



¿Qué funciona en educación?

Evidencias para la mejora educativa

9

mayo de 2018

¿Es la evaluación del alumnado un mecanismo de mejora del rendimiento escolar?

Sheila González Motos

La evaluación ha sido entendida tradicionalmente como la valoración del profesorado sobre los aprendizajes de su alumnado. Actualmente, sin embargo, la evaluación ha dejado de ser un instrumento finalista (concesión de una nota) y se ha convertido en algo más complejo, en un recurso de aprendizaje más. Ahora bien, la amplitud y variedad de instrumentos evaluadores aconseja investigar y reflexionar sobre qué prácticas tienen un mayor impacto. Esta revisión de la evidencia quiere aportar datos que permitan conocer si la evaluación formativa es un instrumento de mejora del rendimiento de los estudiantes, qué mecanismos evaluadores son más efectivos para la mejora de los resultados académicos y qué perfiles de alumnos se benefician más de la implementación de las distintas estrategias de evaluación en el aula.

“La educación se ha basado durante demasiado tiempo en inercias y tradiciones, y los cambios educativos en intuiciones o creencias no fundamentadas. El movimiento ‘Qué funciona’ irrumpe en el mundo de la educación con un objetivo claro: promover políticas y prácticas educativas basadas en la evidencia. Ivàlua y la Fundació Jaume Bofill han creado una alianza para impulsar este movimiento en Cataluña”.



¿Qué funciona en educación?

Evidencias para la mejora educativa

¿Es la evaluación del alumnado un mecanismo de mejora del rendimiento escolar?



Sheila González Motos

Doctora en Ciencias Políticas y Transformación Social. Ha investigado en el ámbito de la educación y la inmigración, con un especial interés en temas de desigualdades sociales. Actualmente es investigadora postdoctoral y profesora en el Departamento de Sociología de la Universidad Autónoma de Barcelona.

Motivación

La evaluación de los aprendizajes forma parte de la tarea de cualquier docente y se convierte en una pieza central en el sistema educativo, dadas las implicaciones que tiene para los alumnos la obtención de determinadas cualificaciones, titulaciones y/o acreditaciones. Al mismo tiempo, y al margen de los instrumentos empleados, se coincide en ubicar la evaluación entre las tareas a las que los docentes dedican más tiempo, circunstancia que aconseja identificar qué formas evaluadoras son más eficientes y suponen una mejor manera de utilizar el tiempo dedicado a la evaluación. Asimismo, en un contexto como el actual, donde ganan fuerza planteamientos pedagógicos que cuestionan la evaluación como pieza central del sistema escolar, es conveniente reflexionar sobre el valor formativo de distintos mecanismos evaluadores.

La evaluación ha sido entendida tradicionalmente como la valoración que hace el profesorado sobre los aprendizajes de su alumnado al final de un periodo más o menos amplio (unidad didáctica, curso), sea a través de pruebas escritas, exámenes orales, trabajos individuales o actividades grupales, entre otros instrumentos. Las distintas corrientes pedagógicas, no obstante, han ido incorporando nuevos elementos, convirtiendo así la evaluación en algo más complejo: esta no siempre se produce

al final de un periodo, sino que puede ser también inicial (diagnóstico) o de progreso (continua); no siempre la realiza el profesorado, sino que también puede recaer en los compañeros (coevaluación) o en organismos externos al colegio (pruebas de competencias básicas, PISA...) y también, incluso, en el propio alumno (autoevaluación); y, lo más importante, la evaluación deja de ser un instrumento finalista (concesión de una nota) y se convierte en un recurso de aprendizaje más.

En este sentido, la OCDE realizó una extensa investigación sobre el impacto de los métodos de evaluación desarrollados en la mayoría de los países, no solo con voluntad finalista, sino también como instrumentos de aprendizaje [1]. Se ha escrito mucho sobre este tema, pero son pocos los estudios rigurosos que permiten señalar una línea de intervención

clara en materia evaluadora. La calidad de la evidencia es, en general, baja y las conclusiones a las que llega no son unívocas [1] [2]. A pesar de ello, algunas investigaciones aportan datos y reflexiones interesantes para el tema que nos ocupa. Esta revisión de la evidencia se enfoca hacia aquellos instrumentos evaluadores que tienen vocación formativa, es decir, que no se limitan a poner una nota, sino que se convierten en parte del proceso de aprendizaje. Así pues, se trata de abordar el impacto sobre el rendimiento escolar y no de encontrar los mecanismos más justos para certificar los conocimientos aprendidos.

Esta revisión de la evidencia se enfoca hacia aquellos instrumentos evaluadores que tienen vocación formativa, es decir, que no se limitan a poner una nota, sino que se convierten en parte del proceso de aprendizaje. Así pues, se trata de abordar el impacto sobre el rendimiento escolar y no de encontrar los mecanismos más justos para certificar los conocimientos aprendidos.



¿De qué métodos evaluadores hablamos?

La evaluación, como decíamos, es un ámbito muy amplio, que incluye multitud de instrumentos y de finalidades. De entrada, y de forma muy genérica, podemos identificar tres tipos de evaluaciones. En primer lugar, la evaluación formativa (*formative assessment*), utilizada por el profesorado y por los alumnos durante el proceso de aprendizaje, para ir ajustando las dinámicas de enseñanza y de estudio. En segundo lugar, la evaluación sumativa o final (*summative assessment*), que tiene lugar una vez que se cierra una unidad, semestre o curso y tiene por objetivo certificar la consecución de un cierto nivel de aprendizaje (en función de diversos criterios). Y, finalmente, la evaluación estandarizada (*interim assessment*), que tiene por objetivo aportar información comparable entre clases, colegios, distritos o países sobre el grado de rendimiento o conocimiento del alumnado, al servicio de los educadores y de los responsables políticos.

