemplo, la definición de la unidad de análisis ***escuela*** se refiere a los aspectos de esa escuela que contribuyen al desempeño en las medidas del grado o año en que tiene lugar la evaluación y, según la estructura del sistema escolar, los grados o año que encabecen la evaluación. Puede argumentarse que juzgar el desempeño escolar por las evaluaciones de conocimientos básicos de aritmética de alumnos de un grado específico coloca un peso desproporcionado en los maestros de matemáticas de una escuela de ese grado. Esto puede ser una decisión política intencional pero es necesario considerar estos asuntos y abordarlos explícitamente. El alcance de las pruebas de alumnos varía considerablemente entre los países miembros de la OCDE. En general, en los niveles educativos inferiores, sólo se aplican pruebas en áreas de aprendizaje básicas, como conocimientos básicos de aritmética y alfabetización. En los últimos años de secundaria se aplican a menudo pruebas de más materias, pero no siempre con instrumentos de evaluación estandarizados. Estas dificultades pueden allanarse mediante la definición de los modelos, pero deben reconocerse en el desarrollo de un sistema basado en modelos de valor agregado. En los sistemas que usan los resultados de valor agregado destinados a la mejora escolar, las medidas múltiples de valor agregado incrementan significativamente la potencia explicativa del análisis del desempeño escolar y la toma de decisiones. Esta toma de decisiones se beneficiaría de una variedad de datos que especifique el desempeño en diferentes áreas temáticas con apoyo en datos contextuales de los alumnos. Al contrario, los sistemas centrados en mejorar la rendición de cuentas escolar o la elección de escuela quizá requieran centrarse en una sola medida de desempeño.

Medidas categóricas y continuas

Dada la elección de evaluaciones en áreas temáticas particulares, otro asunto que necesita abordarse es cómo medir o categorizar el desempeño. Para medir el desempeño de los alumnos se puede emplear una medida continua que lo ubique en un continuo de puntuaciones (no obstante los efectos de techo de los instrumentos de evaluación de alumnos), o una medida categórica o dicótoma. Los instrumentos de evaluación de alumnos individuales también se diseñan para definir mejor los niveles predeterminados de aprovechamiento de los estudiantes. Quizá se prefiera especificar niveles particulares de desempeño que categoricen a los estudiantes de acuerdo con alguna medida, por ejemplo, de capacidad baja, media y alta. Las puntuaciones de valor agregado de las escuelas medirían por tanto su contribución a

esas categorías predeterminadas. Las medidas dicótomas son recomendables si el objetivo es medir la capacidad de la escuela para situar a sus estudiantes en un nivel de desempeño único, o superarlo. Los ejemplos más comunes incluirían habilidades mínimas de aritmética y alfabetización en niveles dados de grado o año. Estas medidas pueden ser el interés principal de las evaluaciones específicas de alumnos o extrapolarse desde medidas continuas. Esto ofrece a las escuelas el incentivo de centrarse en este aspecto del desempeño, lo que se vería como consecuencia positiva. Sin embargo, tiene una posible consecuencia negativa si dicho interés se realiza a expensas de los estudiantes en otros niveles de desempeño (Fitz-Gibbon y Tymms, 2002). La decisión de centrarse en medidas específicas debe alinearse con los objetivos de políticas del desarrollo de modelos de valor agregado e impulsar el desarrollo de programas.

El interés en niveles específicos de desempeño fomenta que los directores y maestros alcancen dichos niveles y también promueve el interés en grupos de estudiantes o materias particulares. En sistemas que no desean centrarse en una medida específica, quizá lo más adecuado sea una variable continua que mida el desempeño de los alumnos y, mediante esto, el valor agregado escolar. Así, las escuelas y otros administradores estarían en posibilidades de analizar una distribución mayor de datos para desarrollar y supervisar el desempeño escolar, así como programas y políticas específicos. Asimismo, ofrecería una distribución más equitativa de incentivos dentro de las escuelas en lugar de centrarse en un nivel específico de habilidades. En algunos casos se elaboran medidas continuas para agruparlas después en categorías predefinidas o estándares mínimos. Esto sería ventajoso si se diseñan los instrumentos adecuados para la evaluación de alumnos individuales.

Las decisiones sobre el desarrollo de evaluaciones de alumnos individuales en los modelos de valor agregado reciben una fuerte influencia de la estructura actual de la evaluación de alumnos, que puede estar ya bien establecidas en un sistema educativo. Debe decidirse si se elaboran más evaluaciones para complementar el marco existente. Pueden surgir complejidades adicionales al asegurar que las nuevas evaluaciones no perturben los objetivos del sistema educativo. Las evaluaciones ya existentes a menudo son determinantes del avance de los estudiantes a lo largo de su educación, y cualquier evaluación nueva podría perturbar el proceso educativo que se pretende evaluar con los exámenes ya existentes. En cambio, si se elaboran nuevas formas de evaluación, ambas formas deben ser complementarias.

La estructura de la variable dependiente afecta a decisiones relacionadas con la elección del modelo, pues determina el tipo de modelos que se pueden elegir. Si la variable dependiente en el modelo de valor agregado es dicotómica (o se reconstruirá de ese modo para aplicaciones particulares), necesita identificarse en una etapa temprana dadas las implicaciones de la elección de modelo. Las variables dependientes dicotómicas tienen distintos requerimientos con respecto al modelo que las variables dependientes continuas. Estos modelos se analizan en profundidad en la Parte II de este informe.

IDENTIFICACIÓN DEL MODELO ADECUADO DE VALOR AGREGADO QUE ABORDE MEJOR LOS OBJETIVOS DE POLÍTICAS

Dados los objetivos de políticas que impulsan el desarrollo de un sistema basado en modelos de valor agregado, es posible establecer las etapas básicas de un proceso mediante el cual se elija el modelo adecuado de valor agregado para su implementación principal. Este proceso

comienza con la identificación de los factores principales que afectarán a la elección del modelo. Estos factores incluyen la forma de aplicarlo y de interpretar sus resultados para alcanzar objetivos de políticas, y, en conexión con esto, la estructura de la medición de desempeño de los estudiantes (variable dependiente) con la cual se estimará el valor agregado. Cada modelo de valor agregado tiene ventajas y desventajas que deben considerarse en el contexto de los objetivos generales y el uso de la información de valor agregado. La segunda etapa del proceso es identificar los criterios estadísticos y metodológicos para elegir el modelo más adecuado de valor agregado. Esto se basará en los resultados de las estimaciones de diversos modelos de valor agregado realizados sobre los datos piloto o con los datos provenientes de las evaluaciones de alumnos ya realizadas en los sistemas educativos.

La especificación de los objetivos de políticas y analíticos establece un marco con el cual evaluar la validez de distintos modelos de valor agregado. El uso de los modelos de valor agregado para promover la rendición de cuentas escolar, mejoras escolares o elección de escuela plantea requerimientos específicos en los modelos de valor agregado y la necesidad de abordar diversos problemas estadísticos y metodológicos. Una distinción básica es si los resultados de los modelos se van a usar de manera interna o si también se van a hacer públicos. Esto guiará la toma de decisiones como la forma de abordar la inestabilidad de las puntuaciones escolares y el error de medida en escuelas más pequeñas, y asimismo ofrecer respuestas a preguntas más amplias sobre análisis adicionales que pueden realizarse con modelos más complejos para estudiar escuelas, estudiantes o programas educativos específicos. Asimismo, es importante recordar que, al elegir entre diversos modelos de valor agregado, debe efectuarse un análisis del impacto potencial para las escuelas que los usen. Por ejemplo, si se van a categorizar a las escuelas de bajo desempeño de este modo, debe analizarse las diferencias en esta categorización (sobre todo con el paso de los años, si es posible con los datos disponibles) con distintos modelos para identificar los diversos impactos en las escuelas y ver cómo se abordarían estas diferencias en la implementación real.

Deben identificarse numerosos criterios estadísticos y metodológicos. La Parte II de este informe identificó varios de estos retos y es posible elegir un modelo con dichos criterios. Deben tenerse en mente los objetivos de políticas generales al elegir estos criterios. Por ejemplo, puede dedicarse un mayor interés a la capacidad de separar de manera significativa el desempeño de diferentes escuelas o de reducir la inestabilidad de las puntuaciones escolares con el paso de los años. Puede decidirse excluir escuelas determinadas (p. ej., escuelas pequeñas, o a las que asisten estudiantes con necesidades especiales de aprendizaje) del análisis principal para obtener el "mejor ajuste" del modelo elegido. Estas decisiones serían beneficiosas si los objetivos de políticas están especificados con claridad así como la manera de usar la información de valor agregado como punto de partida (p. ej., en programas educativos particulares).

Los criterios básicos destacados en la Parte II de este informe se establecen durante la fase de implementación, y después se pondrían a prueba durante la fase piloto, de modo que se pueda tomar una decisión clara sobre el modelo más adecuado de valor agregado. Estos criterios se centrarían en:

- ***La cantidad de varianza y sesgo en diversos modelos.*** Diferentes modelos producirán diferencias en las estimaciones de los errores estándar que acompañan a cada puntuación de valor agregado de las escuelas. Esto tiene consecuencias para la capacidad de efectuar

distinciones estadísticamente significativas entre el desempeño de escuelas, lo cual quizá sea un objetivo clave del modelo. Esto será de particular importancia si las puntuaciones de valor agregado de las escuelas se van a hacer públicas y si las puntuaciones se categorizarán con base en diferencias estadísticamente significativas. Tal vez se prefiera un modelo porque errores estándar más pequeños implican que más escuelas se van a distinguir con precisión del promedio o se van a clasificar como que alcanzaron algún objetivo predefinido.

- ***El uso de datos socioeconómicos contextuales en diferentes modelos de valor agregado.*** Algunos modelos incluyen pocas características contextuales, mientras otros modelos de valor agregado contextualizado abarcan una gran cantidad de medidas socioeconómicas. La cantidad y frecuencia de medidas de aprovechamiento previas y actuales puede afectar a la fuerza explicativa por incluir tales características, y esto se pone a prueba en la fase piloto del proceso de implementación. También debe considerarse el impacto en los incentivos, así como el grado en que estos ajustes del modelo afectan a las acciones que resulten de las puntuaciones de valor agregado de las escuelas. La inclusión de características socioeconómicas afecta asimismo a los errores estándar asociados con los estimadores escolares y la forma en cómo el modelo se sostiene frente a los supuestos subyacentes.
- ***Datos perdidos y cómo se explican en el modelo.*** Como se presentó en la Parte II, algunos modelos de valor agregado están mejor preparados para trabajar con datos perdidos. En otros modelos será evidente el impacto en la fuerza predictiva del modelo y en el nivel de varianza y sesgo de las puntuaciones de valor agregado de las escuelas. Será necesario decidir sobre la exclusión de algunas variables, pero también hay procedimientos en la fase de implementación que reducen el patrón de los datos perdidos al crear incentivos para una alta participación estudiantil, o para desalentarla.
- ***Cambios en los resultados de escuelas pequeñas en diferentes modelos.*** Las muestras pequeñas procedentes de las escuelas de menores dimensiones a menudo producen medidas menos precisas y confiables, así como menos estables en años sucesivos. Los modelos que "concentran" la media de los resultados de valor agregado de las escuelas pequeñas generan resultados más útiles, pero hay problemas claros con este nivel de intervención en los datos. En general, los países participantes consideraron que cohortes de menos de 20 a 30 estudiantes produjeron estimadores de valor agregado con dificultades en la interpretación de resultados. Este problema debe analizarse en la fase piloto de implementación.
- ***Cambios de las puntuaciones de valor agregado de las escuelas con el paso del tiempo.*** También puede analizarse la estabilidad de las puntuaciones escolares con el tiempo, así como medir su efecto en escuelas determinadas. Esto se relacionaría con el tamaño de la varianza y el potencial de sesgo en el modelo. Si la estabilidad de las puntuaciones escolares se considera demasiado baja, se imponen estándares que reduzcan cualquier impacto negativo. Por ejemplo, si la inestabilidad se concentra en escuelas determinadas, éstas se retiran del análisis principal. Para estas escuelas se aplican estimaciones adicionales y, según los principales objetivos de políticas, se introducen iniciativas separadas de rendición de cuentas o de mejoras. Asimismo se aplican estándares para retirar del estudio a las escuelas con cambios anormalmente grandes en varios años. Se podría definir como una proporción del cambio en todas las puntuaciones escolares, o en puntuaciones semejantes. El uso de una media móvil de tres años en la medida del valor agregado suavizaría los

cambios a lo largo del tiempo. Además, pueden ofrecer la oportunidad de efectuar otro análisis de escuelas con cambios anormales en la puntuación escolar de un solo año.

Todos estos temas se evalúan mientras los modelos están en la etapa piloto. Este análisis también ofrece a los tomadores de decisiones una oportunidad de analizar el impacto de aplicar diferentes estándares al uso de los datos, como la inclusión de datos perdidos y de escuelas con muestras más pequeñas. Las diferencias de estos estándares tendrían diversos impactos con diferentes modelos de valor agregado. Para incrementar la transparencia, estos criterios se ponderan para guiar la posterior toma de decisiones. Las decisiones sobre estos asuntos no serán evidentes, pues algunos modelos pueden ser superiores de acuerdo con algunos criterios e inferiores con otros. Las decisiones requerirán juicios del desempeño de cada modelo según los criterios elegidos. Cuando surgen dificultades, vale la pena considerar el análisis de las diferencias entre puntuaciones de valor agregado entre dos modelos y estimar el impacto de dichas diferencias conforme a los objetivos de políticas prescritos (p. ej., escuelas que se identifiquen como de bajo desempeño).

Después de especificar las características básicas de lo que se requiere de los modelos de valor agregado, se emprende un análisis con los datos de las evaluaciones estudiantiles existentes o de los datos obtenidos en la etapa piloto del desarrollo de un sistema basado en modelos de valor agregado. Este análisis evalúa la adecuación de distintos modelos de valor agregado para cumplir con los objetivos del sistema y para abordar los criterios estadísticos y metodológicos predeterminados. Los resultados de este análisis deben presentar las ventajas y desventajas de los diversos modelos de valor agregado y, a partir de aquí, recomendar uno. Más importante aún, deben identificar las implicaciones de la elección del modelo en el uso y aplicación de las puntuaciones de valor agregado en las escuelas y en los objetivos de políticas y programas prescritos. Esto subrayará el impacto para tipos particulares de escuelas, pero también debe identificar el grado en que los distintos modelos cumplirían con los objetivos de políticas prescritos. Para lograr estos fines, es importante, al evaluar la adecuación de distintos modelos de valor agregado, analizar no sólo el modelo en general (p. ej., bondad de ajuste), sino también el impacto de diferentes modelos en escuelas individuales.

ELABORACIÓN DE UNA BASE DE DATOS EFICAZ

Esta sección analiza los aspectos básicos de la elaboración de una base de datos que apoye el desarrollo y administración eficientes de un sistema basado en modelos de valor agregado. Dada la discusión sobre el error de medida y sobre la falta de especificación del modelo en la Parte II de este informe, debe considerarse la calidad de los datos y, de ser posible, mejorarla en la etapa de desarrollo. Este requerimiento afecta al tema clave del alcance del conjunto de datos que ofrece oportunidades para elaborar sistemas de datos más completos con los cuales analizar el valor agregado y aspectos más amplios del sistema educativo. Sin embargo, la ampliación de la base de datos no debe acompañarse de una reducción de su calidad. La siguiente discusión sobre la elaboración de una base de datos integrada para ayudar en la toma de decisiones y el desarrollo de políticas debe considerarse en el contexto de los datos actuales recopilados en cada sistema educativo y de los costos de elaborar una base de datos eficaz, dada la importancia de la calidad de los datos.

