
Incidencia de la forma de evaluar los docentes de Educación Primaria en el rendimiento de los estudiantes en España

The Impact of the Evaluation Methodology of the Primary Education Teachers on the Performance of the Student in Spain

F. JAVIER MURILLO

Universidad Autónoma de Madrid
javier.murillo@uam.es

CYNTHIA MARTÍNEZ-
GARRIDO

Universidad Autónoma de Madrid
cynthia.martinez@uam.es

NINA HIDALGO FARRAN

Universidad Autónoma de Madrid
nina.hidalgo@uam.es

Resumen: La forma en cómo se evalúa a los estudiantes en el aula es un elemento fundamental que incide directamente en su desarrollo cognitivo y socio-afectivo. Esta investigación busca determinar la relación entre las actividades de evaluación que llevan a cabo los docentes de Educación primaria de escuelas españolas y el desarrollo en los alumnos de las siguientes competencias: Lingüística, Matemática, Científica, Social y Ciudadana. Para alcanzarlo se realiza una explotación especial de los datos de la Evaluación de Diagnóstico General (EDG) de 2009 a 28.708 estudiantes, 1.341 docentes, 887 centros españoles mediante el uso de Modelos Multinivel de tres de niveles. Los resultados muestran que los estudiantes cuyos docentes realizan en mayor medida controles después de cada tema o lección, corrigen los deberes y cuadernos y valoran el interés y la participación en clase de los estudiantes obtienen mejores resultados. Y, por el contrario, se ha encontrado una relación inversa entre realizar controles escritos trimestrales y el desarrollo de las competencias de los estudiantes.

Palabras clave: evaluación; actividades de evaluación; desarrollo de competencias; enseñanza eficaz; educación primaria; España.

Abstract: The way teachers assess in classroom impact over cognitive and socio-affective development of the students. This paper aims to determine the relationship between the teacher's evaluation to their students, and the academic achievement of their students in four specific areas: Linguistic Competence, Mathematical Competence, Scientific Competence, and Social and Citizenship Competence. To reach our aim, this paper made a special analysis with the data of the Spanish General Diagnostic Evaluation -2009- to 28,708 students, and 1341 teachers of 887 schools. The use of Multilevel Model Analysis with three levels shows the impact of each evaluation strategy on the four competencies studied. Our study confirms those teachers who prepare exams after each lesson, give homework, review notebooks, and value the interest and participation of students in classroom do better. On the other side, we have found an inverse relationship between prepare quarterly exams and the skill development of the students.

Keywords: teaching effectiveness; assessment; primary education; Spain.

INTRODUCCIÓN

Afirmar que las estrategias de evaluación que el docente utiliza en el aula tienen una incidencia en el aprendizaje de sus estudiantes no parece aportar muchas ideas novedosas. Efectivamente, la investigación que busca determinar qué hacen los docentes que consiguen que sus alumnos aprendan más de lo que sería previsible, la investigación sobre enseñanza eficaz ha sido clara en determinar que la forma en que el docente evalúa a sus alumnos, así como su frecuencia y la manera de devolver la información son factores clave asociados al desempeño escolar (Hunt, Wiseman y Touzel, 2009; Muijs y Reynolds, 2000; Orlich, Harder, Callahan, Trevisany y Brown, 2010; Walberg y Paik, 2000).

A pesar de ello, y frente a voces que piden sistemas de evaluación alternativos (p.e. Baker, 2010; Huerta-Macías, 1995; Mateo y Martínez Olmos, 2005), se generaliza la idea de que el simple hecho de evaluar, con la preocupación casi exclusiva de que se garantice la “calidad” psicométrica de los instrumentos, mejorará automáticamente los resultados de los estudiantes. De esta forma, parece ya no importar el qué, el cómo, el cuándo o el para qué se evalúa, parece solo importar que se haga “objetivamente”.

En ese contexto, es necesario dar un paso atrás y volver a mirar cómo evalúan los y las docentes de Educación Primaria en España, y verificar si esa forma de evaluar tiene incidencia en el desempeño de los estudiantes. De esta manera se estará en condiciones de aportar ideas que hagan que se desarrollen sistemas de evaluación que logren realmente incidir en el desarrollo de los estudiantes.

En esta investigación se realizará un Análisis Multinivel de la base de datos de la Evaluación General Diagnóstica de Educación Primaria del año 2009 para determinar qué prácticas educativas tienen incidencia en el aprendizaje de los estudiantes en cuatro materias diferentes.

REVISIÓN DE LA LITERATURA

La investigación sobre Eficacia Escolar, así como sobre Enseñanza Eficaz, ha determinado que las estrategias de evaluación y seguimiento de los estudiantes es un factor que incide de una forma clara directa y medible en el desempeño académico. Efectivamente, ya en el primer estudio sobre Eficacia Escolar desarrollado por Weber (1971), se encontró que la evaluación de los estudiantes era uno de los factores clave que discriminaba entre las buenas escuelas y las malas. Desde entonces, cientos de trabajos han encontrado análogas conclusiones. Así, por ejemplo,

Edmonds (1979) incluyó la evaluación frecuente como uno de los elementos clave que incidían en el aprendizaje de los estudiantes en su famoso “Modelo de 5-factores” que resumía la primera década de investigación sobre eficacia escolar.

A partir de ese momento, diferentes revisiones de estudios llegaron a análogas conclusiones. De los estudios realizados en la década de los 70 dan cuenta Medley y Crook (1980). Tras una concienzuda revisión de más de 600 estudios, estos autores determinaron que la supervisión del trabajo individual de los estudiantes era uno de los cinco comportamientos propios de los docentes eficaces. Black y Wiliam (1998), por su parte, revisaron los trabajos de la década de los 80, y encontraron que la evaluación de los estudiantes, tanto en Educación Primaria como en Secundaria, aumenta los logros de los estudiantes cuando:

- 1 los resultados de la evaluación se utilizan para adaptar el contenido de la enseñanza a los estudiantes,
- 2 los estudiantes reciben retroalimentación que les lleva a mejorar su trabajo,
- 3 tanto docentes como estudiantes conocen los objetivos del aprendizaje y las criterios de evaluación,
- 4 son los propios estudiantes los que pueden evaluarse (auto-evaluación y evaluación entre pares), y
- 5 los estudiantes toman una postura activa en la evaluación.

