

Criterios y modelos estadísticos de eficacia escolar

Criteria and statistical models of school effectiveness

Luis Lizasoain Hernández

Resumen

El objetivo de este artículo es presentar los criterios y modelos estadísticos empleados en un estudio de eficacia escolar desarrollado en la Comunidad Autónoma del País Vasco empleando como variable criterio los resultados en matemáticas, comprensión lectora en lengua castellana y en lengua vasca, resultantes de las evaluaciones de Diagnóstico aplicadas en cinco años. Se definen cuatro criterios de eficacia escolar: puntuaciones extremas, residuos extremos, crecimiento de puntuaciones y crecimiento de residuos. Para ello se han aplicado técnicas de regresión multinivel empleando modelos jerárquicos lineales. Los resultados permiten una selección de centros tanto de alta como de baja eficacia que se basa en cuatro enfoques distintos y complementarios de la eficacia (o ineficacia) escolar.

Palabras clave: Eficacia y mejora escolar; Buenas prácticas; Evaluación educativa; Modelos jerárquicos lineales.

Abstract

The aim of this paper is to present the statistical criteria and models used in a school effectiveness research carried out in the Basque Country Autonomous Community using the mathematics, Spanish language and Basque language scores as outcome variables. These scores are collected from the Diagnosis Assessments applied for five years. Four school effectiveness criteria are defined: extreme scores, extreme residuals, scores growth and residuals growth. Multilevel regression techniques have been applied using hierarchical linear models (HLM).

Correspondencia: Luis Lizasoain Hernández, luis.lizasoain@ehu.es.

Results allow a selection of both high- and low-effectiveness schools based on four different and complementary school effectiveness approaches.

Keywords: Effective school research; Best practices; Educational assessment; Hierarchical linear modeling.

Dedicatoria

En este artículo se presenta un análisis detallado de los criterios y procedimientos estadísticos empleados en diversos proyectos de eficacia y mejora escolar que he coordinado. Pero los mismos no hubiesen podido llevarse a cabo sin el trabajo y la dedicación de muchas personas con las que he tenido la suerte de compartir tiempo de trabajo y también y sobre todo, de amistad. A todas ellas, y muy especialmente a los miembros del equipo de investigación que conformamos entre la Universidad del País Vasco (UPV/EHU) y el Instituto Vasco de Evaluación e Investigación Educativa (ISEI/IVEI), con profundo agradecimiento, dedico este trabajo.

Introducción

El objetivo de este trabajo es presentar, discutir y analizar los diversos criterios y las técnicas estadísticas que se han empleado en distintos proyectos de eficacia escolar y mejora escolar. Los mismos se han llevado a cabo –o están en curso– en España, en las Comunidades Autónomas del País Vasco (CAPV), Andalucía, La Rioja y las Islas Baleares; y en los estados de Baja California y Aguascalientes en México, en las provincias de Loja y Zamora-Chinchipe en Ecuador, y en el departamento de Cundinamarca en Colombia.

Aunque con las lógicas diferencias inherentes a sociedades a veces muy diversas y a sistemas educativos distintos, el enfoque adoptado en todos ellos es similar, empleándose como variable criterio el rendimiento académico del alumnado medido mediante una prueba estandarizada aplicada a toda la población estudiantil perteneciente a uno o varios cursos académicos.

En las publicaciones que siguen (presentadas por orden descendente de fecha de publicación), se presentan las características comunes y las específicas de este conjunto de estudios: Azkarate, Bartau y Lizasoain, (2019a; 2019b); Angulo y Lizasoain (2019); Azkarate, Bartau, Aierbe, y Oregui, (2018); Oregui, Aierbe, Bartau y Azkarate (2018); Padilla, Guzmán, Lizasoain y García-Medina (2018); Pedroza, Peniche y Lizasoain (2018); Bartau, Azpillaga y Joaristi (2017); Etxeberria, Intxausti y Azpillaga (2017); Martínez Abad, Lizasoain Hernández, Castro Morera y Joaristi Olariaga (2017); Intxausti, Etxeberria y Bartau (2016); Intxausti, Joaristi, y Lizasoain (2016); Lizasoain, Bereziartua y Bartau (2016); Azpillaga, Intxausti, y Joaristi, (2014); Joaristi, Lizasoain y Azpillaga, (2014); Martínez-Abad, Chaparro Lizasoain (2014) y Lizasoain y Angulo (2014).

Este trabajo se centra en el primer estudio realizado que se inició en 2012 en el País Vasco y que se ha desarrollado vinculado a los tres proyectos I+D que se referencian a pie de página¹. El último de ellos está actualmente en curso siendo dirigido por

¹ El estudio de eficacia y mejora escolar en la Comunidad Autónoma del País Vasco se ha llevado a cabo en el contexto de los proyectos siguientes:

las profesoras Azpillaga y Bartau del Departamento de Métodos de Investigación y Diagnóstico en Educación de la UPV/EHU.

La finalidad general de los dos primeros proyectos fue la de identificar y posteriormente estudiar los centros considerados de alta eficacia con objeto de obtener evidencias sobre sus prácticas y líneas de actuación que se consideren asociadas a dicha eficacia y que puedan servir para el diseño de acciones y programas de mejora para los centros cuyos resultados no son los deseables.

Y este fin conllevó la adopción de un enfoque metodológico mixto en la medida en que, en una primera fase, y para identificar los centros de alta eficacia se emplearon procedimientos estadísticos de regresión multinivel mediante modelos jerárquicos lineales (HLM). Posteriormente, en una segunda etapa, una vez definido el subconjunto de tales centros eficaces, se aborda el estudio en profundidad de los mismos, de su praxis y realidad cotidiana, mediante un enfoque de estudio de casos múltiple más comprensivo empleando técnicas cualitativas con objeto de detectar las buenas prácticas y los factores asociados que pueden ser explicativos de tan excelentes resultados.

Con lo hasta aquí expuesto, está claro que este conjunto de estudios se insertan en la línea de la eficacia y mejora escolar, ámbito de una dilatada tradición académica. Existe casi unanimidad entre los autores al señalar a Coleman y el conocido informe que lleva su nombre (Coleman, 1966) si no como el origen, sí como el desencadenante de esta línea de investigación. Y con él, como antecedentes importantes es preciso también citar los trabajos de Jencks y colaboradores (1972).