Como los estimadores de valor agregado son una fuerza poderosa para el cambio, es crítico

que la base de datos se construya y mantenga con el mayor cuidado, de modo que se eviten errores u omisiones que contaminen los resultados. La calidad de los datos para los modelos de valor agregado tiene un claro impacto en la confianza con que se interprete el desempeño escolar. La elaboración de sistemas de datos varía por país debido a razones diversas. La elaboración de una base de datos de alumnos eficaz ha sido crucial para la efectividad del sistema basado en modelos de valor agregado en el **Reino Unido**. En 1997 se promovió el desarrollo de mejores datos en el nivel de alumnos, y en 1999 se introdujo un identificador de alumno único que contribuiría a homologar los datos en todo el sistema educativo. Otro avance clave fue el inicio del censo escolar anual ***de alumnos individuales*** en 2002, que recopila los datos de características de antecedentes que las escuelas registran con fines administrativos. Para aumentar la profundidad y eficacia del análisis, es necesario posteriormente incorporar estos datos individuales a un sistema único de datos que permita a los usuarios analizar los resultados de valor agregado de las escuelas junto con diversos datos contextuales y escolares. Es posible mejorar la eficiencia al consolidar las fuentes de datos en un solo sistema completo de datos.

El primer paso en la elaboración de una base de datos de alta calidad como la requerida es identificar los datos que se usarán para los modelos de valor agregado. Para las autoridades educativas que desean elaborar un sistema basado en modelos de valor agregado con el fin de facilitar la toma de decisiones referentes a mejoras escolares y desarrollo de políticas, conviene elaborar una base de datos amplia que se extienda más allá de los requerimientos mínimos de datos para los modelos de valor agregado. Una decisión clave en esta etapa es si los beneficios de un sistema de datos más completo son mayores que los costos de desarrollo y mantenimiento. Este sistema incluiría datos complementarios de diversas fuentes, pero para los sistemas que no desean complementar sus datos básicos de evaluaciones de alumnos, los recursos pueden concentrarse en garantizar el desarrollo y mantenimiento de una base de datos que produzca estimadores de valor agregado de alta calidad. Si se requiere un sistema de datos más completo, es necesario abordar la cuestión de qué información es la que se necesita recopilar. Se pueden recopilar cuatro tipos principales de datos para su inclusión en el análisis de valor agregado y fomentar el desarrollo de políticas. Los cuatro tipos principales de datos son:

i. ***Datos de la evaluación de los alumnos,*** que abarcan todas las puntuaciones de evaluación de los alumnos para los modelos de valor agregado. Esto incluiría todas las puntuaciones individuales previas y actuales, cruzadas mediante identificadores de cada alumno. También incluiría toda medición compuesta de puntuaciones de evaluaciones combinadas (p. ej., un promedio de puntuaciones de diferentes materias) y mediciones específicas consideradas de importancia para fines de políticas (p. ej., requerimientos mínimos de alfabetización). Pueden calcularse indicadores o variables adicionales como objetivos de desempeño, o pueden emplearse puntuaciones escolares o del alumno para desencadenar acciones específicas. Según el desarrollo de la base de datos con el tiempo, quizá sea útil realizar un seguimiento de los estudiantes para identificar resultados adicionales en su escolaridad y resultados en el mercado de trabajo. Esto sería necesario para el análisis del valor agregado de las escuelas medido con respecto de otros resultados, como porcentaje de alumnos que avanzan a la educación secundaria postobligatoria, y para analizar datos escolares respecto de otros resultados socioeconómicos.

ii. ***Información contextual del alumno,*** que incluye todas las características individuales (p. ej., edad de los estudiantes) y familiares, entre otras, que se consideren necesarias para el análisis en el modelo de valor agregado (en este caso, contextualizado). La elección de estas características se analizó en el Capítulo 6, "Elección de modelo y aspectos metodológicos", y debe plantearse dos objetivos. El primer objetivo es usar estas características contextuales en los modelos de valor agregado, en particular en modelos de valor agregado contextualizado más extensos. Estas son características importantes con las cuales captar el efecto de factores ajenos al control de la escuela que afectan el progreso escolar del alumno. Sin embargo, algunos modelos de valor agregado no las requieren, pues añaden poca fuerza predictiva al modelo y no tienen un gran impacto en los resultados escolares. El segundo objetivo es el empleo de estas características para investigar el valor agregado en escuelas determinadas o grupos específicos de estudiantes. Por ejemplo, puede haber un interés particular en el valor agregado de estudiantes con antecedentes socioeconómicos más bajos, de grupos inmigrantes particulares o de grupos étnicos. El análisis de estos subgrupos requiere los datos contextuales de los alumnos correspondientes.

   Medir la contribución de las escuelas y otros factores al avance estudiantil requiere una base de datos que identifique y defina con precisión los datos de los alumnos individuales. Esto requiere tener identificados normalmente a los estudiantes con alguna forma de número o código de identificación que se distinga en los datos de evaluación de los alumnos y en toda la demás información estudiantil contextual. Los números de identificación del estudiante son necesarios para identificar y registrar a los alumnos cuando ingresan y egresan de la escuela. El problema de la movilidad estudiantil debe abordarse en los modelos de valor agregado tanto por los valores perdidos que se crean en el conjunto de datos como por los problemas de atribuir crecimiento del desempeño de alumnos a las diferentes escuelas. Para un análisis preciso de este problema, debe haber un sistema de información que registre de manera adecuada la movilidad estudiantil entre escuelas, en particular entre los periodos de evaluación estudiantil predeterminados que alimentan a los análisis de valor agregado. En algunos países esto es más difícil que en otros. Algunos países, como **Dinamarca** y **Noruega**, emplean sistemas de información administrativa ya presentes que sistemáticamente asignan a todos los estudiantes un número de identificación y permite su registro eficaz. El establecimiento de estos sistemas es costoso y muy exigente en cuanto a recursos. Se pueden encontrar otras complicaciones si se implican a diferentes jurisdicciones e instituciones. En **Polonia**, el registro de estudiantes se intentó por primera vez con datos de las Juntas de exámenes nacionales (*National Examination Boards*). Sin embargo, los datos requeridos de los alumnos sólo existían en los datos recopilados por las Juntas regionales (*Regional Boards*). Así, se emprendió un proceso para hacer corresponder los datos de las diversas Juntas regionales que se vio obstaculizado por la falta de números de identificación de los estudiantes (sólo se contaba con nombre, género y fecha de nacimiento). Fue un proceso costoso y consumidor de recursos, pero necesario para el desarrollo de un sistema basado en modelos de valor agregado. Como tal, también provocó cambios en el manejo de los datos del sistema, como la introducción de números de identificación del alumno.

iii. ***Información escolar*** considerada necesaria en varios sistemas, que incluye datos del sector escolar y tipo de escuela, además de datos que indiquen si la escuela se ubica en regiones

específicas. El tamaño de la escuela (medido por la cantidad de alumnos) debe ser identificable debido a la inestabilidad asociada a menudo a las puntuaciones de valor agregado de escuelas pequeñas. Asimismo puede ser conveniente recopilar información clave sobre programas y políticas que facilite el análisis de sus relaciones con las puntuaciones de valor agregado. Estos datos ofrecen un ingrediente básico en el control de calidad general en el sistema educativo, y facilitará el desarrollo y supervisión de programas y políticas específicos dirigidos a las mejoras escolares. Esto puede hacerse en el ámbito de la escuela, del distrito o de la región, según la naturaleza del programa. Por ejemplo, Goldhaber y Brewer (2000) analizaron la relación entre la certificación de profesores y las puntuaciones del docente de valor agregado. En el **Reino Unido** se efectuaron análisis internos sobre programas específicos, como los programas de escuelas especializadas que proporcionan fondos adicionales y un currículo extendido en áreas particulares. El diseño del modelo de valor agregado y la información en que se apoyó permitieron una medición de desempeño del impacto de estos programas por desarrollarse.

También puede recopilarse información contextual escolar con el propósito, semejante al interés de la información contextual del alumno, de incluir características en un modelo de valor agregado contextualizado que "nivele el campo de juego" para análisis comparativos de puntuaciones de valor agregado de las escuelas. Puede usarse información escolar en lugar de la del estudiante si no es posible recopilar datos de los estudiantes o si es más fácil hacerlo de las escuelas. Esto puede ser el caso si ya existen datos escolares administrativos que midan lo suficiente los factores contextuales requeridos. Sin embargo, debe garantizarse la confiabilidad de estos datos. En algunos sistemas se emplean varias medidas socioeconómicas en datos administrativos como parte de programas que otorgan recursos adicionales a escuelas en desventaja. Estas medidas quizá no necesariamente midan de forma adecuada los factores que necesiten captarse con el fin de aislar los efectos de la escuela en los modelos de valor agregado, en particular si ofrecen medidas burdas del nivel socioeconómico. Las medidas menos precisas son también menos eficaces en la provisión de datos que faciliten el análisis de grupos de estudiantes particulares y de escuelas cancelando el potencial de estos análisis para el estudio de las diferencias dentro de las escuelas. Otro problema con los datos administrativos es el potencial de sesgo. En algunos sistemas educativos, los datos escolares administrativos sobre el nivel socioeconómico o alguna desventaja de aprendizaje en la escuela provienen de los directores o administradores. Si los directores proporcionan estas medidas sabiendo que pueden afectar a la puntuación de valor agregado de la escuela o al nivel de recursos que recibe, entonces debe considerarse que la provisión de estos datos es una posible fuente de parcialidad. Estos problemas han salido a la luz en varios sistemas educativos y generan dificultades en la interpretación de los modelos de valor agregado contextualizado.

iv. ***Información de la evaluación de la escuela e informes,*** que al ofrecer más información evaluativa referente al desempeño escolar, contribuyen a la interpretación de las puntuaciones de valor agregado y al uso de modelos de valor agregado para el desarrollo de programas, y permiten acciones que mejoren el sistema de evaluaciones escolares. Es importante subrayar que las puntuaciones de valor agregado no proporcionan una panorámica completa del desempeño escolar. Se otorga más confianza a las interpretaciones y acciones provenientes de puntuaciones de valor agregado si se cuenta con información

evaluativa adicional. Si esto forma parte de un sistema de datos completo, vincular la información de valor agregado de las escuelas a la información evaluativa de inspecciones escolares y autoevaluaciones escolares, por ejemplo, representa una fuente valiosa para el desarrollo de iniciativas de mejoras escolares. La información escolar adicional permitiría análisis más detallados de escuelas de bajo y alto desempeño. Además, se gana eficiencia al permitir que las instituciones y actores que evalúan escuelas y programas escolares analicen la información de valor agregado de las escuelas. Esto facilita la orientación de las evaluaciones escolares hacia las áreas pertinentes y permite una evaluación escolar basada en resultados y no en factores de entrada o insumos. Asimismo, esto contribuye al funcionamiento de inspecciones escolares en tanto que permite el análisis de las recomendaciones y juicios de inspectores y de cómo se relacionan con las puntuaciones de valor agregado escolar. Esto facilita en gran medida la calidad del seguimiento dentro y fuera de las inspecciones escolares.

La vinculación de la información de valor agregado de las escuelas a otra información evaluativa también se considera, a la luz del uso de los datos de valor agregado, una mejora de la elección de escuela. La publicación de las puntuaciones de valor agregado de las escuelas es beneficiosa para padres y familias, pues ofrece información para la toma de decisiones sobre la escuela que mejor se adapte a sus necesidades. Por la variedad de necesidades de padres y familias, y los requerimientos que exigen de las escuelas, quizá se considere conveniente proporcionar más información evaluativa que facilite una mejor elección de escuela. Esto puede presentarse en un formato semejante al de los Cuadros de Desempeño Escolar disponibles en el **Reino Unido** o a la información evaluativa escolar ahora disponible para el público en la **Comunidad Flamenca de Bélgica**.

Si bien la creación de una base de datos y métodos de recopilación de datos flexibles tiene el potencial de facilitar en gran medida el uso de modelos de valor agregado para el desarrollo continuo de programas, es conveniente que se identifiquen los datos de los alumnos en el periodo inicial de desarrollo. Un paso importante en la identificación de los datos requeridos es garantizar un conjunto aceptado de definiciones básicas de todas las variables que se recopilarían. En algunos países, las leyes de privacidad pueden limitar el uso de datos contextuales. En **Polonia**, las leyes de privacidad impiden el uso extenso del nivel socioeconómico en sus modelos de valor agregado, y en **Eslovenia** se requieren acuerdos firmados por los padres antes de obtener datos socioeconómicos de los estudiantes. La articulación de los objetivos y acciones específicos vinculados a los modelos de valor agregado es central para el problema de identificar los requerimientos de datos. Esto facilita la identificación de las características e información básicas que se necesita recopilar y saber por anticipado si estos datos se van a usar internamente, si su uso se extenderá a las escuelas y otros interesados educativos o si los va a usar el público en general. Una vez abordados estos aspectos y acordada una estrategia amplia sobre la elaboración de un sistema de datos, es posible revisar los sistemas de datos existentes y las capacidades de los recursos invertidos en ellos. Esto incluiría una consideración de asuntos prácticos, como el *software* en uso y el control de calidad actual, como la garantía de estándares comunes en las recopilaciones de datos. En ese momento será posible determinar si se necesitan más datos, si es necesario diseñar nuevos métodos de recopilación y si se requiere establecer una nueva infraestructura de sistemas de información.

IMPORTANCIA DEL PROGRAMA PILOTO PARA LOS MODELOS DE VALOR AGREGADO

El objetivo del programa piloto es evaluar y desarrollar distintos aspectos del sistema basado en modelos de valor agregado. Esto incluye cuestiones de operación e implementación, decisiones sobre las evaluaciones estudiantiles y la elección del modelo específico de valor agregado, el diseño de estrategias de comunicación y compromiso con los actores clave, y evaluar la forma de interpretar las puntuaciones de valor agregado de las escuelas y otra información, y de cumplir así de la mejor manera con los objetivos de políticas. Estos asuntos se han analizado a lo largo de este informe y necesitan abordarse durante el programa piloto. Por tanto, el programa piloto no debe verse sólo como una prueba del modelo específico de valor agregado a utilizar en un sistema educativo. El análisis en este reporte contó con información de los resultados de los programas desarrollados en los países participantes.

Un programa piloto suele aplicarse en un subconjunto de escuelas, y se considera una operación de prueba antes de la puesta en ejecución real. Debe tratarse de la misma manera que la implementación real de un sistema basado en modelos de valor agregado para crear una evaluación realista y válida. El método con que se selecciona o solicita su participación en el programa piloto a un subconjunto de escuelas varía entre países, pero es importante que la muestra de escuelas que participe pueda aplicarse a una implementación real. Esto requiere obtener una muestra de escuelas representativa de la población escolar mayor y que se comprometa de manera eficaz con el proceso de evaluar el establecimiento de los modelos de valor agregado. Para fomentar el compromiso eficaz en los estudios piloto de la vida real, algunos sistemas educativos destacaron que el estudio piloto no se usara como herramienta para la rendición de cuentas escolar. Al seleccionar al subconjunto de escuelas, vale la pena considerar la idea de que al inicio las escuelas quizá se sientan menos inclinadas a participar en un estudio que las someta a una rendición de cuentas y medición de desempeño adicionales.

En casos donde no sea posible obtener una muestra representativa, conviene incluir una cantidad suficiente de escuelas de diferentes sectores y regiones en el programa piloto, pues esto permitirá un mejor análisis de la presencia de factores específicos en una región particular, por ejemplo, que necesiten tenerse en cuenta para la implementación real. Tal vez se hallen factores específicos que requieran cambiar alguna variable específica en el modelo de valor agregado (p. ej., variables que midan el sector escolar, o la proporción de estudiantes con necesidades especiales de aprendizaje o de antecedentes desfavorables), o problemas de implementación que necesiten considerarse. Por ejemplo, quizá sea necesario modificar las estrategias de comunicación y compromiso con los interesados en escuelas de áreas regionales, rurales o indígenas.

Todos los aspectos referentes a la evaluación de los estudiantes, el uso de sistemas de información para compilar conjuntos de datos y la operación de estimadores de valor agregado deben efectuarse como si fuera la implementación real. Si ya existe una estructura de evaluaciones de alumnos, sería conveniente utilizar estos datos para evaluar la confiabilidad de los sistemas de información en uso y las estimaciones de los modelos de valor agregado. Esto ofrecería una evaluación de cualquier limitación de capacidad del sistema de información a emplear. Asimismo, permitiría un juicio más completo sobre la adecuación de la elección del modelo de valor agregado.