No se trata, sin embargo, de lógicas evaluadoras excluyentes o incompatibles, dado que tanto la evaluación final como la estandarizada pueden ser utilizadas con finalidades formativas o, a la inversa, mecanismos diseñados para la evaluación formativa pierden su valor formativo cuando los docentes se limitan a otorgar una puntuación o nota [3][4].

Tabla 1.
Tipos de evaluaciones

| Tipos de evaluaciones | Objetivo | Destinatario | Momento |
|-----------------------------|--|-----------------------------------|---|
| Evaluación formativa | (Re)orientar el proceso de enseñanza-aprendizaje | Alumnado y docentes | Durante el proceso de aprendizaje |
| Evaluación final o sumativa | Identificar consecución de aprendizaje | Alumnado | Al final del proceso (unidad, semestre, curso...) |
| Evaluación estandarizada | Comparar logros entre clases, centros educativos, distritos, países... | Docentes y responsables políticos | Indiferente |

Fuente: elaboración propia

Como avanzábamos más arriba, esta revisión se centra en la evaluación formativa o el *feedback* formativo.¹ No existe una definición única y completa de lo que se entiende por evaluación formativa. La inexactitud en el uso del término ha generado una ingente literatura, pero también una falta de acuerdo respecto a sus resultados. Podemos afirmar que la vaguedad de las definiciones ha contribuido a la debilidad de los resultados empíricos y ha dificultado la identificación de buenas prácticas [3]. Con todo, existe un consenso para identificar como evaluación formativa aquellos instrumentos utilizados en el ámbito educativo que responden a un doble objetivo: por una parte, mejorar el conocimiento y la adquisición de habilidades y, por otra, incrementar la motivación por el aprendizaje [5]. En este marco se encuadran principalmente los instrumentos evaluadores diseñados con esta doble finalidad, pero también hay espacio para aquellos otros métodos que, partiendo de otros objetivos más propios de la evaluación estandarizada o de la evaluación sumativa, desarrollan su potencial formativo.

Existe un consenso para identificar como evaluación formativa aquellos instrumentos utilizados en el ámbito educativo que responden a un doble objetivo: por una parte, mejorar el conocimiento y la adquisición de habilidades y, por otra, incrementar la motivación por el aprendizaje.



Estos dos objetivos (conocimiento y motivación) pueden alcanzarse a partir de diferentes modelos de evaluación, *a priori* contrapuestos, pero complementarios en la práctica:

- La evaluación puede ser **normativa** (*normative feedback*), cuando compara el rendimiento del alumnado respecto al resto de compañeros del grupo-clase, o **autorreferencial** (*self-reference feedback*), cuando mide el rendimiento respecto a las capacidades propias.
- El mensaje que acompaña al *feedback* permite distinguir entre el **retorno directivo**, es decir, indicaciones al alumno sobre qué necesita revisar, y el **retorno facilitador**, conjunto de comentarios y sugerencias que le ayudan a orientar su propio proceso de mejora.

¹ Aunque hay autores que identifican diferencias entre ambos conceptos, en buena parte de la literatura revisada se utilizan como sinónimos, indistintamente.

- En unos casos, la evaluación se centra en el ejercicio presentado por un alumno, de forma que la información de corrección que facilita el profesorado se dirige a cada alumno sobre su ejercicio concreto. En otros casos, en cambio, el retorno no se produce de forma individual, sino agregada, es decir, se presentan ante el conjunto de la clase las valoraciones sobre el desarrollo general del grupo en una prueba realizada a nivel individual.
- Una cuarta categoría permite diferenciar entre una **evaluación orientada al aprendizaje** (*learning orientation*) que informa al alumno sobre cuán cerca o lejos se encuentra de los objetivos finales marcados para su aprendizaje; y una **evaluación orientada al rendimiento** (*performance orientation*), que se limita a informar de los resultados de las pruebas realizadas.
- A pesar de que la mayor parte de la evaluación en el sistema escolar la realiza el personal docente, existen otras fórmulas evaluadoras como son la **autoevaluación** (*self-assessment*) o la **evaluación entre iguales o coevaluación** (*peer assessment*). La primera se fundamenta en la capacidad del alumnado para entender los objetivos de la enseñanza, evaluar las tareas realizadas, detectar los errores y corregirlos. La evaluación entre iguales comporta la revisión de un trabajo por parte de los compañeros a partir de una serie de criterios evaluadores previamente discutidos en grupo para establecer una perspectiva de evaluación común.

Tabla 2.

