

Evaluación formativa y resultados de aprendizaje en los centros que imparten Educación Secundaria Obligatoria ^[1]

por Javier GIL FLORES
y Eduardo GARCÍA JIMÉNEZ
Universidad de Sevilla

1. Introducción

Una idea ampliamente aceptada en la comunidad académica, y cada vez más por quienes desempeñan tareas docentes en contextos escolares, es la conveniencia de entender la evaluación del alumnado en los centros educativos como un instrumento al servicio del aprendizaje. Así se ha reflejado en la literatura publicada en nuestro país, con obras donde se recogen reflexiones y experiencias de innovación educativa sobre este tema (entre ellas, Ballester y otros, 2000; Álvarez-Méndez, 2003; Santos-Guerra, 2003; Díez, 2007; Mayo, Castillo, Ávila y Zaitegui, 2010), y también ha quedado plasmado en los diseños curriculares oficiales, propuestos por la Administración educativa. En estos documentos, se insiste en que la evaluación ha de contribuir a optimizar el aprendizaje, mejorar la enseñanza y ajustar el propio currículo. La evaluación, por tanto, ha de orientar e impulsar el trabajo de los estudiantes y profesores, para lo cual no puede limitarse a momentos terminales, sino que ha de estar presente durante todo el proceso de aprendizaje, integrada en el mismo, facilitando la información necesaria que permita valorar lo aprendi-

do y adoptar las medidas oportunas cuando resulten convenientes.

En este sentido, se alude a la denominada evaluación formativa. El ampliamente utilizado término *formative assessment* tiene su origen en las aportaciones de Scriven (1967) en el contexto de la evaluación de programas, al diferenciar entre una evaluación sumativa, dirigida a establecer el valor global de un programa, y una evaluación formativa, orientada a la mejora del mismo. Posteriormente esta misma distinción sería trasladada por Bloom (1969) a la evaluación de estudiantes, atribuyendo a la evaluación formativa el propósito de facilitar retroalimentación y favorecer correcciones durante el proceso de enseñanza-aprendizaje, mientras que correspondería a la evaluación sumativa la misión de valorar los aprendizajes logrados por el alumno al término de un curso o de una fase del mismo.

El sentido actual de lo que se entiende por evaluación formativa queda bien reflejado en el trabajo de Black y William (2009:9), para quienes la práctica de evaluación en la clase es formativa en la medida en que:

«la evidencia acerca del rendimiento de los estudiantes se obtiene, interpreta y utiliza por los profesores, los propios estudiantes y los compañeros para tomar decisiones sobre los siguientes pasos a dar en la enseñanza, que probablemente serán más adecuados o estarán mejor fundamentados que los que habrían dado sin contar con dicha evidencia».

Si bien tradicionalmente se ha hecho hincapié solo en la utilización de la información para mejorar el aprendizaje (por ejemplo, la rentabilización de los resultados de pruebas semanales para introducir mejoras en las prácticas desarrolladas), desde una perspectiva constructivista las prácticas de evaluación formativa se entrelazan con las actividades que se realizan de manera continuada en las aulas (Filsecker y Kerres, 2012). Tienen lugar, por tanto, durante el desarrollo del proceso de enseñanza y aprendizaje, favoreciendo el continuo ajuste de éste para la mejora del rendimiento de los estudiantes en relación con los objetivos de aprendizaje (Popham, 2008). La estrategia para llevarla a cabo supondría determinar cuál es la situación de los estudiantes (recogida de evidencias sobre el aprendizaje), cuáles son los objetivos que definen la dirección en la que deben avanzar (clarificar metas y criterios de evaluación) y cómo proporcionar de manera continuada el feedback adecuado que les ayude a conseguirlo.

La atención prestada a la evaluación formativa se ha mantenido durante las últimas décadas, generando un importante volumen de trabajos que han sido objeto de sucesivas revisiones de la literatura. Al-

gunas de estas revisiones (Natriello, 1987; Crooks, 1988, Allal y Mottier, 2005; Hattie y Timperley, 2007) han llegado a conclusiones que afirman sin reservas los efectos positivos de esta modalidad de evaluación. De este modo, se ha vinculado el uso de la evaluación formativa a una mejora de las prácticas educativas, a la identificación de lagunas en el currículum desarrollado en las aulas y a la mejora de los resultados educativos, entre otros beneficios.

De cara al estudio que presentamos en este artículo, nos interesan especialmente las revisiones que han focalizado la atención sobre investigaciones acerca del impacto que el uso de la evaluación formativa en las aulas tendría sobre los resultados de aprendizaje. Este tipo de revisiones suelen recurrir al meta-análisis para presentar una síntesis de los resultados obtenidos en cada estudio individual, atendiendo al tamaño del efecto observado en los mismos. El tamaño del efecto es una medida que refleja la magnitud de la diferencia entre dos grupos, generalmente un grupo de alumnado en cuyas clases se utiliza alguna modalidad de evaluación formativa y un grupo de control donde ésta no se aplica. El tamaño del efecto se expresa de manera estandarizada, por lo que es posible su comparación entre diferentes investigaciones con independencia del modo en que se haya operativizado la medición de los resultados de aprendizaje. Un criterio para valorar el tamaño del efecto podría ser el propuesto a partir de la síntesis realizada por Hattie, citada en Hattie y Timperley (2007), sobre más de 500 meta-análisis referidos a más de 100 factores que influyen en el rendimiento de los estudiantes. La media del tamaño

del efecto en esa síntesis fue de 0.40, cifra que puede ser tomada como estándar para valorar los resultados de cualquier meta-análisis que pretenda calibrar la influencia de un determinado factor sobre el rendimiento. Como mostraremos a continuación, la mayor parte de las revisiones acerca del efecto de la evaluación formativa sobre el rendimiento han arrojado valores que sobrepasan ese umbral.

Uno de los primeros meta-análisis es el realizado por Fuchs y Fuchs (1986), abarcando 23 investigaciones en niveles educativos de preescolar, primaria y secundaria. El promedio para el tamaño del efecto ascendió a 0.70, permitiendo afirmar que el uso de procedimientos de evaluación formativa incrementaba significativamente el rendimiento de los estudiantes. Con algunas matizaciones, ésa fue también la conclusión a la que se llegó en el meta-análisis realizado por Kluger y DeNisi (1996), incluyendo 131 investigaciones focalizadas en el impacto del feedback sobre el rendimiento. En este caso, el promedio para el tamaño del efecto fue de 0.41, aunque un 38% de los efectos medidos resultó ser negativo, o lo que es igual, en esos estudios el rendimiento del alumnado disminuía con el feedback utilizado.