No es adecuado elegir *a priori* el modelo específico a implementar en un sistema educativo. Debe considerarse la etapa piloto como el momento de evaluar el modelo de valor agregado más adecuado para la implementación real. Esta evaluación debe efectuarse respecto de un conjunto de criterios predeterminados, como se presentó anteriormente. Para que un programa piloto sea fructífero, se necesitan datos de varios años para determinar cómo difieren las puntuaciones escolares entre diversos modelos. En algunos sistemas educativos ya existe una estructura de evaluaciones de alumnos desde mucho antes del establecimiento de un sistema basado en modelos de valor agregado. Los datos de evaluaciones de múltiples años por tanto pueden usarse como información para la elección de escuela. En sistemas educativos en donde no exista este marco, la decisión final del modelo más adecuado se puede extender a la ejecución inicial de los modelos de valor agregado en una población escolar más amplia. Esto extendería el periodo del análisis de las puntuaciones de valor agregado escolar a evaluaciones sucesivas, las cuales serían importantes si se descubriese excesiva inestabilidad en puntuaciones específicas de la escuela. Por tanto, quizá sea prudente posponer el empleo de puntuaciones de valor agregado con fines de rendición de cuentas escolar debido a la mayor incertidumbre en la estimación. Según la medida de la inestabilidad y la capacidad de aislarla en un subconjunto particular de escuelas, se podría ver como parte del desarrollo continuo más amplio de los modelos de valor agregado. Cualquier análisis para determinar la forma de mejorar el modelo debe considerarse parte de un proceso continuo.

El programa piloto ofrece una excelente oportunidad para desarrollar una estrategia de compromiso y comunicación con los actores clave. El proceso de compromiso comienza con el reclutamiento de escuelas para el programa piloto y con darles la oportunidad de contribuir a los objetivos del sistema general de modelos de valor agregado. Directores, maestros y otros miembros del personal pueden ayudar a evaluar y diseñar respuestas a cuestiones de operación, del uso eficaz de información de valor agregado, en particular en el ámbito escolar, y de la estrategia de comunicación y compromiso. Así se obtiene más información del personal participante respecto del marco de evaluaciones de alumnos, de la recopilación de datos complementarios, en particular en el ámbito escolar, y para el desarrollo del sistema de información más adecuado. Se considera parte importante de un procedimiento de muestreo que reduce el nivel de inconveniencias y trabajo impuesto a la unidad de muestra (en este caso, la escuela). La retroalimentación durante la etapa piloto incrementaría en gran medida la eficiencia operativa y reduciría el impacto en el trabajo normal de las escuelas.

Un elemento importante del programa piloto, respecto de los procedimientos operativos, es garantizar procedimientos precisos de recopilación de datos. Si se va a reunir información adicional de las escuelas, debe emprenderse un desarrollo adecuado del programa piloto y de los cuestionarios. Si se van a usar datos administrativos, también debe verificarse esto con las escuelas para asegurar la precisión y suficiencia de los datos. El control de calidad de los datos y su recopilación debe formar parte de la implementación real, pero los procedimientos de supervisión pueden desarrollarse y evaluarse durante el programa piloto. Aunque la elección del modelo de valor agregado requiere experiencia estadística y no se ajusta idealmente a las opiniones de todos los interesados, es conveniente obtener información sobre el uso de los datos para el desarrollo de un modelo de valor agregado contextualizado. Los actores clave pueden estar en condiciones de aconsejar sobre la necesidad de incluir factores específicos

que afecten al desempeño estudiantil en los modelos de valor agregado y que quizá afecten también las acciones derivadas de las puntuaciones de valor agregado de las escuelas.

Dado que las escuelas serían el objetivo principal de una estrategia de comunicación, los directores, maestros y otros miembros del personal pueden ofrecer puntos de vista esenciales para la comunicación eficaz con otras escuelas e interesados. Esta información influiría en los objetivos de la estrategia, pero el programa piloto proporciona asimismo una oportunidad de evaluar el valor de materiales específicos de información y guía (p. ej., sobre el uso del sistema de información para analizar los datos de valor agregado de las escuelas), y seminarios y talleres para las escuelas. Esto iría más allá de la interpretación correcta de las puntuaciones de valor agregado escolares al uso de sistemas de información que contengan información de valor agregado de la escuela y del alumno para supervisar el desempeño escolar y diseñar los programas correspondientes de mejoras escolares.

El programa piloto brinda una importante oportunidad de diseñar programas eficaces de formación y asimismo de obtener el compromiso de directores y maestros en el uso de la información de valor agregado para la mejora escolar. Este compromiso debe representar un paso importante en la obtención del apoyo de los interesados para el establecimiento efectivo del sistema basado en modelos de valor agregado. Los directores y maestros ofrecen puntos de vista valiosos para interpretar y presentar mejor la información de valor agregado. Esto incluiría la presentación de la información de valor agregado, incluso la clasificación de puntuaciones específicas, y el empleo de otra información evaluativa. Asimismo se evalúa el valor de varios programas formativos para definir mejor tanto su propósito principal como sus resultados. En algunos países, un aspecto básico fue comprometer a directores y maestros para analizar sus inquietudes sobre las puntuaciones escolares que percibían poco realistas. Los beneficios de este diálogo requirieron formación adicional para los interesados, la cual trascendió las sesiones informativas para desarrollar la capacidad analítica dentro de las escuelas. Las evaluaciones de seguimiento del valor de dicha formación determinan si es posible mejorar aspectos particulares de los modelos de valor agregado o la interpretación de la información de valor agregado.

El programa piloto debe servir para evaluar aspectos de validez y confiabilidad de los instrumentos de evaluación. Las pruebas estandarizadas son el resultado final de un largo proceso de diseño y desarrollo, moldeado por una multitud de metas y limitaciones (Braun, 2000). Al calibrar la validez de los instrumentos de evaluación, es necesario abordar aspectos tanto sustantivos como técnicos. Por ejemplo, los análisis adicionales pueden incluir el grado de articulación entre el contenido real de la prueba y los estándares de contenido que se supone debe aplicar la escuela. Deben analizarse este y otros problemas para garantizar la confiabilidad de los instrumentos de evaluación antes de la implementación real del sistema basado en modelos de valor agregado.

Dados los objetivos de un programa piloto, cabe esperar que surjan problemas. Deben diseñarse planes para documentar y después resolver dichos problemas. Este es un paso básico para cumplir con el objetivo del programa piloto de desarrollar el sistema basado en modelos de valor agregado. Los problemas que surjan en el programa piloto pueden así verse como una oportunidad y no como un fracaso, y deben incorporarse a un sistema de control de calidad que opere durante toda la vida útil del sistema basado en modelos de valor agregado.

Un sistema eficaz de control de calidad debe garantizar que se mantengan procedimientos de alta calidad y que se enfrenten los problemas para asegurar una mejoría continua. Estos procedimientos deben supervisar aspectos del sistema como el marco de las evaluaciones de los alumnos individuales, el modelo para estimar el valor agregado, la interpretación de las puntuaciones de valor agregado de las escuelas y la precisión de los datos en el sistema. Todo problema que necesite atenderse en el programa piloto debe servir como ejemplo de los retos que necesitan supervisarse una vez que el sistema entre en funcionamiento. Esta supervisión debe dedicarse a garantizar que las puntuaciones de valor agregado de las escuelas sean estimadores precisos del desempeño escolar.

CAPÍTULO 8

Desarrollo y uso de los modelos de valor agregado

Uso de resultados de una media móvil de tres años 173
Comunicación y compromiso con actores clave 174
Desarrollo de un programa de formación 178
Presentación y uso de la información de valor agregado 182

La eficacia de un sistema que se base en mediciones de desempeño escolar para emprender acciones descansa en la confianza de los interesados en la confiabilidad de las mediciones de desempeño con el paso del tiempo. Un control de calidad eficaz de los resultados de los modelos y de los datos con que se alimenta dicho análisis es central para el uso eficaz de un sistema basado en modelos de valor agregado. El análisis que aquí se presenta subraya la importancia de supervisar los resultados de valor agregado de las escuelas con el tiempo. Esta supervisión necesita dirigirse a los cambios de los resultados de las escuelas individuales, pues éstos son el interés principal de los actores clave y de los esfuerzos por elevar el desempeño. Dada la necesidad de reducir la variación inestable en los resultados de valor agregado de las escuelas, la discusión destaca la necesidad de calcular y presentar una media móvil de tres años de cada puntuación de valor agregado de las escuelas como el indicador central o el indicador publicado de desempeño escolar. Se presenta entonces un análisis sobre la forma en cómo los sistemas desarrollan estrategias fructíferas de comunicación y compromiso con los actores clave y la formación de los mismos, en particular directores y maestros. El capítulo concluye con un análisis de la manera en que la fase piloto del proceso de implementación alimenta la toma de decisiones respecto de la publicación de las puntuaciones de valor agregado de las escuelas.

La credibilidad de todo sistema analítico reside, en primer lugar, en la integridad de los datos y las operaciones realizadas con esos datos. Así, diseñar y establecer procedimientos eficaces de control de calidad en cada etapa del proceso es un aspecto esencial de un análisis de valor agregado. Parece claro que tanto los datos de las pruebas como las covariantes deben verificarse y editarse con cuidado antes del análisis. Esto implica identificar valores fuera del límite o inusuales, así como características distribucionales inesperadas. En ocasiones son útiles las comparaciones con datos de años anteriores. Los patrones de los datos perdidos también pueden hacer que sea necesario diseñar y llevar a cabo un análisis y las acciones consecuentes. Como ejemplo de una iniciativa de supervisión específica, puede seleccionarse una muestra de escuelas después de dada levantamiento de datos y analizarla con más detalle para garantizar que los datos sean precisos y se interpreten correctamente. En particular, debe incidirse en los cambios sustanciales en la cantidad de estudiantes excluidos de los exámenes (p. ej., a causa de discapacidades) o la cantidad de alumnos ausentes el día del exámen, pues esto señalaría la presencia de sesgos en la estimación de los efectos de la escuela. En algunos países, las escuelas enfrentan graves trabas si los estudiantes no se presentan a los exámenes determinados.

Los cambios en las puntuaciones de valor agregado de las escuelas tenderán a verse como señales de cambios en el desempeño escolar, aunque esto puede no justificarse con bases estadísticas. Las puntuaciones de valor agregado menos estables pueden provocar directa o indirectamente inferencias o acciones incorrectas, y su utilidad potencial quedaría limitada por una impresión de inexactitud. Idealmente, los indicadores de desempeño escolar serían relativamente estables, pero conservan la capacidad de subir y bajar en respuesta a los cambios reales del desempeño escolar. Es poco probable alcanzar esta situación ideal en todos los casos. Por tanto, es necesario analizar los cambios en los resultados de valor agregado de forma extensa durante el programa piloto y efectuar análisis de estos cambios a lo largo del tiempo una vez establecido el sistema. En sistemas educativos que analizan datos pre-existentes, existe la oportunidad de probar más la especificación del modelo y evaluar la estabilidad de las puntuaciones escolares con el tiempo para brindar información con el fin de elegir un modelo y determinar lo adecuado de las evaluaciones de alumnos y los datos empleados en el mismo.

En algunos países participantes se efectuó un análisis de la estabilidad de los resultados escolares. Es aceptable cierta inestabilidad de las puntuaciones de las escuelas en todos los modelos de valor agregado y, desde luego, es deseable cierta inestabilidad. En algunos sistemas educativos fue evidente una mayor inestabilidad, y esto quizá refleje la menor calidad de los sistemas de exámenes en esos países. Por tanto, deben examinarse las evaluaciones de los alumnos con las que se midió el valor agregado si se considera excesiva la inestabilidad de las puntuaciones escolares. La estabilidad de las puntuaciones escolares depende no sólo de la definición de inestabilidad anormal o excesiva, sino también de la categorización de escuelas por su nivel de desempeño. Asimismo se observó que la inestabilidad de las puntuaciones varia con el tamaño de la escuela, el tipo de modelo usado, la cantidad de variables contextuales incluidas, la cantidad de años transcurridos entre las medidas del rendimiento previo y el actual, y la cobertura de la comparación de valor agregado (todas las escuelas en el país o un subconjunto). Estos hallazgos ilustran los beneficios de llevar a cabo más análisis en escuelas con cambios grandes de puntuaciones escolares o cambios de apariencia aleatoria en ciertos años.

Los análisis adicionales de escuelas con puntuaciones menos estables en el tiempo pueden dificultarse debido al complejo objetivo de diferenciar los cambios observados en lo que pueden denominarse componentes "persistentes" y "pasajeros". Los primeros se refieren a cambios estables en el desempeño real, y los últimos, a todos los demás factores. El componente pasajero de inestabilidad puede atenuarse en cierta medida al incorporar más datos (es decir, más años anteriores y más materias) y al promediar los resultados en cohortes sucesivas. Sin embargo, un análisis más detallado de los datos puede revelar la fuente de inestabilidad en las puntuaciones escolares. Por ejemplo, la elaboración de modelos para algunas materias o subgrupos particulares dentro de una escuela puede servir para establecer si parecen verosímiles los cambios en las puntuaciones. También pueden indicar que los cambios en los datos usados o en evaluaciones de alumnos específicas generaron variación en las puntuaciones escolares. Esto ayudaría en un análisis que estimara las diferencias entre efectos escolares persistentes y pasajeros. Si se conocen problemas específicos, como un cambio de clasificación de datos contextuales, esto puede señalarse en las publicaciones y llamar la atención de los inspectores escolares y otros usuarios de los datos. Para la presentación de los resultados, la inestabilidad de los resultados escolares puede verse como otro argumento para presentar intervalos de confianza en torno a estimaciones puntuales. Asociar un intervalo de confianza a cada puntuación de valor agregado de una escuela aminora la probabilidad de una interpretación deficiente.

Se consideró aconsejable en varios sistemas educativos presentar estimaciones puntuales con intervalos de confianza, con la advertencia de que la superposición de intervalos de confianza indica que las estimaciones puntuales correspondientes no presentan diferencias estadísticamente significativas. Cuando se efectúan muchas comparaciones como éstas se corre un riesgo alto de cometer muchos errores de Tipo I. Este peligro se mitiga con técnicas de inferencia simultánea, las más conocidas de las cuales se llaman "métodos Bonferroni". Técnicas más recientes, como las que se basan en el enfoque de la Tasa de Descubrimiento Falso (Benjamini y Hochberg, 2000), son cada vez más populares. Para el público en general, las presentaciones gráficas son muy eficaces. La llamada *caterpillar plot* es particularmente eficaz. Las estimaciones de los efectos de la escuela se ordenan

por su clasificación a lo largo del eje X y por magnitud a lo largo del Y. Además, se coloca un intervalo de confianza para cada efecto de manera vertical y centrado en la estimación puntual. Asimismo, debe explicarse que si bien los intervalos de confianza ilustran mejor las diferencias estadísticamente significativas en los resultados escolares, no son una panacea ni captan la incertidumbre debida al sesgo y a otros cambios seculares. Algunas fuentes potenciales de sesgo pueden incorporarse al modelo, como en el modelo de Ponisciak y Bryk (2005) analizado en la Parte II. En la práctica, las escuelas que manifiestan cambios inusualmente grandes deben estudiarse con cuidado. Si la inestabilidad es importante y se supone que se debe sobre todo a factores pasajeros, conviene la triangulación por medio de evidencias adicionales (p. ej., de inspecciones escolares), en particular si los cambios en las puntuaciones escolares dan lugar a acciones como sanciones o recompensas considerables. Si los resultados se van a usar internamente, debe tenerse la cautela adecuada. Por otra parte, si los resultados se van a hacer públicos, deben emplearse guías para determinar si deben suprimirse estos resultados. Las guías tienen que tomar en cuenta el tamaño de la muestra y otros factores. En países participantes se observó que cohortes de muestra más pequeñas mostraban una inestabilidad mucho más grande de resultados a lo largo de los años. Una posibilidad para las autoridades correspondientes es por tanto no informar de resultados de escuelas que no cumplan con los requisitos mínimos de tamaño muestral, o de escuelas en las que la longitud del intervalo de confianza asociado a una diferencia exceda un umbral predeterminado. Estos dos resultados a menudo se relacionan, y los administradores y los tomadores de decisiones deben considerar cómo tratar a las escuelas pequeñas. El grupo de expertos consultado para este reporte consideró que sería problemático interpretar los resultados de valor agregado de escuelas con menos de 20 a 30 alumnos en una cohorte, pero se reconoce que el tamaño de la escuela difiere sustancialmente por país. Sin embargo, es posible agrupar a las escuelas más pequeñas para obtener muestras más grandes que, al menos desde un punto de vista estadístico, se interpreten mejor. No obstante, pueden surgir problemas para interpretar los resultados de grupos de escuelas más pequeñas si no hay una lógica *a priori* para dicho agrupamiento. En algunos países miembros de la OCDE, por ejemplo, es posible agrupar a las escuelas más pequeñas según regiones geográficas y unidades administrativas específicas. Así, los resultados de valor agregado se analizan para determinar las medidas de desempeño de estas regiones o unidades administrativas. Estas medidas son en particular útiles para análisis de políticas si estas unidades administrativas tienen distintos programas educativos, cuyo impacto después puede informarse mediante las medidas de valor agregado. Sin embargo, la interpretación de estos resultados debe realizarse con cuidado, debido a las diferencias que puedan existir entre las escuelas que dificulten la interpretación de una puntuación única de un grupo de escuelas heterogéneo. Esto reviste particular importancia en sistemas con niveles mayores de autonomía escolar y por ende con una mayor divergencia potencial en políticas y programas educativos.