No es el objetivo de este artículo el realizar una revisión detallada del estado de la cuestión en este tema. Las personas interesadas pueden encontrar excelentes síntesis de la investigación sobre eficacia y mejora escolar en las siguientes tres referencias: en primer lugar, la obra que Javier Murillo publicó en 2005 titulada *La investigación sobre eficacia escolar* (Murillo, 2005). Dos años después, Townsend y Ávalos coordinan el *International Handbook of School Effectiveness and Improvement* (Townsend y Avalos, 2007). Por último, en 2014 Reynolds y colaboradores (Reynolds, Sammons, De Fraine, Van Damme, Townsend, Teddlie y Stringfield, 2014) publicaron una detallada revisión sobre el *estado del arte* de este campo de investigación.

Método

Reynolds y colaboradores en el trabajo antes citado afirman que las dos preguntas básicas de la investigación sobre eficacia escolar son las siguientes: “*What makes a “good” school?*” y “*How do we make more schools good?*” (Reynolds, Sammons, De Fraine, Van Damme, Townsend, Teddlie y Stringfield, 2014, p. 197).

“Caracterización y buenas prácticas de los centros escolares de alto valor añadido”. Proyecto I+D+i del Ministerio de Ciencia e Innovación. Referencia EDU2011-24366. Investigador Principal: Luis Lizasoain Hernández. Período 2012-2014.

“Estudio longitudinal y contextualizado de centros escolares de muy alta y muy baja eficacia: diseño de acciones de mejora escolar”. Proyecto I+D del Ministerio de Economía e Innovación. Referencia EDU2014-53511-P. Investigador Principal: Luis Lizasoain Hernández. Período 2015-2018.

“Mejora escolar desde la perspectiva de género en la Comunidad Autónoma Vasca”. de Proyecto I+D del Ministerio de Ciencia, Innovación y Universidades. Referencia PGC2018-094124-B-100. Investigadoras Principales: Verónica Azpillaga Larrea e Isabel Bartau Rojas. Período 2019-2022.

Dar una adecuada respuesta a qué se entiende por un *buen* centro educativo, por un colegio de alta eficacia, es una cuestión crucial porque de la misma va a depender en gran medida el tipo de resultados que se obtengan. Uno de los primeros trabajos en abordar esta cuestión fue el de Aitkin y Longford que ya en 1986 analizaban y discutían cuestiones relativas a la modelización estadística en este tipo de estudios. Posteriormente Goldstein (1997) en su trabajo sobre los métodos de la investigación en eficacia escolar, y De Maeyer, Van den Bergh, Rymenans, Van Petegem y Rijlaarsdam (2010) en el suyo sobre los criterios de eficacia usados en modelos multivariados, abordaron también detalladamente esta cuestión.

Muestra y variables de resultado

En el caso que aquí se describe, la variable dependiente son los resultados de las Evaluaciones de Diagnóstico (ED) que el ISEI/IVEI² aplicó con carácter censal a todos los estudiantes de 4º de Educación Primaria (EP) y 2º de Educación Secundaria Obligatoria (ESO) de los centros sostenidos con fondos públicos de la CAPV en los años 2009, 2010, 2011, 2013 y 2015. En dichas EDs se evalúan siempre las tres competencias instrumentales básicas (matemáticas y comprensión lectora en lengua vasca y en lengua castellana). Las puntuaciones son estimadas mediante modelos de TRI que se distribuyen con una media de 250 y una desviación típica de 50.

La variable de resultado empleada en los análisis es, a nivel individual del estudiante (nivel 1, N1), la puntuación obtenida en cada una de las tres competencias instrumentales básicas; y al nivel del centro (N2), el promedio de las mismas. En las líneas anteriores está explícitamente formulado un enfoque multinivel (estudiantes y centros). Esta perspectiva multinivel es una característica reseñable de los análisis efectuados y se basa en la consideración de que los datos educativos presentan muy frecuentemente una estructura anidada, una jerarquía. Desde el punto de vista estadístico, este enfoque multinivel se lleva a cabo mediante modelos jerárquicos lineales (HLM).

El hecho de emplear como variable criterio los resultados de las EDs tiene implicaciones importantes. La primera de ellas es que el foco de atención se pone exclusivamente en el componente *instructivo* del proceso educativo. Y además, únicamente, en lo relativo a las tres competencias instrumentales básicas. Por tanto, el enfoque es forzosa e inevitablemente parcial puesto que la actividad educativa de un centro escolar va mucho más allá de lo meramente instructivo. Y los aspectos emocionales y afectivos son cada vez más frecuentemente incorporados en este tipo de estudios. Véase un ejemplo en el trabajo realizado por Murillo y Hernández (2011).

Pero, una vez dicho esto y asumiendo como variable criterio básica las puntuaciones de las EDs, volvamos sobre la pregunta inicial: ¿qué se entiende por un *buen* centro?, ¿cuándo se considera que un centro es de *alta* eficacia?

2 Aunque en la dedicatoria ya se ha hecho referencia, estos proyectos no hubieran podido salir adelante sin el apoyo y las facilidades prestadas por el Instituto Vasco de Evaluación e Investigación Educativa (ISEI/IVEI), por lo que expreso públicamente el agradecimiento.

Criterios de eficacia basados en puntuaciones y residuos

A esta pregunta se le pueden dar varias respuestas, y la primera sería la de emplear las *puntuaciones medias brutas* de cada centro como indicador de su eficacia, como indicador de los resultados que obtiene.

En el caso de la CAPV, el análisis descriptivo de los datos de las EDs muestra que, a nivel 2, las puntuaciones medias de los centros se distribuyen con una media cercana a 250 y una desviación típica de entre 18 y 20 puntos. Aunque la gran mayoría de los centros se agrupan alrededor de la media con pequeñas diferencias arriba o abajo, entre el centro que obtiene la puntuación máxima (alrededor de 300 puntos) y el que obtiene la mínima (alrededor de 170) hay –como se ve– importantes diferencias.

Por tanto, una primera respuesta a la pregunta planteada podría basarse en establecer un punto de corte arbitrario de, por ejemplo, 280 puntos en la puntuación en matemáticas. De esta manera, aquellos centros cuya puntuación media fuese igual a superior a 280 puntos podrían ser considerados los mejores centros, al menos en lo que a la enseñanza-aprendizaje de las matemáticas concierne. Al respecto, ha de tenerse en cuenta que 280 son casi 35 puntos más que la media, lo que supone casi 2 desviaciones típicas. Es decir, una puntuación ciertamente muy alta. Y para un centro educativo, lograr una puntuación media de tal calibre no es tarea fácil, implica que *todos* sus estudiantes obtengan puntuaciones alrededor de dicho valor. Y eso es complicado, y tanto más difícil cuanto mayor sea el número de estudiantes.