Sin duda, la revisión más citada en la literatura es la que llevaron a cabo Black y William (1998), abarcando más de 250 estudios realizados en diferentes países. El tamaño del efecto de la evaluación formativa sobre los resultados escolares para estos trabajos se situó entre 0.40 y 0.70, tamaños que pueden ser considerados más elevados que los habitualmente encontrados para otras intervenciones

educativas (Shepard, 2010). En consecuencia, se concluía a partir de tales resultados que el aprendizaje alcanzado por los estudiantes es mayor en las aulas donde tienen lugar prácticas de evaluación formativa. En este sentido, Black y William (1998:17) llegan a afirmar:

«la evaluación formativa puede conducir a logros significativos en el aprendizaje. Aunque no haya garantía de que tales logros se alcancen con independencia del contexto y el enfoque concreto adoptados, no hemos encontrado reseñas de efectos negativos que se deriven de un aumento de práctica formativa».

El más reciente meta-análisis de Kingston y Nash (2011) se ha basado en únicamente 13 estudios publicados en las últimas dos décadas. Esta reducida cifra se debe a la exigente selección de los mismos; entre otros criterios, se requirió una suficiente calidad de los trabajos, que éstos se ocuparan de estudiantes de duodécimo grado (equivalente en nuestro sistema educativo a cuarto de la ESO), o que incluyeran una clara justificación del carácter formativo de la intervención realizada. Además, por el diseño adoptado o por la información aportada, no todos los trabajos se prestan a calcular el tamaño del efecto, por lo que la cifra final de investigaciones analizadas no fue más elevada. En este riguroso estudio, el tamaño del efecto se quedó en un modesto valor medio ponderado de 0.20, que para algunos supone disipar el mito de que a la evaluación formativa pueden atribuirse efectos de entre 0.40 y 0.70 sobre el rendimiento (Briggs, Ruiz-Primo, Furtak, Shepard y Yin, 2012).

No es este último resultado lo único que podría rebajar el entusiasmo por la eficacia atribuida a la evaluación formativa. Si bien conceptualmente pueden aportarse argumentos para apoyarla (Nichols, Meyers y Burling, 2009; Popham, 2008), la base empírica que sustenta el poder de la evaluación formativa para la mejora del rendimiento ha venido suscitando críticas. Shute (2008) ha señalado que en las revisiones realizadas por diferentes autores suelen encontrarse más discrepancias que patrones consistentes de resultados, por lo que no queda claro el efecto positivo que se ha atribuido a la evaluación formativa. Dunn y Mulvenon (2009), tras analizar los trabajos incluidos en algunas de las revisiones más ampliamente difundidas, consideran limitado el cuerpo de evidencias científicas que respalda una contribución directa de la evaluación formativa al logro de resultados de aprendizaje positivos, exponiendo sus dudas sobre los enfoques metodológicos adoptados para demostrarlo.

También Bennett (2011) ha cuestionado que la investigación apoye inequívocamente las prácticas de evaluación formativa y ha salido al paso de estudios que habitualmente han sido citados como base para afirmar su eficacia. En particular, respecto a la amplia revisión de Black y William (1998), el más citado de ellos, las críticas vertidas por los referidos autores consideran que las investigaciones incluidas en su meta-análisis se centraban sobre experiencias de evaluación formativa muy dispares (feedback proporcionado a los estudiantes, orientación de los estudiantes hacia la consecución de objetivos, utilización de estrategias de autoevaluación o evaluación por pares, utilización de diferentes técnicas

para recogida de evidencias, etc.), lo que dificultaría la interpretación de los índices estadísticos mediante los que se sintetizan sus resultados. En opinión de Bennett (2011:11), los estudios sobre el efecto de la evaluación basados en el meta-análisis «pueden considerarse sospechosos» porque:

- «Los estudios abordan tópicos demasiado diferentes entre sí para poder hacer un resumen que resulte significativo;

- A menudo los diversos efectos que se describen provienen del mismo estudio o de estudios hechos por los mismos autores, sin que se hayan tenido en cuenta dichas violaciones a la independencia;

- No se han considerado características del estudio tales como la calidad técnica o la falta de consistencia de los datos o

- Los meta-análisis realizados no han sido publicados, de modo que los métodos utilizados para realizarlos escapan a la revisión crítica».

2. Objetivo del estudio

Una de las conclusiones que emanan de las revisiones críticas realizadas sobre la literatura acerca del impacto positivo de la evaluación formativa es la necesidad de realizar más estudios sobre el tema, procurando recoger suficientes datos de múltiples contextos para hacer creíble el argumento de la eficacia (Dunn y Malevon, 2009; Bennett, 2011; McMillan, Venable y Varier, 2013). Incluso los autores de la última re-

visión aquí citada (Kingston y Nash, 2011) concluyen su trabajo reclamando estudios de mayor calidad, que permitan incrementar la base empírica en la que apoyar la eficacia de las buenas prácticas de evaluación formativa. Sumando a esto la prácticamente nula existencia de trabajos realizados en nuestro país sobre el tema, nos hemos propuesto como objetivo estudiar la contribución de la evaluación formativa a la explicación de los resultados de aprendizaje alcanzados en los centros que imparten Educación Secundaria Obligatoria (ESO).

3. Método

Tras revisar críticamente la literatura sobre evaluación formativa y rendimiento, Dunn y Mulvenon (2009) señalaban la existencia de diferentes formas de operativizar el concepto de evaluación formativa y cuestionaban los métodos empleados para demostrar su impacto. Del mismo modo, Bennett (2011) sugería la necesidad de mejorar los estudios que vienen realizándose sobre este tema, por ejemplo, incrementando la calidad de las fuentes de información o asegurando la validez de las medidas de aprendizaje utilizadas como criterio.

En el presente estudio, hemos tratado de atender a este tipo de críticas y sugerencias. En lugar de configurar situaciones experimentales en contextos específicos, como en la mayor parte de los estudios precedentes, hemos recurrido a un diseño descriptivo y correlacional, identificando en los centros de educación secundaria el grado en que se dan determinadas prácticas que podrían ser características de una evaluación formativa. Al utilizar métodos de encuesta, hemos podido tomar como

f fuente de información una muestra más amplia de contextos educativos (centros), que aportan resultados representativos de nuestra realidad escolar. Atendiendo a los problemas derivados de diferentes formas de operativizar la evaluación formativa, hemos preferido descartar la idea de que esta se identifica con un único tipo de práctica, considerando en su lugar diferentes aspectos o componentes de la misma. Por último, frente a las carencias de validez en las medidas de aprendizaje, hemos construido una medida del rendimiento escolar apoyada en un abanico de indicadores que tienen en cuenta evaluaciones internas y externas, realizadas a lo largo de diferentes cursos escolares.