Teddlie y Reynolds (2000), y posteriormente Sutton (2009), complementaron las características señaladas, apuntando a las razones de los efectos positivos de la evaluación sobre el desarrollo de los estudiantes:

- Genera efectos positivos para el alumno en su autoestima y motivación, dado que siente que el profesor está interesado en su progreso,
- Ayuda a que docentes y estudiantes se centren en los objetivos fundamentales que están siendo evaluados,
- Permite ser utilizado para analizar el progreso del estudiante.
- Contribuye a la mejora de la intervención, aportando datos para valorar su impacto y calidad.

De esta forma, el "para qué evaluar", la finalidad que persiga es un elemento que afecta al rendimiento de los estudiantes. Así lo comprueba el estudio realizado por Fuchs y Fuchs (1986) en el que, a través de un meta-análisis de 21 estudios previos elaborados con estudiantes de Educación Infantil, Primaria y Secundaria, muestra

que en aquellos casos en los que se diseñaba la evaluación con el fin de obtener, sintetizar e interpretar información para facilitar la toma de decisiones orientadas a ofrecer retroalimentación a los estudiantes se genera una mejora estadísticamente significativa sobre su rendimiento académico.

El trabajo elaborado por Crooks (1988) destaca otro elemento clave: la planificación de la evaluación. Así, a través de la revisión de 14 investigaciones previas realizadas a estudiantes de Educación Primaria, Crooks encuentra que la influencia de la evaluación sobre el logro académico, la motivación y el desarrollo de estrategias de aprendizaje en los estudiantes no siempre es positiva, depende de cuán concienzuda haya sido planificada e implementada la evaluación de aula.

La frecuencia con la que se evalúa el progreso de los estudiantes es también un factor que incide en su aprendizaje. Bangert-Drowns, Kulik y Kulik (1991) muestran, a través de un meta-análisis en el que revisa 35 estudios, la relación directa entre la frecuencia de evaluación y la mejora del desempeño de los estudiantes de Educación Primaria. Concretamente, encuentran los efectos positivos de las evaluaciones continuas, aunque también señalan que el aumento de la frecuencia de evaluación no es proporcional al rendimiento, sino que crece muy levemente.

La investigación sobre la estrategia de evaluación más adecuada, sin embargo, no ha dado resultados tan concluyentes, quizá simplemente porque no existe esa "más adecuada". Las investigaciones de Muijs y Reynolds (2001) y de House y Howe (2003), entre otras, sostienen que para potenciar al máximo el desarrollo de los estudiantes y contribuir a la equidad en el aula es preferible recurrir a la combinación de estrategias de evaluación más que al uso de una determinada estrategia en particular. Esa visión de la equidad en el aula es especialmente interesante, dado que no sólo interesa que aprendan los alumnos, sino que lo hagan todos. Resultado análogos fueron hallados por Weurlander, Söderberg, Scheja, Hult y Wernerson (2012). En esa lógica, Stiggins y DuFour (2009) defienden que la elaboración de un sistema evaluador que se ajuste a todos los niveles de aprendizaje de todos los estudiantes requiere de cuatro condiciones: i) diseñada en torno a objetivos claros de aprendizaje; ii) comprometida con la enseñanza basada en el logro de los estándares de aprendizaje; iii) que represente con fiabilidad los objetivos de aprendizaje, y iv) que sus resultados se entreguen de manera oportuna y comprensible.

Un último elemento abordado por la investigación y que queremos destacar es la incidencia de la evaluación no solo en los resultados cognitivos, también en los resultados socio-afectivos de los estudiantes. El reciente estudio longitudinal de cinco años de duración realizado por Kyriakides, Creemers y Antoniou (2009) con

2.505 estudiantes, de 108 aulas localizadas en 50 escuelas de Educación Primaria, clasifica en cinco tipos los comportamientos de los docentes en función de las variaciones encontradas en los resultados escolares (cognitivos y afectivos) de los estudiantes. Una de las principales conclusiones de este estudio es que la evaluación de aula impacta no sólo sobre el desarrollo académico de los estudiantes, sino también en su desarrollo afectivo y psicomotor.

En España, un estudio del Instituto Nacional de Evaluación Educativa (2007) identifica diez factores de la práctica docente asociados al desempeño de los estudiantes. Entre ellos destaca el uso de diferentes formas de evaluación del alumnado de Educación primaria. Así, los docentes den mucha importancia a la evaluación de la alumna o alumno por sus compañeros de grupo, a la autoevaluación del alumnado y a las pruebas escritas sobre un tema es un factor asociado al rendimiento de los estudiantes. Concretamente llega a explicar el 3% de la varianza. Otros factores relacionados son la Programación de la planificación y evaluación (explican el 6% de la varianza), la revisión de cuadernos y trabajos realizados en casa o en clase y a la corrección individualizada de las actividades o ejercicios (con el 5% de la varianza); el uso del libro de texto y corrección en grupo (el 4% de la varianza) y la observación de las intervenciones del alumnado (5%). De esta forma, se valida la incidencia de la evaluación en el aprendizaje de los estudiantes.

Como complemento de estos datos, en un reciente trabajo elaborado por Vázquez Cano (2012) en 110 centros de Primaria y de Secundaria se encuentra que el 96,5% del profesorado únicamente realiza exámenes de contenido teórico, y que el 92% no devuelve al alumnado una nota desglosada del grado de adquisición de las competencias en la evaluación final. No obstante, los resultados también muestran que el 92% del profesorado no le otorga el peso final de la asignatura a una sola prueba de evaluación, sino que establece porcentajes a las diferentes actividades o instrumentos de evaluación que utiliza en la materia.

Con todo esto se concluye que aunque son numerosas las investigaciones que han evidenciado el impacto de la evaluación sobre el rendimiento académico de los estudiantes a nivel internacional, no son tantos los trabajos que aportan datos a este respecto dentro del panorama nacional español. Esta investigación lleva a cabo un minucioso estudio que busca responder a la siguiente pregunta: ¿las diferentes actividades de evaluación que desarrollan los docentes de Primaria de escuelas españolas inciden en el desarrollo de las competencias académicas del estudiante? ¿qué actividades son las que influyen y qué capacidades son influidas?

METODOLOGÍA

Con esta investigación se pretende determinar la relación entre la forma de evaluar por parte de los docentes de Educación Primaria en España y el desempeño académico de sus estudiantes en cuatro áreas: Competencia Lingüística, Competencia Matemática, Competencia Científica, y Competencia Social y Ciudadana.

