
DetECCIÓN Y CARACTERIZACIÓN DE LOS CENTROS ESCOLARES DE ALTA EFICACIA DE LA COMUNIDAD AUTÓNOMA DEL PAÍS VASCO MEDIANTE MODELOS TRANSVERSALES CONTEXTUALIZADOS Y MODELOS JERÁRQUICOS LINEALES¹

Detection and Characterization of Highly Effective Schools in the Autonomous Community of the Basque Country Using Contextualized Cross-Sectional Attainment Models and Hierarchical Linear Models

LUIS JOARISTI

Universidad del País Vasco/ Euskal Herriko Unibertsitatea, UPV/EHU
luis.joaristi@ehu.es

LUIS LIZASOAIN

Universidad del País Vasco/ Euskal Herriko Unibertsitatea, UPV/EHU
luis.lizasoain@ehu.es

VERÓNICA AZPILLAGA

Universidad del País Vasco/ Euskal Herriko Unibertsitatea, UPV/EHU
veronica.azpillaga@ehu.es

1 Este trabajo se encuadra en el contexto del proyecto “Caracterización y buenas prácticas de los centros escolares de alto valor añadido” financiado por la Dirección General de Investigación y Gestión del Plan Nacional de I+D+i del Programa Nacional de Proyectos de Investigación Fundamental, con referencia EDU2011-24366. Investigador Principal, Luis Lizasoain Hernández del Departamento de Métodos de Investigación y Diagnóstico en Educación de la Universidad del País Vasco/Euskal Herriko Unibertsitatea, UPV/EHU. El IP pertenece a la Unidad de Formación e Investigación “Educación, Cultura y Sociedad” (UFI 11/54) de la UPV/EHU.

Resumen: El objetivo de este trabajo es la identificación y caracterización de los centros escolares de alta eficacia de la Comunidad Autónoma del País Vasco. Como criterio se emplean las puntuaciones obtenidas en lengua española, lengua vasca y matemáticas correspondientes a las Evaluaciones Diagnósticas de los años 2009 y 2010. Se consideran centros de alta eficacia aquellos que obtienen una puntuación media superior a la esperable una vez controlados mediante modelos jerárquicos lineales los efectos de las variables contextuales. Como resultado, de las casi 600 escuelas analizadas, se han identificado 32 cuyas características reflejan la diversidad del sistema educativo vasco.

Palabras clave: escuelas eficaces; evaluación diagnóstica; modelos transversales contextualizados; modelos jerárquicos lineales.

Abstract: The aim of this work is the identification and characterization of highly effective schools in the Basque Autonomous Community. As criterion variables the scores in 2009 and 2010 Diagnostic Assessments in Spanish language, Basque language and Mathematics are used. Schools are considered as highly effective when their average performance is higher than the expected score once the effects of contextual variables have been controlled using hierarchical linear models. As a result, of the nearly 600 analyzed schools, 32 were identified whose characteristics reflect the diversity of the Basque educational system.

Keywords: effective schools; diagnosis evaluation; transversal contextualized models; hierarchical linear models.

INTRODUCCIÓN

La finalidad última del proyecto en que se enmarca el presente trabajo es la de elaborar un catálogo de buenas prácticas de los centros escolares de alta eficacia de la Comunidad Autónoma del País Vasco (CAPV), de forma que el mismo pueda ser empleado en los planes de mejora de aquellos centros cuyos resultados no alcanzan los niveles deseables. Como criterio se emplean los resultados en lengua española, lengua vasca y matemáticas correspondientes a las Evaluaciones Diagnósticas (ED) de los años 2009 y 2010. De esta manera se trata de dar cumplimiento al carácter formativo y de mejora que toda evaluación debe tener. Así, en el preámbulo de la Ley Orgánica 2/2006, 3 de mayo, de Educación, se puede leer que la evaluación diagnóstica “tendrá carácter formativo y orientador, proporcionará información... y permitirá adoptar las medidas pertinentes para mejorar”. (Ley Orgánica 2/2006, 3 de mayo, de Educación BOE N° 106 de 4 de Mayo. Preámbulo, página 17162 del BOE).

La idea básica es que la praxis de los centros que obtienen buenos resultados en dicha evaluación puede ser considerada como ejemplo y guía para las acciones y planes de mejora. Pero es importante matizar que por buenos resultados se consideran aquellos que suponen un alto valor añadido, en el sentido de que su puntuación media en algún aspecto educativo sea superior a la que cabría esperar una vez que se han desdoblado o controlado los efectos de las variables contextuales. Por tanto, aquí se está empleando el término valor añadido en su sentido más lato

a diferencia de la recomendación de la OECD (2008) (versión en español, OCDE, 2011) que propone restringir el uso de este término a estudios longitudinales centrados en el cambio y el crecimiento. Si no se da esa perspectiva longitudinal, este organismo sugiere emplear la expresión modelo transversal contextualizado para referirse a este tipo de estudios. En este caso, dado que la ED es una evaluación transversal que se realiza en cuarto curso de primaria y segundo de secundaria obligatoria, no se trata obviamente de un estudio longitudinal de ahí que se haya incorporado la recomendación de la OCDE en el mismo título del trabajo. Pero a lo largo del texto se emplea en ocasiones el término de valor añadido con el enfoque antes apuntado tal y como, por ejemplo, es usado por Gaviria y Castro (2005) o Martínez Arias (2009).

Dicho de otra manera, sea un estudio longitudinal o transversal, la idea fundamental es que los buenos resultados no se corresponden directamente con las puntuaciones de los estudiantes en las distintas pruebas de las ED, sino que es más riguroso y equitativo basarlos en el valor añadido de los centros una vez traído o controlado el efecto de las variables contextuales.

Estos centros han sido detectados a partir de los resultados de las ED realizadas en la CAPV en los años 2009 y 2010 aplicando técnicas estadísticas de modelización lineal multinivel, en que se han controlado factores individuales así como contextuales para aislar los que son meramente escolares.

De ello se deriva que el objetivo general es detectar, para su posterior estudio, los centros escolares con los mejores resultados condicionados al efecto escolar, controlando los previsibles efectos de las variables contextuales. Para ello, como estrategia de análisis estadístico, se han utilizado los Modelos Jerárquico Lineales. Estos modelos asumen que los datos se presentan bajo una estructura multinivel y permiten el análisis conjunto de las variables de cada uno de los niveles, estudiante y centro escolar en este caso (Goldstein, 2003; Raudenbush y Bryk, 2002; Snijders y Bosker, 1999).