3.1. Muestra

La población estudiada es la constituida por los 1.246 centros andaluces públicos y concertados que impartían docencia en la ESO, durante el curso 2012-2013. Mediante muestreo estratificado fueron seleccionados 250 centros educativos (aproximadamente el 20% del total), representando la proporción de centros de titularidad pública o privada radicados en Andalucía, así como su distribución por provincias (ver TABLA 1). Si bien nuestro interés se focaliza sobre los centros, la información necesaria para llevar a cabo el estudio habría de ser proporcionada por el profesorado de los mismos. En este sentido, la elección de la muestra de centros determinó automáticamente el colectivo de profesorado invitado a participar en el estudio, constituido por la totalidad del claustro en los centros seleccionados. La cifra de profesorado que finalmente aceptó participar, respondiendo al cuestionario, se elevó a 2.255 docentes.

TABLA 1: Centros seleccionados y profesorado participante en el estudio.

Provincia	Población de centros (n=1.246)		Centros seleccionados (n=250)		Profesorado participante
	Concertados	Públicos	Concertados	Públicos	
Almería	17	76	3	15	119
Cádiz	70	128	14	26	365
Córdoba	43	88	8	18	255
Granada	53	93	11	19	263
Huelva	19	59	4	12	106
Jaén	35	85	7	17	284
Málaga	64	136	13	27	470
Sevilla	94	186	19	37	393
TOTAL	395	851	79	171	2.255

3.2. Variables y recogida de datos

Un primer bloque de variables son las referidas a la presencia de prácticas de evaluación formativa en los centros. Para recoger información sobre las mismas hemos utilizado un cuestionario presentado por vía telemática. La administración del cuestionario fue completada en febrero de 2013, tras enviar a los centros de la muestra cartas informando sobre el estudio e invitando al profesorado a participar en la encuesta, para lo cual habrían de acceder al sitio en Internet donde había sido alojado el instrumento.

En el cuestionario se incluían nueve ítems relativos a prácticas de evaluación susceptibles de estar presentes en los centros: «a partir de la evaluación inicial se introducen medidas de refuerzo o adaptaciones curriculares», «el alumnado conoce desde el inicio del curso los criterios y procedimientos de evaluación de cada materia», «a diario se lleva un registro de las actuaciones del alumnado», «se ano-

ta en un registro acumulativo el nivel de adquisición de las competencias básicas», «cuando se detecta que el alumnado no alcanza los niveles esperados, se intenta conocer las causas», «una vez corregidos, se devuelven los exámenes con comentarios sobre cómo deberían haberse respondido las preguntas», «el profesorado comenta al alumnado los resultados de la evaluación, señalando las principales carencias», «en las sesiones de evaluación se proponen actividades de recuperación para el alumnado que lo necesita» y «en los claustros trimestrales, a partir de los resultados de la evaluación, se proponen medidas de refuerzo o apoyo para el alumnado». Para cada una de las prácticas reflejadas en los ítems, el profesor debía expresar la frecuencia con que se dan en su centro, utilizando una escala de respuesta de seis puntos en la que los valores extremos correspondían a «poco frecuente» (valor 1) y «muy frecuente» (valor 6). Finalmente, a cada centro se ha atribuido un valor en cada una de las prácticas de evaluación,

obtenido como promedio de las respuestas dadas por el profesorado perteneciente a ese centro. De este modo, la estimación del grado de presencia que tienen en el centro las referidas prácticas no depende de las valoraciones realizadas por un único profesor, que podrían estar influenciadas por apreciaciones subjetivas sobre lo que ocurre en la institución o por experiencias singulares en su desempeño docente.

Los elementos de la evaluación formativa reflejados en los nueve ítems se encuadran entre los que suelen identificarse en la literatura; las características de una evaluación formativa eficaz que mayor consenso suscitan son precisamente la interacción momento a momento, la comunicación de criterios de evaluación, la recogida de información sobre el aprendizaje, el feedback y los ajustes correctivos en el proceso de enseñanza y aprendizaje (McMillan, Venable y Varier, 2013). La validez de contenido para el cuestionario se ha asegurado además sometándolo al juicio de miembros del Servicio de Evaluación de la Consejería de Educación y Ciencia de la Junta de Andalucía, que corroboraron la idoneidad de los ítems para recoger información sobre las prácticas de evaluación formativa en los centros. Además, se aseguró la validez facial o aparente, realizando una aplicación piloto a profesorado de ESO, que opinó sobre la adecuación del instrumento para recoger opiniones del colectivo al que va destinado. Por último, la validez de constructo ha sido respaldada por los resultados que mostramos en apartados posteriores de este trabajo, donde el análisis de componentes principales refleja la existencia de una estructura subyacente de contenido configurada en torno a tres di-

mensiones que caracterizan la evaluación formativa: la recogida continua de evidencias sobre el aprendizaje, la retroalimentación al alumnado y la adopción de medidas correctoras. Tras la aplicación del instrumento a la muestra de profesorado participante en el estudio, también se ha estimado la fiabilidad del mismo, recurriendo para ello a la medición de su consistencia interna a través del coeficiente alfa de Cronbach, que alcanzó el valor 0,819.

El segundo bloque de variables lo integran aquellas que corresponden a los resultados de aprendizaje logrados en los centros. Los datos para estas variables fueron proporcionados por la Administración educativa. En este bloque se incluyen las *tasas de promoción* en cada uno de los cursos primero, segundo y tercero de la ESO, definidas como el porcentaje de alumnado del centro que pasó al curso siguiente de esta etapa, y la *tasa de titulación*, que informa sobre el porcentaje de alumnado que obtiene el título de Graduado en ESO. Junto a estas variables, procedentes de la evaluación interna de los aprendizajes, se han tenido en cuenta los niveles medios alcanzados por el centro en las *competencias de matemática y en comunicación lingüística*, medidas para el alumnado que finaliza segundo de la ESO a través de las pruebas anuales de la evaluación de diagnóstico, que lleva a cabo la Consejería de Educación en todos los centros de la Comunidad Autónoma. Estas puntuaciones se expresan en una escala con media 500, que se hace corresponder con el valor medio para toda Andalucía, y una desviación típica 100.