Éste es el primer criterio de eficacia escolar, el de las *puntuaciones altas y positivas*.

Pero, dicho esto, la pregunta subsiste porque, desde que Coleman publicase el informe antes citado (Coleman, 1966), se sabe que la actividad educativa, y más específicamente la instructiva, depende en alto grado del contexto socioeconómico en el que la misma se desarrolla.

El ISEI/IVEI al realizar la ED, no solo desarrolla y aplica las pruebas de rendimiento en cada una de las competencias que se evalúan, sino que también hace lo propio con un cuestionario para las familias en el que se recaba información sobre diferentes cuestiones contextuales. Entre otras, cabe citar a diversos indicadores que permiten calcular un Índice Socioeconómico y Cultural familiar (ISEC).

Si se quiere cuantificar –y en caso controlar– dichos efectos contextuales, se requiere un enfoque analítico diferente que permita dimensionar con mayor propiedad y equidad el efecto de los centros. Se trata de un enfoque de la eficacia escolar en el que se tengan en cuenta dichos efectos contextuales. En palabras de Murillo (2005), se entiende por una escuela eficaz “...aquella que consigue un desarrollo integral de todos y cada uno de sus estudiantes, más allá de lo que sería previsible teniendo en cuenta su rendimiento previo y la situación socio-económica y cultural de sus familias” (Murillo, 2005, p. 25).

Como han evidenciado los estudios de Willms (2010), las variables contextuales operan a todos los niveles. Por ejemplo, el citado ISEC es una variable con capacidad explicativa de la variabilidad observada en el rendimiento académico al nivel 1 de los estudiantes. Pero el nivel socioeconómico medio de los centros tiene similar poder explicativo con respecto al rendimiento medio de los centros (nivel 2). Y planteamiento similar podría hacerse con variables como si el estudiante pertenece o no

a una familia inmigrante y se ha incorporado tardíamente al sistema educativo. Al nivel 1 del estudiante es una variable dicotómica que siempre resulta significativa. Al nivel 2, se trata de la proporción de estudiantes de estas características que se escolarizan en cada centro.

En definitiva, hay numerosas variables contextuales relevantes, y éstas operan y se dan en los diferentes niveles. Desde esta perspectiva de la eficacia escolar, se requiere un procedimiento estadístico que se ajuste a un mundo real que es complejo, multivariado y multinivel, una técnica analítica que se ajuste a la estructura anidada de los datos (Gaviria y Castro, 2005).

Dicho procedimiento es la regresión múltiple multinivel realizada mediante Modelos Jerárquicos Lineales, que considera y estudia simultáneamente el efecto de múltiples variables a diferentes niveles. La idoneidad de este enfoque analítico para la investigación educativa fue ya señalada por De la Orden (1999). A su vez, los ya citados Gaviria (1999) y Murillo (1999) hicieron lo propio con respecto a la evaluación de sistemas educativos el primero, y a la propia investigación sobre eficacia escolar el segundo.

La notación general para dichos modelos es la siguiente:

Nivel 1 de los estudiantes:

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

Nivel 2 de los centros:

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

$$\beta_{qj} = \gamma_{q0} + \sum_{s=1}^S \gamma_{qs} W_{sj} + u_{qj}$$

En la gran mayoría de los modelos que se han elaborado en este proyecto, se ha comprobado que, si bien los centros varían significativamente en intersecciones, no ocurre lo mismo con las pendientes asociadas a las diferentes covariables, de ahí que la última ecuación adopte la forma de:

Es decir, al no haber variabilidad en lo que a las pendientes se refiere, el parámetro asociado a las mismas es constante. Además, en la literatura se señala que los modelos en los que se *permite* variar tanto intersecciones como pendientes, son apropiados para los estudios diferenciales de la eficacia. Es decir, cuando el objetivo es estudiar el efecto diferencial de alguna o algunas covariables en los diferentes centros (Clarke, Crawford, Steele y Vignoles, 2010). Strand (2010) realiza un estudio diferencial centrado

en etnicidad, género, pobreza y rendimiento previo. Palardy (2008), analiza los efectos diferenciales por centro relativos a la clase social empleando análisis multinivel de curva latente de crecimiento.

Por tanto, cobra aquí especial importancia el parámetro u_{0j} de la primera ecuación del nivel 2. Es el residuo de cada centro asociado a la intersección, una vez controlados todos los factores individuales y de centro incluidos en el modelo. Se trata de la diferencia entre la media observada en cada centro y la gran media de todos. Es el indicador de eficacia para cada centro en la medida en que representa cuánto se aparta la puntuación media de cada centro de la estimada por el modelo una vez que se ha tenido en cuenta el efecto de todas las covariables a ambos niveles. (Raudenbush y Bryk, 2002, capítulo 5, páginas 152 y siguientes).

En el caso de las investigaciones realizadas en la CAPV, y en función de la información disponible en los cuestionarios de contexto del ISEI/IVEI, las covariables consideradas han sido las siguientes:

- Nivel 1 de los estudiantes:
 - o Modelo lingüístico en que el estudiante cursa sus estudios (A, B, D).
 - o Sexo.
 - o Si el euskara es o no la lengua de uso habitual en la familia.
 - o Condición de inmigrante de incorporación tardía al sistema educativo.
 - o Índice socioeconómico y cultural de la familia (ISEC).
 - o Rendimiento previo (calificación del año anterior en el área o materia informada por el propio estudiante).
 - o Si el estudiante es o no repetidor (edad normativa).
- Nivel 2 de los centros:
 - o Red (pública-concertada).
 - o Variables agregadas del N1:
 - ISEC medio del centro.
 - Tasa de inmigrantes.
 - Tasa de euskara casa.
 - Tasa de repetidores (1- tasa de idoneidad).
 - Rendimiento previo medio.

Estos modelos permitieron detectar y seleccionar aquellos centros que funcionan muy bien (de alta eficacia). Para ello, se calcula el residuo para cada centro (u_{0j}) una vez que se ha controlado el efecto de este grupo de variables contextuales relevantes. Y cuanto mayor –y positivo- sea el residuo, mejor el funcionamiento del centro; más se separa de la puntuación que le correspondería en función de su contexto.