Un primer objetivo específico ha sido el de determinar qué criterio es adecuado utilizar en la búsqueda de estos centros escolares de alta eficacia, pues de ello dependía que el desarrollo de las posteriores fases basadas en análisis detallados bajo metodología cualitativa fuera factible.

Un segundo objetivo específico partía de la identificación de tales centros y consistía en la caracterización de los mismos, en función de la información disponible sobre ellos y basada en variables que forman parte de los cuestionarios que se aplican en la propia ED.

Una vez caracterizados los centros, la atención en fases posteriores se centrará en la elaboración de un catálogo de buenas prácticas, y aquí la aproximación a la

realidad de la labor de estos centros es cualitativa, basada en entrevistas a tres grupos de informantes clave: inspectores, asesores de formación y equipos directivos.

De esta manera, más allá de la concreción del presente trabajo, la finalidad última es que este catálogo pueda ser una herramienta útil para el diseño y aplicación de acciones y programas de mejora de las escuelas de la CAPV que no alcanzan el nivel de resultados deseable.

De enfoque análogo al expuesto, se destaca el trabajo de Gritter (2006); en el mismo se estudian los factores determinantes del alto rendimiento y está centrado en el estudio comparado de las escuelas de más alto rendimiento del programa *Just For The Kids*, que es parte del *National Center for Educational Accountability*. Huffman, Pankake y Munoz (2006) utilizan la metodología de los modelos multinivel para analizar datos en tres niveles, de manera que a los habituales de estudiante y centro escolar se añade el nivel de distrito. El trabajo de Ray, McCormack y Evans (2009) se centra en la utilización de modelos de valor añadido orientados a la mejora de las escuelas británicas. En el marco de las disposiciones del acta *No Child Left Behind*, Wright (2009), estudia las escuelas que sin cumplir el estándar de progreso anual adecuado (*Adequate Yearly Progress*, AYP) establecen diferentes estrategias para mejorar.

El equipo de Bryk (Bryk, Sebring, Allensworth, Luppescu y Easton, 2010) realiza un estudio longitudinal desarrollando un modelo propio de valor añadido aplicado a las escuelas públicas de Chicago. A una escala más reducida, Jackson y Lunenburg (2010) llevaron a cabo un estudio de 24 centros escolares de educación secundaria que presentaban niveles de rendimiento medio muy diferentes, agrupándose en centros ejemplares, reconocidos, aceptables y no aceptables.

Entre los trabajos de autores más próximos y con similitud al presente, cabe mencionar la obra de Marchesi y Martín (2002) en que se evalúa la Educación Secundaria bajo los niveles habituales de estudiante y escuela. Gaviria, Martínez Arias y Castro (2004) estudian los factores de eficacia escolar en el sistema educativo brasileño mediante la validación de un modelo de tres niveles: estudiantes, profesores y estados federados brasileños.

Murillo y Román (2011), utilizan un modelo multinivel de tres niveles, constituido por estudiantes, escuelas y países, para analizar el efecto del centro escolar. Para ello, teniendo presente el nivel cultural y socioeconómico de las familias, disponen de la base de datos del Segundo Estudio Regional Explicativo y Comparativo (SERCE), de la UNESCO/OREALC realizado en 2009, que emplea una muestra de estudiantes de 3º y 6º de Educación primaria a los que se han medido el rendimiento en Matemáticas y Lectura. Incidiendo en los niveles de logro socio-afectivos de los estudiantes, Murillo y Hernández-Castilla (2011a, 2011b)

realizan su análisis mediante modelos multinivel de cuatro niveles (estudiante, aula, centro y país).

MÉTODO

Participantes

Los datos que han sido analizados proceden de los resultados de las ED realizadas en los años 2009 y 2010 por el Instituto Vasco de Evaluación e Investigación Educativa (ISEI-IVEI)¹ a todos los estudiantes de 4º curso de Educación Primaria y de 2º de Educación Secundaria Obligatoria en todos los centros educativos de la CAPV de titularidad pública o privada concertada. Se trata, por tanto, de un estudio censal. Los archivos de datos obran en poder del citado instituto y, por imperativo legal, son de difusión limitada y su tratamiento ha de regirse por las pautas del secreto estadístico. Las variables criterio son los resultados en las tres competencias instrumentales básicas: comunicación en Lengua vasca, en Lengua española y Matemáticas. En 2009 se evaluó además la competencia en Cultura científica, tecnológica y de la salud y en 2010 lo relativo a la competencia Social y ciudadana.

En la tabla 1 aparece el número de total de centros escolares de cada etapa en los que se aplicó la ED en cada año. De este total, se procedió a no considerar los centros muy pequeños, siendo considerados como tales los que para los cursos seleccionados tuviesen menos de 13 estudiantes en el nivel evaluado; la reducción en el número de centros puede observarse en dicha tabla.

Tabla 1. Distribución de los centros escolares eliminados

AÑO Y ETAPA	Nº DE ALUMNOS	Nº DE CENTROS	%
2009	< 13	95	18.34
Ed. Primaria	≥ 13	423	81.66
	Total centros	518	

1 El Instituto Vasco de Investigación y Evaluación Educativa (ISEI-IVEI) es un organismo dependiente del Departamento de Educación, Política Lingüística y Cultura del Gobierno Vasco y su tarea se centra en la evaluación general del sistema educativo no universitario (<http://www.isei-ivei.net>).

AÑO Y ETAPA	Nº DE ALUMNOS	Nº DE CENTROS	%
2010 Ed. Primaria	< 13	88	16.86
	≥13	434	83.14
	Total centros	522	
2009 Ed. Secundaria Obligatoria	< 13	6	1.79
	^a ≥ 13	329	98.21
	Total centros	335	
2010 Ed. Secundaria Obligatoria	< 13	3	0.90
	≥ 13	330	99.10
	Total centros	333	

A este respecto se hace notar que la proporción de centros pequeños, y por ende no considerados, fue mucho mayor en la etapa de Primaria que en la de Secundaria, lo que probablemente se explique por razones de concentración de los centros de Secundaria en poblaciones de mayor tamaño. De cualquier forma, la cuestión de los centros de pequeño tamaño puede tener su importancia en lo que a la estabilidad o volatilidad de las puntuaciones de valor añadido se refiere.

Asimismo, dadas las metas perseguidas, se excluyeron del estudio aquellos centros que sólo participaron en uno de los dos años de aplicación de la ED. Así, el colectivo definitivo de centros estudiados constaba de 409 centros de Educación Primaria y 324 de Secundaria. De ellos, dado que hay centros concertados que imparten en ambas etapas, finalmente fueron 591 las diferentes unidades escolares de las que se disponía de datos. En el nivel de los estudiantes, las cifras son las siguientes. De la ED del año 2009 se disponía de datos de 16976 estudiantes de 4º de Primaria y 16446 de 2º de ESO. En 2010 fueron respectivamente 18418 y 17995.