Los valores para todas las variables sobre resultados de aprendizaje fueron

calculados como promedio de los registros para el período comprendido entre los años escolares 2006-2007 y 2011-2012. Al considerar la media en un período de 6 años, se atenuaban las posibles fluctuaciones que pudieran darse entre diferentes cohortes de alumnado debidas a factores individuales, permitiéndonos una mejor caracterización del rendimiento de los centros.

Por último, se ha incluido en el análisis una variable de especial relevancia en la explicación del rendimiento y que hemos tenido en cuenta al valorar la relación entre éste y las prácticas de evaluación formativa. Concretamente, se ha considerado la variable *nivel socioeconómico y cultural*, que informa sobre el contexto familiar del alumnado atendido en los centros. Se utiliza el índice elaborado en el marco de la evaluación de diagnóstico que tiene lugar anualmente en los centros andaluces, basado en la información recogida mediante cuestionarios cumplimentados por sus padres o tutores legales (Consejería de Educación de la Junta de Andalucía, 2008). Se trata de una puntuación estandarizada, obtenida para cada alumno, que permite el cálculo de un valor medio en el centro educativo. Los datos para cada centro de la muestra, correspondientes a esta variable, también han sido proporcionados por la Administración educativa.

3.3. Análisis de datos

Hemos comenzado el análisis describiendo la frecuencia con que se dan en los centros las diferentes prácticas de evaluación. Para ello, hemos calculado los

estadísticos media y desviación típica en cada uno de los ítems correspondientes. Usando la técnica del análisis de componentes principales, las nueve variables han sido reducidas a tres factores, que pueden ser identificados con otros tantos componentes de la evaluación formativa realizada por el profesorado. Siguiendo el mismo procedimiento, las variables relativas a resultados de aprendizaje han permitido construir un índice que informa sobre el rendimiento escolar en el centro. Posteriormente, mediante análisis de regresión lineal múltiple hemos valorado la importancia de los diferentes componentes de la evaluación formativa en la explicación de los resultados de aprendizaje, incluyendo en el modelo una variable habitual en la explicación del rendimiento, como es el nivel socioeconómico y cultural de la población atendida por el centro educativo. El objetivo es controlar el efecto de esta variable. El uso de modelos de regresión es habitual en los estudios centrados en la explicación del rendimiento educativo (tomado como variable criterio) a partir de otras variables presumiblemente vinculadas (variables predictoras), si bien presenta como limitación la imposibilidad de derivar relaciones causales partiendo de los resultados obtenidos. La adopción de este modelo de análisis es adecuada, teniendo en cuenta tanto el propósito del estudio como el nivel de medida (escala de intervalo) de las variables que hemos utilizado.

Tras identificar las modalidades de evaluación formativa que se relacionan con el rendimiento, se ha explorado esta vinculación recurriendo a técnicas des-

criptivas. Para ello, tomando como criterio los cuartiles de la distribución, se ha dividido la muestra de centros en cuatro grupos que corresponden a cuatro grados de presencia de las prácticas de evaluación formativa. Para cada grupo pueden calcularse los niveles de rendimiento medio logrados, representándolos mediante diagramas de barras.

4. Resultados

4.1. Prácticas de evaluación formativa en los centros

De acuerdo con los resultados mostrados en la TABLA 2, las prácticas de evaluación formativa más frecuentes en los centros están referidas a la comunicación con el alumnado. En la escala de frecuencia de seis puntos, se aproximan al valor máximo las medias logradas para

los ítems que hacen referencia a este ámbito de actuación. Así, son especialmente frecuentes tanto la información al alumnado sobre los criterios y procedimientos de evaluación (media 5,62), como la información sobre los resultados de la evaluación, señalando las carencias detectadas (media 5,44). En el extremo opuesto, la menor frecuencia corresponde al registro acumulativo sobre la adquisición de competencias básicas (media 3,50), probablemente por la dificultad intrínseca de este objeto de evaluación que resulta novedoso en el actual marco curricular. Junto a esta cuestión, los elementos de la evaluación formativa con menor presencia en los centros tienen que ver con el análisis de las causas de los bajos niveles de aprendizaje detectados (media 4,69) y la adopción de medidas paliativas en las sesiones de evaluación (4,68) y en los claustros trimestrales (media 4,79).

TABLA 2: *Estadísticos descriptivos para la frecuencia de las prácticas de evaluación formativa.*

	Media	Desv. Típica
Introducción de medidas educativas a partir de la evaluación inicial	5,08	,686
Información al alumnado sobre criterios y procedimientos de evaluación	5,62	,378
Registro diario de las actuaciones del alumnado	5,06	,560
Registro acumulativo de la adquisición de competencias básicas	3,50	,879
Análisis de las causas de niveles de competencia por debajo de lo esperado	4,69	,690
Devolución de exámenes con comentarios sobre las respuestas erróneas	5,18	,555
Información al alumnado sobre carencias detectadas en la evaluación	5,44	,406
Propuesta de actividades de recuperación en las sesiones de evaluación	4,68	,780
Propuesta de medidas de refuerzo para el alumnado, derivadas de los resultados de las sesiones de evaluación, en los claustros trimestrales	4,79	,891

Asumiendo que la evaluación formativa implica distinto tipo de actuaciones, hemos identificado dimensiones singulares dentro de la misma, recurriendo al análisis de componentes principales. La matriz de correlaciones para los nueve ítems ha sido sometida a un análisis previo que contraste su idoneidad de cara a ser sometida a la extracción de componentes. En este sentido, el test de esfericidad de Barlett arrojó un valor de 707,41 ($p=0,000$) y la medida

de adecuación de muestreo de Kaiser-Meyer-Olkin alcanzó niveles elevados, situándose en 0,806. Hemos procedido por tanto a la extracción de factores, que dio lugar a un modelo tridimensional con el que se explica un 70,89% de la varianza total. Tras efectuar una rotación varimax, los pesos obtenidos por cada ítem en los tres factores son los recogidos en la Tabla 3, donde para clarificar la revisión hemos suprimido los valores inferiores a 0,30.