Para alcanzar este objetivo se realizó una explotación especial de los datos de la Evaluación de Diagnóstico General (EDG) de Educación Primaria realizada en 2009 por el actual Instituto Nacional de Evaluación Educativa, del Ministerio de Educación de España (Instituto de Evaluación, 2010). Junto con las pruebas de rendimiento en cuatro competencias básicas, se aplicaron una serie de cuestionarios de contexto que aportaron información que ayuda no sólo a contextualizar los resultados de desempeño obtenidos, también permite profundizar en el conocimiento de las escuelas, los docentes y los estudiantes españoles de Educación Primaria.

Las variables utilizadas en el estudio fueron de tres tipos:

1. *Variables de producto.* Desempeño de los estudiantes en cuatro áreas: Competencia en Comunicación Lingüística, Competencia Matemática, Competencia en el Conocimiento y la interacción con el Mundo Físico, y Competencia Social y Ciudadana. Todas ellas estimadas mediante Teoría de Respuesta al Ítem (TRI) y con una puntuación en una escala de media 500 puntos y desviación típica 100.
2. *Variables explicativas.* Forma de evaluar por parte de los docentes, compuesto por ocho variables, todas ellas son variables de escala tipificadas:
 - a. Realizar una prueba sobre conocimientos previos antes de cada tema.
 - b. Realizar controles después de cada tema o lección.
 - c. Realizar controles escritos trimestrales.
 - d. Corregir los deberes y cuadernos.
 - e. Revisar las actividades que hacen en clase.
 - f. Realizar exámenes orales.
 - g. Valorar la autoevaluación del alumnado.
 - h. Valorar el interés y la participación en clase de estudiantes.
3. *Variables de ajuste.* Características del estudiante: Índice socio-económico y cultural de la familia del estudiante (variable tipificada), Género del estudiante (hombre/mujer, variable *dummy*); y país de origen del estudiante (nativo/extranjero, variable *dummy*).

Este estudio hace una explotación especial de la base de datos completa de Evaluación de Diagnóstico General de 2009 para Educación Primaria. Ello significa que utilizó datos de 1.341 docentes de dicha etapa que enseñan en 887 centros educativos, y sus 28.708 estudiantes de 4º curso. El muestreo utilizado por el INE para su obtención, en línea con otras evaluaciones internacionales, fue estratificado por conglomerados, con una estratificación constante de aproximadamente 50 centros en cada una de las Comunidades Autónomas.

Varios fueron los instrumentos que se aplican en la EDG y fueron utilizados por esta investigación. En primer lugar, cuatro pruebas de desempeño una para cada una de las cuatro competencias evaluadas: Competencia en Comunicación Lingüística (fiabilidad de 0,861), Competencia Matemática (0,800), Competencia en el Conocimiento y la Interacción con el Mundo Físico (0,854) y Competencia Social y Ciudadana (0,856). En segundo lugar, un cuestionario dirigido a estudiantes y otro a familias, a partir del cual se elaboró el índice ISEC. Por último, se utilizó un cuestionario para docentes, del cual se obtuvieron los datos sobre su forma de evaluar.

Como estrategia de análisis de datos se utilizaron Modelos Multinivel de tres niveles (estudiante, aula y escuela), como alternativa metodológica más adecuada para abordar datos jerarquizados (Goldstein, 1995; Ma, Ma y Bradley, 2008; Murillo, 2008). Así, se realizaron cuatro de procesos de modelamiento multinivel con el software MLwiN 2.26, uno para cada una de las variables de producto, con los siguientes pasos:

1. Modelo nulo, sólo con la variable producto y el intercepto.
2. Modelo ajustado, con ISEC, género y país de origen.
3. Modelo ajustado con las variables de evaluación introducidas de forma individual.
4. Modelo final, con todas las variables incorporadas de forma simultánea, análogo al siguiente:

$$y_{ijk} = \beta_{0jk} + \beta_{1jk} ISEC_{ijk} + \beta_{2jk} Género_{ijk} + \beta_{3jk} Origen_{ijk} + \beta_{4k} Eval1_{jk} + \dots \\ \dots + \beta_{10k} Eval8_{jk} + e_{ijk}$$

$$\beta_{0jk} = \beta_0 + v_{0k} + u_{0k}$$

$$\beta_{1jk} = \beta_1 + v_{1k} + u_{1k} \dots \beta_{3jk} = \beta_3 + v_{3k} + u_{3k}$$

$$\beta_{4k} = \beta_4 + u_{4k} \dots \beta_{0k} = \beta_0 + u_{0k}$$

Con:

$$[e_{ijk}] \sim N(0, \Omega_e) : \Omega_e = [\sigma_{e_0}^2]$$

$$[u_{jk}] \sim N(0, \Omega_u) : \Omega_u = [\sigma_{u_0}^2]$$

$$[v_k] \sim N(0, \Omega_v) : \Omega_v = [\sigma_{v_0}^2]$$

Donde, para cada i= estudiante en cada j= aula, y k=escuela:

y_{ijk} es la cada una de las cuatro variables de producto: Competencia Lingüística, Competencia Matemática, Competencia Científica y Competencia Social y Ciudadana.

$ISEC_{ijk}$, nivel socio-económico y cultural de la familia del estudiante.

$Género_{ijk}$, género del estudiante (hombre/mujer).

$Origen_{ijk}$, país de origen del estudiante (nativo/inmigrante).

$Eval1_{ij}, \dots, Eval8_{ij}$, cada una de las ocho variables de evaluación del docente estudiadas.

e_{ijk} es el residual para cada estudiante.

u_{jk} el residual de cada aula.

v_k el residual de cada escuela.

RESULTADOS

Tal y como se ha señalado, el proceso de modelización necesario para estudiar la incidencia de las estrategias de evaluación sobre el desempeño de los estudiantes está conformado por cuatro fases consecutivas. En primer lugar, se estiman los cuatro modelos nulos, uno para cada una de las variables de producto (tabla 1). Los mismos sirven como línea base y posibilitan valorar la calidad del modelo.

Tabla 1. Resultados de los modelos multinivel nulos

	COMPETENCIA LINGÜÍSTICA	COMPETENCIA MATEMÁTICA	COMPETENCIA CIENTÍFICA	COMPETENCIA SOCIAL Y CIUDADANA
	β (ee)	β (ee)	β (ee)	β (ee)
PARTE FIJA				
Intercepto	500,31 (1,43)	500,78 (1,40)	500,31 (1,57)	500,49 (1,45)

PARTE ALEATORIA				
Entre escuelas	1386,35 (87,21)	1314,27 (84,92)	1779,13 (102,01)	1462,34 (89,77)
Entre aulas	36,15 (30,26)	43,12 (32,08)	0,00 (0,00)	8,53 (27,27)
Entre estudiantes	8356,74 (77,02)	8529,30 (78,62)	7958,33 (72,76)	8300,08 (76,50)

Fuente: Elaboración propia.