Instrumentos y Variables

Junto con las puntuaciones de los estudiantes en las cuatro competencias evaluadas, se disponía también de información obtenida mediante la aplicación del cuestionario de contexto familiar y de un cuestionario relativo al centro escolar respondido por las direcciones de los mismos.

El cuestionario de contexto familiar focaliza su atención en los aspectos del entorno familiar del estudiante que pueden estar asociados a su aprendizaje. Y así, junto con los indicadores habituales del nivel socioeconómico y cultural familiar, aparecen cuestiones relativas a la lengua o lenguas de uso cotidiano, lugar de nacimiento, años de residencia en la CAPV, etc.

El cuestionario de centros recoge información sobre cuestiones como el clima del centro, la composición del claustro (edad, experiencia, titulación), los recursos disponibles, la participación en programas de innovación, modelos de gestión y organización del centro, etc.

Por todo ello, se dispuso en principio de un conjunto importante de variables susceptibles de ser incorporadas a los modelos lineales jerárquicos para obtener los residuos de cada escuela en cada año (2009 y 2010), etapa (Primaria y Secundaria) y competencia evaluada (tres más su promedio), por tanto, 16 modelos ($2 \times 2 \times 4$).

La primera decisión importante adoptada fue la de incluir en dichos modelos únicamente aquellas variables contextuales en las que el centro no puede intervenir y que no dependen de políticas o acciones emprendidas por las escuelas. Por ejemplo, una variable como el clima escolar es un factor muy relevante sobre el que hay en la literatura evidencia acumulada suficiente en cuanto a su relación con el rendimiento académico de los estudiantes. Pero el que en un centro determinado se dé o no un clima adecuado para el normal desenvolvimiento de la labor educativa e instructiva es una cuestión que tiene un componente contextual, pero que también, al menos en un cierto grado, depende de la política del centro, de su práctica, organización, estrategias de resolución de conflictos, etc. Por el contrario, es menos discutible que el nivel socioeconómico y cultural medio de las familias de los estudiantes que acuden al centro no esté en función de las prácticas pedagógicas del mismo. Se trata de una variable que podemos considerar como estrictamente contextual.

En consecuencia y con el fin de que el residuo a obtener se calcule únicamente mediante el control estadístico de las variables específicamente contextuales en las que el centro no puede intervenir, se incluyeron variables como el Índice Socioeconómico y Cultural de la familia (ISEC), la condición de inmigrantes o no de los estudiantes, el modelo lingüístico, la lengua familiar, la red escolar, la tasa de idoneidad asociada a la repetición de algún curso anterior, el rendimiento previo, etc.

Una vez aplicado este primer criterio de selección de variables, se empleó también el de la relevancia estadística de las mismas. Como ha quedado dicho, estaba disponible toda la información de todos los elementos y preguntas de los cuestionarios, pero no todas tenían la capacidad de discriminación estadística suficiente como para ser consideradas. Para ello se llevó a cabo una fase inicial de análisis exploratorio de los datos mediante técnicas de minería de datos, en concreto empleando árboles estadísticos de decisión, y más específicamente los algoritmos CART y CHAID. Estas técnicas exploratorias son de gran utilidad a la hora de seleccionar variables relevantes para métodos más robustos (Lizasoain y Joaristi, 2010). Bryk y colaboradores (2010) apuntan en esta misma dirección al afirmar que

el procedimiento CHAID resultó de gran utilidad al estudiar posibles efectos de interacción entre los indicadores considerados en su estudio que realizaron sobre las escuelas de Chicago.

Las covariables del Nivel 1 se refieren a los estudiantes y fueron las siguientes:

- Modelo lingüístico en que estudia. Está analizado a través de dos variables booleanas denominadas Modelo A y Modelo B que son indicadores de si el estudiante cursa sus estudios en el modelo lingüístico A o en el B respectivamente. El sistema educativo de la Comunidad Autónoma del País Vasco incorpora las dos lenguas oficiales (vasco y español) en el contexto educativo mediante tres modelos lingüísticos diferentes: A, B y D. En el modelo lingüístico A la enseñanza es en español, con el vasco como asignatura, en todos los niveles, etapas y modalidades. En el modelo lingüístico D el proceso de enseñanza-aprendizaje se desarrolla íntegramente en lengua vasca, salvo la asignatura de Lengua española. El modelo lingüístico B es bilingüe y en el mismo hay materias que se imparten en vasco y otras en español según la enseñanza, ciclo o etapa.
- Género del estudiante.
- Lengua familiar. Se refiere a si el idioma empleado habitualmente en el entorno familiar del estudiante es o no la lengua vasca.
- Condición de inmigrante. Se dispone de datos facilitados por el ISEI-IVEI sobre aspectos de la inmigración; se considera que el estudiante es inmigrante si lo es él mismo o alguno de sus progenitores. Es una variable indicadora que toma el valor 1 en caso de que el estudiante sea inmigrante y 0 en caso contrario.
- Índice Socioeconómico y Cultural familiar (ISEC). Ha sido obtenido mediante análisis de componentes principales de los indicadores incluidos en el cuestionario de contexto. Sus valores están estandarizados de manera que su media es 0 y su desviación típica 1; sus valores extremos son -3.85 y 2.32.
- Condición de repetidor (edad normativa). Se determina si el estudiante ha repetido curso o no a lo largo de sus estudios en función de su año de nacimiento. Es una variable indicadora que toma el valor 1 en caso de que el estudiante haya repetido y 0 en caso contrario.
- Rendimiento en Lengua vasca, Lengua española y Matemáticas obtenidos el curso anterior. En esta variable, al igual que en las que siguen, la información es proporcionada por el propio estudiante.
- Expectativas respecto del nivel máximo de estudios que espera alcanzar.
- Si tiene o no deberes que realizar en casa.

- Número de horas que dedica a realizar los deberes.
- Ayuda externa que recibe para realizar los deberes.
- Número de actividades extraescolares que realiza a lo largo de la semana.