La inestabilidad en las puntuaciones de valor agregado de las escuelas quizá no sólo surja por cambios en el desempeño escolar o por problemas en los estimadores de valor agregado. Las puntuaciones escolares también se ven afectadas por cambios en el modelo de valor agregado que se utilizó. Con el tiempo, es inevitable que el modelo, los datos o ambas cosas sufran cambios en respuesta a la inspección continua de los análisis de valor agregado o a exigencias externas. Conviene confirmar periódicamente que el modelo aún es adecuado para fines de

políticas actuales y considerar las implicaciones de cambios en los datos disponibles. Si bien estos cambios deben ser mínimos de modo que no anulen por completo la comparabilidad de resultados con el tiempo, sólo sería natural suponer que las estimaciones estadísticas se alterarán y mejorarán ligeramente con desarrollos posteriores del sistema. Estos cambios también pueden provenir de otros cambios en las políticas que deseen centrarse en aspectos distintos de desempeño escolar o en estimaciones de valor agregado contextualizado más extensas. Estos cambios deben ponerse a prueba para determinar el impacto en todas las puntuaciones de valor agregado de las escuelas (no sólo el modelo general), y es importante que se analicen con los actores relevantes para garantizar que la interpretación del valor agregado permanezca constante con el tiempo.

En el **Reino Unido**, los cambios se han mantenido al mínimo, pero hubo una diferencia entre la especificación del modelo de valor agregado contextualizado de escuelas secundarias de 2005 y 2006 y entre los modelos de escuelas primarias de 2006 y 2007.[1] Si se comparan o promedian las puntuaciones de valor agregado con el tiempo para juzgar el desempeño escolar, es importante que se tome en cuenta todo cambio en el modelo subyacente. Con algunos cambios, sería posible calcular las puntuaciones de valor agregado con las bases antiguas y nuevas, pero en otros casos, como el de la inclusión de nuevos datos, no lo sería. El efecto de los años anteriores se estimaría basándose en el cambio del año más reciente, aunque esto no necesariamente ofrecería un estimador sólido para los años anteriores. Cuando sea posible recalcular los años previos con la nueva base o, al contrario, obtener un estimador para el nuevo año con la base antigua, se dan dos conjuntos de cifras y se calcula una tendencia o promedio con una base consistente. Sin embargo, esto no evita las dificultades de que la puntuación anterior de valor agregado de una escuela pudiera ser diferente si se calculara con la nueva base, sobre todo si se emprendieron acciones y la escuela sufrió consecuencias específicas basadas en los resultados obtenidos con el modelo anterior. También hay estimaciones importantes por establecer respecto a lo significativo que debe ser un cambio para que garantice el cálculo y divulgación de cifras anteriores o nuevas revisadas con la base antigua. Estos juicios dependerán de la cantidad de escuelas afectadas, el tamaño del impacto y los recursos requeridos para calcular las cifras alternas.

USO DE RESULTADOS DE UNA MEDIA MÓVIL DE TRES AÑOS

Dado el potencial de la inestabilidad excesiva de los resultados de valor agregado de algunas escuelas en años consecutivos, es conveniente que las acciones derivadas de los resultados de valor agregado de las escuelas se basen en un promedio de puntuaciones móvil de tres años. Se considera que es necesario tener cuidado al interpretar los datos de sólo uno o dos años. Por tanto, surge la pregunta de cómo utilizar los datos provisionales al establecer un sistema basado en modelos de valor agregado y cómo garantizar respuestas oportunas que pudieran no mostrarse tan pronto en los análisis de media móviles de tres años. Dadas las dificultades en la interpretación de los resultados de un solo año, sería adecuado que las acciones con consecuencias mayores para las escuelas (y maestros y directores) se matizaran o pospusiesen hasta que se contara con datos de años siguientes y se obtuviese un promedio de tres años, o los resultados se apoyaran en otra información. Definir con exactitud las consecuencias menores y mayores es un cálculo subjetivo, y no es posible hacerlo con precisión en este informe dados la amplitud de las acciones de políticas y el grado al que los parámetros pertinentes difieren

entre países y sistemas educativos. Desde una perspectiva de políticas, puede distinguirse entre acciones más alineadas con la rendición de cuentas escolar y las más alineadas con las mejoras escolares. Las acciones en sistemas de rendición de cuentas escolar tendrían en potencia consecuencias mayores (percibidas dentro de las escuelas) en comparación con el uso de puntuaciones de valor agregado con fines de mejoras escolares internas, pero esto no es sostenible para todas las acciones e intervenciones en estos sistemas.

Al esperar tres años para obtener una puntuación precisa de valor agregado se reconoce que es ineficiente no emplear los datos de alguna manera, y que este retraso es perjudicial para los estudiantes si no se emprenden acciones para mejorar las escuelas de bajo desempeño. Las puntuaciones de valor agregado que indiquen bajo desempeño pueden desencadenar más análisis de datos y de los procesos escolares existentes. Estos análisis tendrían el propósito de identificar indicadores adicionales de bajo desempeño para efectuar una evaluación más completa sobre la cual basar las acciones correctivas adecuadas. Pueden recopilarse y analizarse más datos (aunque es obvio que hay limitaciones de recursos para recoger esos datos) que incluyan un análisis del desempeño de los alumnos, como puntuaciones brutas de pruebas, retención estudiantil y tasas de aprobación, y más análisis de datos de características de ingreso de los alumnos y otros datos administrativos, como movilidad del alumnado. Esto podría ofrecer otros indicadores de cambios dentro de la escuela (como cambios en la composición estudiantil) o cambios en ese desempeño de estudiantes que confirmara (o desmintiera) el resultado de valor agregado de un año único. Se puede efectuar otro análisis de indicadores escolares adicionales. Las tasas de movilidad de los docentes pueden ser una señal de un problema en la escuela o de que los cambios quizás ocurrieron con la llegada de nuevos maestros. Un cambio de director puede generar cambios en los programas escolares o en la organización escolar que serían de importancia en el contexto del resultado del modelo de valor agregado. La información sobre los procesos escolares también sería valiosa como apoyo de la información del resultado del modelo de valor agregado de un solo año. La información de modelos de valor agregado y la información sobre los procesos escolares son complementarias, no sustitutivas, y la combinación de múltiples indicadores ofrece mayor confianza en las decisiones orientadas a emprender acciones específicas.

Puede ser aconsejable no publicar los resultados de los modelos de valor agregado hasta obtener una media móvil de tres años. La mayor inestabilidad de las puntuaciones de las escuelas en estos primeros años puede generar problemas si se publican estos resultados. Los interesados perderían confianza con rapidez en un sistema con tal inestabilidad, sobre todo si la publicación de resultados escolares es algo novedoso en un sistema educativo. Por tanto, en los años iniciales, se considera que hay beneficios en empezar con un proceso que se centre en medidas de mejoras escolares y, si se desea, desarrollarlo para convertirlo en un sistema con acciones más significativas basadas en los resultados del modelo de valor agregado, como la publicación de resultados. Por otro lado, los resultados pueden publicarse de manera provisional, y emplearse información adicional como apoyo de los resultados de valor agregado de uno a dos años.

COMUNICACIÓN Y COMPROMISO CON ACTORES CLAVE

Numerosos actores se pueden beneficiar con un sistema que emplee modelos de valor agregado. Sin embargo, también, se reconoce que si se emplean resultados de valor agregado como base para acciones, éstas podrían suponer un impacto negativo en individuos y organizaciones

particulares (p. ej., al sancionar a escuelas). Esto puede provocar una mala respuesta a la introducción de nuevos sistemas que midan, entre otras cosas, el desempeño de individuos u organizaciones. Esta reacción puede ser en particular visible si los modelos de valor agregado se introducen como parte de un programa más amplio de rendición de cuentas de escuelas o de docentes. Por estos problemas potenciales, sería pertinente hacer partícipes a los interesados en la elaboración, implementación y uso continuado de los modelos de valor agregado. El compromiso eficaz se lograría mediante una extensa estrategia de comunicación que complemente programas amplios de formación. Estas medidas reconocerían y facilitarían el desarrollo de escuelas como organizaciones de aprendizaje eficaces, y se analizan más adelante.

Maestros, directores y otros miembros del personal escolar son los principales interesados cuyo trabajo se vería afectado por la implementación de un sistema basado en modelos de valor agregado. Las percepciones de desconfianza, mayor presión, frustración y el temor de perder autonomía son reacciones comunes a la implantación de un sistema que supervise el desempeño (Saunders, 2000). Aunque muchos esfuerzos por impulsar reformas deben afrontar estos problemas, los que dependen de análisis de valor agregado enfrentan ciertas dificultades específicas. En primer lugar, los modelos de valor agregado pueden verse particularmente carentes de méritos si los modelos son lo bastante complejos para parecer opacos a muchos interesados. En segundo lugar, la información generada está en el ámbito escolar, pero toda mejora real depende de los cambios en los ámbitos tanto de la escuela como de los docentes. Así, un reto es cómo generar y presentar información comprensible para los docentes; otro es desarrollar la capacidad para que los maestros, directores y otros miembros del personal empleen la información con eficacia. Desarrollar capacidad implica más guía y formación para los maestros, directores y otros interesados (Saunders, 2000); también requerirá inversiones en el personal de las oficinas centrales y recursos analíticos. Una estrategia de comunicación y compromiso de los interesados debe centrarse en la mayor precisión inherente a los modelos de valor agregado de mediciones del desempeño escolar. Se ha visto que esto representa un gran beneficio, pues los interesados se inclinan por los modelos de valor agregado en tanto ofrezcan una medición más precisa, y por ende más justa, del desempeño escolar que otros indicadores que se hayan empleado en los sistemas educativos (Dudley, 1999). Por ejemplo, Fitz-Gibbon (1997) destacó las opiniones favorables de los directores en el **Reino Unido** respecto de la introducción de los modelos de valor agregado y, como detallamos más adelante, Jakubowski (2007) observó que los maestros también eran favorables al empleo de los modelos de valor agregado para medir el desempeño escolar (Recuadro 8.1). Un beneficio importante del compromiso efectivo de los actores principales es que debe reducir la posibilidad de un comportamiento que en potencia sesgue los datos de los modelos de valor agregado. Como se presenta en la Parte I, varios sistemas pueden padecer de comportamientos adversos que sesguen la evaluación de alumnos y los datos escolares, así como crear incentivos para una conducta de los docentes y de la escuela deficientes. Para vencer estos problemas es necesario que maestros y directores confíen en que el sistema es justo y puede reflejar el verdadero desempeño escolar. También debe aclararse que estas medidas requieren un compromiso de largo plazo que quizá altere las relaciones entre la autoridad central y las escuelas, y la dinámica dentro de las escuelas.

Las estrategias de comunicación fructíferas en varios sistemas educativos implicaron el compromiso de los actores relevantes en diversas facetas de la implementación del sistema. Estos sistemas trascendieron la mera comunicación de detalles del modelo de valor agregado

en desarrollo para animar a los interesados a utilizar la información de valor agregado en su propio beneficio. El compromiso eficaz conlleva la comunicación por múltiples canales en el desarrollo y operación de los modelos de valor agregado y del sistema que emplea puntuaciones de valor agregado escolar como base para sus acciones (Saunders, 2000). Esto es de particular importancia si los modelos de valor agregado van a servir en un sistema de medidas para mejoras escolares que requiera la interpretación de resultados escolares y la formulación de acciones derivadas de dichas interpretaciones en el ámbito escolar.

La comunicación eficaz abarca cada etapa del proceso. Cada etapa necesita comunicarse con eficacia a los actores relevantes, y deben emprenderse iniciativas para comprometerlos y obtener su apoyo. Esto incluye los objetivos y razones del sistema, el desarrollo y elección del modelo de valor agregado a emplear, la implementación del sistema, sobre todo el sistema de evaluaciones de alumnos, y el uso de la información de valor agregado por parte de diversos interesados. Estas estrategias formaron parte integral del éxito de los modelos de valor agregado en sistemas educativos de varios países participantes, y se analizan más adelante.

Como sucede con el desarrollo del sistema en general, es necesario explicar con claridad a los interesados los objetivos de introducir un sistema basado en modelos de valor agregado. Los elementos principales de los objetivos básicos de políticas se analizaron en la Parte I de este informe, ya que es importante identificar y considerar con cuidado su impacto en directores, maestros y otros miembros del personal escolar. Hay beneficios en la expresión clara sobre cómo se van a usar las puntuaciones de valor agregado para medir el desempeño escolar. De particular importancia para los interesados puede ser la unidad de análisis en los modelos de valor agregado, y cómo se usarán y presentarán los resultados. La unidad de análisis puede variar para centrarse en regiones, unidades administrativas, escuelas y maestros. Este informe se centró en el valor agregado del ámbito escolar, pero la unidad de análisis debe abordarse de manera explícita, incluso con discusiones para saber si se identificará explícitamente a las escuelas en algún material publicado.

Dar a conocer las puntuaciones de valor agregado de las escuelas al público en general es central para el objetivo de promover la elección de escuela. Independientemente de la intención de la publicación de los resultados de valor agregado de las escuelas, los maestros, directores y otros actores relevantes pueden percibir esto como una forma de rendición de cuentas escolar. Como se presentó en la Parte I, la publicación de resultados puede crear percepciones negativas entre las escuelas y aumentar la sospecha de los motivos de la introducción de un sistema basado en modelos de valor agregado. El diseño de una estrategia de comunicación que aborde estas necesidades debe ser constructivo. En algunos sistemas educativos se efectuaron visitas escolares y se repartieron materiales de divulgación para transmitir la forma de presentar los resultados escolares de valor agregado. Esto a menudo complementó las iniciativas de educación y formación dirigidas a aumentar la comprensión de los modelos de valor agregado y el uso de su información. La estrategia de comunicación puede incluir publicaciones para las escuelas y sesiones informativas con ejemplos explícitos de publicación de puntuaciones escolares de valor agregado, con cuadros y diagramas ilustrativos. Esto asimismo explicaría cómo interpretar estos cuadros y diagramas, sobre todo la interpretación estadística de las puntuaciones escolares de valor agregado, y, de ser pertinente, los intervalos de confianza y cómo usarlos para clasificar las diferencias significativas en el

desempeño escolar. De nuevo, incluir a maestros y directores en las decisiones referentes a la presentación de resultados escolares y otra información (p. ej., en un perfil escolar) es una estrategia eficaz de compromiso y mejora la calidad general del sistema.

Al elaborar este sistema, la mayoría de los gobiernos diseñará una estrategia de medios para la publicación de mediciones escolares de valor agregado y una explicación para interpretarla. Los datos de valor agregado son complejos y multidimensionales, y una simple clasificación de escuelas puede ser confusa si no está alineada con los objetivos de políticas y prácticas específicos. Deben tomarse medidas para garantizar que la presentación de la información de valor agregado en los medios no oculte los aspectos positivos del desarrollo del sistema. En varios países participantes se consideró que aunque el objetivo no fuera convertir las puntuaciones escolares de valor agregado en clasificaciones de escuelas, los medios sí lo harían. Una presentación de resultados más indeseada se manifestaría como un interés de los medios en puntuaciones brutas de las pruebas. Tal vez sea prudente generar información sobre el desempeño de la escuela y del alumno que incluya puntuaciones brutas de pruebas, puntuaciones de valor agregado y puntuaciones de valor agregado contextual para ofrecer una panorámica más completa para un análisis por parte de las familias y docentes. Si bien es difícil controlar las notas de los medios, pueden adoptarse medidas tanto para educar a los medios sobre la interpretación de la información de valor agregado como para ofrecer declaraciones explícitas sobre lo que puede o no interpretarse a partir de las puntuaciones de valor agregado y otras informaciones. Además, es posible destacar aspectos particulares de las medidas de desempeño. Por ejemplo, junto con la información presentada, puede generarse una sola clasificación escolar basada en las puntuaciones de valor agregado contextualizado de las escuelas si se considera que ésta es la medida más precisa.