El segundo paso consiste en incluir en cada uno de los modelos nulos las tres variables de ajuste consideradas: ISEC, género y origen del estudiante. En la tabla 2 se presentan los resultados de su estimación. El objetivo de su introducción es controlar su efecto para que no influya en la relación que estamos analizando, es decir, garantizar la validez interna del estudio. Como muestra la tabla 2, el ISEC y el origen del estudiante influyen sobre el desempeño (el origen en tres de las cuatro variables de producto) por lo que es adecuado controlar su efecto.

Además, de estos modelos ajustados se extraen algunas ideas que, aunque alejadas de los objetivos del estudio, resultan interesantes:

- El Índice socio-económico y cultural (ISEC) incide en el desempeño de los estudiantes en las cuatro áreas estudiadas. Así, por cada desviación típica en ISEC por encima (o por debajo) de la media, estudiante obtendrá 28,5 puntos más (o menos) en Competencia Lingüística; 26,5 en Matemáticas; 26,6 en Competencia Científica, y 25,1 en Competencia Social y Ciudadana. Dado que estas variables están escaladas en una media de 500 y una desviación típica de 100, ello significa que por cada desviación típica que suba el índice ISEC de la familia del estudiante, disminuirá o aumentará más de 0,25 desviaciones típicas el desempeño del estudiante en las cuatro medidas de producto.

- El género no incide en el desempeño en ninguna de las cuatro variables analizadas (su aporte no es significativo con un $\alpha=0,05$). Es decir, no hay diferencias significativas en el desempeño de los estudiantes en ninguna de las cuatro variables estudiadas en función de su género.

- Por el contrario, sí existen diferencias significativas en función del Origen del estudiante, de tal forma que los niños y niñas que han nacido fuera de España obtienen resultados más bajos que los estudiantes nativos españoles. En concreto, los datos apuntan a que los estudiantes inmigrantes obtienen 4,11 puntos menos en el desarrollo de la Competencia Lingüística, 4,83 puntos menos en Competencia Científica y 5 puntos menos en Competencia Matemática. Así mismo, los resultados muestran que el origen de los estudiantes no influye sobre el desarrollo de la

Competencia Social y Ciudadana. Por tanto, resulta indiferente ser nativo o inmigrante para el desarrollo Social y Ciudadano del estudiante.

Tabla 2. Resultados de los modelos multinivel ajustados

	COMPETENCIA LINGÜÍSTICA β (ee)	COMPETENCIA MATEMÁTICA β (ee)	COMPETENCIA CIENTÍFICA β (ee)	COMPETENCIA SOCIAL Y CIUDADANA β (ee)
PARTE FIJA				
Intercepto	501,33 (1,16)	501,82 (1,16)	501,32 (1,38)	500,97 (1,27)
ISEC	28,48 (0,64)	26,46 (0,65)	26,58 (0,63)	25,07 (0,65)
Género (niño/niña)	NS	NS	NS	NS
Origen (nativo/ inmigrante)	-4,11 (2,05)	-5,01 (2,01)	-4,83 (2,03)	NS
PARTE ALEATORIA				
Entre escuelas	778,44 (57,35)	786,51 (58,63)	1303,19 (77,86)	1013,48 (67,49)
Entre aulas	33,52 (28,94)	37,63 (30,06)	0,00 (0,00)	7,13 (27,27)
Entre estudiantes	7857,27 (72,42)	8104,70 (74,70)	7487,14 (68,45)	7899,72 (72,81)

Nota: NS Aportación no significativa con $\alpha=0,05$.

Fuente: Elaboración propia.

De los modelos ajustados también es posible estimar el efecto escolar a través del cálculo del Coeficiente de Correlación Intraclase. El coeficiente de correlación intraclase se calcula a partir de las varianzas a nivel escuela, a nivel de aula y a nivel estudiante obtenidas en el proceso de modelización para cada una de las variables de producto según la siguiente fórmula:

$$\rho = \frac{\sigma_{v_0}^2}{\sigma_{v_0}^2 + \sigma_{\mu_0}^2 + \sigma_{\varepsilon_0}^2}$$

Dónde:

$\sigma_{v_0}^2$, varianza entre escuelas.

$\sigma_{\mu_0}^2$, varianza entre aulas.

$\sigma_{\mu_0}^2$, varianza entre estudiantes.

Los resultados indican que el efecto escolar es de 8,98% para Competencia Lingüística; 8,81% para Competencia Matemática; 14,82% para Competencia Científica, y 11,36% para Competencia Social y Ciudadana. Cifras análogas se han encontrado en otros estudios (p.e. Murillo, 2005) lo que confirman la adecuación del proceso seguido.

El siguiente paso (tabla 3) es elaborar los modelos multinivel introduciendo individualmente las ocho variables explicativas en cada uno de los cuatro modelos ajustados que se han mostrado en la tabla 2. Con ellos se determinan cuáles son las actividades de evaluación que impactan de manera significativa en el desarrollo de los estudiantes.

TABLA 3. RESULTADOS DE LOS MODELOS MULTINIVEL CONSTRUIDOS A PARTIR DE LA INTRODUCCIÓN DE CADA VARIABLE EXPLICATIVA EN LOS MODELOS AJUSTADOS

	COMPETENCIA LINGÜÍSTICA β (ee)	COMPETENCIA MATEMÁTICA β (ee)	COMPETENCIA CIENTÍFICA β (ee)	COMPETENCIA SOCIAL Y CIUDADANA β (ee)
Realizar una prueba sobre conocimientos previos antes de cada tema	NS	NS	NS	NS
Realizar controles después de cada tema o lección	NS	NS	4,20 (1,61)	4,67 (1,57)
Realizar controles escritos trimestrales	NS	-1,72 (0,80)	NS	-1,68 (0,82)
Corregir los deberes y cuadernos	3,32 (1,16)	NS	4,58 (1,74)	3,92 (1,72)
Revisar las actividades que hacen en clase	NS	NS	4,87 (2,00)	NS
Realizar exámenes orales	NS	NS	NS	NS

Valorar la autoevaluación del alumnado	NS	NS	NS	NS
Valorar el interés y la participación en clase de estudiantes	NS	NS	NS	4,29 (2,08)

Nota: NS Aportación no significativa con $\alpha=0,05$.

Fuente: Elaboración propia.