Tipos de evaluaciones formativas

| Tipo de evaluación |
|---|
| Evaluación normativa / Evaluación autorreferencial |
| Evaluación directiva / Evaluación facilitadora |
| Evaluación individual / Evaluación de grupo |
| Evaluación orientada al aprendizaje / Evaluación orientada al rendimiento |
| Evaluación estándar / Autoevaluación / Evaluación entre iguales |

Fuente: elaboración propia

Al margen de estos modelos, hay que tener en cuenta que el abanico de objetos para evaluar y sobre los que se fundamenta el retorno formativo es muy amplio. Hattie & Timperley [6] identifican tres niveles sobre los que se puede aplicar el *feedback* formativo:

- **Tarea concreta desarrollada por el alumnado:** se trata de indicar el grado de logro del estudiante en un ejercicio concreto, indicando las respuestas correctas e incorrectas.
- **Proceso de aprendizaje:** en este nivel, la evaluación sobrepasa la valoración del ejercicio concreto y se dirige al proceso de aprendizaje que conecta las distintas tareas que hay que realizar, aportando estrategias para la detección de errores y para la búsqueda de soluciones.

- **Capacidad de autorregulación:** implica incluir en la evaluación aspectos vinculados al compromiso y a la capacidad de organización del alumno, a partir de indicaciones sobre cómo mejorar la autonomía, la disciplina y el autocontrol de cara a lograr el objetivo de aprendizaje acordado.

En la literatura revisada se recoge este amplio abanico de opciones en lo que respecta a la evaluación. Además, una parte importante de ella incorpora en el análisis otras variables con las que interacciona la evaluación, como el nivel académico de los estudiantes, su edad o el tipo de estudios.

Preguntas que guían la revisión

La evaluación ha sido tradicionalmente uno de los ejes organizativos del sistema escolar y, por tanto, de la tarea docente. En los últimos años se están produciendo dos procesos paralelos y antagónicos: por una parte, las instituciones internacionales y los poderes públicos reclaman disponer de más información sobre el rendimiento en los centros escolares a través de la introducción de nuevas pruebas diagnósticas y, por otro lado, han ganado fuerza movimientos pedagógicos que cuestionan la utilidad de las evaluaciones (tanto externas como internas). Esta revisión de la evidencia pretende aportar datos a este debate respondiendo a las preguntas siguientes:

- 1 ¿Puede ser la evaluación formativa un instrumento de mejora del rendimiento escolar de los estudiantes?
- 2 ¿Qué mecanismos evaluadores se muestran más efectivos para la mejora de los resultados académicos? ¿Y de otras competencias?
- 3 ¿Qué perfiles se benefician más de la implementación de distintas estrategias de evaluación en el aula?
- 4 en clave de intervención pública, ¿es recomendable apostar por la inversión en evaluación? ¿En función de qué parámetros?

Revisión de la evidencia

Revisiones y estudios considerados

A pesar de la ingente producción académica que se ha generado desde distintos ámbitos (educación, psicología, sociología...) acerca de la evaluación en el ámbito escolar, es poca la evidencia robusta que permite señalar conclusiones claras al respecto. El presente estudio se fundamenta en las revisiones, metanálisis y estudios experimentales recogidos en la [Tabla 3](#), elaborados principalmente en el mundo anglosajón —si bien muchos de los estudios incluidos se han desarrollado en contextos más cercanos, como es el caso de España— y que se caracterizan por fundamentar sus aportaciones en aproximaciones metodológicas sólidas.

La revisión está integrada fundamentalmente por nueve informes que revisan en profundidad un número elevado de estudios de diseño experimental o procedimiento metanalítico, y por cinco estudios experimentales que han realizado aportaciones especialmente relevantes al tema que nos ocupa. Buena parte de las revisiones incluyen, junto con la evidencia referida a etapas educativas obligatorias, algunas investigaciones y metanálisis sobre etapas formativas superiores (universidad), sin diferenciar claramente sus efectos netos en cada etapa. La inclusión de los estudios experimentales en nuestra revisión nos permite centrar la mirada en la educación primaria y la secundaria, que es el objeto de esta revisión.

Hay que tener en cuenta que la metodología varía entre los distintos estudios revisados. Aunque la mayor parte de la literatura consultada aborda la comparativa entre mecanismos de evaluación finalista e instrumentos de *feedback* formativo, en algunos casos hay una mayor profundización que permite comparar distintos tipos de *feedback* formativo para identificar los de mayor impacto. Además, en algún caso la comparación incluye también grupos de control sin evaluación ni *feedback* formativo, con el objetivo de medir el impacto diferencial de este y de la evaluación finalista respecto a la falta de evaluación.

Tabla 3.
Estudios revisados

| Autores | Tipo de documento |
|--|----------------------|
| Klute <i>et al.</i> (2017) [7] | Revisión |
| Elliot <i>et al.</i> (2016) [2] | Revisión |
| Dunn & Mulvenon (2009) [3] | Revisión |
| Hattie & Timperley (2007) [6] | Revisión |
| Shute (2007) [5] | Revisión |
| Ross (2006) [8] | Revisión |
| Dochy <i>et al.</i> (1999) [9] | Revisión |
| Black & Wiliam (1998) [10] | Revisión |
| Kluger & DeNisi (1996) [11] | Revisión |
| Meusen-beekman <i>et al.</i> (2016) [12] | Estudio experimental |
| De Marcos <i>et al.</i> (2010) [13] | Estudio experimental |
| Wiliam <i>et al.</i> (2004) [14] | Estudio experimental |
| Fuente: elaboración propia | |
| McDonald & Boud (2003) [15] | Estudio experimental |
| Schunk (1996) [16] | Estudio experimental |

¿Puede mejorar la evaluación los resultados educativos de los alumnos?