Las covariables del Nivel 2 corresponden a características de centro y para ello se calcularon para cada escuela las tasas, como proporciones, y en su caso los promedios de las puntuaciones de los estudiantes de cada centro, segmentando por años (2009 y 2010) y etapas (Educación Primaria y Educación Secundaria). Las covariables disponibles fueron:

- La red escolar a que pertenece el centro. Se trata de una variable booleana y es un indicador de si el estudiante cursa sus estudios en un centro de la red concertada o no.
- Tasa de alumnado vascoparlante del centro (como proporción del que utiliza habitualmente la lengua vasca en el entorno familiar).
- Tasa de inmigrantes del centro.
- ISEC medio del alumnado del centro.
- Tasa de alumnado repetidor del centro.
- Promedio de los rendimientos obtenidos el curso anterior por el alumnado del centro.
- Tasa de alumnado del centro con deberes a realizar en casa.
- Promedio de horas que dedica el alumnado del centro a realizar los deberes.
- Promedio de la ayuda para realizar los deberes que recibe el alumnado del centro.
- Promedio del número de actividades extraescolares que realiza a lo largo de la semana el alumnado del centro.

Procedimiento

Como ha quedado patente, los datos presentan una estructura multinivel en que las puntuaciones en cada competencia están asociadas al estudiante, que constituye el elemento básico del Nivel 1; a su vez, los estudiantes están agrupados en distintos centros educativos, los cuales constituyen los elementos básicos del Nivel 2.

Además, como ya se ha mencionado, no se trata de una estructura longitudinal, pues no se dispone de ningún tipo de identificador del estudiante que se mantenga a lo largo de las etapas y de sus ciclos.

Las variables criterio corresponden a las puntuaciones obtenidas por los estudiantes en las ED de los años 2009 y 2010 en las competencias de Lengua vasca, Lengua española y Matemáticas. Además, junto con las tres competencias instrumentales básicas, en 2009 se evaluó la competencia en Cultura científica, tecnológica y de la salud y en 2010 la competencia Social y ciudadana, se considera la puntuación promedio de las cuatro competencias con el fin de tomar en consideración también estas competencias adicionales y una información sintética.

La ecuación de regresión lineal múltiple del modelo multinivel del Nivel 1 refleja la variación de la puntuación en cada variable criterio dentro de cada centro y se representa bajo el siguiente conjunto de ecuaciones:

$$\text{Parte estructural: } Y_{ij} = \beta_{0j} + \sum_{q=1}^Q \beta_{qj} X_{qij} + r_{ij}$$

$$\text{Parte probabilística: } r_{ij} \sim N(0, \sigma^2)$$

Los distintos elementos indican lo siguiente:

Y_{ij} : puntuación obtenida en cada una de las tres competencias y su promedio por el estudiante i de la escuela j de forma independiente, es decir, no multivariante.

β_{0j} : es la puntuación media de cada centro en la competencia correspondiente.

β_{qj} : refleja la influencia lineal de la covariable X_q del alumnado.

X_{qij} : es la puntuación del estudiante i de la escuela j en la covariable X_q . La lista de variables se encuentra en el apartado correspondiente a las variables del Nivel 1 del modelo lineal jerárquico.

r_{ij} : es el residuo para el estudiante i de la escuela j , es decir, cuánto se separa cada estudiante de lo esperado o estimado por la ecuación en su escuela.

Como hipótesis de partida, se considera que la influencia de las covariables del Nivel 1 es análoga en todos los centros; por lo tanto, ya en el Nivel 2, el modelo estructural que corresponde a cada asignatura y que refleja la variación entre los centros es:

$$\text{Parte estructural: } \beta_{0j} = \gamma_{00} + \sum_{s=1}^S \gamma_{0s} W_{sj} + u_{0j}$$

$$\text{Parte probabilística: } u_{0j} \sim N(0, \tau_{00}), \text{ en que}$$

β_{0j} : es el rendimiento medio de cada centro (j).

γ_{00} : representa el efecto común a todos los centros.

γ_{0s} : es el efecto lineal de la covariable W_s en el rendimiento medio de los centros.

W_{sj} : es el valor que toma la escuela j en la covariable de escuela W_s , cuya lista se puede encontrar en el apartado correspondiente a las covariables del Nivel 2 del modelo lineal jerárquico,

u_{0j} : representa la variación residual entre escuelas una vez controlados todos los factores individuales y de centro incluidos en el modelo.

El modelo mixto o combinado consiste en:

$$Y_{ij} = \gamma_{00} + \sum_{s=1}^S \gamma_{0s} W_{sj} + \sum_{q=1}^Q \beta_{qj} X_{qij} + u_{0j} + r_{ij}$$

Como se ha dicho, esto dará lugar a una ecuación para cada competencia, año y etapa. Como u_{0j} es el residuo del centro j tras controlar los efectos de las covariables contextuales de estudiante y centro, el mismo es considerado como un indicador de lo que cada centro se separa de lo esperado. Es la diferencia entre en el rendimiento medio obtenido y el esperado una vez controlado el efecto de las variables contextuales.

Los análisis de las siguientes fases de la investigación, las relacionadas con las buenas prácticas, se basarán en este residuo. Bajo este punto de vista, se supone que los centros cuyo residuo sea alto y positivo están funcionando por encima de lo esperable tras controlar los efectos de las variables contextuales. Son, por tanto, centros de alta eficacia a los que se realizará un estudio en profundidad para detectar las buenas prácticas que desarrollan y que explican esos buenos resultados, ese alto nivel de eficacia. Este tipo de estudio no longitudinal en el que se controlan y detraen los efectos contextuales para aislar los efectos de centro es, como antes se ha apuntado, lo que la OCDE (2011) recomienda denominar como modelos transversales contextualizados para diferenciarlos de aquellos longitudinales cuyo objetivo es el estudio del cambio o del crecimiento (incorporen o no variables contextuales) y para los que se sugiere reservar la denominación de estudios de valor añadido en sentido estricto.

RESULTADOS

Antes de iniciar la exposición de este apartado, es preciso dejar claro que el objetivo de este trabajo no es el de elaborar, comparar o interpretar modelos explicativos del rendimiento de los estudiantes. Focalizar la atención en este planteamiento es ciertamente fructífero vista la numerosa presencia de este tipo de estudios en la literatura especializada. Pero en esta investigación es otro el fin que se persigue pues se trata de realizar el control de las covariables pertinentes y significativas ($\alpha = .05$) para, como antes se ha señalado, identificar los centros que más se separan en sentido positivo de los modelos generales, los de mayor residuo en cada competencia, año y etapa.

La profesora Martínez Arias lo resume perfectamente al señalar que estos modelos “pueden resultar muy útiles en los procesos de diagnóstico y mejora de las escuelas... Combinados con otras metodologías de naturaleza más cualitativa como observaciones, entrevistas, portafolios, etc., ayudan a la identificación de buenas prácticas que se pueden potenciar en las reformas educativas”. (Martínez Arias, 2009, p. 239).