Del análisis de la tabla 3 se extraen una serie de interesantes ideas para cada una de las áreas evaluadas:

- *Competencia Lingüística*. De las ocho variables explicativas estudiadas relativas a la evaluación, sólo una parece tener efecto en el desempeño en competencia lingüística –descontando el efecto del ISEC y el origen–, el hecho de que el docente corrija los deberes y cuadernos. Concretamente, los datos indican que por cada punto que aumente (o disminuya) sobre la media la variable Corregir los deberes y cuadernos, el desarrollo de la Competencia Lingüística de los estudiantes aumenta (o disminuye) en 3,3 puntos.

- *Competencia Matemática*. De nuevo, una sola variable explicativa parece realizar una aportación estadísticamente significativa al modelo ajustado para esta variable producto: Realizar controles trimestrales. Y, curiosamente, el índice es negativo, lo que significa que el hecho de que el docente realice controles trimestrales tiene una incidencia negativa en el desarrollo de la Competencia Matemática. Expresado en números, por cada punto que el docente aumente (o disminuya) la frecuencia de la realización de controles trimestrales, el desempeño en la variable producto Competencia Matemática de los estudiantes disminuye (o aumenta) 1,72 puntos.

- *Competencia Científica*. En tres de los ocho modelos multinivel calculados con esta variable de producto, uno para cada variable explicativa introducida individualmente, la variable de evaluación hace una aportación estadísticamente significativa, en concreto, con las variables: Corregir los deberes y cuadernos, Realizar controles después de cada tema, y Revisar las actividades realizadas en clase. Ello implica que estas tres variables inciden en el desempeño en la Competencia Científica. Específicamente, los datos muestran que por cada punto que aumente (o disminuya) la frecuencia con la que el docente revisa las actividades realizadas en clase, el estudiante aumenta (o disminuye) en 4,87 puntos su desarrollo en esta competencia, 4,58 por la frecuencia en corregir los deberes y cuadernos, y 4,2 puntos por la frecuencia en realizar controles después de cada tema.

• *Competencia Social y Ciudadana*. Son cuatro las variables explicativas que, según los datos, influyen de manera significativa en el desarrollo de esta competencia: Realizar controles después de cada tema o lección, Realizar controles escritos trimestrales, Corregir los deberes y cuadernos, y Valorar el interés y la participación en clase de estudiantes. Todos ellos con una relación positiva; excepto realizar controles escritos trimestrales que, al igual que ocurría con el desempeño en Competencia Matemática, tiene una relación negativa. En los tres casos positivos la incidencia es de más 4 puntos, y en el caso del coeficiente negativo de 1,7.

Una vez discriminado las variables que realizan una aportación significativa individualmente a los modelos ajustados, para cada variable de producto se estima un modelo con todas ellas de forma simultánea (el modelo final) con el fin de obtener un mayor estimador de máxima verosimilitud en el modelo. El estudio de la variación de la razón de verosimilitud refleja que prácticamente no varía para ninguna de las variables. Así pues, el último paso es el cálculo de los cuatro modelos multinivel final, con las variables de ajuste y todas las variables explicativas de forma simultánea (tabla 4). Para las variables producto Competencia Lingüística y Competencia Matemática, al tener solo una variable explicativa con aportación significativa en los modelos ajustados, estos coinciden con los modelos finales. Luego para estas dos variables estos modelos finales no aportarían más información. Para las otras dos variables de producto, sin embargo, los modelos finales aportan información sobre las variables que inciden más en el desarrollo de las competencias, y las que hacen una aportación diferente.

Tabla 4. Resultados de los modelos multinivel finales

	COMPETENCIA LINGÜÍSTICA β (ee)	COMPETENCIA MATEMÁTICA β (ee)	COMPETENCIA CIENTÍFICA β (ee)	COMPETENCIA SOCIAL Y CIUDADANA β (ee)
PARTE FIJA				
Intercepto	492,33 (4,73)	503,57 (1,45)	481,13 (1,60)	479,34 (6,96)
ISEC	28,64 (0,69)	26,44 (0,65)	26,63 (0,63)	25,12 (0,68)
Origen (nativo-inmigrante)	-4,10 (2,01)	-5,04 (2,10)	-4,75 (2,03)	NS
VARIABLES DE PROCESO				
Realizar controles después de cada tema o lección			3,73 (1,62)	4,32 (1,16)

Corregir los deberes y cuadernos	3,32 (1,16)		4,26 (2,00)	
Realizar controles escritos trimestrales		-1,67 (0,80)		-1,18 (0,82)
Valorar el interés y la participación en clase de estudiantes				4,34 (2,10)
PARTE ALEATORIA				
Entre escuelas	774,88 (57,18)	785,66 (58,46)	1277,23 (76,56)	987,30 (66,48)
Entre aulas	33,35 (28,93)	35,84 (29,94)	0,00 (00,00)	5,94 (27,16)
Entre estudiantes	7856,91 (72,42)	8104,61 (74,70)	7488,80 (68,46)	7878,70 (72,80)

Nota: NS Aportación no significativa con $\alpha=0,05$.

Fuente: Elaboración propia.

Analizando la tabla 4 se observa que, para Competencia Científica, de las tres variables que tuvieron una aportación significativa individualmente se “cae” la variable Revisar las actividades que realizan en casa. Para la variable producto Competencia Social y Ciudadana de las cuatro variables que realizan aportaciones en los modelos individuales, y tomadas conjuntamente solo tres de ellas realiza una aportación significativa; se pierde, por tanto, la variable corregir los deberes.

Con todo ello, se puede concluir que las actividades de evaluación que realiza el docente en el aula tiene una incidencia en el desarrollo global del estudiante controlando el efecto del nivel socio-económico y el origen (tabla 5). Más exactamente, tres son las actividades evaluativas que han mostrado su relación positiva con el desempeño en alguna o en varias de las competencias estudiadas:

- Realizar controles después de cada tema o lección, que incide sobre el desarrollo de su Competencia Científica y en el desarrollo de su Competencia Social y Ciudadana.
- Corregir los deberes y cuadernos, que impacta sobre el desempeño de los estudiantes en Competencia Lingüística y en Competencia Científica.
- Valorar el interés y la participación de los estudiantes en clase, que afecta de manera positiva sobre el desarrollo de su Competencia Social y Ciudadana.

Por el contrario, llevar a cabo controles escritos trimestrales afecta negativamente en el desarrollo de los estudiantes en Competencia Matemática y en Competencia Social y Ciudadana.

Por último, hay tres actividades de evaluación que según estos datos no parecen estar directamente asociadas con el desarrollo de ninguna competencia:

- Realizar una prueba sobre conocimientos previos antes de cada tema.
- Realizar exámenes orales.
- Valorar la autoevaluación del alumnado.