TABLA 3: *Matriz de componentes rotados para la solución de tres factores.*

	Componente		
	1	2	3
Información al alumnado sobre criterios y procedimientos de evaluación	,824		
Información al alumnado sobre carencias detectadas en la evaluación	,811		
Devolución de exámenes con comentarios sobre las respuestas erróneas	,734		
Propuesta de medidas de refuerzo para el alumnado, derivadas de los resultados de las sesiones de evaluación, en los claustros trimestrales		,879	
Propuesta de actividades de recuperación en las sesiones de evaluación		,818	
Introducción de medidas educativas a partir de la evaluación inicial		,677	,412
Registro acumulativo de la adquisición de competencias básicas			,867
Análisis de las causas de niveles de competencia por debajo de lo esperado		,316	,787
Registro diario de las actuaciones del alumnado	,520		,579

Los ítems con mayor peso en el primer factor son los que aluden a la información proporcionada al alumnado sobre criterios y procedimientos de evaluación (0,824), así como sobre sus carencias (0,811) y errores (0,734). Este factor se identifica con actuaciones evaluadoras centradas en la *información y feedback al alumnado*, de tal modo que un centro con puntuación alta en el mismo estará caracterizado por

una elevada presencia de este tipo de tareas en la evaluación que practica.

En el segundo factor, las variables con mayores saturaciones hacen referencia a las intervenciones derivadas de la evaluación, en forma de propuesta de medidas de refuerzo para el alumnado (0,879), actividades de recuperación (0,818) y otras medidas consideradas tras la evaluación

inicial (0,677). Estas tareas permiten identificar el segundo factor con la *adopción de medidas correctoras*.

Finalmente, el tercer factor correspondería a las prácticas de evaluación formativa que se concretan en el *registro continuo y análisis del aprendizaje*. Los centros con un nivel elevado en este factor se caracterizan por tareas de evaluación reflejadas en los ítems con pesos más altos: registro acumulativo sobre la adquisición de competencias (0,867), registro diario de las actuaciones del alumnado (0,579) y análisis de las causas de un bajo rendimiento (0,787).

4.2. Resultados de aprendizaje en los centros

Al analizar los indicadores de rendimiento (ver TABLA 4), se comprueba que los porcentajes de alumnado que promociona o titula en los centros de la muestra, promediados para los cursos 2006-2007 a 2011-2012, registran valores muy próximos entre sí, situados en torno al 80%. En cuanto a la competencia matemática y la competencia en comunicación lingüística, la media se aproxima al valor 500, que es el valor alcanzado por el parámetro media para ambas competencias en la población de centros andaluces.

TABLA 4: *Estadísticos descriptivos para los indicadores de rendimiento en los centros, promediados para los cursos 2006-2007 a 2011-2012.*

Indicador	Media	Desv. Típica
Porcentaje de promoción en 1ª de ESO	81,83	7,13
Porcentaje de promoción en 2ª de ESO	79,05	8,31
Porcentaje de promoción en 3ª de ESO	80,64	7,66
Porcentaje de titulación en ESO	78,97	9,84
Puntuación en competencia matemática	500,05	39,54
Puntuación en competencia en comunicación lingüística	500,53	41,62

El rendimiento actúa en el presente estudio como variable dependiente, que se pretende explicar a partir de las prácticas evaluadoras desarrolladas en los centros. Con el fin de reducir el número de variables criterio, nuevamente se ha aplicado un análisis factorial por el método de componentes principales, comprobando previamente la idoneidad de la matriz de correlaciones (medida de Kaiser-Meyer-Olkin 0,83 y test de Barlett 1083,18,

con $p=0,000$). La solución factorial arroja un único factor que explica el 68,81% de la varianza, con pesos factoriales por encima de 0,70 para los porcentajes de promoción y las puntuaciones en competencias básicas, y cercano al 0,50 para el porcentaje de titulación en ESO.

El factor resultante puede ser tomado como índice de rendimiento del centro, en función del cual caracterizar los resulta-

dos de aprendizaje en cada uno de los centros de la muestra. Las puntuaciones en el factor vienen expresadas como valores estandarizados. Para facilitar su manejo, las hemos trasladado a una escala de media 500 y desviación típica 100.

4.3. Explicación del rendimiento a partir de las prácticas de evaluación formativa

Para valorar el papel que los componentes de la evaluación formativa identi-

ficados en los centros juegan en la explicación de los resultados de aprendizaje, hemos llevado a cabo una regresión del índice de rendimiento del centro sobre las prácticas de evaluación formativa. Con el fin de contrastar la importancia de los componentes de la evaluación formativa como variables predictoras, hemos incluido en la ecuación de regresión el nivel socioeconómico y cultural atribuido a las familias del alumnado que es atendido en los centros. El modelo de regresión resultante se muestra en la TABLA 5.

TABLA 5: *Coefficientes y estadístico T para la regresión del rendimiento sobre las prácticas de evaluación formativa.*

	Coeficientes		T	Sig.
	B	β		
(Constante)	509,11		97,83	,082
Información y feedback al alumnado	9,32	,092	1,82	,071
Adopción de medidas correctoras	11,25	,110	2,19	,030
Registro continuo y análisis del aprendizaje	4,22	,042	,83	,409
Índice de nivel socioeconómico y cultural	134,58	,685	13,50	,000

La proporción de varianza explicada por el modelo de regresión queda reflejada en el valor de R^2 , que se ha situado en 0,517. De acuerdo con los valores para los coeficientes β , la mayor contribución a la explicación del rendimiento corresponde, como cabría esperar, al nivel socioeconómico y cultural ($\beta=0,685$). Sin embargo, aún en presencia de este poderoso predictor, el efecto de las prácticas de evaluación formativa es apreciable en el caso de la información y feedback al alumnado ($\beta=0,092$) y la adopción de medidas correc-