Los estudios realizados acerca del impacto que tiene la evaluación sobre el rendimiento escolar de los alumnos tienden a identificar una importante mejora en los resultados tanto de lectura como de matemáticas o de capacidad de retención de la información. Si bien algunas investigaciones concluyen que el alumnado objeto de evaluaciones y de *feedback* formativos de calidad puede llegar a incrementar su progreso académico en torno a ocho meses respecto al rendimiento logrado en un curso anual, los metanálisis más recientes sitúan en tres meses el progreso adicional, o en cuatro meses si el profesorado ha recibido formación específica para la implementación de esta evaluación formativa [17].

Con mayor o menor intensidad en función de variables que abordamos en apartados posteriores, la evaluación formativa tiene impacto sobre el rendimiento y los resultados escolares [7] [10]. Entre las áreas curriculares consideradas, la evaluación formativa muestra un mayor efecto sobre las matemáticas, mientras que el impacto sobre la lectura y la escritura es menor, si bien es observable y significativo [7].

Los estudios realizados acerca del impacto que tiene la evaluación sobre el rendimiento escolar de los alumnos tienden a identificar una importante mejora en los resultados tanto de lectura como de matemáticas o de capacidad de retención de la información.



Se ha prestado una especial atención en la literatura al efecto de la autoevaluación sobre el aprendizaje, coincidiendo mayoritariamente en que esta forma de evaluación incide positivamente, aunque en algunos casos el diseño de la investigación aconseja ser cautelosos con los resultados [3] [16] [8]. No obstante, los datos apuntan, en general, a una mejora de los resultados académicos de los grupos que utilizan mecanismos de autoevaluación [15], aunque son inferiores a la evaluación dirigida por docentes [7].

En lo que respecta a la evaluación entre compañeros, los resultados son también positivos. Este tipo de evaluación se muestra a menudo más inteligible para los alumnos, dado que el lenguaje y la expresión del corrector-alumno son más cercanos que los del profesor. Además, este tipo de evaluación no solo permite al alumno evaluar el trabajo de un compañero, sino que también le aporta información sobre su propio rendimiento, a través de la comparación, revirtiendo así en su desarrollo. En este caso, sin embargo, algunas investigaciones han detectado también que este tipo de evaluación tiene efectos negativos, como puedan ser la ansiedad o el estrés, por la tensión que produce evaluar a los demás, pero también por la comparación con los resultados propios [12].

Tanto la autoevaluación como la evaluación entre compañeros tienen efectos diferenciales en función de la competencia trabajada. Así, mientras que la autoevaluación o la coevaluación se muestran efectivas para el trabajo de las matemáticas, en otras áreas como la escritura o la lectura los resultados son menos claros [7].

Buena parte de la investigación reciente ha puesto su atención en la aportación de las nuevas tecnologías a la educación (*m-learning*), en general, y a la evaluación,

especialmente a la autoevaluación, en particular. Aunque la mayor parte de esta se centra en contextos de aprendizaje *online*, algunas investigaciones experimentales han sido desarrolladas en entornos educativos presenciales. En estos casos, las herramientas digitales se configuran como complemento del aprendizaje presencial, ya sea facilitando las pruebas de diagnóstico inicial como abriendo posibilidades de autoevaluación o ampliando el *feedback* entre educadores y estudiantes [13]. Los datos muestran que los grupos que utilizan sistemas de nuevas tecnologías para la autoevaluación obtienen mejores notas finales que el grupo donde no existe evaluación, con intensidades distintas según la edad y el tipo de estudio, como veremos más adelante.

Como ya apuntábamos, a pesar de la cantidad de investigaciones, las conclusiones son generales y los mecanismos de mejora no están siempre claros. La mayor parte de la literatura concluye que la evaluación y el *feedback* formativo comportan un incremento del rendimiento y del aprendizaje de los alumnos [5] [7] [10], pero también hay estudios que no identifican ningún efecto o que incluso obtienen efectos negativos [5] [11]. Estos efectos contrarios a la norma general se explican desde dos lógicas. En primer lugar, por la existencia de otras variables que interfieren en los efectos de la evaluación, como la calidad docente o el nivel de aprendizaje previo del alumno [14]. En este sentido, diversas investigaciones han apuntado a la interacción entre los niveles de conocimiento previo y las pedagogías en el aula, de forma que los alumnos con niveles de aprendizaje previo distintos requieren estrategias de aprendizaje diferenciadas y responden de forma distinta a las estructuras y contenidos. De aquí la importancia de poder evaluar este conocimiento previo, en tanto que condicionante del aprendizaje.




En segundo término, los errores en el diseño de la evaluación podrían explicar su falta de efectividad e, incluso, su impacto negativo sobre el aprendizaje. A modo de ejemplo, parece existir acuerdo en identificar la evaluación basada en notas y no en retorno cualitativo como una práctica perjudicial, aunque también se observan dinámicas no aconsejables según momentos, edades o tipos de estudio, como veremos a continuación [5].

¿Cuáles son las características de la evaluación efectiva?