Por todo ello, es necesario aclarar que en los resultados no se incluyen las estimaciones de los parámetros. A continuación se presenta la Tabla 2, correspondiente a los modelos de la Educación Primaria con las covariables de los dos niveles que constituyen los modelos lineales jerárquicos por competencia y año.

Tabla 2. Covariables significativas para los dos niveles en Educación Primaria

EDUCACIÓN PRIMARIA	MODELO JERÁRQUICO	AÑO 2009	AÑO 2010
Lengua vasca	Nivel 1	Modelo A	Modelo A
		Modelo B	Modelo B
		Sexo	Sexo
		Idioma familiar	Idioma familiar
		Inmigrante	Inmigrante
		ISEC familiar	ISEC familiar
		Repetidor	Repetidor
		Rendimiento en Lengua vasca el curso anterior	Rendimiento en Lengua vasca el curso anterior

Lengua vasca	Nivel 2	Tasa de alumnado vasco parlante del centro	Tasa de alumnado vasco parlante del centro
		Tasa de inmigrantes del centro	Tasa de inmigrantes del centro
		ISEC medio del centro	ISEC medio del centro
		Rendimiento del centro en Lengua vasca el curso anterior	Rendimiento del centro en Lengua vasca el curso anterior
Lengua española	Nivel 1	Sexo	Modelo A
		ISEC familiar	Idioma familiar
		Repetidor	ISEC familiar
		Rendimiento en Lengua española el curso anterior	Repetidor Rendimiento en Lengua española el curso anterior
Lengua española	Nivel 2	Tasa de alumnado vasco parlante del centro	Tasa de alumnado vasco parlante del centro
		ISEC medio del centro	ISEC medio del centro.
		Rendimiento del centro en Lengua española el curso anterior.	Rendimiento del centro en Lengua española el curso anterior
			Red escolar del centro
Matemáticas	Nivel 1	Modelo A	Modelo B
		Modelo B	Inmigrante
		Idioma familiar	ISEC familiar
		ISEC familiar	Repetidor
		Repetidor Rendimiento en Matemáticas el curso anterior	Rendimiento en Matemáticas el curso anterior
Matemáticas	Nivel 2	Tasa de alumnado vasco parlante del centro	Tasa de alumnado vasco parlante del centro
		ISEC medio del centro	ISEC medio del centro
		Tasa repetidores centro	Rendimiento del centro en Matemáticas el curso anterior
		Rendimiento del centro en Matemáticas el curso anterior	

Promedio	Nivel 1	Modelo A	Modelo A
		Sexo	Sexo
		Inmigrante	Inmigrante
		ISEC familiar	ISEC familiar
		Repetidor	Repetidor
		Rendimiento promedio el curso anterior	Rendimiento promedio el curso anterior
Promedio	Nivel 2	Tasa de alumnado vascoparlante del centro	ISEC medio del centro.
		ISEC medio del centro	Tasa de repetidores del centro
		Tasa de repetidores del centro	Rendimiento promedio del centro el curso anterior
		Rendimiento promedio del centro el curso anterior	

* Rendimiento promedio del centro el curso anterior, es el promedio de los rendimientos de las tres competencias obtenidos el curso anterior por el alumnado del centro.

De la misma manera, en la Tabla 3 se resumen los modelos jerárquicos lineales de la Educación Secundaria.

Tabla 3. Covariables significativas para los dos niveles en Educación Secundaria Obligatoria

EDUCACIÓN SECUNDARIA OBLIGATORIA	MODELO JERÁRQUICO	AÑO 2009	AÑO 2010
Lengua vasca	Nivel 1	Modelo A	Modelo A
		Modelo B	Modelo B
		Sexo	Sexo
		Inmigrante	Idioma familiar
		ISEC familiar	Inmigrante.ISEC familiar
		Repetidor	Repetidor
		Rendimiento en Lengua vasca el curso anterior	Rendimiento en Lengua vasca el curso anterior

DETECCIÓN Y CARACTERIZACIÓN DE LOS CENTROS ESCOLARES DE ALTA EFICACIA

Lengua vasca	Nivel 2	Tasa de alumnado vasco parlante del centro ISEC medio del centro Rendimiento del centro en Lengua vasca el curso anterior	ISEC medio del centro Tasa repetidores centro Rendimiento del centro en Lengua vasca el curso anterior
Lengua española	Nivel 1	Modelo A Idioma familiar Inmigrante ISEC familiar Repetidor Rendimiento en Lengua española el curso anterior	Modelo A Sexo Idioma familiar Inmigrante ISEC familiar Repetidor Rendimiento en Lengua española el curso anterior
Lengua española	Nivel 2	Tasa de alumnado vasco parlante del centro ISEC medio del centro Tasa repetidores centro Rendimiento del centro en Lengua española el curso anterior	Tasa de alumnado vasco parlante del centro ISEC medio del centro Rendimiento del centro en Lengua española el curso anterior Red escolar del centro
Matemáticas	Nivel 1	Modelo B Inmigrante ISEC familiar Repetidor Rendimiento en Matemáticas el curso anterior	Modelo A Idioma familiar Inmigrante ISEC familiar Repetidor Rendimiento en Matemáticas el curso anterior
Matemáticas	Nivel 2	Tasa de alumnado vasco parlante del centro ISEC medio del centro Tasa de repetidores del centro Rendimiento del centro en Matemáticas el curso anterior	Tasa inmigrantes centro. ISEC medio del centro. Tasa de repetidores del centro. Rendimiento del centro en Matemáticas el curso anterior

Promedio	Nivel 1	Modelo A	Modelo A
		Idioma familiar	Sexo
		Inmigrante	Inmigrante
		ISEC familiar	ISEC familiar
		Repetidor	Repetidor
		Rendimiento promedio el curso anterior	Rendimiento promedio el curso anterior
Promedio	Nivel 2	Tasa de alumnado vasco parlante del centro	Tasa de inmigrantes del centro
		ISEC medio del centro	ISEC medio del centro
		Tasa de repetidores del centro	Tasa de repetidores del centro
		Rendimiento promedio del centro el curso anterior	Rendimiento promedio del centro el curso anterior
			Red escolar del centro

* Rendimiento promedio del centro el curso anterior, es el promedio de los rendimientos de las tres competencias obtenidos el curso anterior por el alumnado del centro.