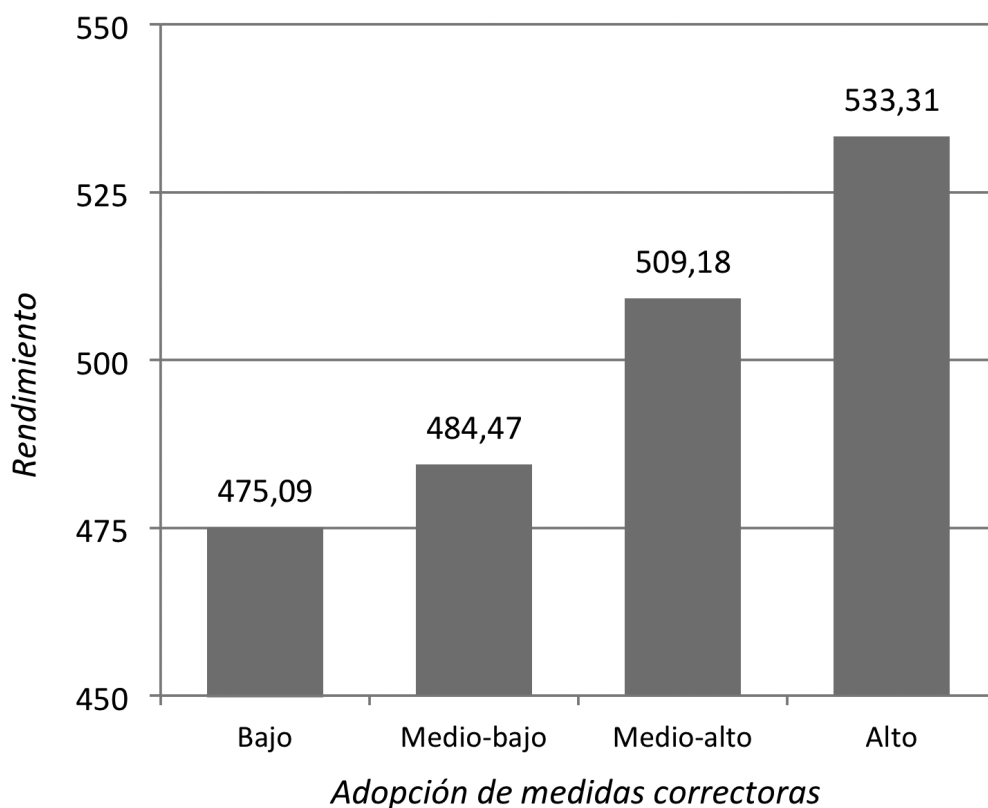
toras ($\beta=0,113$). Este último componente de la evaluación formativa llega a ser estadísticamente significativo en el modelo de regresión, con $t=2,188$ ($p \leq 0,05$). Por tanto, la adopción de medidas tras la evaluación inicial del alumnado y la propuesta de actividades de recuperación o refuerzo derivadas de la evaluación se asocian a una mejora de los resultados de aprendizaje.

Con el propósito de clarificar el papel de la adopción de medidas correctoras en la explicación del rendimiento, hemos di-

vidido la muestra en grupos con distinta frecuencia en esta práctica de evaluación. Partiendo de los cuartiles para la distribución de las puntuaciones en el factor adopción de medidas correctoras, se han diferenciado cuatro grupos de centros, que hemos caracterizado por un nivel bajo, medio-bajo, medio-alto y alto en la adopción de medidas correctoras. Los va-

lores medios alcanzados para el índice de rendimiento del centro, revelan un crecimiento constante de los resultados de aprendizaje a medida que se incrementa el nivel de adopción de tales medidas, con una distancia de más de 58 puntos entre los centros con alta o baja presencia de este tipo de prácticas evaluadoras formativas (ver GRÁFICO 1).

GRÁFICO 1: *Resultados de aprendizaje en función de la adopción de medidas correctoras.*



5. Conclusiones

El presente estudio se ha centrado en generar argumentos empíricos adicionales sobre los que fundamentar la idea comúnmente aceptada de que la evaluación formativa favorece los resultados de aprendizaje.

La eficacia de las actuaciones evaluadoras de carácter formativo, idea firmemente consolidada en la literatura a partir de los resultados obtenidos en el trabajo de Black y William (1998), ha sido cuestionada en los últimos años (Shute, 2008; Dunn y

Mulvenon, 2009; Bennett, 2011; McMillan, Venable y Varier, 2013). Más que negarla, se ha puesto en entredicho la adecuación de los procedimientos seguidos para contrastar su veracidad. En nuestro caso, nos hemos apartado de los diseños de investigación de corte experimental, planteando una aproximación correlacional al problema, lo que nos ha permitido aportar una base más amplia de datos para enjuiciar la existencia de una posible relación.

Los resultados del trabajo que presentamos en estas páginas han confirmado la contribución de la evaluación formativa a la explicación del rendimiento logrado en los centros, pero con ciertas matizaciones. En primer lugar, esa contribución es relativamente baja si la comparamos con el efecto de otras variables tradicionalmente asociadas al rendimiento. Es el caso del nivel socioeconómico y cultural característico del entorno familiar de donde procede el alumno, que la literatura ha asociado reiteradamente al rendimiento escolar (White, 1982; Sirin, 2005; Van Ewijk y Slegers, 2010) y aquí ha constituido, con diferencia, el factor más importante en la explicación de los resultados de aprendizaje. Y en segundo lugar, la evaluación formativa no puede entenderse como una práctica única, sino como un conjunto de actuaciones de distinta naturaleza, aunque todas ellas dirigidas al objetivo último de mejorar el aprendizaje. Habiendo identificado diferentes componentes de la evaluación formativa, se constata en nuestro caso que no todos ellos desempeñan el mismo papel, y eso explicaría en parte las discrepancias apreciadas en algunas de las revisiones referidas (por ejemplo, Shute, 2008) en la introducción a este trabajo.

De acuerdo con nuestros resultados, son las actuaciones que implican la adopción de medidas correctoras a partir de la evaluación las que significativamente contribuyen al rendimiento. Este tipo de prácticas apuntan al aspecto central del concepto de evaluación formativa, que es precisamente el ajuste de la enseñanza o el aprendizaje a partir de las evidencias recogidas durante la evaluación (Popham, 2008; McMillan, Venable y Varier, 2013). Sin embargo, en conjunto, la adopción de medidas correctoras son el tipo de actuaciones relativamente menos frecuente en los centros andaluces que imparten ESO, donde son más habituales las tareas de información y feedback al alumnado y las tareas de registro de evidencias y análisis de las causas del aprendizaje. Ninguna de éstas ha resultado aquí relevante de cara a explicar los resultados escolares. Por tanto, una recomendación clara habría de derivarse del presente estudio: incrementar en los centros el uso formativo que hace el profesorado a partir de los resultados de la evaluación, dando lugar a la adopción de medidas que contribuyan a la consecución de los objetivos de aprendizaje.