Quizá convenga diseñar una estrategia de medios junto con maestros, directores y otros interesados, pues las escuelas son a menudo las protagonistas de las notas de los medios. Además, los interesados del sector educativo pueden ser eficientes para comunicar un mensaje común de la forma de interpretar las mediciones de valor agregado, y puede suponerse que la información obtenida de directores, maestros y otros interesados contribuye al desarrollo de una estrategia de medios más completa. Esta estrategia puede facilitar la implementación del sistema y reducir la probabilidad de notas mediáticas engañosas que oculten las ventajas de la introducción de un sistema basado en modelos de valor agregado. Es importante en todo contexto organizativo que los individuos se sientan reconocidos en su lugar de trabajo, sobre todo si en él se experimentan iniciativas de cambios organizativos, como la introducción de un sistema de medida de desempeño (O'Day, 2006). Otorgar a directores, maestros y otros miembros del personal escolar no sólo una mayor comprensión de los modelos de valor agregado sino capacitación para interpretar y analizar la información de valor agregado con fines de mejoras escolares facilita la implementación eficaz de un sistema basado en modelos de valor agregado. Además, brindar formación para analizar los datos, elaborar programas escolares y supervisar el avance estudiantil puede aminorar las sospechas e ilustrar los beneficios tangibles para los interesados. Asimismo sería ventajoso asignar recursos para acciones particulares de mejoras escolares derivadas de los modelos de valor agregado. Con el fin de promover el uso de los sistemas basados en modelos de valor agregado para mejoras escolares puede asignarse una suma determinada de recursos de modo que los usen

las escuelas que analicen la información de valor agregado con la cual elaboren programas específicos dirigidos a elevar el desempeño estudiantil. Esto representaría tanto un incentivo para emprender estos análisis como un mensaje para los interesados de que el sistema se establece con fines de mejoras escolares y no es tan sólo una capa adicional de burocracia o rendición de cuentas escolar. Asimismo destacaría la convicción de tomar decisiones basadas en datos para elevar el desempeño en todo el sector educativo.

DESARROLLO DE UN PROGRAMA DE FORMACIÓN

La inspección cercana de datos relacionados con la escuela como ejercicio básico del desarrollo escolar es un fenómeno relativamente nuevo. Muchos educadores no cuentan con la formación apropiada en medición o estadística, y algunos quizá no confíen en la interpretación de la información de valor agregado. En consecuencia, la introducción de indicadores de desempeño escolar basados en análisis de valor agregado debe diseñarse y establecerse con cuidado, y la formación tiene que considerarse un requerimiento básico en la introducción de mediciones cuantitativas de desempeño (Yang *et al.,* 1999).

Los programas de formación han demostrado su eficacia en el **Reino Unido** con el análisis de resultados de valor agregado a cargo de directores y socios de mejoras escolares. Esta delegación de facultades requiere estrategias eficaces de comunicación y formación de modo que los directores y maestros utilicen mejor la información de valor agregado con fines de mejoras escolares. Es más probable que un sistema reciba apoyo si las escuelas perciben y disfrutan beneficios más tangibles. Con este fin, puede ser conveniente comprometer a los directores, maestros y otros actores relevantes en el desarrollo del sistema mediante el cual se efectúe el análisis de la información de valor agregado en el ámbito escolar. Esto puede abarcar decisiones acerca de las fuentes de información para incluir en los modelos y de la clase de análisis útiles para las escuelas. La consulta y participación de estos actores relevantes también puede contribuir a establecer una interfaz accesible para el sistema de información y el *software* con que se efectúe el análisis en el ámbito escolar. Al contrario, un sistema que implique poco uso de datos en el ámbito escolar se vería como una imposición en las escuelas y como una estrategia administrativa "de la cima a la base" (Wikeley, 1998).

Los esfuerzos para incorporar a los actores reacios al sistema requerirían una planificación cuidadosa que refleje la comprensión de cómo la información de valor agregado afectaría a directores y maestros. La experiencia con este tipo de formación en países participantes revela la necesidad de comunicar la información estadística básica sobre la estimación de los resultados de valor agregado. Si bien esta formación puede considerarse para cubrir sólo los aspectos básicos de los modelos de valor agregado (la formación no pretende dotar a los interesados de las habilidades de elaborar sus propios modelos de alto nivel), se recibió retroalimentación positiva del personal que trabajaban con maestros sobre la forma de interpretar las puntuaciones de valor agregado y aspectos como intervalos de confianza y el cálculo de diferencias estadísticamente significativas entre los resultados escolares. Esta formación puede extenderse al análisis de la estabilidad de puntuaciones escolares en años consecutivos y el impacto de dicha inestabilidad en el uso de las puntuaciones de valor agregado de las escuelas para alcanzar objetivos de políticas establecidos. La formación también debe permitir el análisis de datos de alumnos que ilustren la variación del desempeño tanto en una escuela como en grupos particulares de

alumnos. Esto permitiría a las escuelas identificar puntuaciones de valor agregado en diferentes materias y en diferentes grupos de edad, así como el análisis de grupos particulares de alumnos, por ejemplo, según nivel socioeconómico, género, origen étnico o situación familiar. Las escuelas con esta capacidad de análisis deben identificar mejor a los alumnos que presenten bajo desempeño, diseñar acciones adecuadas para elevarlo y supervisar el impacto de éstas. Asimismo, esto debe facilitar el aprendizaje organizativo extenso en el ámbito escolar sobre la eficacia de diversos enfoques conforme las escuelas se beneficien de decisiones basadas en datos, y las escuelas y directores busquen mejorar sus métodos con base en una comprensión precisa del desempeño escolar.

Algunos países participantes que diseñaron programas de formación informaron que, al tomar los cursos de formación, fue visible para los individuos que los modelos de valor agregado ofrecen datos más precisos del desempeño escolar que el análisis de las puntuaciones brutas. En **Polonia**, por ejemplo, se optó por un proceso de formación que se presenta en el Recuadro 8.1. Al comprender mejor cómo interpretar la información de valor agregado y lo que de verdad

Recuadro 8.1 Programas de formación en Polonia

En **Polonia** se puso en marcha un programa extenso de formación junto con la introducción de un sistema basado en modelos de valor agregado. Efectuado en 2006, se estableció una estructura escalonada mediante la cual se educó a los docentes formadores de manera central y luego formaron a docentes en centros locales de formación. Los objetivos de la formación se centraron en:

- Interpretar puntuaciones de valor agregado.
- Ilustrar cómo se pueden usar los modelos de valor agregado para evaluar el progreso de los alumnos y facilitar así los programas de mejoras escolares.
- Crear un grupo de maestros, directores, inspectores y consejeros docentes capaces de enseñar a los demás y promover el uso válido de la información de valor agregado.

El programa de formación consistió en una combinación de conferencias, clases con ejercicios y sesiones abiertas. Las conferencias de apertura introdujeron la idea de la evaluación de valor agregado de las escuelas y explicaron los aspectos teóricos de los modelos de valor agregado. Se creía que esta estrategia, aunque fuese muy exigente, eliminaría la sensación de que un pequeño grupo de expertos imponía métodos no transparentes para el público.

A las conferencias siguieron ejercicios en aula en pequeños grupos. Todos los asistentes recibieron cuadros con puntuaciones de exámenes de escuelas secundarias determinadas a partir de una regresión sobre puntuaciones de escuelas primarias. Además se presentaron coeficientes de variables ficticias estimadas en el modelo (p. ej., género, estudiantes disléxicos). Los maestros posteriormente calcularon los residuos de la regresión (restaron las puntuaciones reales de cada estudiante de las puntuaciones predichas).

• • • •

Con los residuos, los participantes calcularon el valor agregado de las escuelas como el promedio de los residuos por estudiante en una escuela determinada. Además, se enseñó a los maestros a calcular intervalos de confianza para la media de residuos que se usaron después para comparar escuelas. Se explicó que este enfoque sólo se usaba como herramienta heurística y no era por completo válido desde un punto de vista estadístico. Se destacó que la evaluación de valor agregado efectuada de esta forma sencilla era una herramienta útil para verificar si había diferencias significativas en desempeño escolar y para crear hipótesis preliminares que después interpretaría el personal escolar con más conocimientos sobre una escuela, maestros y estudiantes.

Debe notarse que se prefirió este modelo simple de valor agregado respecto de modelos más complicados porque podía usarse internamente en las escuelas y era relativamente fácil de explicar. La virtud de este modelo sencillo se exploró más durante la formación. Los expertos mostraron a los maestros y directores la forma cómo podían calcular puntuaciones de valor agregado para grupos definidos dentro de las escuelas mediante la comparación de dichas puntuaciones entre niñas y niños o entre clases. Estos ejercicios simples se efectuaron con una hoja de cálculo Excel®, muy común en las escuelas de **Polonia**.

Después se dictó una conferencia que resumió las ventajas y desventajas de la evaluación de valor agregado en **Polonia** con base en investigaciones para demostrar la validez externa de la estrategia de valor agregado. A la conferencia siguió una sesión abierta en la que los asistentes pudieron plantear preguntas, y los expertos, aclarar malentendidos y explicar algunos tecnicismos. Por último, se aplicó una breve encuesta entre los participantes, quienes evaluaron la formación y la introducción de un sistema basado en modelos de valor agregado de manera más general.

Los participantes no sólo se beneficiaron con la formación de tres días, sino también recibieron materiales para formar a otros maestros. Los materiales se imprimieron como manuales, con una descripción técnica del modelo de valor agregado establecido en **Polonia** así como todos los ejercicios que se desarrollaron y enseñaron en la formación. Además, se incluyeron ejercicios en Excel y se dieron a los participantes en CD para facilitar la formación siguiente. Al seminario siguió una sesión de formación de cinco horas efectuada el mes siguiente en cada uno de los 50 centros de formación de profesores regionales y locales. Por añadidura, los representantes de las inspecciones escolares (*Kuratoria*) participaron en esta formación y se diseñaron sesiones de formación especial adicionales donde se presentaron modelos de valor agregado y se analizaron como herramienta potencial para supervisar la calidad de la enseñanza en escuelas secundarias. Por último, se lanzó en 2007 un sitio de Internet de "calculadora de valor agregado" con información para usar esta nueva herramienta e incorporarla a los programas de formación de profesores en centros locales. Muchos centros de formación de profesores locales desde entonces respondieron al creciente interés en el valor agregado al incorporar cursos de valor agregado a sus programas de formación.

miden las puntuaciones de valor agregado de las escuelas, a les quedan más claros los beneficios de este análisis. Muchos actores clave en el **Reino Unido** los actores clave recibieron bien la introducción de los modelos de valor agregado porque se percibió como una medición de desempeño mucho más justa que el análisis de puntuaciones brutas de pruebas. La formación destacaría más estos beneficios y por ende incrementarían la probabilidad de la aceptación de los modelos de valor agregado por parte de los actores relevantes.

El análisis de los modelos de valor agregado necesitaría definir los modelos de valor agregado contextualizado y la interpretación de las puntuaciones y coeficientes de valor agregado de las escuelas con las variables contextuales incluidas. Esto implicaría estudiar si se emplearían modelos de valor agregado contextualizado, las razones de dicha decisión y poner a prueba el modelo en la fase piloto. Según la estructura de las evaluaciones de alumnos, la fuerza predictiva de los modelos de valor agregado tal vez no aumente en gran medida con las características contextuales; aun así, pueden ser importantes para propósitos de desarrollo de políticas y un compromiso eficaz de los actores relevantes. El uso de modelos de valor agregado contextualizado puede calmar las inquietudes de que los modelos sólo miden las condiciones de entrada de los alumnos en lugar del desempeño escolar. Aunque los modelos con una mayor cantidad de medidas previas pueden eliminar la necesidad de medidas de características de antecedentes estudiantiles, no es tan sencillo transmitir esto a los actores que tal vez estén menos versados en análisis estadístico. Asimismo es conveniente consultar a los actores relevantes sobre la inclusión de datos contextuales en un modelo de valor agregado contextualizado y sobre la información escolar adicional que complementaría tales datos. Los actores relevantes pertinentes tienen una considerable experiencia con las características de los alumnos y escolares que afectan el desempeño estudiantil y pueden ofrecer puntos de vista valiosos sobre la forma de agregar estos datos al desarrollo de políticas. Además, esta es otra oportunidad de incluir a los actores relevantes en el proceso de desarrollo y comprometer su apoyo para el empleo de dichos modelos.

Diseñar programas de formación eficaces y estrategias de comunicación efectivas son actividades que consumen muchos recursos. Por fortuna, estos retos producen las recompensas correspondientes. Las estimaciones de los efectos escolares, cuando se acompañan de otras informaciones contextuales y comparativas, representan un punto de partida para conversaciones entre y dentro de las escuelas. Al fragmentar los resultados por diversas características de los alumnos se desvela una panorámica detallada de los puntos fuertes y débiles de los programas escolares. Estos análisis suelen efectuarse con regularidad en el **Reino Unido** y en algunos sistemas educativos de **EUA**, como Dallas, Texas, y varios distritos de Tennessee (Braun, 2005a). Tomar decisiones más eficaces basadas en datos con información de valor agregado abarca el desarrollo de sistemas de información más completos dentro de las escuelas. Es necesario reconocer que el desarrollo de sistemas eficaces de información en organizaciones complejas, como son las escuelas, requiere algo más que sólo formación y capacidades analíticas (O'Day, 2002). Es conveniente destacar la comunicación y colaboración eficaces con las escuelas para garantizar que la toma de decisiones referentes al desarrollo y supervisión de programas escolares sea efectiva en toda la escuela y no quede confinada a la administración superior. Si se considera beneficioso acentuar más una estrategia para toda la escuela de toma de decisiones basada en datos, puede fomentarse la formación para promover la colaboración entre colegas y el desarrollo de programas escolares por parte de equipos de docentes.

PRESENTACIÓN Y USO DE LA INFORMACIÓN DE VALOR AGREGADO

La puntuación de valor agregado de una escuela será un número que refleje su desempeño en relación con otras escuelas. La interpretación de esta puntuación requiere una evaluación valorativa que debe usarse como base para acciones que promuevan los objetivos de políticas declarados. En la Parte I de este informe se ofrecieron numerosos ejemplos de presentación de la información de valor agregado tanto para uso interno como para conocimiento público. El análisis se mantiene para los aspectos pertinentes de la fase de implementación. Se centra en la evaluación del método adecuado para publicar información de valor agregado, su uso interno y dentro de las escuelas, y cómo se traduce en acciones específicas.

La publicación de los resultados escolares debe alinearse con los objetivos de políticas deseados. Es conveniente elaborar estas publicaciones con los datos piloto y recibir retroalimentación sobre estas publicaciones de los actores relevantes. Esta retroalimentación ayuda al desarrollo general de las publicaciones mismas y también para destacar áreas que los actores clave consideren particularmente sensibles. Se brinda así información para las decisiones relacionadas con la publicación de puntuaciones de valor agregado en la implementación real. El uso de la información de valor agregado dentro de las escuelas y para el desarrollo de políticas internas requiere formación, el desarrollo del *software* pertinente y las reflexiones respecto a qué información debe quedar disponible para el análisis y en qué forma. Todo esto se estructura en la fase piloto con análisis de los actores relevantes al crear retroalimentación que debe después servir de información para la planificación de la implementación real.