De esta manera, los centros así seleccionados, pueden ser ejemplos y fuente de buenas prácticas pues están obteniendo unos resultados superiores a lo esperado, superiores a los centros parecidos a ellos que operan en condiciones y contextos similares. Son, por tanto, centros de alta eficacia en función del segundo criterio: el criterio de los *residuos altos y positivos*. Se trata de un criterio de eficacia escolar equitativo pues toma en cuenta el contexto socioeconómico y cultural del centro.

Este enfoque es lo que durante algún tiempo se denominó como *modelos de valor añadido*. Pero en 2011 la OCDE recomendó que dicha expresión se emplease *solo* para estudios de corte longitudinal cuyo objetivo fuese el estudio y modelización del cambio y el crecimiento (OCDE, 2011). De cualquier forma, independientemente de la denominación de los diferentes tipos de estudio, el concepto de valor añadido resulta de utilidad. En palabras de Gaviria y Castro, "...de alguna forma estos residuos son indicadores de valor añadido" (Gaviria y Castro, 2005, p. 106). Este *valor añadido* es, en definitiva, lo que el modelo estadístico no explica, el límite hasta donde la estadística llega.

El siguiente texto de Martínez Arias (2009) ilustra muy bien el enfoque y planteamiento global del enfoque mixto al que antes se hacía referencia:

Los modelos de valor añadido 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).

Eficacia e ineficacia escolar

Pero, así formulado, este diseño de la investigación conllevaba un riesgo que amenazaba seriamente la validez interna de las conclusiones a las que se pudiese llegar después de realizado dicho estudio cualitativo. Supongamos que, una vez realizado el estudio cualitativo de todos los centros de alta eficacia, identificados y realizados los análisis pertinentes, se obtiene una relación de prácticas y factores asociados más o menos comunes a todos estos centros. Cabría pensar que dichos factores serían los explicativos de la alta eficacia. Pero, ¿qué ocurriría si se realizase el mismo tipo de estudio cualitativo en los centros de *baja* eficacia y los resultados fuesen los mismos o muy similares?, ¿si se observase que sus prácticas son más o menos las mismas? Evidentemente, la validez de las conclusiones se vería muy seriamente comprometida por no decir que casi anulada.

Si en los centros eficaces se piensa en buenas prácticas como factores explicativos de sus buenos resultados, en estos cabe pensar en situaciones de gran dificultad, en problemas serios no resueltos, y también en malas prácticas. Por tanto, para evitar el riesgo de formular conclusiones erróneas, es imprescindible estudiar tanto los centros de alta eficacia (en adelante CAEF) como los de baja eficacia (CBEF) y contrastar las evidencias que ambos conjuntos proporcionan. En palabras de Reynolds y colaboradores (2014, p. 214): "In any science, it is important – if not always popular – to make the implicit explicit. A few studies have been conducted that have formally addressed the topics of school ineffectiveness, school decline, and processes that kill school improvement efforts". Algunos de esos pocos estudios son los siguientes: Hernández-Castilla, Murillo y Martínez-Garrido (2013); Van De Grift, y Houtveen, (2006; 2007).

La aproximación a estos centros debe realizarse con la máxima delicadeza y extremando los estándares y normas deontológicas. No hay que olvidar nunca que una cosa

es que, con respecto a las puntuaciones de las EDs, su nivel de eficacia sea bajo, y otra muy distinta considerarlos como centros de baja eficacia en general. En este texto, y por razones de economía expresiva, se emplea esta notación, pero ha de quedar claro (y además se dispone de abundante evidencia al respecto) que un centro escolar puede ser muy eficaz en estas competencias y hacerlo muy mal en otras o en otros aspectos de la labor formativa (por ejemplo, el cuidado y mejora de la convivencia dentro del centro). Y de la misma manera pero al revés, un centro puede tener una eficacia baja en matemáticas o lengua y trabajar excelentemente las competencias artísticas, por citar dos ejemplos reales.

Dicho lo cual, y volviendo sobre el modelo estadístico, el procedimiento concreto seguido fue el siguiente:

En primer lugar, para cada curso, competencia y año se obtiene un modelo HLM que incorpora como covariables, las variables contextuales de N1 y N2 que en cada caso hayan resultado estadísticamente significativas. En dichos modelos se dejan variar *solo* las intersecciones. Y posteriormente, para cada centro se calcula su residuo de nivel 2 (u_{0j}) en cada competencia y año.

De esta forma se dispone para cada centro de 15 medidas (3 residuos en cada uno de los 5 años). A continuación se promedian primero los 3 residuos de cada año, y luego los 5 promedios. Luego se ordena la lista de centros y se seleccionan los casos extremos de forma que los centros de mayor residuo promedio serán considerados los más eficaces (CAEFs) y los de menor serán los más ineficaces (CBEFs).

El factor tiempo suele conllevar el hecho de que este tipo de puntuaciones de valor añadido, estos residuos, en ocasiones adolecen de una cierta inestabilidad (Newton, Darling-Hammond, Haertel y Thomas, 2010). Por tanto, son relativamente frecuentes situaciones en las que el residuo de un centro un año es muy alto y positivo, pero al siguiente pasa a ser simplemente medio. Por ello, para contrarrestar esta inestabilidad, se procedió a promediar para cada centro su conjunto de residuos.

Esto supone que el foco de atención se pone en aquellos centros cuyos resultados son muy notables en *todos los años y en todas las competencias*. Esto incrementa la robustez de los resultados y, por ende, del modelo pero conlleva que los valores promedio no sean exageradamente grandes.

Considerando este procedimiento de detección de centros *tanto de alta como de baja eficacia*, conviene renombrar los dos criterios hasta ahora abordados de forma que el criterio 1 es el de las *puntuaciones extremas* y el segundo el de los *residuos extremos*.

Criterios de eficacia basados en el crecimiento

Dos párrafos más arriba se ha introducido una cuestión de gran relevancia como es el factor tiempo. Se ha dispuesto de los datos de 5 aplicaciones de la ED (2009, 2010, 2011, 2013 y 2015). Por tanto, cabe plantearse un enfoque centrado en el cambio, en el crecimiento. No se trata obviamente de un estudio longitudinal *stricto sensu* a nivel de los estudiantes porque las EDs, al aplicarse en 4º de EP y 2º de ESO, suponen diferentes cohortes. Pero sí es posible pensar en un enfoque temporal al nivel de los centros ya que se dispone de 5 puntos temporales. Estas comparaciones temporales son posibles de manera directa en la medida en que todas las EDs tienen la misma métrica (media

de 250 y desviación típica de 50) y las pruebas disponen de ítems de anclaje lo que permite la comparabilidad. Un enfoque similar, aunque en el campo de la Psicología clínica es el adoptado en el trabajo de Spoth, Redmond, Shin y Azevedo (2004).