El amplio abanico de lógicas y métodos evaluadores hace difícil establecer un consenso en torno al impacto general de la evaluación formativa sobre el rendimiento escolar, pero, al mismo tiempo, la gran cantidad de instrumentos evaluados permite a las investigaciones revisadas identificar algunos factores evaluadores que contribuyen a mejorarlo:

- **Los distintos instrumentos de evaluación presentan un impacto diferencial sobre el rendimiento del alumnado.** El aprendizaje resulta más elevado cuando la corrección básica (correcto/incorrecto) va acompañada de *feedback* facilitador, es decir, de pistas para que el alumnado encuentre la respuesta correcta, dado que la incertidumbre ante el fracaso puede causar frustración y reducir la motivación por aprender. En este sentido, explicar el error concreto sin utilizar frases genéricas ni remitirse únicamente a la norma general mejora también notablemente el

aprendizaje. En cambio, instrumentos como los de verificación (correcto/incorrecto) o la repetición del ejercicio hasta que sea correcto plantean impactos más reducidos o nulos. Los *feedbacks* basados en premios o castigos tienen poco efecto. Parece que los efectos varían en función del tipo de aprendizaje: si se trata de mejorar el mismo ejercicio, los mecanismos de repetición se muestran más efectivos; el efecto se traslada a la mejora de tareas diferentes cuando los instrumentos de evaluación incorporan un mayor acompañamiento [5] [6].

- **El *feedback* autorreferencial se muestra más efectivo, dado que ubica al alumno en su propio proceso de aprendizaje.** La asignación de notas, como el resto de mecanismos de *feedback* normativo, resulta un mecanismo de menor impacto y comporta, a menudo, efectos negativos. En el caso concreto de la asignación de notas, se observa una reducción del impacto de la evaluación, dado que centra la preocupación del alumnado en la puntuación y no en los comentarios formativos que acompañan a la corrección [2] [5].

- **El *feedback* debe adecuarse al alumnado al que se dirige.** Cuando no hay coherencia entre mensaje y receptor, no se observa impacto del *feedback* sobre el rendimiento del alumnado. A modo de ejemplo, el alumnado de nivel elevado se beneficia de menores niveles de detalle en el *feedback*, mientras que el más motivado obtiene un mayor rendimiento a partir de mecanismos más complejos de evaluación. Asimismo, los alumnos de bajo rendimiento se benefician más del retorno/corrección inmediata, mientras que parece que los de alto rendimiento se benefician más del retorno en diferido [5].
- **La investigación no es concluyente respecto al grado de elaboración y complejidad aconsejable como retorno de la evaluación.** La literatura revisada concluye que cualquier retorno evaluador tiene más efecto que la simple corrección, y que la complejidad del mecanismo empleado tiene un efecto diferencial en función de lo que se está evaluando. Sin embargo, no existen conclusiones claras respecto a qué nivel de complejidad comporta mayores ganancias, y algunos estudios determinan que no existe ningún efecto diferencial según el grado de complejidad [5].

- **El *feedback* tiene un impacto mayor cuando los objetivos del aprendizaje están muy claros.** La evaluación se muestra así como un complemento de las dinámicas formativas desarrolladas en el aula, enfatizando sus efectos cuando el conjunto de la actividad educativa se organiza de acuerdo con unos objetivos establecidos [11].

- **Los efectos de la evaluación sobre el rendimiento escolar son mayores cuando la complejidad de la tarea es baja.** A medida que la actividad gana en complejidad, se reduce la aportación del retorno formativo al rendimiento escolar [11].

La asignación de notas, como el resto de mecanismos de *feedback* normativo, resulta un mecanismo de menor impacto y comporta, a menudo, efectos negativos.

La literatura revisada concluye que cualquier retorno evaluador tiene más efecto que la simple corrección, y que la complejidad del mecanismo empleado tiene un efecto diferencial en función de lo que se está evaluando.

El *feedback* tiene un impacto mayor cuando los objetivos del aprendizaje están muy claros. La evaluación se muestra así como un complemento de las dinámicas formativas desarrolladas en el aula.

- **La evaluación orientada al aprendizaje que no se limita a la valoración del rendimiento en cada prueba tiene un mayor impacto.** La evaluación y el *feedback* permiten una mayor mejora del rendimiento cuando se enfocan al aprendizaje en su conjunto y no cuando se limitan a la corrección del rendimiento en cada prueba. En este último caso el impacto puede ser negativo o inconsistente [5].
- **La evaluación y el *feedback* pueden ser inmediatos o en diferido.** Algunas investigaciones identifican que el *feedback* inmediato evita la memorización de errores, mientras que otras consideran que puede interferir antes de tiempo, sin dejar al alumno espacio para procesar y corregir el error por sí mismo. Vinculado a esto, diversos estudios aconsejan variar el momento de corrección en función del perfil del alumnado: mientras que el de bajo rendimiento se beneficia más de la corrección inmediata, el más avanzado mejora más claramente cuando el *feedback* se realiza de forma diferida [5].
- **Hay consenso en señalar que la evaluación formativa que se produce de forma continua durante el proceso de aprendizaje es la evaluación que obtiene un mayor impacto.** Si bien la evaluación final y la diagnóstica son también relevantes, diversos estudios sobre la evaluación del progreso o la evaluación continua señalan una importante reducción del impacto de factores individuales sobre el rendimiento a partir de la realización de pruebas de progreso con el correspondiente retorno a lo largo del curso [9].
- A pesar de que tanto la autoevaluación como la coevaluación y la evaluación dirigida por docentes se muestran efectivas, **el impacto sobre el rendimiento escolar es mayor en los casos en que el profesorado es quien conduce** y monitoriza el progreso del alumnado [7].

A pesar de que tanto la autoevaluación como la coevaluación y la evaluación dirigida por docentes se muestran efectivas, el impacto sobre el rendimiento escolar es mayor en los casos en que el profesorado es quien conduce y monitoriza el progreso del alumnado.