Deben establecerse guías para la interpretación de las puntuaciones de valor agregado que ayuden al desarrollo de las acciones e interpretaciones adecuadas por parte de los actores relevantes. En varios países esto se ha centrado en la clasificación de resultados como indicadores en categorías específicas de desempeño (p. ej., escuelas de bajo y de alto desempeño). Estas guías deben elaborarse y después evaluarse mediante la interacción con los actores pertinentes durante la fase piloto y a lo largo de la implementación. Identificar de manera explícita la forma de interpretar las puntuaciones de valor agregado y de usarlas para generar acciones específicas incrementa el nivel de transparencia y eficiencia interna. Los actores clave necesitan conocer estas acciones con el fin de confiar en el sistema y también de diseñar las medidas apropiadas para elevar el desempeño. Los actores e instituciones (p. ej., inspectores, ministros, departamentos y escuelas) que lleven a cabo las acciones predeterminadas también pueden planear y desarrollar intervenciones mejor para elevar el desempeño escolar. Por ejemplo, una escuela clasificada como de bajo desempeño puede desencadenar una inspección escolar y un periodo de evaluación más intensa. Si los criterios para esta clasificación y la acción resultante están definidos con claridad, se ponen en marcha los procedimientos que permitan a las escuelas e inspecciones escolares (o una institución adecuada) preparar y desarrollar mejor un marco evaluativo que responda de manera eficiente a la clasificación. Esto puede incluir un análisis de los datos de valor agregado para aplicar una evaluación escolar que aborde las necesidades de cada escuela específica. Esto permitiría diseñar con más eficiencia las estrategias dirigidas y dotar a los inspectores escolares, directores y maestros con más información para el análisis y posteriormente aumentar el desempeño escolar.

En la fase piloto y la implementación preliminar del sistema basado en modelos de valor agregado, es posible analizar los datos piloto o la evaluación previa de alumnos para calibrar el impacto de las puntuaciones de valor agregado y las acciones resultantes. Por ejemplo, se efectúa un análisis de la proporción de las escuelas que recibirían recompensas y sanciones, de las que recibirían una inspección escolar, de las que se pondría en periodo de prueba y de las que se clasificaría como de alto o bajo desempeño. Este análisis brinda información para decidir dónde situar los "puntos de corte" en la distribución de las puntuaciones de valor agregado y por consecuencia, las acciones que de ellas se desprendan.

NOTA

1. Estos cambios tuvieron lugar entre el estudio piloto de los Cuadros de Desempeño Escolar y la publicación nacional de las puntuaciones de valor agregado contextualizado para todas las escuelas.

Bibliografía

Aitkin, M. y **N.T. Longford** (1986), "Statistical Modelling Issues in School Effectiveness Studies", en *Royal Statistical Society*, Serie A, 149 (1), pp. 1-43.

Amato, P. y **B. Keith** (1991), "Parental Divorce and Adult Well-Being: A Meta-Analysis", en *Journal of Marriage and Family,* 53 (1), pp. 43-58.

Antelius, J. (2006), "Value-Added Modelling in Sweden", informe de antecedentes para el OECD Project on the Development of Value-Added Models in Education Systems, Skolverket.

Armein-Beardsley, A. (2008), "Methodological Concerns About the Education Value-Added Assessment System", en *Educational Measurement,* 37 (2), pp. 65-75.

Atkinson Review (2005), *Final Report: Measurement of Government Output and Productivity for the National Accounts*, Palgrave McMillan.

Ballou, D. (2001), "Pay for Performance in Public and Private Schools", en *Economics of Education Review*, febrero, pp. 51-61.

Ballou, D. (2009), "Test Scaling and Value-Added Measurement", en *Education Finance and Policy,* 4 (4), pp. 351-383.

Ballou, D., W. Sanders y **P. Wright** (2004), "Controlling for Student Background in Value-Added Assessment of Teachers", en *Journal of Educational and Behavioral Statistics*, 29.

Bandeira de Mello, V., C. Blankenship y **D. McLaughlin** (2009), *Mapping State Proficiency Standards onto NAEP Scales: 2005-2007* (NCES 2010-456), National Center for Education Statistics, Washington DC.

Becker, G. (1964), *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, Columbia University Press, Nueva York, NY.

Bejar, I.I. (2009), *Standard Setting: What Is It? Why Is It Important? R&D Connections,* ETS, Princeton, NJ.

Benjamini, Y. y **Y. Hochberg** (2000), "The Adaptive Control of the False Discovery Rate in Multiple Hypotheses Testing", en *Journal of Behavioural Education Statistics*, 25, pp. 60-83.

Betebenner, D. (2008), "Toward a Normative Understanding of Student Growth", en K.E. Ryan y L.A. Shepard (eds.), *The Future of Test-based Educational Accountability*, Taylor & Francis, Nueva York, NY, pp. 155-170.

Betebenner, D. (2009a), *Growth, Standards and Accountability*, The Center for Assessment, Dover, NH.

Betebenner, D. (2009b), "Norm and Criterion-Referenced Student Growth", en *Educational Measurement: Issues and Practices*, 28 (4), pp. 42-52.

Betebenner, D. (2009c), *SGP: Student Growth Percentile and Percentile Growth Projection/Trajectory Functions* (Manual del programa de cómputo, en versión 0.0-4).

Betebenner, D. y **R.L. Linn** (2010), *Growth in Student Achievement: Issues of Measurement, Longitudinal Data Analysis, and Accountability*, Educational Testing Service, Princeton, NJ.

Bethell, G. (2005), *Value-Added Indicators of School Performance: The English Experience Anglia Assessment,* Battisford, Suffolk, Inglaterra (sin publicarse).

Borjas, G. (1995), "Ethnicity, Neighborhoods, and Human-Capital Externalities", en *American Economic Review*, 85, pp. 365-390.

Borjas, G. (2001), "Long-Run Convergence of Ethnic Skill Differentials, Revisited", en *Demography*, 38 (3), pp. 357-361.

Bourque, M.L. (2005), "The History of No Child Left Behind", en R.P. Phelps (ed.), *Defending Standardized Testing* (pp. 227-254), Lawrence Erlbaum Associates, Hillsdale, NJ.

Branch, G.F., E.A. Hanushek y **S.G. Rivkin** (2009), "Estimating Principal Effectiveness", Working Paper 32, National Center for Analysis of Longitudinal Data in Education Research, The Urban Institute, Washington DC.

Braun, H.I. (2000), "A Post-Modern View of the Problem of Language Assessment", en A.J. Kunnan (ed.), *Studies in Language Testing 9: Fairness and Validation in Language Assessment. Selected Papers from the 19th Language Testing Research Colloquium*, Local Examinations Syndicate, University of Cambridge, Cambridge, pp. 263-272.

Braun, H.I. (2005a), "Value-Added Modelling: What Does Due Diligence Require?", en R. Lissitz, *Value-Added Models in Education: Theory and Applications*, JAM Press, Maple Grove, MN.

Braun, H.I. (2005b), *Using Student Progress to Evaluate Teachers: A Primer on Value-Added Models*, Policy Information Perspective, Educational Testing Service, Princeton, NJ.

Braun, H.I. (2006a), *Background Paper: The Use of Value-Added Models for School*, OCDE, París.

Braun, H.I. (2006b), "Empirical Bayes", en J.G. Green (ed.), *Complementary Methods for Research in Education*, American Educational Research Association, Washington DC.

Braun, H.I. (2009), "Discussion: with Choices Come Consequences", en *Educational Measurement: Issues and Practices*, 28 (4), pp. 52-55.

Braun, H.I., Y. Qu y **C.S. Trapani** (2008), *Robustness of Value-Added Analysis of School Effectiveness*, ETS RR-08-22, Educational Testing Service, Princeton, NJ.

Briggs, D. (2008), "The Goals and Uses of Value-Added Models", informe preparado para el Commitee on Value-Added Methodology for Instructional Improvement, National Research Council and the National Academy of Education, Washington DC.

Briggs, D., J.P. Weeks y **E. Wiley** (2008), "Vertical Scaling in Value-Added Models for Student Learning", en *National Conference on Value-Added Modeling*, WACER, Madison, WI.

Briggs, D. y **D. Betebenner** (2009), "Is Growth in Student Achievement Scale Dependent?", informe para la reunión anual del National Council for Measurement in Education, San Diego, CA.

Briggs, D. y **J.P. Weeks** (2009), "The Impact of Vertical Scaling Decision on Growth Interpretation", en *Educational Measurement: Issues and Practices*, 28 (4), pp. 3-14.

Briggs, D. y **J.P. Weeks** (2009), "The Sensitivity of Value-Added Modeling to the Creation of a Vertical Score Scale", en *Education Finance and Policy,* 4 (4), pp. 384-414.

Brooks-Gunn, J., G. Duncan, P. Klebanov y **N. Sealand** (1993), "Do Neighborhoods Influence Child and Adolescent Development?", en *American Journal of Sociology*, 99, pp. 353-93.

Bryk, A., Y. Thum, J. Easton y **S. Luppescu** (1998), *Academic Productivity of Chicago Public Elementary Schools, Technical Report*, The Consortium on Chicago School Research, Chicago, IL.

Burgess, S., C. Propper, H. Slater y **D. Wilson** (2005), "Who Wins and Who Loses from School Accountability? The Distribution of Educational Gain in English Secondary Schools", Working Paper Series No. 05/128, CMPO, The University of Bristol.

Burgess, S. y **A. Briggs** (2010), "School Assignment, School Choice and Social Mobility", en *Economics of Education Review,* 29 (4), pp. 639-649.

Burstein, L. (1980), "The Analysis of Multi-Level Data in Educational Research and Evaluation", en *Review of Research in Education*, pp. 158-233.

Caldwell, B. (2002), "Autonomy and Self-managment: Concepts and Evidence", en T. Bush y L. Bell, *The Principles and Practice of Educational Management*, Paul Chapman, Londres, pp. 34-48.

Caldwell, B. y **J. Spinks** (1998), *Beyond the Self-Managing School*, Falmer Press, Londres.

Carlsson, G. (1958), *Social Mobility and Class Structure*, Gleerup, Lund, Suecia.

Castro, M., C. Ruíz y **E. López** (2009), "Forma básica del crecimiento en los modelos de valor añadido: Vías para la supresión del efecto de regresión", en *Revista de Educación*, 348, pp. 111-136.

Choi, K. y **M. Seltzer** (2005), *Modelling Heterogeneity in Relationships Between Initial Status and Rates of Change: Latent Variable Regression in a Three-Level Hierarchical Model*, marzo, National Center for Research on Evaluation, Standards and Student Testing/UCLA, Los Angeles, CA.

Choi, K., P. Goldschmidt y **K. Yamashiro** (2005), "Exploring Models of School Performance: From Theory to Practice", en J.H. (eds.), *Yearbook for the National Society for the Study of Education*, 104 (2), Blackwell, Malden, MA.

Cizek, J. (ed.) (2005), "Adapting Testing Technology to Serve Accountability Aims: The Case of Vertically-Moderated Standard Setting", en *Applied Measurement in Education*, edición especial, 18, pp. 1-9.

Cizek, G.J. y **M.B. Bunch** (2007), *Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Tests*, Sage, Thousand Oaks, CA.

Clark, D., F. Martorell y **J.E. Rockoff** (2009), "School Principals and School Performance", Working Paper 38, National Center for Analysis of Longitudinal Data in Education Research, The Urban Institute, Washington DC.

Coates, H. (2009), "What's the Difference? A Model for Measuring the Value-Added by Higher Education in Australia", en *Higher Education Management and Policy*, 21(1), pp. 69-88.

Coleman, J. (1966), *Equality of Educational Opportunity*, US Department of Health, Education, and Welfare, Washington DC.

Corcoran, M., R. Gordon, D. Laren y **G. Solon** (1992), "The Association Between Men's Economic Status and Their Family and Community Origins", en *Journal of Human Resources*, 27 (4), pp. 575-601.

De Leeuw, J. y **E. Meijer** (eds.) (2008), *Handbook of Multilevel Analyisis*, Springer, Nueva York, NY.

Dette, H. y **S. Volgushev** (2008), "Non-Crossing Non-parametric Estimates of Quantile Curves", en *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70 (3), pp. 609-627.

Dixit, A. (2002), "Incentives and Organisations in the Public Sector: An Interpretive Review", en *Journal of Human Resources*, 37 (4), pp. 696-727.

Doeringer, P. y **M. Piore** (1985), *Internal Labour Markets and Manpower Analysis*, Armonk, Nueva York, NY.

Doran, H.C. y **L.T. Izumi** (2004), *Putting Education to the Test: A Value-Added Model for California*, Pacific Research Institute, San Francisco, CA.

Doran, H. y **J. Cohen** (2005), "The Confounding Effects of Linking Bias on Gains Estimated from Value-Added Models", en R. Lissitz, *Value-Added Models in Education: Theory and Applications*, JAM Press, Maple Grove, MN.

Doran, H. y **T. Jiang** (2006), "The Impact of Linking Error in Longitudinal Analysis: An Emprical Demonstration", en R. Lissitz, *Longitudinal and Value-Added Models of Student Performance*, JAM Press, Maple Grove, MN, pp. 210-229.

Dorans, N., M. Pommerich y **P. Holland** (2007), *Linking and Aligning Scores and Scales (Statistics for Social and Behavioral Sciences)*, Springer, Nueva York, NY.

Dudley, P. (1999), "Using Data to Drive Up Standards: Statistics or Psychology?", en C. Conner, *Assessment in Action in the Primary School*, Falmer Press, Londres.

Dunn, J.L. y **J. Allen** (2008), "The Interaction of Measurement, Model, and Accountability: What are the NCLB Growth Models Measuring?", informe presentado en la reunión anual del National Council on Measurement in Education, marzo, Nueva York, NY.

Dunn, J. y **J. Allen** (2009), "Holding Schools Accountable for the Grothw of Non-Proficiente Students: Coordinating Measurement and Accountability", en *Educational Measurement: Issues and Practices*, 28 (4), pp. 3-14.

Dyer, H., R. Linn y **M. Patton** (1969), "A Comparison of Four Methods of Obtaining Discrepancy Measures Based on Observed and Predicted School System Means on Achievement Tests", en *American Educational Research Journal*, 6, pp. 591-606.

Eurostat (2001), *Handbook on Price and Volume Measures in National Accounts*, European Communities, Luxemburgo.

Ferrão, M.E., P. Costa, V. Dias y **M. Dias** (2006), "Medição da competência dos alunos do ensino básico em Matemática: 3EMat, uma proposta [*Measuring math skills of students in compulsory education: 3EMat, a proposal*]". Actas da XI Conferência Internacional de Avaliação Psicológica (*Proceedings of the XI International Conference on Psychological Evaluation*), Braga, Portugal.

Ferrão, M. (2007a), "Sensitivity of VAM Specifications: Measuring Socio-Economic Status", informe de antecedentes para el OECD Project on the Development of Value-Added Models in Education Systems, Varsovia.

Ferrão, M.E. (2009), "Sensibilidad de las especificaciones del modelo de valor añadido: Midiendo el estatus socioeconómico", en *Revista de Educación*, (348), pp. 137-152.

Ferrão, M.E. y **H. Goldstein** (2008), "Adjusting for Measurement Error in the Value-Added Model: Evidence from Portugal", en *Quality and Quantity*, 43 (6), pp. 951-963.

Fielding, A., M. Yang y **H. Goldstein** (2003), "Multilevel Ordinal Models from Examination Grades", (3), pp. 127-153.

Figlio, D. y **L. Kenny** (2006), "Individual Teacher Incentives and Student Performance", NBER Working Paper, No. 12627.

Fitz-Gibbon, C. (1997), *The Value Added National Project Final Report: Feasibility Studies for a National System of Value-Added Indicators*, School Curriculum and Assessment Authority, Londres.

Fitz-Gibbon, C. y **P. Tymms** (2002), "Technical and Ethical Issues in Indicator Systems: Doing Things Right and Doing Wrong Things", en *Education Policy Analysis Archives*, 10 (6).

Fitzpatrick, A.R. (2008), "The Impact of Anchor Test Configuration on Student Proficiency Rates", informe para la reunión anual del National Council on Measurement in Education, marzo, Nueva York, NY.

Friedman, T. (2005), *The World is Flat: A Brief History of the 21st Century*, Farrar, Strauss y Giroux, Nueva York, NY.

Ginther, D., R. Haveman y **B. Wolfe** (2000), "Neighborhood Attributes as Determinants of Children's Outcomes: How Robust are the Relationships?", en *Journal of Human Resources*, 35 (4), pp. 603-642.