En las gráficas de la figura 1, se observa que centro de la línea verde muestra una clara tendencia creciente en las puntuaciones medias obtenidas, mientras que el centro rojo tiene la tendencia inversa.

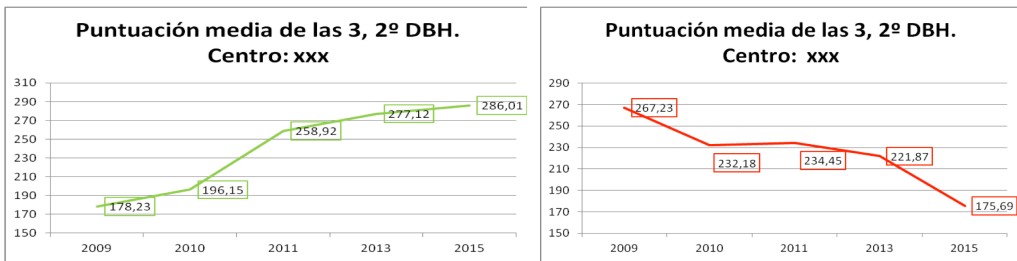


Figura 1. Ejemplos de crecimiento (o decrecimiento) de puntuaciones.

El centro verde es un centro que en 2009 comenzó con una puntuación media extremadamente baja y a lo largo de las diferentes ediciones de las EDs ha ido mejorando notablemente sus puntuaciones alcanzando en 2015 una puntuación de 286. Con este perfil, no hubiera sido considerado de alta eficacia bajo ninguno de los dos criterios anteriormente expuestos.

Pero esta clara tendencia a la mejora debe ser reconocida. Se trata de un centro que ha demostrado que es capaz de mejorar notablemente su rendimiento medio a lo largo del tiempo. Y esto ha de ser considerado como otra manera de ser altamente eficaz. Máxime si se piensa que el objetivo final es la mejora de los centros. Y centros como éste *ya han demostrado* que son capaces de hacerlo. Éste es el tercer criterio: el del *crecimiento (o decrecimiento) de las puntuaciones*.

Para describir el cuarto y último criterio, la figura 2 muestra el perfil de un centro cuyas puntuaciones obtenidas (línea amarilla) muestran una clara tendencia creciente. La puntuación media en 2009 es de 220,32 y la serie acaba con los 276,34 puntos de 2015. Pero lo que hace muy interesante a este centro es la línea roja de las puntuaciones estimadas, pues la misma se mantiene prácticamente paralela al eje de las abscisas con valores para los 5 años en torno a los 245 puntos. Esto ha de interpretarse en el sentido de que el contexto de este centro se mantiene estable. Pero lo importante emerge al comparar ambas líneas ya que lo que se detecta es que al crecer las puntuaciones obtenidas manteniéndose más o menos constante el contexto (las puntuaciones estimadas), los residuos muestran una acusada tendencia creciente.

En 2009 el residuo era negativo (alrededor de -27 puntos). En 2010 sigue siendo negativo pero su magnitud es mucho menor (alrededor de -8). Y ya en la ED de 2011 la tendencia se invierte y el residuo pasa a ser positivo (+6). En 2013 alcanza un valor de +24 y la serie finaliza en 2015 con un residuo de +32 puntos.

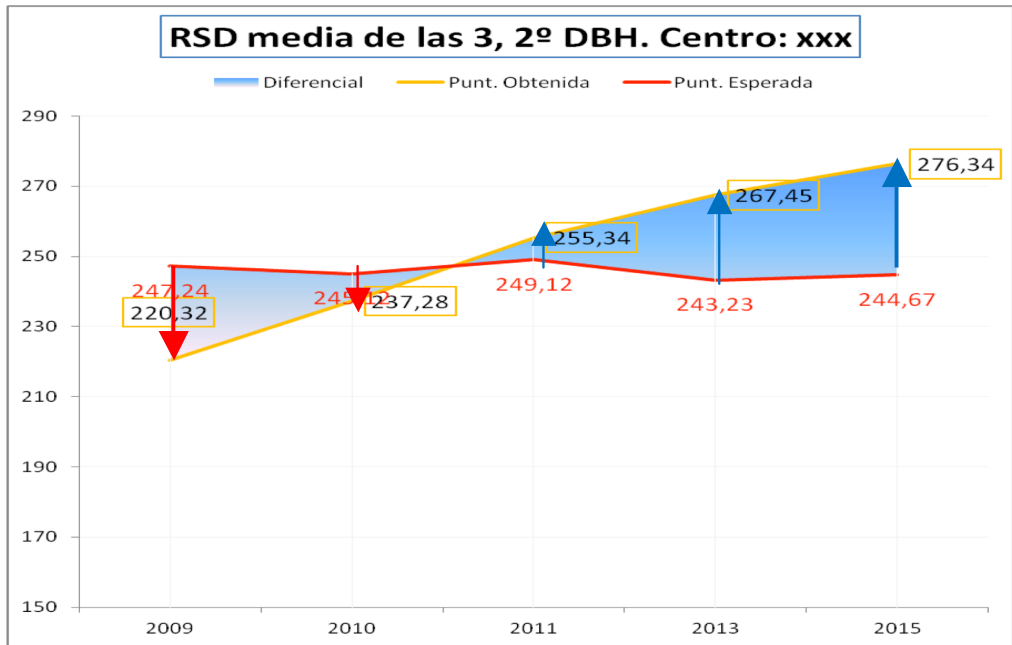


Figura 2. Ejemplo 1 de crecimiento de residuos.

Por tanto, este centro sí puede ser considerado de alta eficacia desde los dos criterios: crecen las puntuaciones y *también crecen los residuos (cuarto criterio)*. Es un centro con una clara tendencia a la mejora. Pero el ejemplo del centro de la figura 2 no es la única posibilidad.

En la figura 3 aparece representado un centro cuyas puntuaciones obtenidas se mantienen más o menos constantes a lo largo de las 5 EDs. Si solo se tuvieran en cuenta sus puntuaciones brutas, este centro no entraría en ninguna de las selecciones de alta eficacia. Pero si, como siempre, se comparan las puntuaciones obtenidas con las estimadas, la cosa cambia.