Lo fundamental de estos resultados no son tanto los modelos en sí, sino su aplicación en la búsqueda y selección de los centros escolares que más destacan por su residuo positivo. De cualquier forma, sí se puede señalar que estos resultados son básicamente coincidentes con estudios similares en lo relativo al efecto diferencial de las distintas variables consideradas (Santiago, Lukas, Joaristi, Lizasoain y Moyano, 2008); el efecto de los modelos lingüísticos es muy importante en la competencia en Lengua vasca y menor en Lengua española o Matemáticas (Joaristi, Lizasoain, Lukas y Santiago, 2009) y el efecto de algunas variables contextuales es menor en Educación Secundaria que en Primaria, por lo que parece que algunos de estos efectos tienden a amortiguarse.

De cualquier forma, una vez diseñados, validados y aplicados estos modelos al conjunto de los datos considerados, se calculó el residuo u_{0j} de cada centro como indicador de la diferencia entre el rendimiento medio observado y el esperado tras controlar las variables contextuales. El paso que lógicamente se sigue es seleccionar aquellos centros en los que tal residuo sea positivo y más elevado, ya que ello denota una elevada eficacia en el desempeño de su labor de enseñanza-aprendizaje.

Al analizar las distribuciones de frecuencias de estos residuos para cada competencia, año y etapa se detectan centros con valores muy altos (superiores incluso

a 2 unidades típicas). Pero el problema surge al constatar que estos indicadores de valor añadido adolecen de una cierta inestabilidad, por lo que es relativamente frecuente que centros que en un año y asignatura determinados obtienen un residuo muy alto, no les ocurra lo mismo en otras competencias o, sobre todo, en el año siguiente.

Esta relativa inestabilidad de las medidas de valor añadido ha sido estudiada por diversos autores, por ejemplo, Newton, Darling-Hammond, Haertel y Thomas (2010). La OCDE en su obra sobre las mejores prácticas para estudiar el valor añadido de los centros escolares apunta que uno de los factores de dicha inestabilidad es la alta volatilidad de los resultados de las escuelas pequeñas (OCDE, 2011, p. 137). Pero como ha quedado dicho, en este trabajo las mismas no han sido consideradas. Se trata en cualquier caso de una cuestión que demanda un estudio específico más detallado.

Pero tomando en consideración esta cuestión, en esta fase del estudio era necesario establecer un criterio que permitiese seleccionar aquellos centros escolares de alta eficacia global para su posterior estudio en profundidad. Y, dado que el foco de interés está centrado en los centros de alta eficacia en general, se optó por emplear un criterio que primase por una parte la obtención de altos residuos en todas (o casi todas) las competencias evaluadas y, por otra, que tuviesen una cierta estabilidad en el tiempo de manera que estos altos residuos debían darse en los dos años de aplicación de la ED.

El procedimiento finalmente adoptado para la selección de los centros de más alto valor añadido global consistió en reordenar por rangos los residuos calculados para cada competencia y año y en función de ello seleccionar aquellos centros cuyos residuos se situasen en los dos años por encima del centil 80, bien en las tres competencias instrumentales básicas o bien en dos de ellas y en el promedio de las cuatro evaluadas cada año.

La elección del centil 80 como punto de corte es una decisión de carácter heurístico basada en dos consideraciones. De una parte es un punto de corte elevado que garantiza que el tamaño del residuo es lo suficientemente grande como para que el supuesto de la alta eficacia sea plausible; puntos de corte más bajos podrían quizá comprometer tal asunción. En segundo lugar, el hecho de que esta condición deba darse en las tres competencias instrumentales básicas (o en dos de ellas y en el promedio de las cuatro) y en los dos años hace que el establecimiento de puntos de corte más elevados sea excesivamente restrictivo implicando que sean muy pocos los centros que cumplan dichas condiciones. El criterio del centil 80 dio como resultado una selección compuesta por un número de centros manejable de forma

que el estudio cualitativo posterior era abordable en los plazos establecidos en función de las posibilidades del equipo investigador.

Se trata, como se ha afirmado, de un criterio heurístico en el que se consideran las condiciones en las que la investigación se desarrolla, pero que también lleva aparejado un elevado nivel de exigencia que garantiza que la hipótesis de la alta eficacia no resulte descabellada. En cualquier caso, en la literatura suele ser habitual en este tipo de estudios considerar como subgrupos extremos de una ordenación las unidades integrantes, bien del primer y último cuartil, bien de los quintiles extremos. Por ejemplo, Yeh y Ritter (2009) emplean el criterio de los cuartiles en un estudio de costo-eficacia aplicando técnicas de valor añadido. Mientras que Galbraith, Merrill y Kline (2012) usan el de los quintiles extremos en ordenaciones de docentes en su estudio sobre la eficacia docente a nivel universitario.

En el caso del trabajo que se presenta, se ha optado por una aproximación al criterio del quintil por las razones antedichas. Como resultado de la aplicación del mismo, inicialmente fueron 42 los centros seleccionados y su distribución por etapas y redes se muestra en la Tabla 4.

Tabla 4. Distribución de los centros seleccionados

ETAPA	CENTROS PÚBLICOS	CENTROS CONCERTADOS	TOTAL
Educación Primaria	9	11	20
Educación Secundaria	10	12	22
Total	19	23	42

Sin embargo, dada la inestabilidad observada en algunas puntuaciones y para triangular estos resultados y asegurar así la selección más idónea de centros de cara al desarrollo de las siguientes fases, se procedió a un filtrado ulterior. Las siguientes fases del proyecto precisarán de un trabajo cualitativo para detectar las buenas prácticas que explican o se asocian a estos resultados positivos y que se centrarán en la selección de los informantes clave y los datos que de manera directa éstos aporten, la determinación de los conceptos básicos respecto a términos como centro eficaz o buenas prácticas, la definición, estructuración del discurso y del marco categorial y la adecuada recogida de información, así como su posterior análisis. Por todo ello se procedió a examinar con más detalle la situación de aquellos centros cuyos residuos tenían mayor variabilidad y recurriendo a la información aportada por la Inspección y por el personal técnico asesor de formación e innovación do-

cente. Esto condujo a eliminar a 10 centros, de forma que la selección final fue de 32, cuyas características básicas se presentan en las tablas 5 y 6.

Tabla 5. Red, etapa y modelo lingüístico de los centros seleccionados

RED	MODELO LINGÜÍSTICO	EDUCACIÓN PRIMARIA	EDUCACIÓN SECUNDARIA	TOTAL CENTROS
Centros concertados	A	1	3	4
	B	3	1	4
	D	1	2	3
	A-B	1	2	3
	B-D	2	0	2
	A-B-D	0	0	0
	Total concertados	8	8	16
Centros públicos	A	1	1	2
	B	0	1	1
	D	6	4	10
	A-B	1	1	2
	B-D	0	0	0
	A-B-D	0	1	1
Total públicos	8	8	16	
Total centros		16	16	32

Así, con objeto de caracterizar adecuadamente este subconjunto de centros, se los compara con los que no fueron seleccionados. En lo que se refiere a las variables contextuales asociadas a los centros escolares (las que se han controlado al obtener los residuos indicadores del valor añadido), las características estadísticas se muestran en la Tabla 6.