Tabla 5. Resumen del proceso de modelado multinivel

	COMPETENCIA LINGÜÍSTICA	COMPETENCIA MATEMÁTICA	COMPETENCIA CIENTÍFICA	COMPETENCIA SOCIAL Y CIUDADANA
Realizar una prueba sobre conocimientos previos antes de cada tema				
Realizar controles después de cada tema o lección			XX	XX
Realizar controles escritos trimestrales		XX*		XX*
Corregir los deberes y cuadernos	XX		XX	X
Revisar las actividades que hacen en clase			X	
Realizar exámenes orales				
Valorar la autoevaluación del alumnado				
Valorar el interés y la participación en clase de estudiantes				XX

Nota: X variables con una aportación significativa en los modelos con variables introducidas individualmente. XX variables con una aportación significativa en los modelos finales

(*) aportación negativa

Fuente: Elaboración propia.

Vista la incidencia de algunas actividades de evaluación sobre el desarrollo de las diferentes competencias, puede ser interesante analizar qué es lo que hacen los docentes de Educación Primaria en España en las otras variables sobre diferentes acti-

vidades de evaluación. De esta forma será posible valorar en qué medida las prácticas evaluativas que se han mostrado especialmente eficaces son desarrolladas por los maestros y las maestras en España. Los datos globales aparecen en la tabla 6.

Tabla 6. Distribución de los docentes según la frecuencia de realización de diferentes actividades de evaluación. Datos en porcentaje

	NUNCA O CASI NUNCA	ALGUNAS VECES	CASI SIEMPRE	SIEMPRE
Realizar una prueba sobre conocimientos previos antes de cada tema	31,15	37,78	22,01	9,06
Realizar controles después de cada tema o lección*	,90	4,51	23,67	70,92
Realizar controles escritos trimestrales*	45,57	22,52	11,96	19,95
Corregir los deberes y cuadernos*	,08	2,18	23,98	73,76
Revisar las actividades que hacen en clase	,08	1,13	24,00	74,79
Realizar exámenes orales	12,80	62,58	17,92	6,70
Valorar la autoevaluación del alumnado	12,76	42,14	28,40	16,70
Valorar el interés y la participación en clase de alumnas/os*	,00	1,20	16,69	82,11

Notas: en negrita las modas.

(*) variables asociadas con el desempeño

Fuente: Elaboración propia a partir de los datos de le EGD 2009.

Los datos aportados por los 1.341 docentes estudiados muestran con claridad las actividades que desarrollan respecto a la evaluación de sus estudiantes. En una rápida mirada es posible agrupar las prácticas en tres grandes grupos en función de su frecuencia de uso.

Por una parte están las prácticas de evaluación desarrolladas por la mayoría de los docentes de Primaria, que serían cuatro: Realizar controles después de cada tema o lección, Corregir los deberes y cuadernos, Revisar las actividades hechas en clase, y Valorar para la calificación el interés y la participación en clase de los estudiantes. En todas ellas, más del 70% de los maestros y maestras afirman hacerlo siempre (el 70,9% Realizar controles después de cada tema o lección, el 73,7% Corregir los deberes y cuadernos, y el 74,8% valorar para la calificación el interés y la partici-

pación en clase de los estudiantes), y más del 96% de ellos dicen hacerlo siempre o casi siempre (Realizar controles después de cada tema o lección el 96,6%, Corregir los deberes y cuadernos el 97,7%, Revisar las actividades hechas en clase el 98,8%, y Valorar el interés y la participación en clase de los estudiantes el 98,8%).

En el segundo grupo estarían las actividades de evaluación realizadas de forma poco habitual. Allí se encuentran tres de las actividades analizadas: Realizar pruebas sobre conocimientos previos antes de cada tema, Realizar pruebas orales, y Tener en consideración la propia autoevaluación del alumnado. En los tres casos, la puntuación modal estaría en la categoría “algunas veces” y en todos ellos hay más docentes que afirman que no las han desarrollado “nunca” o “algunas veces” que aquellos que dicen desarrollarlas “casi siempre” o “siempre”. Concretamente, el porcentaje de docentes que afirma Desarrollar pruebas sobre conocimientos previos antes de cada tema es del 63,9%, realizar exámenes orales del 75,3%, y Valorar la autoevaluación del alumnado un 55,0%.

Por último se encuentra la actividad Realizar controles escritos trimestrales, en la que el 45,6% de los maestros y maestras dicen no realizarla nunca. Esto es interesante porque es la actividad que se había encontrado que tenía un efecto negativo sobre el aprendizaje de los estudiantes.

DISCUSIÓN Y CONCLUSIONES

Los resultados obtenidos en esta investigación muestran, en primer lugar, que la forma de evaluar por parte de los y las docentes de Educación Primaria en España está directamente relacionada con el desempeño académico de los estudiantes en las cuatro áreas analizadas: Competencia Lingüística, Competencia Matemática, Competencia Científica, y Competencia Social y Ciudadana. Estos resultados, por sí solos, tienen la valía de confirmar –con datos actuales, en nuestra realidad y con una muestra estadísticamente representativa para todo el país– un hecho ya demostrado por la experiencia y la investigación (p.e. Black y Wiliam, 1998; Stiggins y DuFour, 2009; Sutton, 2009; Teddlie, 2000): el cómo son evaluados incide en cuánto aprenden los estudiantes. Este hallazgo, sin embargo, no por conocido es menos importante; subraya la necesidad de mejorar los sistemas de evaluación que el docente usa para que los estudiantes aprendan más y mejor.

El análisis detallado acerca de qué actividades de evaluación están asociadas al desempeño aporta resultados más interesantes, novedosos y útiles. De las ocho actividades evaluativas estudiadas por la Evaluación de Diagnóstico General (EDG), que desarrollan los docentes para Primaria, cinco han mostrado su relación con el

desempeño en alguna competencia.

En primer lugar, se ha observado que realizar controles después de cada tema o lección incide positivamente en el aprendizaje de los estudiantes. Este resultado hace referencia a la importancia de tener la enseñanza bien estructurada en lecciones, y en las que en cada una estén perfectamente delimitados los objetivos, los contenidos, las actividades y las actividades de evaluación de carácter formativo (Bangert-Drowns, Kulik y Kulik, 1991; Hunt, Wiseman y Touzel, 2009; Killen, 2005; Murillo, Martínez-Garrido y Hernández-Castilla, 2011).