La propuesta de introducir medidas correctoras requiere una reflexión acerca del modo en que habría de realizarse. Si la evaluación desarrollada por el profesor ha de actuar formativamente, es necesario que se utilice para modificar de alguna manera la enseñanza y el aprendizaje, por lo que resulta crucial plantear el modo en que debe realizarse este tipo de ajustes (Black y Wiliam, 2005). Es preciso reconocer que los ajustes y medidas correctoras reflejadas en el presente estudio podrían remitirnos a una concepción limitada de

lo que supone una evaluación formativa. En un planteamiento de enseñanza próximo a las propuestas de Bloom (1969), una unidad de enseñanza estaría dividida en varias partes, las cuales constan de actividades orientadas a alcanzar los objetivos de la unidad. En este esquema, tras completar las actividades, la evaluación se concreta en pruebas escritas que generen feedback para profesores y alumnos, permitiendo la adopción de medidas de refuerzo dirigidas a quienes no alcanzaron los objetivos. Con una dimensión temporal más amplia, en este mismo esquema encajaría la adopción de medidas tras las sesiones de evaluación o tras el análisis de resultados realizado en las reuniones trimestrales del claustro de profesores. Aun siendo una estrategia que aquí se ha revelado eficaz, supone mantener un concepto limitado y primitivo de lo que representa la evaluación formativa. La superación de este concepto requeriría la integración de la evaluación en las actividades cotidianas realizadas en clase, permitiendo modificar de manera continua la actuación de profesores y alumnos en el día a día, y no tras la culminación de paquetes o períodos de instrucción delimitados. Ello implica la utilización de una variedad de instrumentos y técnicas de evaluación, apoyándose en las actividades de aprendizaje desarrolladas en clase, en la observación, en los diálogos mantenidos con el alumnado, en los debates entre alumnos, que ofrecen oportunidades para valorar el aprendizaje logrado y para adaptar en el momento las actividades a las dificultades de aprendizaje constatadas. Es decir, la regulación retroactiva, posibilitada por la valoración de los resultados de pruebas escritas periódicas y la adopción de medidas correc-

toras derivadas de ellos, habría de completarse con una regulación interactiva, más unida al trabajo cotidiano en el aula.

Aunque hayamos encontrado una vinculación significativa entre el rendimiento escolar y solo uno de los componentes de la evaluación formativa aquí identificados, no debería interpretarse que otras formas de evaluación formativa han de ser necesariamente poco adecuadas o escasamente productivas. Es el caso de las prácticas basadas en la participación del alumnado, a través de fórmulas de autoevaluación o evaluación por pares (Topping, 2003; Fontana y Fernandes, 2004; Ross, 2006; Andrade, Du y Wang, 2008), que no han sido analizadas en este trabajo. La evaluación y regulación de su propio proceso de aprendizaje exige al alumnado no solo conocer, comprender y aceptar los criterios de evaluación, sino también reconocer sus propios logros y dificultades, plantearse el modo de avanzar en el aprendizaje y comprometerse en ello. Precisamente, saliendo al paso de la dificultad que representa para el profesorado la revisión y retroalimentación al alumno cuando se trabaja con un elevado volumen de evidencias de aprendizaje, cabría pensar en un mayor espacio para las estrategias de autoevaluación o evaluación por pares, que harían de la evaluación continua y formativa un proceso sostenible.

Además, al margen de los posibles efectos que pudieran manifestarse en términos de resultados de aprendizaje, cabe pensar en otros beneficios derivados de la puesta en práctica de la evaluación formativa. Por ejemplo, se ha constatado en trabajos anteriores que la implementación de este tipo

de evaluación por parte del profesorado propicia cambios profundos en las percepciones de éste respecto a su rol en relación con los estudiantes y respecto a sus prácticas en el aula (Black y Wiliam, 2005).

Antes de finalizar, nos detendremos en señalar algunas fortalezas y limitaciones del trabajo realizado. Entre las primeras, podríamos mencionar la amplia muestra de centros incluidos en el estudio o la importante base empírica sobre la que se ha construido la variable que informa sobre el rendimiento de los centros, abarcando diferentes indicadores medidos a lo largo de un período de seis cursos escolares. Como limitaciones, indicamos las inherentes a los propios estudios correlacionales, los cuales permiten avanzar en el análisis de la relación entre variables, pero sin llegar a establecer relaciones causales. Junto a ello, la recogida de información mediante técnicas de autoinforme implica la obtención de información filtrada a través de la subjetividad del profesorado, sin que hayamos observado directamente las prácticas evaluadoras desarrolladas en los centros. No obstante, este problema queda atenuado al contar con las aportaciones de un conjunto de profesores en cada centro, evitando así los sesgos que pudieran afectar a la percepción de un único individuo.

Dirección para la correspondencia: Javier Gil Flores. Departamento MIDE. Facultad de Ciencias de la Educación. C/ Pirotecnia, s/n, 41.013, Sevilla. Email: jflores@us.es.

Fecha de recepción de la versión definitiva de este artículo: 3. V. 2014.

Nota

- [1] Este trabajo se apoya en datos recogidos en el marco del estudio sobre la «Cultura Evaluadora en los Centros Andaluces que imparten ESO», en el que hemos participado junto al Servicio de Evaluación de la Consejería de Educación de la Junta de Andalucía, en virtud del contrato firmado entre ésta y la Universidad de Sevilla.

Bibliografía

- ALLAL, L. y MOTTIER, L. (2005) Formative assessment in learning: a review of publications in French, en OCDE (eds.), *In Formative assessment: improving learning in secondary classrooms* (Paris, OECD Publications), pp. 241-264.
- ÁLVAREZ MÉNDEZ, J. M. (2003) *La evaluación a examen: ensayos críticos* (Madrid, Miño y Dávila).
- ANDRADE, H. L., DU, Y. y WANG, X. (2008) Putting rubrics to the test: The effect of a model, criteria generation, and rubric-references self-assessment on elementary school students' writing, *Educational Measurement: Issues and Practice*, 27:2, pp. 3-13.
- BALLESTER, M. y Otros (2000) *Evaluación como ayuda al aprendizaje* (Barcelona, Graó).
- BENNET, R. E. (2011) Formative assessment: a critical review, *Assessment in Education: Principles, Policy and Practice*, 18:1, pp. 5-25.
- BLACK, P. y WILIAM, D. (1998) Assessment and Classroom Learning, *Assessment in Education*, 5, pp. 7-71.
- BLACK, P. y D. WILIAM, D. (2005) *Changing teaching through formative assessment: research*