Tabla 6. Diferencias entre los centros no seleccionados y seleccionados

COVARIABLES	NO SELECCIONADOS N = 561 CENTROS			SELECCIONADOS N = 32 CENTROS		
	M (D.T.)	MÍN	MÁX	M (D.T.)	MÍN	MÁX
COVARIABLES DEL NIVEL 2						
Nº de estudiantes del centro	45.65 (26.90)	13.00	176.00	53.05 (29.03)	13.00	144.00

Red escolar (% de centros concertados)	51.68			50.00		
ISEC promedio del centro	-.13 (.53)	-3.05	1.38	-.11 (.52)	-1.37	.99
Tasa de inmigrantes (%)	8.14 (11.97)	.00	84.62	7.58 (14.53)	.00	80.00
Tasa de alumnado vasco parlante (%)	24.01 (24.61)	.00	100.00	22.77 (28.45)	.00	86.67
Tasa de repetidores (%)	15.79 (14.13)	.00	88.24	16.25 (15.04)	.00	66.13
Promedio de rendimiento previo en Lengua vasca	2.40 (.46)	.40	4.00	2.31 (.45)	1.26	3.31
Promedio de rendimiento previo en Lengua española	2.58 (.48)	1.07	4.00	2.47 (.47)	1.30	3.61
Promedio de rendimiento previo en Matemáticas	2.53 (.46)	.89	4.00	2.47 (.46)	1.34	3.52
Promedio de la media de los 3 rendimiento previos	2.51 (.38)	1.20	4.00	2.42 (.40)	1.47	3.21
Promedio de horas de deberes	1.56 (.24)	.50	2.37	1.56 (.26)	1.10	2.39
Promedio de ayuda recibida	.87 (.25)	.06	1.74	.88 (.26)	.38	1.79
Promedio de numero de actividades extraescolares	1.34 (.33)	.29	2.50	1.33 (.37)	.59	2.16

Tras realizar la prueba t de Student de comparación de medias de dos grupos, al 5% no hay diferencias significativas entre los centros seleccionados y los no seleccionados, y lo mismo ocurre en las pruebas de comparación de las medias del grupo de centros seleccionados con las del total de los centros, donde las diferencias tampoco resultan significativas.

El modelo lingüístico es una covariable asociada al estudiante, no al centro, pues hay centros (Tabla 5) en que se imparte la docencia en más de un modelo. Por lo tanto los resultados de la Tabla 7 se refieren a número de estudiantes. En ella se observa que hay una asociación entre ambos aspectos ($Chi^2 = 399.886$; $p = .000$) y que los centros seleccionados tienen mayor tendencia a los modelos A y B en comparación con los no seleccionados.

Tabla 7. Diferencias en el modelo lingüístico entre los centros seleccionados y los no seleccionados

MODELO LINGÜÍSTICO	CENTROS NO SELECCIONADOS	CENTROS SELECCIONADOS
A	7880 (12.07 %)	984 (21.57 %)
B	19417 (29.75 %)	1434 (31.43 %)
D	37976 (58.18 %)	2144(47.00 %)

DISCUSIÓN Y CONCLUSIONES

Para dar cumplimiento al objetivo de detectar los centros escolares con los mejores resultados condicionados al efecto escolar, los Modelos Jerárquicos Lineales, en el contexto de los estudios transversales contextualizados, constituyen una herramienta adecuada, pues permiten detraer el efecto de las variables contextuales de forma que el residuo u_{0j} positivo y alto es un indicador adecuado para la selección de los centros de alta eficacia.

Así, de los resultados de los Modelos Jerárquicos Lineales se extrae que es necesario controlar una serie de variables contextuales tanto del estudiante como del centro, siendo más correcto proceder a la selección de los centros no por su puntuación directa promedio sino por el residuo.

El criterio empleado de incluir sólo aquellas variables estrictamente contextuales, es decir, aquellas que sean independientes de la políticas y estrategias adoptadas por los centros, resulta ser una garantía de equidad para la selección de los centros, en la medida en que así se detrae sólo el efecto de lo meramente contextual.

Los 32 centros finalmente seleccionados no se diferencian de forma significativa de los no seleccionados con respecto a ninguna de las variables caracterizadoras empleadas. Esto supone que reflejan adecuadamente la diversidad del sistema educativo de la CAPV y que por tanto la alta eficacia no está asociada a ningún subgrupo específico de centros.

En concreto, sus características básicas son las siguientes:

- se distribuyen de forma análoga a los centros no seleccionados,
- son centros tan heterogéneos en los diversos aspectos considerados como los no seleccionados: red, etapa, ISEC medio (8 muy alto, 17 medio y 7

- bajo), proporción de estudiantes inmigrantes (desde ninguno hasta 80%),
- la mayoría son centros de núcleos urbanos grandes o medianos aunque también los hay en zonas eminentemente rurales,
 - respecto al modelo lingüístico se observa una cierta tendencia al alza del modelo A en comparación con la población, pudiéndose explicar tal circunstancia sabiendo que las puntuaciones en Lengua vasca de los estudiantes del modelo A son muy bajas en comparación con los de los otros modelos, lo cual hace que la puntuación esperada de un centro con modelo A sea muy baja y que, por tanto, sea más probable obtener residuos altos. Como sea, esta apreciación es una hipótesis a probar.

Por lo tanto, esta variedad que se encuentra en los centros seleccionados, es decir, los centros excelentes, indica que la alta eficacia no está asociada a ningún perfil definido o a una tipología específica de centro escolar. Un centro puede ser de alta eficacia independientemente de la red a la que pertenezca, de la etapa que imparta, del modelo lingüístico en el que desarrolle la enseñanza, o de la procedencia geográfica, social, cultural, económica de los estudiantes a quienes escolariza.

Esta conclusión nos parece que es el hallazgo más relevante de este trabajo y es una cuestión que habrá que contrastar con los resultados del análisis en profundidad mediante metodología cualitativa actualmente en curso. De confirmarse esta hipótesis, este carácter transversal de la elevada eficacia podría coadyuvar a facilitar la transferibilidad de las buenas prácticas que finalmente se detecten.