Recuadro 1.

Formación en autoevaluación e impactos educativos. Experimento en Barbados

Betty McDonald y David Boud [13] realizaron un experimento en diez centros de educación secundaria en la isla de Barbados. Con el objetivo de evaluar el impacto de las habilidades de autoevaluación sobre el rendimiento escolar de los alumnos diseñaron un programa de formación en autoevaluación para alumnos y profesores.

Se seleccionaron diez institutos según los resultados obtenidos en las pruebas de evaluación externa realizadas anualmente por el Caribbean Examinations Council y en las pruebas de acceso a la educación secundaria que todo el alumnado debe realizar para escoger centro (Barbados Secondary Schools Entrance Examination). Según estos resultados, los diez institutos respondían a una muestra de colegios de rendimiento alto, medio y bajo.

En cada centro se seleccionaron dos grupos-clase del último curso de la educación secundaria (equivalente a 4.º de la ESO), de entre 30 y 40 alumnos. Los investigadores revisaron, centro por centro, las evaluaciones de los cursos anteriores de los alumnos de las dos clases para asegurar que no se producía ninguna agrupación por niveles que pudiera sesgar la muestra. En cada colegio se aleatorizó cuál de las dos clases quedaría asignada al grupo de tratamiento y cuál al grupo de control. Así, un total de 256 alumnos recibieron formación en habilidades de autoevaluación, mientras que en el grupo de control (259 alumnos) no recibieron esta formación.

El grupo experimental recibió formación en autoevaluación durante el curso escolar y como parte del currículum escolar ordinario a partir de 12 módulos diseñados por los investigadores. De igual forma, los docentes fueron instruidos para facilitar al alumnado esta formación. El grupo de control realizó el currículum ordinario sin incluir esta formación.

Al final del curso, todos los alumnos realizaron el Caribbean Examinations Council (CxC). Los resultados entre los dos grupos difieren de manera significativa. El grupo tratado (con formación en autoevaluación) obtiene unas medias superiores en todas las áreas curriculares evaluadas (Economía, Humanidades, Ciencias y Tecnología).

Los investigadores apuntan a la necesidad de tener en cuenta los efectos no buscados que ha podido producir la formación en autoevaluación sobre otras variables también explicativas del rendimiento escolar, como puedan ser la motivación de los docentes implicados en el proceso o la motivación por parte del alumnado del grupo de tratamiento.

Más información en:

McDonald, B. & Boud, D. (2003). "The Impact of Self-assessment on Achievement : The effects of self-assessment training on performance in external examinations". *Assessment in Education. Principles, Policy and Practice*, 10(2), pp. 209–220. [15]

¿Qué alumnado es más sensible a la evaluación formativa?

Más que señalar qué alumnado es más sensible a la evaluación formativa, habría que preguntarse qué perfiles se benefician más de los distintos instrumentos de evaluación formativa. En otras palabras, cuáles son los efectos de estos instrumentos sobre los distintos tipos de alumnado:

- El alumnado de alto rendimiento obtiene mejores resultados con los mecanismos menos intensos (como la simple verificación: correcto/error). En cambio, el de bajo rendimiento experimenta un mayor aprendizaje con mecanismos más intensivos de *feedback* facilitador, que incluyan una explicación detallada del error y la posibilidad de corrección por parte del alumno. Por su parte, los alumnos de nivel medio no muestran grandes diferencias de rendimiento en función del mecanismo de *feedback* empleado. En cualquier caso, los tres perfiles de alumnos muestran mejoras sustanciales en comparación con los que no han recibido ningún retorno [5].

El alumnado de alto rendimiento obtiene mejores resultados con los mecanismos menos intensos (como la simple verificación: correcto/error). En cambio, el de bajo rendimiento experimenta un mayor aprendizaje con mecanismos más intensivos de *feedback* facilitador.



- El alumnado de bajo rendimiento se beneficia más del retorno/corrección inmediatos, mientras que el de alto rendimiento se beneficia más del retorno en diferido [5].
- El alumnado de bajo rendimiento es más sensible al tipo de retorno que recibe. Cuando el *feedback* es normativo (comparación respecto al conjunto de la clase), los alumnos de bajo perfil tienden a atribuir sus resultados a una falta de habilidad y no esperan mejorarlos en las pruebas siguientes, lo que incrementa la desmotivación. En cambio, el *feedback* autorreferencial (comparación respecto a las capacidades del propio alumno) vincula directamente los resultados al esfuerzo realizado, lo que hace que el alumnado centre su atención en su progreso particular [5].

El alumnado de bajo rendimiento es más sensible al tipo de retorno que recibe. Cuando el *feedback* es normativo, los alumnos de bajo perfil tienden a atribuir sus resultados a una falta de habilidad y no esperan mejorarlos en las pruebas siguientes, incrementándose así la desmotivación.