La línea roja de las puntuaciones estimadas muestra una clara tendencia *decreciente*, es decir el contexto de este centro ha ido empeorando con el paso del tiempo (debido muy probablemente a los efectos de la crisis). Lo que esto supone es que los residuos se van incrementando año a año: el de 2009 es de casi 0, de 9 puntos en 2010, 16 en 2011, 35 en 2013 y 36 en 2015.

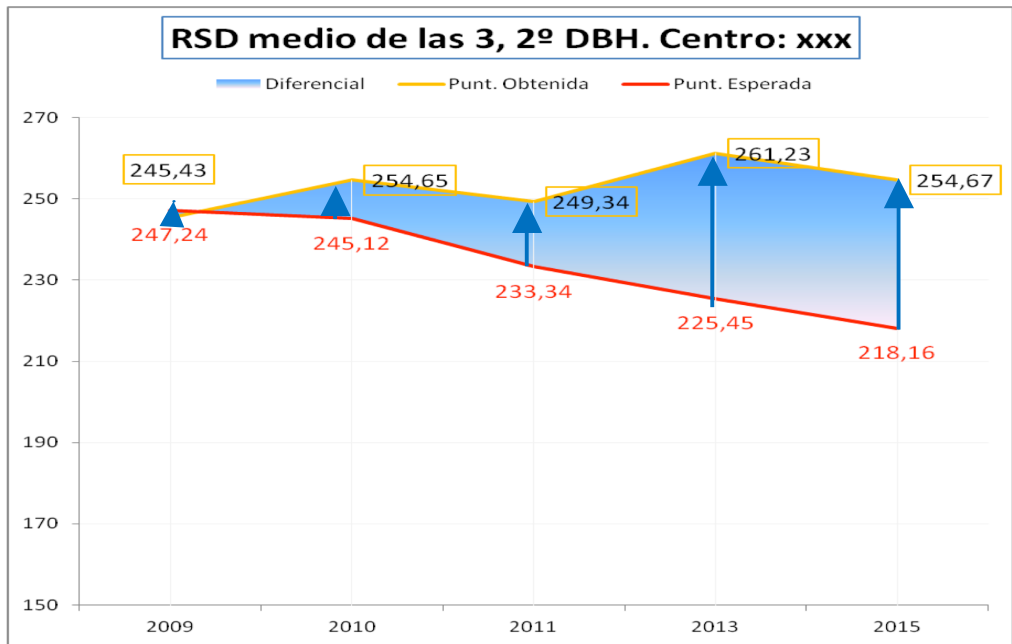


Figura 3. Ejemplo 2 de crecimiento de residuos.

Es decir, este centro consigue mantener un nivel bastante aceptable en lo que a las puntuaciones se refiere, a pesar de que el contexto en el que desarrolla su labor se va haciendo cada vez más complicado y difícil. En este caso nos encontramos también con un caso claro de crecimiento de residuos (flechas azules), aunque el mismo refleja una situación socioeconómica bien distinta. Se trata, por tanto, también de un centro de alta eficacia según el *criterio 4 de crecimiento de residuos*.

En conclusión, el enfoque longitudinal permite definir dos nuevos criterios de eficacia escolar:

- Crecimiento de puntuaciones: un centro se considera de alta eficacia si el promedio de sus puntuaciones muestra a lo largo de los 5 años una clara tendencia creciente y si *además* sus residuos crecen también (y se considera de baja eficacia si la tendencia es decreciente).
- Crecimiento de residuos: un centro se considera de alta eficacia si el promedio de sus residuos muestra a lo largo de los 5 años una clara tendencia creciente (y se considera de baja eficacia si la tendencia es decreciente).
- El procedimiento estadístico empleado tiene como objetivo detectar los centros que más crecen (y decrecen) bien en puntuaciones, bien en residuos, y para ello se emplea de nuevo un enfoque multinivel mediante HLM. Estos modelos son una herramienta muy versátil, y si antes se formuló una estructura jerárquica de datos de forma que los estudiantes (N1) se anidaban en centros (N2), ahora se puede emplear un enfoque similar considerando otra estructura jerárquica

de datos en la que ahora sean los tiempos (años de aplicación de las EDs) los que conformen el nivel 1 (N1), y a su vez, estos se aniden en centros (N2). En el capítulo 6 de la obra de Raudenbush y Bryk (2002), y en los capítulos 5 y 6 de la de Heck y colaboradores (2014), se puede encontrar información detallada al respecto.

El modelo es el que aparece en la tabla 1:

Tabla 1

Modelos jerárquicos para los criterios longitudinales

	Para los residuos	Para las puntuaciones
N1 (tiempos)	$RSD_{ij} = \pi_{0j} + \pi_{1j} A\tilde{N}O_{ij} + e_{ij}$	$PUNT_{ij} = \pi_{0j} + \pi_{1j} A\tilde{N}O_{ij} + e_{ij}$
N2 (centros)	$\pi_{0j} = \beta_{00} + u_{0j}$ $\pi_{1j} = \beta_{10} + u_{1j}$	$\pi_{0j} = \beta_{00} + u_{0j}$ $\pi_{1j} = \beta_{10} + u_{1j}$

En este caso se estima la variable dependiente -la puntuación media o el residuo medio- en cada uno de los 5 años (t) para cada centro (j). La misma se estima mediante dos parámetros π_{0j} y π_{1j} . La notación aquí es distinta y se suele usar la letra pi, pero estos parámetros son los equivalentes a los de los modelos anteriores de forma que:

- B_{00} es la gran media (intersección) de todos los centros en puntuaciones o residuos al inicio (año 0, ED de 2009).
- B_{10} es la pendiente media de todos los centros.
- U_{0j} es el residuo de cada centro relativo a la intersección.
- U_{1j} es el incremento anual en pendiente de cada centro.

Aquí se ha destacado el parámetro U_{1j} pues es lo que la pendiente de cada centro (π_{1j}) se aparta de la gran pendiente (de la pendiente media de todos los centros, B_{10}). Es ahora el estimador de eficacia, de la misma manera que en el caso de los residuos extremos lo era U_{0j} . Este último indicaba lo que la media de cada centro se apartaba de la gran media. Aquí el razonamiento es análogo pero referido a las pendientes: como el criterio es el crecimiento, el indicador de eficacia es ahora la diferencia entre la pendiente de cada centro y la gran media de todas las pendientes.