- and practice. Ver <http://www.oecd.org/educeri/34488335.pdf> (Consultado el 18. XI. 2013).
- BLACK, P. y WILIAM, D. (2009) Developing a theory of formative assessment, *Educational Assessment, Evaluation and Accountability*, 21:1, pp. 5-31.
- BLOOM, B. S. (1969) Some theoretical issues relating to educational evaluation, en TYLER, R.W. (ed.), *Educational evaluation: New roles, new means. The 63rd yearbook of the National Society for the Study of Education*, 69:2 (Chicago, IL, University of Chicago Press), pp. 26-50.
- BRIGGS, D. C., RUIZ-PRIMO, M., FURTAK, E., SHEPARD, L. y YIN, Y. (2012) Meta-analytic methodology and inferences about the efficacy of formative assessment, *Educational Measurement: Issues and Practice*, 31:4, pp. 13-17.
- BROWN, G. T. L. y HIRSCHFELD, G. H. F. (2008) Student` conceptions of assessment: Links to outcomes, *Assessment in Education: Principles, Policy & Practice*, 15:1, pp. 3-17.
- CONSEJERÍA DE EDUCACIÓN DE LA JUNTA DE ANDALUCÍA (2008) *Evaluación de Diagnóstico. Informe*. Ver http://www.juntadeandalucia.es/educacion/agaev/c/document_library/get_file?uuid=731deb36-16c1-4b05-ad9a-0fe038c890da&groupId=35690 (Consultado el 18. XI. 2013).
- CROOKS, T. J. (1988) The impact of classroom evaluation practices on students, *Review of Educational Research*, 58:4, pp. 438-481.
- DÍEZ, P. (2007) *La evaluación como instrumento de aprendizaje. Técnicas y estrategias* (Madrid, MEC).
- DUNN, K. E. y MULVENON, S. W. (2009) A critical review of research on formative assessments: The limited scientific evidence of the impact of formative assessments in education, *Practical Assessment & Research and Evaluation*, 14:7, pp. 1-11.
- FILSECKER, M. y KERRES, M. (2012) Repositioning formative assessment from an educational assessment perspective: a response to Dunn & Mulvenon (2009), *Practical Assessment, Research & Evaluation*, 17:16, pp. 2-9.
- FONTANA, D. y FERNANDES, M. (1994) Improvements in math performance as a consequence of self-assessment in Portuguese primary school pupils, *British Journal of Educational Psychology*, 64, pp. 407-417.
- FUCHS, L. S. y FUCHS, D. (1986) Effects of systematic formative evaluation: a meta-analysis, *Exceptional Children*, 53, pp. 199-208.
- HATTIE, J. y TIMPERLEY, H. (2007) The power of feedback, *Review of Educational Research*, 77:1, pp. 81-112.
- KINGSTON, N. y NASH, B. (2011) Formative assessment: a meta-analysis and a call for research, *Educational Measurement: Issues and Practice*, 30:4, pp. 28-37.
- KLUGER, A. N. y DeNISI, D. A. (1996) The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory, *Psychological Bulletin*, 119:2, pp. 254-84.
- MAYO, I., CASTILLO, I., ÁVILA, M. y ZAITIGUI, N. (2010) *La evaluación como proceso sistemático para la mejora educativa* (Madrid, Centro de Publicaciones del Ministerio de Educación).

- McMILLAN, J. H., VENABLE, J. C. y VARIER, D. (2013) Studies of the effect of formative assessment on student achievement: So much more is needed, *Practical Assessment, Research & Evaluation*, 18:2, p. 15.
- NATRIELLO, G. (1987) The impact of evaluation processes on students, *Educational Psychologist*, 22:2, pp. 155-175.
- NICHOLS, P. D., MEYERS, J. L. y BURLING, K. S. (2009) *Educational Measurement: Issues and Practice*, 28:3 pp. 14-23.
- POPHAM, W. J. (2008) *Transformative assessment* (Alexandria, VA, Association for Supervision and Curriculum Development).
- ROSS, J. A. (2006) The reliability, validity, and utility of self-assessment, *Practical Assessment, Research and Evaluation*, 11:10, pp. 1-13.
- SANTOS GUERRA, M. A. (2003) *Una flecha en la diana: la evaluación como aprendizaje* (Madrid, Narcea).
- SCRIVEN, M. (1967) The methodology of evaluation, en TYLER, R. W., GAGNE, R. M. y SCRIVEN, M. (eds.) *Perspectives of curriculum evaluation* (Chicago, IL, Rand McNally), pp. 39-83.
- SHEPARD, L. A. (2010) Commentary: Evaluating the validity of formative and interim assessment, *Educational Measurement: Issues and Practice*, 28:3, pp. 32-37.
- SHUTE, V. J. (2008) Focus on formative feedback, *Review of Educational Research*, 78:1, pp. 153-89.
- SIRIN, S. R. (2005) Socioeconomic status and academic achievement: a meta-analytic review of research, *Review of Educational Research*, 75:3, pp. 417-453.
- TOPPING, K. J. (2003) Self and peer assessment in school and university: Reliability, validity and utility, en SEGERS, M. S. R., DOCHY, F. J. R. C. y CASCALLAR, E. C. (eds.), *Optimizing new modes of assessment: In search of qualities and standards* (Dordrecht, Kluwer Academic Publishers), pp. 55-87.
- VAN EWIJK, R. y SLEEGERS, P. (2010) The effect of peer socioeconomic status on student achievement: a meta-analysis, *Educational Research Review*, 5:2, pp. 134-150.
- WHITE, K. R. (1982) The relation between socioeconomic status and academic achievement, *Psychological Bulletin*, 91:3, pp. 461-481.

Resumen: Evaluación formativa y resultados de aprendizaje en los centros de Educación secundaria

En los últimos años, algunos autores han cuestionado las evidencias recogidas sobre los efectos positivos que el uso de la evaluación formativa en las aulas tendría sobre las prácticas y resultados educativos, y han señalado la necesidad de realizar nuevos estudios. El presente trabajo analiza la relación entre la evaluación formativa practicada en los centros de educación secundaria y el rendimiento logrado. La muestra ha estado constituida por 250 centros andaluces, participando como fuente de datos un total de 2.255 profesores que informaron sobre las prácticas evaluadoras. Los resultados confirman una contribución significativa de la

evaluación formativa a la explicación del rendimiento, aunque relativamente baja si se compara con el efecto de otras variables, como es el caso del nivel socioeconómico y cultural de las familias. Además, no todos los componentes de la evaluación formativa juegan el mismo papel, destacando aquellas actuaciones que implican la adopción de medidas correctoras.

Descriptor: Evaluación formativa, educación secundaria, rendimiento, estatus socioeconómico y cultural.

Summary:
Formative assessment and learning outcomes in Secondary Education centres

In recent years, some authors have been questioning evidences on positive effects of formative assessment in the classrooms

practices and educational outcomes. They have remarked the need to do additional research on this topic. This work analyses the relationship between formative assessment in Secondary Education centres and the achievement attained. The sample has been constituted by 250 Andalusian centres, where 2,255 teachers have informed about their assessment practices. Results confirm a significant contribution to the achievement explanation coming from formative assessment. However, it is relatively low if compared to the effect of other variables, such as the socio-economical and cultural status of families. Furthermore, not every component in formative assessment plays the same role. Practices that imply the adoption of corrective measures are the more relevant.

Key Words: Formative assessment, Secondary Education, achievement, socioeconomic and cultural status.