- Las investigaciones centradas en alumnado discapacitado muestran conclusiones más claras que en el caso del alumnado ordinario. Así pues, a pesar de la falta de acuerdo entre algunas investigaciones respecto al impacto diferencial en comparación con el conjunto de la población, se observa unanimidad entre aquellos estudios que indican que el alumnado de educación especial se beneficia de forma intensa de la introducción de métodos evaluadores formativos [3] [10].
- El nivel socioeconómico del alumnado ha sido poco abordado por la literatura revisada, pero los estudios centrados en colegios de entornos sociales más empobrecidos identifican impactos por encima del conjunto de los estudios, lo que podría indicar una mejora en el rendimiento de este alumnado superior a la del conjunto [3] [10]. Se confirma así que el alumnado empobrecido es claramente sensible a los efectos de la evaluación.
- Algunas investigaciones apuntan al hecho de que el *feedback* formativo solo tiene impacto o lo tiene en mayor medida en el alumnado más joven. De todas formas, la identificación de estas edades varía en función de la investigación. Existe consenso, en cambio, a la hora de señalar que la evaluación formativa requiere de un aprendizaje, de forma que cuanto antes se inicie, mayores serán los impactos. En el caso de la evaluación basada en métodos digitales se apunta también a la mayor familiaridad con las nuevas tecnologías entre los más jóvenes como un factor de motivación que incrementa el impacto [3] [13].

Además del rendimiento, ¿qué otras competencias puede mejorar la evaluación?

La investigación sobre evaluación y *feedback* formativo centra su atención en la aportación de estos instrumentos a la mejora del aprendizaje (presente y futuro), y no únicamente de los resultados escolares. En este sentido son tres los aspectos abordados frecuentemente por las revisiones y los estudios analizados: la motivación por el aprendizaje, la capacidad de autorregulación o de autoorganización del aprendizaje por parte del alumnado y su percepción de autoeficacia [10] [15]. Avanzamos que la literatura observa un impacto positivo de la evaluación formativa solamente en los dos primeros.

- **Motivación por el aprendizaje:** las actitudes hacia el aprendizaje se incrementan a partir de la evaluación formativa. La confianza generada por la capacidad de evaluación del alumnado revierte en un incremento de la motivación por aprender. Hay que apuntar también un incremento de la motivación por el aprendizaje a través de mecanismos de *m-learning* en los grupos más jóvenes, seguramente por el hecho de que estos muestran mayor familiaridad con las nuevas tecnologías, mientras que en estudiantes de edades más avanzadas no se observan variaciones [13].

- **Capacidad de autoorganización del aprendizaje:** la autorregulación del aprendizaje no es tarea sencilla y requiere de multitud de habilidades, desde la capacidad de planificación y de establecimiento de objetivos hasta la aptitud para reflexionar y poner en

La búsqueda de estrategias que ayuden a la organización de sus trabajos durante el curso escolar se ha convertido en parte de la tarea del profesorado de educación primaria y, entre estas estrategias, la evaluación formativa se consolida como una herramienta.



marcha las estrategias adecuadas para el aprendizaje.² Valorada tradicionalmente en etapas educativas superiores (educación secundaria, universidad...), la capacidad de autogestión y de organización del aprendizaje por parte del alumnado es, cada vez más, un objetivo también de la etapa primaria. La búsqueda de estrategias que ayuden a la organización de sus trabajos durante el curso escolar se ha convertido en parte de la tarea del profesorado de educación primaria y, entre estas estrategias, la evaluación formativa se consolida como una herramienta. Aunque no hay una edad para desarrollar la capacidad de autogestión, la investigación pone de manifiesto que, con metodologías adaptadas a etapas tempranas, las estrategias de autorregulación del aprendizaje durante la educación primaria tienen efectos positivos en etapas posteriores. Y la evaluación formativa parece contribuir al desarrollo de estas habilidades, especialmente cuando esta perspectiva se incluye en el conjunto de actividades en el aula [12]. La aportación de la evaluación formativa radica en la información sobre la distancia entre el rendimiento del alumno y el esperado, pero también, y especialmente, en las herramientas que otorga para la mejora del rendimiento, al tiempo que señala cuáles son las mejores estrategias a seguir. Es en este punto donde la autorregulación se ve incrementada, dado que el alumnado recibe un amplio abanico de opciones para incorporar y

² Para obtener más detalles sobre el concepto de autorregulación y sobre la efectividad de las intervenciones educativas que intentan fomentarla, podéis consultar el artículo de Gerard Ferrer-Esteban en el número 5 de la serie "¿Qué funciona en educación: <https://goo.gl/ZMsEp1>.

desarrollar en el futuro y, por tanto, se le permite ser responsable de su aprendizaje [14]. El impacto es mayor sobre la autorregulación cuando la autoevaluación es el mecanismo evaluador, aunque la evaluación entre iguales también incrementa la capacidad de autorregulación del alumnado [12].

- **Percepción de autoeficacia:** la percepción del alumnado respecto a su propia capacidad para ejecutar tareas y/o lograr los objetivos fijados es un elemento relevante del proceso de aprendizaje. Aunque se trata de un aspecto muy cercano a la capacidad de autorregulación y está a menudo relacionado con ella, la investigación no observa efectos significativos entre el método evaluador y la percepción del alumno respecto a sus capacidades [12].

Resumen

Del conjunto de instrumentos y lógicas evaluadoras existentes, en esta revisión hemos centrado la atención solamente en aquellas con una voluntad formativa explícita, aunque en algún momento los estudios han utilizado técnicas evaluadoras finalistas, no formativas, en el grupo de control. La revisión de la evidencia ha mostrado unos impactos significativos del *feedback* formativo sobre el rendimiento académico y también sobre cuestiones más amplias de aprendizaje, como la capacidad de autoorganización o la motivación por aprender. En cambio, las percepciones sobre las capacidades propias no parecen ser sensibles a los métodos evaluadores.