LIMITACIONES

Observando que las comparaciones se han realizado entre los centros seleccionados y los no seleccionados, salta a la vista una considerable carencia. La validez de las conclusiones se vería notablemente incrementada de realizar la comparación con los centros de residuos más bajos (de emplear un criterio paralelo al usado en la selección de los de residuo alto, en este caso el punto de corte se tendría que establecer en el centil 20). Desde el punto de vista estadístico esto no supone mayor problema, pero el desarrollo de la segunda fase del presente trabajo, considerada la fundamental, puede comportar restricciones deontológicas serias. Por tanto se considera que lo más apropiado para realizar el trabajo futuro deberá basarse en la realización de entrevistas y cuestionarios abiertos a muestras de centros de los tres tipos en función de su residuo ajustado y en una escala constituida por los valores altos, medios y bajos. De esta forma se podrá comparar la práctica educativa y diferenciar, en su caso, las buenas de las malas prácticas.

Junto a esto, paralelamente a la fase correspondiente al análisis cualitativo de las características de los centros seleccionados, se incorporarán al análisis estadístico los nuevos resultados de la ED correspondientes al año 2011, siendo uno de sus fines fundamentales el estudio de la estabilidad de los logros académicos de los estudiantes.

Fecha de recepción del original: 28 de mayo de 2013

Fecha de aceptación de la versión definitiva: 5 de diciembre de 2013

REFERENCIAS BIBLIOGRÁFICAS

- Bryk, A. S., Sebring, P. B., Allensworth, E., Luppescu, S. y Easton, J. Q. (2010). *Organizing Schools for Improvement: Lessons from Chicago*. Chicago: University of Chicago Press.
- Galbraith, C. S., Merrill, G. B., y Kline, D. M. (2012). Are student evaluations of teaching effectiveness valid for measuring student learning outcomes in business related classes? A neural network and bayesian analyses. *Research in Higher Education*, 53(3), 353-374.
- Gaviria, J. L., Martínez-Arias, R. y Castro, M. (2004). Un Estudio Multinivel Sobre los Factores de Eficacia Escolar en Países en Desarrollo: El Caso de los Recursos en Brasil. *Education Policy Analysis Archives*, 12(20). Extraído el 4 de octubre de 2013 <http://epaa.asu.edu/epaa/v12n20/>
- Gaviria, J. L. y Castro, M. (2005). *Modelos Jerárquicos Lineales*. Madrid: La Muralla.
- Goldstein, H. (2003). *Multilevel Statistical Models*, 3ª ed. New York: Hodder Arnold.
- Gritter, A. (2006). Great Schools: Identifying Higher-Performing Schools. *Center for Education Policy, Applied Research, and Evaluation*. Extraído el 3 de octubre de 2013 http://www.usm.maine.edu/sites/default/files/Center%20for%20Education%20Policy,%20Applied%20Research,%20and%20Evaluation/Great_Schools.pdf
- Huffman, J. B., Pankake, A. y Munoz, A. (2006). The Tri-Level Model in Action: Site, District, and State Plans for School Accountability in Increasing Student Success. *Journal of School Leadership*, 16(5), 569-582.
- Jackson, S. A. y Lunenburg, F. C. (2010). School Performance Indicators, Accountability Ratings, and Student Achievement. *American Secondary Education*, 39(1), 27-44.
- Joaristi, L., Lizasoain, L., Lukas, J. F. y Santiago K. (2009): Trilingualism (Spanish, English and Basque) in the Educational System of the Basque

- Country. *International Journal of Multilingualism*, 6(1), 105-126. DOI: 10.1080/14790710802650001
- Lizasoain, L. y Joaristi, L. (2010). Estudio Diferencial del Rendimiento Académico en Lengua Española de Estudiantes de Educación Secundaria de Baja California (México). *Revista Iberoamericana de Evaluación Educativa*, 3(3), 115-134.
- Marchesi, A. y Martín, E. (2002). *Evaluación de la educación secundaria: fotografía de una etapa polémica*. Madrid: Fundación Santa María.
- Martínez Arias, R. (2009). Usos, aplicaciones y problemas de los modelos de valor añadido en educación. *Revista de Educación*, (348), 217-250.
- Murillo, F. J. y Hernández-Castilla, R. (2011a). Efectos escolares de factores socio-afectivos. Un estudio Multinivel para Iberoamérica. *Revista de Investigación Educativa*, 29(2), 407-428.
- Murillo, F. J. y Hernández-Castilla, R. (2011b). Factores escolares asociados al desarrollo socio-afectivo en Iberoamérica. *RELIEVE*, 17(2), 1-23. Extraído el 3 de octubre de 2013 http://www.uv.es/RELIEVE/v17n2/RELIEVEv17n2_2.htm
- Murillo, F. J. y Román, M. (2011). ¿La escuela o la cuna? Evidencias sobre su aportación al rendimiento de los estudiantes de América Latina. Estudio multinivel sobre la estimación de los efectos escolares. *Revista de Curriculum y Formación de Profesorado*, 15(3), 27-53.
- Newton, X., Darling-Hammond, L., Haertel, E. y Thomas, E. (2010). Value-Added Modeling of Teacher Effectiveness: An exploration of stability across models and contexts. *Educational Policy Analysis Archives*, 18(23). Extraído el 4 de octubre de 2014 <http://epaa.asu.edu/ojs/article/view/810>
- OCDE (2011). *La medición del aprendizaje de los alumnos: Mejores prácticas para evaluar el valor agregado de las escuelas*. Paris: OCDE.
- OECD (2008). *Measuring improvements in learning outcomes: best practices to assess the value-added of schools*. Paris: OECD.
- Raudenbush, S. W. y Bryk, A. S. (2002). *Hierarchical linear models: applications and data analysis methods*. Thousand Oaks: Sage Publications.
- Ray, A., McCormack, T. y Evans, H. (2009). Value Added in English Schools. *Education Finance and Policy*, 4(4), 415-438.
- Santiago, K., Lukas, J. F., Joaristi, L., Lizasoain, L. y Moyano, N. (2008). A Longitudinal Study of Academic Achievement in Spanish: The Effect of Linguistic Models. *Language, Culture and Curriculum*, 21(1), 48-58.
- Snijders, T. A. B. y Bosker, R. J. (1999). *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. Thousand Oaks: Sage Publications.

- Yeh, S. S. y Ritter, J. (2009). The cost-effectiveness of replacing the bottom quartile of novice teachers through value-added teacher assessment. *Journal of Education Finance*, 34(4), 426-451.
- Wright, R. J. (2009). Methods for Improving Test Scores: The Good, the Bad, and the Ugly. *Kappa Delta Pi Record*, 45(3), 116-121.