Un docente que realiza evaluaciones al final de cada tema es un docente que hace evaluaciones frecuentes. Y la investigación ha encontrado relación entre la evaluación frecuente y el desarrollo cognitivo del estudiante (Muijs, Chapman, Collins y Armstrog, 2010; Murillo, 2007a, 2007b), pero también con el no desarrollo socio-afectivo (Kyriakides, Creemers y Antoniou, 2009; Murillo y Hernández-Castilla, 2011). En esa lógica puede también entenderse otro de los hallazgos de este estudio: los estudiantes de los docentes que realizan controles trimestrales escritos obtienen peores resultados en comparación con los que no los realizan. Seguramente porque los maestros y maestras que realizan controles al final del trimestre no realizan estas evaluaciones frecuentes de las que hablamos.

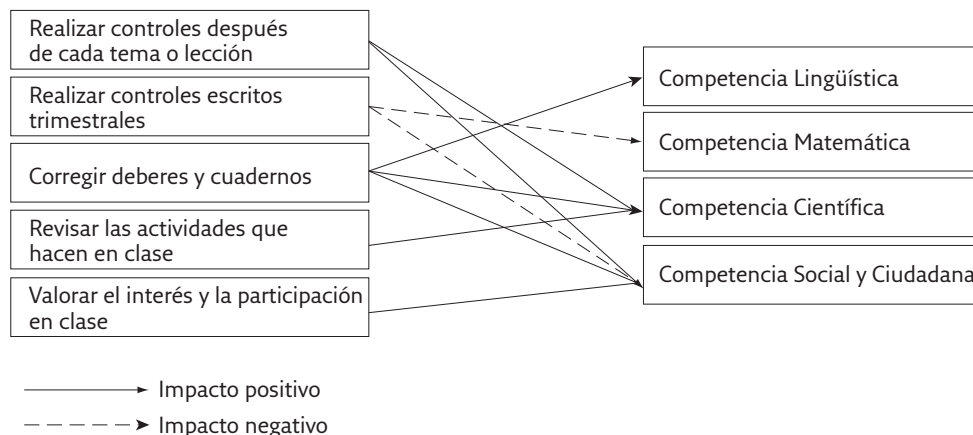
Una actividad evaluativa más que, según esta investigación, incide positivamente en el desempeño de los estudiantes, es que el docente corrija los deberes y cuadernos; al igual que revisar las actividades que hacen en clase. Estos resultados son coherentes con los encontrados por Carlson y Francis (2002). Estos investigadores encontraron que cuando los docentes corrigen las actividades realizadas por los estudiantes y se las devuelven hay un aprendizaje significativamente mayor de éstos.

Especialmente interesante es el hallazgo que establece una relación directa y medible entre los la valoración del interés y la participación en clase de los estudiantes y su aprendizaje. Otro estudio anterior, de carácter experimental, realizado por Vey (2005), refleja la mejora del rendimiento de los estudiantes cuando la evaluación considera la colaboración y comunicación entre los propios estudiantes.

Esta investigación, sin embargo, no ha encontrado relación entre Realizar una prueba sobre conocimientos previos antes de cada tema, Realizar exámenes orales y Valorar la autoevaluación del alumnado, y el desarrollo de competencias por parte de los estudiantes. Estos resultados no son coincidentes con otras investigaciones desarrolladas en el ámbito internacional (ver, por ejemplo, Muijs et al., 2010). Habrá que seguir investigando las razones de esas discrepancias.

En el gráfico 1 se ha representado las relaciones encontradas entre actividades de evaluación desarrolladas y las competencias en las que impactan, una de las preguntas de investigación formuladas.

Gráfico 1. Interrelación entre la actividad de evaluación que desarrolla el docente y la competencia en la que impacta



Fuente. Elaboración propia.

El estudio descriptivo de los que hacen (o dicen hacer) los maestros y maestras de Educación Primaria respecto a esos elementos estudiados permite una valoración de hasta qué punto se están haciendo bien o mal las cosas. Y los datos indican que las cosas se hacen mayoritariamente bien. Una importante mayoría de los docentes realiza controles después de cada tema, corrige deberes y cuadernos, valora el interés y la participación en clase de los estudiantes y revisa las actividades en clase. Por el contrario, son proporcionalmente pocos los que hacen controles escritos trimestrales, elemento que parece tener una incidencia negativa en el aprendizaje.

En todo caso, la principal lección aprendida de esta investigación es la necesidad de cuidar la evaluación como un elemento fundamental del proceso de enseñanza y aprendizaje. Así, si queremos mejorar el desarrollo integral de los estudiantes, no sólo es importante evaluar, sino cómo es esa evaluación. Una evaluación inadecuada, como hemos visto, puede incluso entorpecer el desarrollo global de los estudiantes. De ahí que sea necesario ponerlo en primera línea de atención en la formación tanto inicial como permanente de los docentes.

La principal fortaleza metodológica de esta investigación es el tamaño de la muestra (1.341 docentes de Educación Primaria) y la estrategia de muestreo utilizada (estratificado por conglomerados), lo que permite garantizar la validez externa de

los resultados; así como las técnicas de análisis de datos, los Modelos Multinivel, sin duda los más adecuados para contrastar datos de diferentes niveles (escuela, aula y estudiante), lo que contribuye a garantizar la validez interna, así como a obtener las estimaciones más ajustadas. Las limitaciones vienen de la fuente de datos: los cuestionarios para docentes. En dos sentidos, por un lado, al no haber sido contruidos específicamente para estudiar las actividades de evaluación de los docentes, hay muchas actividades de las que no tenemos información, por ejemplo el uso de diferentes estrategias metodológicas o la frecuencia de evaluación. Y, por otro, la limitación propia de saber sólo lo que dice el docente, no lo que ocurre realmente en el aula.

Todo ello lleva inexorablemente a plantearse la necesidad de más investigaciones sobre el tema. En esa idea sería importante que el Instituto de Evaluación siga mantenido esas preguntas en futuros estudios, pero también que lo complete con más variables relevantes, por ejemplo las ya comentadas sobre la frecuencia de la evaluación y la retroalimentación, y sobre las estrategias de evaluación seguidas y sobre su utilización. En todo caso, parece importante realizar un estudio observacional para ver la incidencia de las estrategias de evaluación en el desempeño de los estudiantes.

Stiggins y DuFour (2009) defendieron la evaluación como una de las herramientas educativas más poderosas que tanto maestros como escuelas tienen a su disposición para alcanzar el logro de los estudiantes. Los resultados de este estudio así lo confirman, de tal forma que podemos hablar de la evaluación como el “combustible que permite iniciar el camino hacia la transformación y mejora de las escuelas. Un camino cada día más necesario.