Glass, D. (1954), *Social Mobility in Britain*, Routledge & Paul, Londres.

Glenn, C. y **de J. Groof** (2005), *Balancing Freedom, Autonomy and Accountability in Education*, Wolf Legal Publishers, Nijmegan, NL.

Goldhaber, D. y **D. Brewer** (2000), "Does Teacher Certification Matter? High School Teacher Certification Status and Student Achievement", en *Educational Evaluation and Policy Analysis*, 22 (2), pp. 129-145.

Goldhaber, D. y **M. Hansen** (2010), "Assessing the Potential Using of Value-Added Estimates of Teacher Job Performance for Making Tenure Decisions", Working Paper 31, National Center for Analysis of Longitudinal Data in Education Research, The Urban Institute, Washington DC.

Goldhaber, D. y **M. Hansen** (2010), "Is It Just a Bad Class? Assessing the Stability of Measured Teacher Performance", CRPE Working Paper, University of Washington, WA.

Goldstein, H. (1987), "Multilevel Covariance Component Models", en *Biometrika*, 74, pp. 430-431.

Goldstein, H., J. Rasbash, M. Yang, G. Woodhouse, H. Pan, D. Nuttall y **S. Thomas** (1993), "A Multilevel Analysis of School Examination Results", en *Oxford Review of Education*, 19 (4), pp. 425-433.

Goldstein, H. y **D.J. Spiegelhalter** (1996), "League Tables and their Limitations: Statistical Issues in Comparison of Institutional Performance", en *Journal of Royal Statistical Society,* Serie A, Parte 3, pp. 385-443.

Goldstein, H. (1986), "Multilevel Mixed Linear Model Analysis Using Iterative Generalised Least Squares", en *Biometrika*, 73, pp. 43-56.

Goldstein, H. (1997), "Methods in School Effectiveness Research", en *School Effectiveness and School Improvement*, 8, pp. 369-95.

Goldstein, H., D. Kounali y **A. Robinson** (2008), "Modelling Mesurement Errors and Category Misclassifications in Multilevel Models", en *Statistical Modelling*, 8 (3), pp. 243-261.

Gorard, S., J. Fitz y **C. Taylor** (2001), "School Choice Impacts: What Do We Know?", en *Educational Researcher*, 30 (7), pp. 18-23.

Gray, J., D. Jesson, H. Goldstein, K. Hedger y **J. Rasbash** (1995), "A Multilevel Analysis of School Improvement: Changes in Schools' Performance Over Time", en *School Effectiveness and School Improvement*, 6 (2), pp. 97-114.

Hægeland, T. (2006), "School Performance Indicators in Norway", informe de antecedentes para el OECD Project on the Devleopment of Value-Added Models in Education Systems.

Hægeland, T., L. Kirkebøen, O. Raaum y **K. Salvanes** (2005), "School Performance Indicators for Oslo", en *Reports 2005/36*, Statistics Norway.

Hægeland, T. y **L. Kirkebøen** (2008), "School Performance and Value-Added Indicators – What is the Importance of Controlling for Socioeconomic Background?", informe de antecedentes para el OECD Project on the Development of Value-Added Models in Education Systems.

Hambleton, R.K. y **M.J. Pitoniak** (2006), "Setting Performance Standards", en R. Brennan, *Educational measurement*, 4a. ed., American Council on Education, Washington DC, pp. 433-470.

Haney, W. y **A. Raczek** (1993), *Surmounting Outcomes Accountability in Education*, US Congress Office of Technology Assessment, Washington DC.

Hanushek, E.A. y **M.E. Raymond** (2004), "The Effect of School Accountability Systems on the Level and Distribution of Student Achievement", en *Forthcoming European Economic Review*.

Hanushek, E.A. (2009), "Teacher Deselection", en Dan Goldhaber y Jane Hannaway (eds.), *Creating a New Teaching Profession*, The Urban Institute Press, Washington DC, pp. 165-180.

Hanushek, E.A. y **S.G. Rivkin** (2010), "Generalizations about Using Value-Added Measures of Teacher Quality", en *American Economic Review,* 100 (2).

Harris, D., A. Hendrickson, Y. Tong, S-H. Shin y **C-Y. Shyu** (2004), "Vertical Scales and the Measurement of Growth", informe presentado en la reunión del National Council on Measurement in Education, abril de 2004, San Diego, CA.

Harris, D.H. (2009), "Would Accountability Based on Teacher Value Added Be Smart Policy? An Examination of the Statistical Properties and Policy Alternatives", en *Education Finance and Policy,* 4 (4), pp. 319-350.

Haveman, R. y **B. Wolfe** (1995), "The Determinants of Children's Attainments: A Review of Methods and Findings", en *Journal of Economic Literature*, 33, pp. 1829-1878.

Hill, R. *et al.* (2005), "Using Value Tables to Explicitly Value Student Growth", *www.nciea.org/cgi-bin/pubspage.cgi?sortby=pub_date*, consultado en enero 10 de 2006.

Ho, A.D., D.M. Lewis y **J.L. Farris** (2009), "The Dependence of Growth Model Results on Proficiency Cut Scores", en *Educational Measurement: Issues and Practices*, 28 (4), pp. 15-26.

Hoxby, C. (2003), "The Economics of School Choice", en *National Bureau of Economic Research Conference Report*, University of Chicago Press, IL.

IGE (2001), *Avaliação Integrada das escolas. Relatório Nacional. Ano lectivo 1999-2000*, Inspecção Geral da Educação, Ministério da Educação.

Ishii, J. y **S.G. Rivkin** (2009), "Impediments to the Estimation of Teacher Value Added", en *Education Finance and Policy,* 4 (4), pp. 520-536.

Jacob, B. (2002), "Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools", NBER Working Paper, No. 8968, Cambridge, MA.

Jacob, B.A. (2010), "The Effect of Employment Protection on Worker Effort: Evidence from Public Schooling", Working Paper 15655, National Bureau of Economic Research, Cambridge, MA.

Jakubowski, M. (2007), "Volatility of Value-Added Estimates of School Effectiveness: A Comparative Study of Poland and Slovenia", informe presentado al Robert Shurman Centre for Advanced Studies, European University, Florencia.

Jakubowski, M. (2008), "Implementing Value-Added Models of School Assessment", RSCAS Working Papers 2008/06, European University Institute.

Kane, T.J. y **D.O. Staiger** (2002), "Volatility in School Test Scores: Implications for Test-Based Accountability Systems", en D. Ravitch (ed.), *Brookings Papers on Education Policy*, Brookings Institution, Washington DC, pp. 235-269.

Kane, T.J. y **D.O. Staiger** (2008), "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation", Working Paper 14607, National Bureau of Economic Research, *www.dartmouth.edu/~dstaiger/Papers/w14607.pdf*, consultado en noviembre 10 de 2009.

Koedel, C. y **J.R. Betts** (2009), "Does Student Sorting Invalidate Value-Added Models of Teacher Effectiveness? An Extended Analysis of the Rothstein Critique", Working Papers 09-02, Department of Economics, University of Missouri, MO.

Koedel, C. y **J. Betts** (2010), "Value Added to What? How a Ceiling in the Testing Instrument Influences Value-Added Estimation", en *Education Finance and Policy*, 5 (1), pp. 54-81.

Kohn, A. (2000), *The Case Against Standardized Testing: Raising the Scores, Ruining the Schools*, Heineman, Portsmouth, NH.

Kolen, M. y **R. Brennan** (2004), *Test Equating, Scaling and Linking: Methods and Practices*, Springer Science and Business Media, Nueva York, NY.

Koretz, D. (2005), "Alignment, High Stakes, and the Inflation of Test Scores", en J. Herman and E. Haertel (eds.), *Uses and Misuses of Data for Educational Accountability and Improvement*, NSSE, Malden, MA, pp. 99-118.

Kreft, I. y **J. De Leeuw** (1998), *Introducing Multilevel Modelling*, Sage Publications, Londres, Thousand Oaks y Nueva Delhi.

Ladd, H.F. y **R.P. Walsh** (2002), "Implementing Value-Added Measures of School Effectiveness: Getting the Incentives Right", en *Economics of Education Review*, 21, pp. 1-17.

Lavy, V. (2002), "Evaluating the Effects of Teachers' Group Performance Incentives on Pupil Achievement", en *Journal of Political Economy*, 110, pp. 1286-1317.

Lazear, E.P. (2000), "The Future of Personnel Economics", en *The Economic Journal*, 110 (467), F611-F639.

Leckie, G. y **H. Goldstein** (2009), "The Limitations of Using School League Tables to Inform School Choice", en *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 172 (4), pp. 835-851.

Levacic, R. (2001), "An Analysis of Competition and its Impact on Secondary School Examination Performance in England", Occasional Paper, No. 34, septiembre, National Centre for the Study of Privatisation in Education, Teachers College, Columbia University, NY.

Linn, R.L. (2005), "Conflicting Demands of 'No Child Left Behind' and State Systems: Mixed Messages about School Performance", en *Education Policy Analysis Archives*,13 (33).

Linn, R.L. (2004), "Rethinking the No Child Left Behind Accountability System", informe presentado en el Center for Education Policy Forum, Washington DC, *www.ctredpol.org*.

Linn, R.L. (2008), "Measurement Issues Associated with Value-Added Methods", en Comitee on Value-Added Methodology for Instructional Improvement, *Program Evaluation and Educational Accountability*, National Academy of Education, Washington DC.

Lissitz, R., H. Doran, W. Schafer y **J. Willhoft** (2006), "Growth Modelling, Value-Added Modelling and Linking: An Introduction", en R. Lissitz, *Longitudinal and Value-Added Models of Student Performance*, JAM Press, Maple Grove, MN, pp. 1-46.

Little, R.J.A. y **D.B. Rubin** (1987), *Statistical Analysis with Missing Data*, Wiley, Nueva York, NY.

Lizasoain, L. y **L. Joaristi** (2009), "Análisis de la dimensionalidad en modelos de valor añadido: Estudio de las pruebas de matemáticas empleando métodos no paramétricos basados en TRI", en *Revista de Educación*, 248 (enero-abril), pp. 175-194.

Lockwood, J.R., D.F. McCaffrey, L. Hamilton, B. Stecher, V. Le y **F. Martinez** (2007) "The Sensitivity of Value-Added Teacher Effect Estimates to Different Mathematics Achievement Measures", en *Journal of Educational Measurement*, 44 (1), pp. 45-65.

Lockwood, J.R. y **D.F. McCaffrey** (2007), "Controlling for Individual Level Heterogeneity in Longitudinal Models, with Applications to Student Achievement", en *Electronic Journal of Statistics*, 1, pp. 223-252.

Lockwood, J.R. y **D.F. McCaffrey** (2009), "Exploring Student-Teacher Interactions in Longitudinal Achievement Data", en *Education Finance and Policy,* 4 (4), pp. 439-467.

Lucas, R. (1988), "On the Mechanics of Economic Development", en *Journal of Monetary Economics*, 22 (1), pp. 3-42.

Madaus, G., P.W. Airasian y **T. Kellaghan** (1980), *School Effectiveness: A Reassessment of the Evidence*, McGraw-Hill, Nueva York, NY.

Mante, B. y **G. O'Brien** (2002), "Efficiency Measurement of Australian Public Sector Organisations: The Case of State Secondary Schools in Victoria", en *Journal of Educational Administration*, 30 (7), pp. 274-291.

Martineau, J.A. (2006), "Distorting Value-Added: The Use of Longitudinal, Vertically Scaled Student Achievement Data for Growth-Based, Value-Added Accountability", en *Journal of Educational and Behavioral Statistics*, 31 (1), pp. 35-62.

Martineau, J. (2009), "Measuring Student Achievement Growth at the High School Level", en L.M. Pinkus, *Meaningful Measurement: The Role of Assessments in Improving High School Education in the Twenty-Frist Century*, Alliance for Excellent Education, Washington DC, pp. 119-142.

Martínez-Arias, R., J.L. Gaviria y **M. Castro** (2009), "Concepto y evolución de los modelos de valor añadido en educación", en *Revista de Educación*, (348), pp. 15-34.

Martínez-Arias, R. (2009), "Usos, aplicaciones y problemas de los modelos de valor añadido en educación", en *Revista de Educación*, (348), pp. 217-250.

Mayer, C. (1996), "Does Location Matter?", en *New England Economic Review,* mayo/junio, pp. 26-40.

McCaffrey, D.F., J.R. Lockwood, D.M. Koretz y **L.S. Hamilton** (2003), *Evaluating Value-Added Models for Teacher Accountability*, The RAND Corporation, Santa Monica, CA.

McCaffrey, D.M., J.R. Lockwood, D. Koretz, T.A. Louis y **L. Hamilton** (2004), "Models for Value-Added Modelling of Teacher Effects", en *Journal of Educational and Behavioral Statistics,* 29 (1), pp. 67-101.

McCaffrey, D.F., J.R. Lockwood, L.T. Mariano y **C. Setodji** (2005), "Challenges for Value-Added Assessment of Teacher Effects", en R. Lissitz (ed.), *Value-Added Models in Education: Theory and Practice,* JAM Press, Maple Grove, MN.

McCaffrey, D.F. y **J.R. Lockwood** (2008), "Value-Added Models: Analytic Issues", informe presentado en el seminario acerca de Modelos de Valor-Agregado, de la National Research Council and National Academy of Education, Board on Testing and Accountability, Washington DC.

McCaffrey, D.F., J.R. Lockwood, T.R. Sass y **K. Mihaly** (2009), "The Inter-Temporal Variability of Teacher Effect Estimates", en *Education Finance and Policy,* 4 (4), pp. 439-467.

McCall, M.S., G.G. Kingsbury y **A. Olson** (2004), *Individual Growth and School Success,* Northwest Evaluation Association, Lake Oswego, OR.

McGuinn, P. (2010), *Ringing the Bell for K-12 Teacher Tenure Reform*, Center for American Progress, Washington DC.

McKewen, N. (1995), "Accountability in Education in Canada", en *Canadian Journal of Education,* 20 (1).

Messick, S. (1989), "Validity", en R. Linn. (ed.), *Educational Measurement,* American Council on Education, Washington DC.

Meyer, R. (1997), "Value-Added Indicators of School Performance: A Primer", en *Economics of Education Review*, 16 (3), pp. 283-301.

Ministry of National Education, Higher Education and Research, Direction de l'évaluation, de la performance et de la prospective (2006), "Lycée Performance Indicators: 2005 General, Technological and Vocational Baccalauréats*",* informe de antecedente para el OECD Project on the Development of Value-Added Models in Education Systems.

Mintrop, H. y **G.L. Sunderman** (2009), *Why High Stakes Accountability Sounds Good but Doesn't Work – and Why We Keep on Doing It Anyway*, The Civil Rights Project/Proyecto Derechos Civiles en UCLA, Los Angeles, CA.

Mortimore, P., P. Sammons, L. Stoll, D. Lewis y **R. Ecob** (1988), *School Matters: The Junior Years*, Open Books, Wells.

NASBE (2005), *Evaluating Value-Added: Findings and Recommendations from the NASBE Study Group on Value-Added Assessments*, National Association of State Boards of Education, Alexandria, VA.

Nichols, S.L. y **D.C. Berliner** (2005), *The Inevitable Corruption of Indicators of Educators through High-Stakes Testing,* Education Policy Research Unit, Arizona State University, Tempe, AZ.

O'Day, J. (2002), "Complexity, Accountability, and School Improvement", en *Harvard Educational Review*, 72 (3), pp. 293-329.

Odden, A. y **C. Busch** (1998), *Financing Schools for High Performance*, Jossey-Bass, San Francisco, CA.

Organización para la Cooperación y el Desarrollo Económicos (OCDE) (1994), *The OECD Jobs Strategy: Evidence and Explanations*, OCDE, París.

OCDE (1996), *Lifelong Learning for All*, OCDE, París.

OCDE (2001), *The New Economy: Beyond the Hype*, OCDE, París.

OCDE (2004), *Learning for Tomorrow's World: First Results from PISA 2003*, OCDE, París.

OCDE (2005), *Teachers Matter: Attracting, Developing and Retaining Effective Teachers*, OCDE, París.

OCDE (2006), *Demand Sensitive Schooling? Evidence and Issues*, OCDE, París.