Y al igual que en casos anteriores, una vez ejecutado el modelo, se guarda para cada centro su parámetro U_{1j} (uno para los residuos y otro para las puntuaciones) y se ordena la relación de centros en función de los mismos. Por último, se seleccionan los casos extremos por arriba y por abajo que serán respectivamente los CAEFs que más crezcan en residuos o puntuaciones, y los CBEFs que más decrezcan.

Discusión y conclusiones

Como se ha expuesto, hay distintos enfoques a la hora de considerar a un centro como de alta o baja eficacia. En este estudio se han tenido en cuenta 3 ejes: los tiempos, las puntuaciones y los residuos. La combinación de los mismos ha dado lugar a los 4 criterios de eficacia (o ineficacia) analizados:

- Criterio 1: de las puntuaciones extremas (modelo de efecto *techo* o *suelo*).
- Criterio 2: de los residuos extremos (modelo transversal contextualizado).
- Criterio 3: del crecimiento (o decrecimiento) de las puntuaciones (modelo longitudinal).
- Criterio 4: del crecimiento (o decrecimiento) de los residuos (modelo longitudinal contextualizado).

Pero aunque como ha sido dicho, estos cuatro criterios representan cuatro maneras o enfoques de considerar la eficacia de un centro, no es menos cierto que no son todos igualmente relevantes. Desde el punto de vista empleado en estos proyectos, los modelos contextualizados son más interesantes porque implican una selección de centros más equitativa. Por ello, los centros seleccionados según estos criterios son más variados en lo que a sus tipologías y características se refiere.

Hasta aquí es donde *llega* la estadística. Para tratar de conocer en profundidad los factores explicativos de tales resultados se requiere una aproximación a la realidad y praxis cotidiana de estos centros, y para ello se emplearon técnicas cualitativas. Del contraste entre las buenas prácticas de los centros de alta eficacia (CAEF) y los problemas, dificultades y eventuales malas prácticas de los de baja (CBEF) emerge una relación comprensiva de factores asociados a la eficacia y la mejora escolar que facilita el diseño y puesta en marcha de programas e intervenciones de mejora escolar más precisos y focalizados.

Para ello, se ha llevado a cabo un proceso muy robusto y fiable de detección de centros, realizado a partir de un estudio censal y atendiendo a diversidad de facetas y perspectivas de lo que se considera alta (o baja) eficacia escolar tomando en cuenta tanto la equidad como la excelencia, asumiendo que en un sistema educativo de calidad éstas no deben ser categorías contrapuestas, sino que se puede y se debe tratar de incrementarlas simultáneamente mediante procesos de mejora continua.

Referencias

- Aitkin, M. & Longford, N. (1986). Statistical Modelling Issues in School Effectiveness Studies. *Journal of the Royal Statistical Society*, 149(1), p1-43.
- Angulo, A. & Lizasoain, L. (2019). La eficacia escolar en el País Vasco. *Revista Dyle. Revista sobre Dirección y Liderazgo Educativo*. Nº 2.
- Azkarate, A., Bartau, I., Aierbe, A., & Oregui, E. (2018). Medios Digitales y Eficacia Escolar en Educación Primaria. En I. Aguaded, R. Mancinas-Chávez, & P. de-Casas-Moreno (Eds.), *Comunicación móvil y generación smartphone: Retos y prospectivas*. Libro

- resúmenes del III Congreso Internacional Comunicación y Pensamiento (pp. 79). Sevilla: EGREGIUS.
- Azkarate, A., Bartau, I., & Lizasoain, L. (2019a). Hábitos y valores del alumnado en centros de primaria de alta eficacia escolar. *Estudios sobre Educación*, 37, 199-222. DOI: 10.15581/004.37.199-222.
- Azkarate, A., Bartau, I., & Lizasoain, L. (2019b). Educación en valores y formación del profesorado de secundaria: el caso de la UPV/EHU. *Profesorado. Revista de Currículum y Formación del Profesorado*, 23(3), 493-516. doi: 10.30827/profesorado.v23i3.11240
- Azpillaga, V., Intxausti, V., & Joaristi, L. (2014). Implicación de las familias en los centros escolares de alta eficacia en la Comunidad Autónoma Vasca. *Bordón*, 66(3), 27-37. doi: 10.13042/Bordon.2014.66302
- Bartau Rojas, I., Azpillaga Larrea, V., & Joaristi Olariaga, L. M. (2017). Metodología de enseñanza en centros eficaces de la Comunidad Autónoma del País Vasco. *Revista De Investigación Educativa*, 35(1), 93-112. doi: 10.6018/rie.35.1.225141
- Clarke, P., Crawford, C., Steele, F., & Vignoles, A. (2010). The choice between fixed and random effects models: Some considerations for educational research, *Institute for the Study of Labor (IZA). Discussion Papers*, 5287.
- Coleman, J. S., et al. (1966). *Equality of Educational Opportunity*. Washington D.C.: UD Dept. of Health, Education and Welfare. National Center for Educational Statistics. Us Gov. Printing Office. doi: 10.3886/ICPSR06389.v3
- De Maeyer, S., Van den Bergh, H., Rymenans, R., Van Petegem, P., & Rijlaarsdam, G. (2010). Effectiveness criteria in school effectiveness studies: Further research on the choice for a multivariate model. *Educational Research Review* 5, 81–96. doi: <https://doi.org/10.1016/j.edurev.2009.09.001>
- De la Orden, A. (1999). Los modelos jerárquicos lineales y su aplicación a la investigación educativa. *Revista de Investigación Educativa*, 17(2), 425-427.
- Etxeberria, F., Intxausti, N., & Azpillaga, V. (2017). School Climate in Highly Effective Schools in the Autonomous Region of the Basque Country (Spain). *REICE: Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 15(4), 5-26. doi: 10.15366/reice2017.15.4.001
- Gaviria, J.L. (1999). El enfoque multinivel en la evaluación de sistemas educativos. *Revista de Investigación Educativa*, 17(2), 429-443.
- Gaviria, J. L., & Castro, M. (2005). *Modelos jerárquicos lineales*. Madrid: La Muralla.
- Goldstein, H. (1997). Methods in School Effectiveness Research, *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, 8(4), 369-395, doi: 10.1080/0924345970080401
- Heck, R. H., Thomas, S. L., & Tabata, L. N. (2014). *Multilevel and longitudinal modeling with IBM SPSS*. New York: Taylor & Francis Group, Routledge.
- Hernández-Castilla, R., Murillo, F. J., & Martínez-Garrido, C. (2013). Factores de ineficacia escolar. *REICE*, 12(1), 103-118.
- Intxausti, N., Etxeberria, F., & Bartau, I. (2016). Effective and inclusive schools? Attention to diversity in highly effective schools in the Autonomous Region of the Basque Country. *International Journal of Inclusive Education*, 21(1), 14-30. doi: 10.1080/13603116.2016.1184324
- Intxausti, N., Joaristi, L., & Lizasoain, L. (2016). Educational leadership as best practice in highly effective schools in the Autonomous Region of the Basque Country