Fecha de recepción del original: 14 de mayo de 2013

Fecha de aceptación de la versión definitiva: 14 de febrero de 2014

REFERENCIAS

- Baker, E. (2010). *What Probably Works in Alternative Assessment*. Los Ángeles: CRESST, UCLA.
- Bangert-Drowns, R. L., Kulik, J. A. y Kulik, C. C. (1991). Effects of frequent classroom testing. *Journal of Educational Research*, 85, 89-99.
- Black, P. y Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5, 7-24.
- Carlson, C. D. y Francis, D. J. (2002). Increasing the reading achievement of at-

- risk children through direct instruction: evaluation of the Rodeo Institute for Teacher Excellence. *Journal of Education for students placed at risk*, 7(2), 141-166.
- Crooks, T. J. (1988). The Impact of Classroom Evaluation Practices on Students. *Review of Educational Research*, 58(4), 438-481.
- Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership*, 37(1), 15-24.
- Fuchs, L. S. y Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*, 53(3), 199-208.
- Goldstein, H. (1995). *Multilevel Statistical Models*. Nueva York: Willey.
- House, E. y Howe, K. (2003). Democratic deliberative evaluation. En T. Kellaghan, D. Stufflebeam y L. Wingate (Eds.), *The International Handbook of Evaluation Studies* (pp. 79-100). Dordrecht: Kluwer.
- Huerta-Macías, A. (1995). Alternative Assessment: Responses to Commonly Asked Questions. *TESOL Journal*, 5(1), 8-11.
- Hunt, G. H., Wiseman, D. G. y Touzel, T. J. (2009). *Effective teaching: preparation and implementation*. Springfield: Charles C. Thomas.
- INEE-Instituto Nacional de Evaluación Educativa (2007). *Evaluación del Sistema Educativo en España en Educación Primaria*. Madrid: INEE.
- Instituto de Evaluación (2010). *Evaluación General de Diagnóstico 2009*. Madrid: Instituto de Evaluación.
- Killen, R. (2005). *Programming and assessment for quality teaching and learning*. Sidney: Thomsom.
- Kyriakides, L., Creemers, B. P. M. y Antoniou, P. (2009). Teacher Behaviour and Student Outcomes: Suggestions for Research on Teacher Training and Professional Development. *Teaching and Teacher Education*, 25, 12-23.
- Ma, X., Ma, L., y Bradley, K. D. (2008). Using multilevel modeling to investigate school effects. En A. A. O'connell y B. McCoach (Eds.), *Multilevel Modeling of Educational Data* (pp. 59-110). Nueva York: Information Age Publishing Inc.
- Mateo, J. y Martínez Olmo, F. (2005). *L'avaluació alternativa dels aprenentatges*. Barcelona: ICE Universitat de Barcelona.
- McMillan, J. (1997). *Classroom Assessment. Principles and Practices for Effective Instruction*. Needham Heights: Allyn & Bacon.
- Medley, D. y Crook, P. R. (1980). Research in Teaching Competency and Teaching Tasks. *Theory into practice*, 19(4), 293-316.
- Muijs, D. y Reynolds, D. (2000). School Effectiveness and Teacher Effectiveness in Mathematics. *School Effectiveness and School Improvement*, 11(3), 273-303.
- Muijs D. y Reynolds D. (2001). *Effective teaching. Evidence and practice*. Londres: Paul Chapman Publication.

- Muijs, D., Chapman, C., Collins, A. y Armstrog, P. (2010). *Maximum Impact Evaluation*. Manchester: University of Manchester.
- Murillo, F. J. (2005). ¿Importa la escuela? Una estimación de los efectos escolares en España. *Tendencias Pedagógicas*, 10, 29-45.
- Murillo, F. J. (2007a). School Effectiveness Research in Latin America. En T. Townsend (Ed.), *International Handbook of School Effectiveness and Improvement* (pp. 75-92). Nueva York: Springer.
- Murillo, F. J. (2007b). Investigación iberoamericana sobre *eficacia escolar*. Bogotá: Convenio Andrés Bello.
- Murillo, F. J. (2008). Los Modelos Multinivel como herramienta para la investigación educativa. *Magis. Revista Internacional de Investigación Educativa*, 1(1), 17-34.
- Murillo, F. J. y Hernández-Castilla, R. (2011). Factores escolares asociados al desarrollo socio-afectivo en Iberoamérica. *RELIEVE*, 17(2), art 2.
- Murillo, F. J., Martínez-Garrido, C. y Hernández-Castilla, R. (2011). Decálogo para una enseñanza eficaz. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 9(1), 6-27.
- Orlich, D. C., Harder, R. J., Callahan, R. C., Trevisan, M. S. y Brown, A. H. (2010). *Teaching Strategies: A Guide to Effective Instruction*. Boston: Wadsworth.
- Reynolds, D. y Teddlie, C. (2000). *The international handbook of school effectiveness research*. Londres: Falmer Press.
- Stiggins, R. y DuFour, R. (2009). Maximizing the power of formative assessment. *Phi Delta Kappan*, 90(9), 640-644.
- Sutton, R. (2009). Formative Assessment the Way the School Does Business: The impact and Implications of formative assessment for teachers, students and school leaders. En A. Hargreaves, A. Lieberman, M. Fullan y D. Hopkins (eds.), *Second International Handbook of Educational Change* (pp. 883-899). Londres: Springer.
- Teddlie, C. y Reynolds, D. (2000). *The International Handbook on School Effectiveness Research*. Nueva York: Routledge Falmer.
- Vázquez Cano, E. (2012). La evaluación del aprendizaje en primaria y secundaria: los indicadores de evaluación. *Espiral: Cuadernos del Profesorado*, 5(10), 30-41.
- Vey, L. D. (2005). *Enhancing the relationship between learning and assessment*. Tesis doctoral. Canberra: University of Canberra. Extraído el 26 de junio de 2014 de http://www.canberra.edu.au/researchrepository/file/edc0ca80-acb7-25b9-4fa8-6e03a01d4e27/1/full_text.pdf
- Weber, G. (1971). *Inner-city children can be taught to read: four successful schools*. Washington: Council for Basic Education.

- Weurlander, M., Söderberg, M., Scheja, M., Hult, H. y Wernerson, A. (2012). Exploring formative assessment as a tool for learning: students experiences of different methods of formative assessment. *Assessment and Evaluation in Higher Education*, 37(6), 747-460.
- Walberg, J. H. y Paik, D. J. (2000). *Effective educational practices*. Ginebra: IBE/UNESCO.