OCDE (2007a), *Education at a Glance 2007: OECD Indicators*, OCDE, París.

OCDE (2007b), *Learning for Tomorrow*, OCDE, París.

OCDE (2007c), *No More Failures: Ten Steps to Equity in Education*, OCDE, París.

OCDE (2007d), *PISA 2006: Science Competencies for Tomorrow's World*, OCDE, París.

OCDE (2008a), *Going for Growth*, OCDE, París.

OCDE (2008b), *Education at a Glance 2008: OECD Indicators,* OCDE, París.

OCDE (2009), *Evaluación y reconocimiento de la calidad de los docentes: Prácticas internacionales,* OCDE, París.

OCDE (2010a), *The High Cost of Low Educational Performance: The Long-Run Economic Impact of Improving PISA Outcomes,* OCDE, París.

OCDE (2010b), *Establishing a Framework for Evaluation and Teacher Incentives: Considerations for Mexico,* OCDE, París.

OCDE (2010c), *Education at a Glance 2010: OECD Indicators,* OCDE, Paris.

Patz, R. (2007), *Vertical Scaling in Standards-Based Educational Assessment and Accountability Systems*, The Council of Chief State School Officers, Washington DC.

Ponisciak, P.M. y **A.S. Bryk** (2005), "Value-Added Analysis of the Chicago Public Schools: An Application of Hierarchical Models", en R. Lissitz (ed.), *Value-Added Models in Education: Theory and Applications*, JAM Press, Maple Grove, MN.

Rasbash, J., G. Leckie, R. Pillinger y **J. Jenkins** (2010), "Children's Educational Progress: Partitioning Family, School and Area Effects", en *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 173 (3), pp. 657-682,

Raudenbush, S. y **J.D. Willms** (1995), "The Estimation of School Effects", en *Journal of Educational and Behavioral Statistics*, 20, pp. 307-335.

Raudenbush, S. y **A. Bryk** (2002), *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2a. ed., Sage Publications, Newbury Park, CA.

Raudenbush, S.W. (2004), *Schooling, Statistics, and Poverty: Can We Measure School Improvement?*, Educational Testing Service, Princeton, NJ.

Raudenbush, S.W. (2009), "Adaptive Centering with Random Effects: An Alternative to the Fixed Effects Model for Studying Time-Varying Treatments in School Settings", en *Education Finance and Policy,* 4 (4), pp. 468-491.

Ray, A. (2006), "School Value-Added Measures in England", informe de antecedentes para el OECD Project on the Development of Value-Added Models in Education Systems.

Ray, A. (2007), "The Volatility of Value-Added Scores*",* informe de antecedentes para el OECD Project on the Development of Value-Added Models in Education Systems.

Ray, A., H. Evans y **T. McCormack** (2009), "El uso de los modelos nacionales de valor añadido para la mejora de las escuelas británicas", en *Revista de Educación*, 348, pp. 47-66.

Ray, A., T. McCormack y **H. Evans** (2009), "Value-Added in English Schools", en *Education, Finance and Policy*, 4 (4), pp. 415-438.

Reardon, S.F. y **S.W. Raudenbush** (2009), "Assumptions of Value-Added Models for Estimating School Effects", en *Education Finance and Policy,* 4 (4), pp. 492-519.

Romer, P. (1994), "Endogenous Economic Growth", en *Journal of Economic Perspectives*, 8 (1), pp. 3-22.

Rothstein, J. (2009), "Student Sorting and Bias in Value-Added Estimation: Selection on Observables and Unobservables*",* en *Education Finance and Policy,* 4 (4), pp. 537-571.

Rothstein, J. (2010), "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement", en *Quarterly Journal of Economics,* 125 (1), pp. 175-214.

Rowan, B., R. Correnti y **R.J. Miller** (2002), "What Large-Scale Survey Research Tells Us About Teacher Effects on Student Achievement: Insights from the Prospects Study of Elementary Schools", en *Teacher College Record*, 104, pp. 1525-1567.

Rubin, D.B. (1976), "Inference and Missing Data", en *Biometrika*, 63, pp. 581-592.

Rubin, D., E. Stuart y **E. Zanutto** (2004), "A Potential Outcomes View of Value-Added Assessment in Education", en *Journal of Educational and Behavioural Statistics*, pp. 103-116.

Ryan, K.E. y **L.A. Shepard** (eds.) (2008), *The Future of Test-Based Educational Accountability*, Routledge, Nueva York, NY.

Ryska, R. (2006), "Value-Added Modelling in the Czech Republic*",* informe de antecedentes para el OECD Project on the Development of Value-Added Models in Education Systems.

Sammons, P.T. (1997), *Forging Links: Effective Schools and Effective Departments*, Paul Chapman Publishing Lda.

Sammons, P., S. Thomas, P. Mortimore, C. Owen y **H. Pennell** (1994), *Assessing School Effectiveness: Developing Measures to Put School Performance in Context*, Office for Standards in Education, Londres.

Sanders, W., A. Saxton y **B. Horn** (1997), "The Tennessee Value-Added Assessment System: A Quantitative Outcomes-Based Approach to Educational Assessment", en J. Millman (ed.), *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluational Measure?*, Corwin Press, Inc, Thousand Oaks, CA, pp. 137-162.

Sanders, W.L. y **S.P. Wright** (2008), "A Response to Amrein-Beardsley (2008) 'Methodological Concerns about the Education Value-Added Assessment System'", White Paper, SAS Institute.

Sanders, W.L., S.P. Wright, J.C. Rivers y **J.G. Leandro** (2009), "A Response to Criticisms of SAS® EVAAS®", White Paper, SAS Institute.

Sass, T. y **D. Harris** (2007), "The Effects of NBPTS-Certified Teachers on Student Achievement*",* CALDER Working Paper, No. 4.

Sass, T.R. (2008), "The Stability of Value-Added Measures of Teacher Quality and Implications for Teacher Compensation Policy", en *Policy Brief 4*, National Center for Analysis of Longitudinal Data in Education Research, The Urban Institute, Washington DC.

Saunders, L. (2000), "Understanding Schools Use of 'Value Added' Data: The Psychology and Sociology of Numbers", en *Research Papers in Education*, 15 (3), pp. 241-258.

SCAA (1994), *Value Added Performance Indicators for Schools*, School Curriculum and Assessment Authority, Londres.

Senge, P. (2000), *Schools that Learn: A Fifth Discipline Fieldbook for Educators, Parents, and Everyone Who Cares About Education*, Doubleday, Nueva York, NY.

Snijders, T.A.B. y **R.J. Bosker** (1999), *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modelling*, Sage, Londres.

Taylor, J. y **N.A. Nguyen** (2006), "An Analysis of the Value Added by Secondary Schools in England: Is the Value Added Indicator of Any Value?", en *Oxford Bulletin of Economics and Statistics*, 68 (2), pp. 203-224.

Tekwe, C., R. Carter, C. Ma, J. Algina, M. Lucas y **J. Roth** (2004), "An Empirical Comparison of Statistical Models for Value-Added Assessment of School Performance", en *Journal of Educational and Behavioral Statistics*, 29 (1), pp. 11-36.

Thomas, S. y **P. Mortimore** (1996), "Comparison of Value-Added Models for Secondary School Effectiveness", en *Research Papers in Education*, 11 (1), pp. 5-33.

Thomas, S., W-J. Peng y **J. Gray** (2007), "Value-Added Trends in English Secondary School Performance Over Ten Years", en *Oxford Review of Education*, 33 (3), próxima publicación.

Thum, Y.M. (2009), "No Child Left Behind: Retos metodológicos y recomendaciones para la medida del progreso anual adecuado", en *Revista de Educación*, 348, pp. 67-90.

Tymms, P. y **C. Dean** (2004), "Value-Added in the Primary School League Tables", informe presentado a la National Association of Head Teachers, mayo, CEM Centre, University of Durham, Durham.

Van de Grift, W. (2007), "Reliability and Validity in Measuring the Added Value of Schools", informe de antecedentes para el OECD Project on the Development of Value-Added Models in Education Systems.

Vicente, P. (2007), "O plano amostral do projecto 3EM. Actas do XIV Congresso Anual da Sociedade Portuguesa de Estatística", en M. N. Ferrão, *Proceedings of the XIV Annual Conference of the Portuguese Statistical Society*, SPE, Lisboa.

Vignoles, A., R. Levacic, J. Walker, S. Machin y **D. Reynolds** (2000), *The Relationship Between Resource Allocation and Pupil Attainment: A Review*, Centre for the Economics of Education, School of Economics, Londres.

Wang, S., H. Jiao, Y. Jin y **Y.M. Thum** (2010), "Investigating Effect of Ignoring Hierarchical Data Structures on Accuracy of Vertical Scaling Using Mixed-Effects Rasch Model", informe presentado en la reunión anual del National Council on Measurement in Education (NCME), abril 30-mayo 5, Denver, CO.

Webster, W.J. (2005), "The Dallas School-Level Accountability Model: The Marriage of Status and Value-Added Approaches", en R. Lissitz (ed.), *Value-Added Models in Education: Theory and Applications*, JAM Press, Maple Grove, MN.

Webster, W. y **R. Mendro** (1997), "The Dallas Value-Added Accountability System", en J. Millman (ed.), *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?*, Corwin Press, Thousand Oaks, CA, pp. 81-99.

Wikeley, F. (1998), "Dissemination of Research as a Tool for School Improvement", en *School Leadership and Management*, 18 (1), pp. 59-73.

Willms, J. y **S. Raudenbush** (1989), "A Longitudinal Hierarchical Linear Model for Estimating School Effects and their Stability", en *Journal of Educational Measurement*, 26 (3), pp. 209-232.

Willms, J.D. (2009), "Value-Added Models in Education", en *Policy Brief*, Canadian Research Institute for Social Policy.

Wilson, D. (2004), "Which Ranking? The Impact of a 'Value-Added' Measure of Secondary School Performance", en *Public Money and Management*, enero, pp. 37-45.

Wright, S., W. Sanders y **J. Rivers** (2006), "Measurement of Academic Growth of Individual Students toward Variable and Meaningful Academic Standards", en R. Lissitz, *Longitudinal and Value-Added Models of Student Performance*, JAM Press, Maple Grove, MN, pp. 385-406.

Wright, S.P. (2010), "An Investigation of Two Non-parametric Regression Models for Value-Added Assessment in Education", White Paper, SAS Institute.

Wright, S.P., J.T. White, y **W.L. Sanders** (2010), "SAS® EVAAS® Statistical Models", White Paper, SAS Institute.

Yang, M., H. Goldstein, T. Rath y **N. Hill** (1999), "The Use of Assessment Data for School Improvement Purposes", en *Oxford Review of Education*, 25 (4), pp. 469-483.

Yen, W.M. (2009), *Growth Models for the NCLB Growth Model Pilot*, Educational Testing Service, Princeton, NJ.

Zieky, M.J., M. Perie y **S. Livingston** (2008), *Cutscores: A Manual for Setting Standards of Performance on Educational and Occupational Tests*, Educational Testing Service, Princeton, NJ.

Zvoch, K. y **J. Stevens** (2006), "Successive Student Cohorts and Lonigtudinal Growth Models: An Investigation of Elementary School Mathematics Performance", en *Education Policy Analysis Archives*, 14 (2).

ORGANIZACIÓN PARA LA COOPERACIÓN Y EL DESARROLLO ECONÓMICOS

La OCDE constituye un foro único en su género, donde los gobiernos trabajan conjuntamente para afrontar los retos económicos, sociales y medioambientales que plantea la globalización. La OCDE está a la vanguardia de los esfuerzos emprendidos para ayudar a los gobiernos a entender y responder a los cambios y preocupaciones del mundo actual, como el gobierno corporativo, la economía de la información y los retos que genera el envejecimiento de la población. La Organización ofrece a los gobiernos un marco en el que pueden comparar sus experiencias políticas, buscar respuestas a problemas comunes, identificar buenas prácticas y trabajar en la coordinación de políticas nacionales e internacionales.

Los países miembros de la OCDE son Alemania, Australia, Austria, Bélgica, Canadá, Chile, Corea, Dinamarca, España, Estados Unidos de América, Estonia, Finlandia, Francia, Grecia, Hungría, Irlanda, Islandia, Israel, Italia, Japón, Luxemburgo, México, Noruega, Nueva Zelanda, Países Bajos, Polonia, Portugal, Reino Unido, República Checa, República de Eslovenia, República Eslovaca, Suecia, Suiza y Turquía. La Comisión de la Comunidad Europea participa en el trabajo de la OCDE.

Las publicaciones de la OCDE aseguran una amplia difusión de los trabajos de la Organización. Éstos incluyen los resultados de la compilación de estadísticas, los trabajos de investigación sobre temas económicos, sociales y medioambientales, así como las convenciones, directrices y los modelos desarrollados por los países miembros.

EDICIONES OCDE, 2, rue André-Pascal, 75775 PARIS CEDEX 16
IMPRESO EN FRANCIA
(96 2008 10 4 P) ISBN 978-92-64-09016-3 – No. 57540 2011

La medición del aprendizaje de los alumnos

MEJORES PRÁCTICAS PARA EVALUAR EL VALOR AGREGADO DE LAS ESCUELAS

Existe un creciente interés en los países miembros y no miembros de la OCDE por mejorar los resultados del aprendizaje de los alumnos. Sin embargo, los sistemas educativos enfrentan el reto de cómo medir de manera confiable y precisa el desempeño de las escuelas en su contribución al aprendizaje de los estudiantes. Los resultados puntuales de alumnos en pruebas estandarizadas, a simple vista, tienden a reflejar las características y condiciones de éstos: los niveles socioeconómicos y en algunos casos, aspectos de idioma y pertenencia a algún grupo étnico, por ejemplo. En respuesta a ello y por su enfoque en *el desarrollo del aprendizaje de los alumnos a lo largo del tiempo,* los modelos de valor agregado pueden ofrecer una base cuantitativa de mayor precisión y confiabilidad para procesos de mejora escolar, el desarrollo de políticas y programas educativos, y para establecer mecanismos eficaces de rendición de cuentas.

Modelos de valor agregado se refiere a una clase de modelos estadísticos que producen una estimación de la contribución de las escuelas al logro de los alumnos para alcanzar objetivos educativos claramente definidos (mayor desempeño cognitivo, por ejemplo), con base en mediciones tomadas en por lo menos dos momentos distintos. Sin contar con una medida confiable y comparativa de desempeño, la equidad y la eficacia de políticas y programas educativos pueden disminuir si los recursos no se canalizan hacia donde más se necesiten. De la misma manera, no será posible mejorar políticas y programas educativos sin lograr identificar aquellas prácticas que claramente muestren buenos resultados. Para ello, los modelos de valor agregado resultan ser un instrumento invaluable al ofrecer una medición más confiable del desempeño escolar, aun para aquellas escuelas que brindan servicios a estudiantes de niveles socioeconómicos desfavorecidos.

Esta publicación ofrece un análisis detallado de los beneficios y los retos en el diseño de modelos de valor agregado, y presenta algunos lineamientos para la puesta en marcha de sistemas de evaluación escolar basados en ellos. El libro cubre ejemplos y prácticas de 13 países, y ha sido traducido y actualizado al 2010 de su versión original en inglés.

Publicaciones relacionadas:
Establishing a Framework for Evaluation and Teacher Incentives: Considerations for Mexico (OCDE, 2011)
Mejorar las escuelas: Estrategias para la acción en México (OCDE, 2010)
Evaluating and Rewarding the Quality of Teachers: International Practices (OCDE, 2009)

Esta obra forma parte del *Acuerdo Para Mejorar la Calidad de la Educación de las Escuelas en México,* firmado por el gobierno de México y la OCDE.

Por favor, cite esta publicación de la siguiente manera:
OCDE (2011), *La medición del aprendizaje de los alumnos: Mejores prácticas para evaluar el valor agregado de las escuelas,* OECD Publishing.
http://dx.doi.org/10.1787/9789264090163-es
Este trabajo está publicado en **OECDiLibrary**, plataforma que reúne todos los libros, publicaciones periódicas y base de datos de la OCDE. Visite *www.oecd-ilibrary.org* y no dude en contactarnos para más información.

2011

OECD*publishing*
www.oecd.org/publishing

ISBN 978-92-64-09016-3
96 2008 10 4P