- (Spain). *Educational Management Administration & Leadership*, 44(3) 397–419. doi: 10.1177/1741143214558570
- Jencks, C. S., Smith, M., Ackland, H., Bane, M. J., Cohen, D., Gintis, H., & Michelson, S. (1972). *Inequality: A reassessment of the effect of the family and schooling in America*. New York, NY: Basic Books.
- Joaristi, L., Lizasoain, L., & Azpillaga, V. (2014). Detección y caracterización de los centros escolares de alta eficacia de la Comunidad Autónoma del País Vasco mediante Modelos Transversales Contextualizados y Modelos Jerárquicos Lineales. *Estudios Sobre Educación*, 27, 37-61. doi: 10.15581/004.27
- Lizasoain, L., Bereziartua, J., & Bartau, I. (2016). La formación permanente del profesorado en centros educativos de alta eficacia. *Bordón*, 68(2), 199-218. doi: 10.13042
- Lizasoain, L., & Angulo, A. (2014). Buenas prácticas de escuelas eficaces del País Vasco. Metodología y primeros resultados. *Participación Educativa. Segunda época*, 3(4), 17-28.
- Martínez Arias, R. (2009). Usos, aplicaciones y problemas de los modelos de valor añadido en educación. *Revista de Educación*, 348, 217-250.
- Martínez-Abad, F., Chaparro, A. A., & Lizasoain, L. (2014). The Socioeconomic Index in the Analysis of Large-Scale Assessments: Case Study in Baja California (Mexico). En F. García Peñalvo (Ed.) *Proceedings TEEM' 14. Technological Ecosystems for Enhancing Multiculturality* (461-467). Salamanca: ACM. doi: 10.1145/2669711.2669940
- Martínez Abad, F., Lizasoain Hernández, L., Castro Morera, M., & Joaristi Olariaga, L. (2017). Selección de escuelas de alta y baja eficacia en Baja California (México). *REDIE, Revista Electrónica de Investigación Educativa*, 19(2), 38-53. doi: 10.24320/redie.2017.19.2.960
- Murillo, F.J. (1999). Los modelos jerárquicos lineales aplicados a la investigación sobre eficacia escolar. *Revista de Investigación Educativa*, 17(2), 453-458.
- Murillo, F.J. (2005). *La investigación sobre eficacia escolar*. Barcelona: Octaedro.
- Murillo Torrecilla, F. J., & Hernández Castilla, R. (2011). Efectos escolares de factores socio-afectivos. Un estudio Multinivel para Iberoamérica. *Revista de Investigación Educativa*, 29(2), 407-427.
- Newton, X., Darling-Hammond, I., Haertel, E., & Thomas, E. (2010). Value-Added Modeling of Teacher Effectiveness: An exploration of stability across models and contexts. *Educational Policy Analysis Archives*, 18(23), 1-23. doi: <https://doi.org/10.14507/epaa.v18n23.2010>
- OCDE (2011). *La medición del aprendizaje de los alumnos: Mejores prácticas para evaluar el valor agregado de las escuelas*. OECD Publishing.
- Oregui, E., Aierbe, A., Bartau, I., & Azkarate, A. (2018). Dispositivos Móviles y Buenas Prácticas en Educación Secundaria. En I. Aguaded, R. Mancinas-Chávez, & P. de Casas-Moreno (Eds.), *Comunicación móvil y generación smartphone: Retos y perspectivas. Libro resúmenes del III Congreso Internacional Comunicación y Pensamiento* (pp. 44). Sevilla: EGREGIUS.
- Padilla, L., Guzmán, C., Lizasoain, L., & García-Medina, A. M. (2018). Eficacia escolar y aspiraciones educativas en el bachillerato. Un estudio longitudinal contextualizado en Aguascalientes. *Revista Mexicana de Investigación Educativa RMIE*, 23(78), 687-709.
- Palardy, G. J. (2008). Differential school effects among low, middle, and high social class composition schools: a multiple group, multilevel latent growth curve analysis.

- School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, 19(1), 21-49. doi: 10.1080/09243450801936845
- Pedroza, H., Peniche, R., & Lizasoain, L. (2018). Criterios para la identificación y selección de escuela eficaces de nivel medio superior. *Revista Electrónica de Investigación Educativa*, 20(1), 14-25. <https://doi.org/10.24320/redie.2018.20.1.2170>
- Raudenbush, S. W. & Bryk, A. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. London: Sage.
- Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C. y Stringfield, S. (2014). Educational effectiveness research (EER): a state-of-the-art review, *School Effectiveness and School Improvement*, 25(2), 197-230, doi: 10.1080/09243453.2014.885450
- Spoth, R., Redmond, C., Shin, C., & Azevedo, K. (2004). Brief Family Intervention Effects on Adolescent Substance Initiation: School-Level Growth Curve Analyses 6 Years Following Baseline. *Journal of Consulting and Clinical Psychology*, 72(3), 535-542
- Strand, S. (2010). Do some schools narrow the gap? Differential school effectiveness by ethnicity, gender, poverty, and prior achievement. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, 21(3), 289-314, doi: 10.1080/09243451003732651
- Townsend, T., & Avalos, B. (2007). *International handbook of school effectiveness and improvement*. Berlin: Springer.
- Van De Grift, W.J.C.M., & Houtveen, A.A.M. (2006). Underperformance in primary schools. *School Effectiveness and School Improvement*, 17(3), 255-273.
- Van De Grift, W.J.C.M., & Houtveen, A.A.M. (2007). Weaknesses in Underperforming Schools. *International Perspectives in At-Risk Education Research*, 12(4), 383-403.
- Willms, J.D. (2010). School Composition and Contextual Effects on Student Outcomes. *Teachers College Record*, 112(4), 2010, 1008-1037

Fecha de recepción: 9 de marzo de 2020.

Fecha de revisión: 10 de marzo de 2020.

Fecha de aceptación: 23 de marzo de 2020.