La amplitud y la variedad de instrumentos evaluadores dificultan la obtención de resultados generalizables, aunque permite, al mismo tiempo, la identificación de algunas prácticas de mayor impacto. Se observa la simple verificación (correcto/incorrecto) o las dinámicas de premios y castigos como instrumentos de poco impacto sobre el rendimiento del alumnado. En cambio, acompañar la corrección de *feedback* facilitador, la autoevaluación por parte de los alumnos o la evaluación entre compañeros se configuran como opciones de interés para la mejora del rendimiento.

Acompañar la corrección de un *feedback* facilitador, la autoevaluación por parte de los alumnos o la evaluación entre compañeros se configuran como opción de interés para la mejora del rendimiento.



En cambio, acompañar la corrección de *feedback* facilitador, la autoevaluación por parte de los alumnos o la evaluación entre compañeros se configuran como opciones de interés para la mejora del rendimiento.

Hay que destacar, además, que nos encontramos en un ámbito con impactos claramente diferenciales según el perfil del alumnado. Así, mientras que el alumnado de bajo rendimiento se beneficia de sistemas de retorno formativo más complejos y más inmediatos, el alumnado más avanzado obtiene mejores resultados con *feedbacks* más simples y realizados en diferido. Estos dos perfiles también se muestran diferencialmente sensibles a los tipos de orientación del *feedback*. En conjunto, el alumnado más vulnerable —ya sea por capacidades o por clase social— parece mostrar una mejora más clara de su rendimiento escolar a partir de la aplicación de instrumentos de evaluación formativa, especialmente cuando esta se adapta a sus características.

Tabla 4.
Argumentos a favor y limitaciones de los instrumentos de evaluación formativa



| Puntos a favor | Limitaciones |
|--|---|
| <ul style="list-style-type: none"> • La evaluación formativa tiene mayores efectos sobre el rendimiento que la evaluación finalista. | <ul style="list-style-type: none"> • La evaluación debe contemplar la vocación de formación y no limitarse a la valoración de los aprendizajes para tener efectos sobre el rendimiento. |
| <ul style="list-style-type: none"> • La evaluación facilitadora contribuye más al aprendizaje que la evaluación directiva. | <ul style="list-style-type: none"> • La evaluación debe evitar la comparación respecto al resto de alumnos y centrar la valoración respecto al potencial del propio alumno. |
| <ul style="list-style-type: none"> • Hay que establecer una conexión clara entre los criterios de evaluación y los objetivos de las unidades de aprendizaje para hacer de la evaluación un instrumento formativo. | <ul style="list-style-type: none"> • La asignación de notas reduce el potencial efecto formativo de la evaluación. |
| <ul style="list-style-type: none"> • La identificación de los errores debe ir acompañada de pistas para su resolución o mejora. | <ul style="list-style-type: none"> • Debe evitarse el retorno desmotivante centrado en los errores; por el contrario, hay que enfatizar las recomendaciones o el apoyo para mejorar. |
| <ul style="list-style-type: none"> • La evaluación debe dirigirse al aprendizaje en su conjunto. | <ul style="list-style-type: none"> • La evaluación formativa no puede limitarse a la valoración del rendimiento en una prueba concreta. |
| <ul style="list-style-type: none"> • El <i>feedback</i> debe organizarse en unidades comprensibles para el alumno. | <ul style="list-style-type: none"> • La evaluación no puede ser compleja, dado que hay que asegurarse de que el alumno la entiende y dotar a este de herramientas para resolver errores previos. |
| <ul style="list-style-type: none"> • Para incentivar un aprendizaje general, no limitado a la cuestión evaluada, el retorno en diferido muestra mejores efectos. | <ul style="list-style-type: none"> • El retorno en diferido no se aconseja para tareas muy complejas ni en el caso del alumnado con dificultades de aprendizaje. |

Fuente: elaboración propia

Implicaciones para la práctica

Los datos aquí mostrados ponen de manifiesto el importante impacto de la evaluación formativa sobre el rendimiento del alumnado y, al mismo tiempo, su fragilidad, en vista de la multitud de factores que pueden reducir o enfatizar estos efectos. En otras palabras, los beneficios de la evaluación no están garantizados en cualquier circunstancia, sino que se necesitan distintos requisitos, algunos de los cuales dependen de los responsables políticos. A continuación se exponen algunas recomendaciones en este sentido:

- Es importante que los esfuerzos de evaluación que realiza el personal docente se hagan de la forma más eficiente y eficaz posible. Por este motivo, hay que formar al profesorado en instrumentos de evaluación formativa y acompañar a los centros educativos en el diseño de sus planes de evaluación con el objetivo de incrementar el rendimiento del alumnado.

- El contexto se ha mostrado muy relevante para el éxito de la evaluación formativa. Esta debe plantearse en el marco de un proceso de aprendizaje no finalista, es decir, orientado al aprendizaje continuo. Buena parte del éxito de la evaluación deriva de la identificación de objetivos claros para el alumnado, de forma que, junto con los criterios de corrección, será necesario clarificar los objetivos del aprendizaje para que estos hagan comprensible el sentido de las tareas y las evaluaciones. Se trata de un proceso largo, que debe iniciarse con la planificación de la actividad y continuar hasta la evaluación y el retorno. Así pues, hay que dotar de tiempo al personal docente para planificar la evaluación como parte del proceso de aprendizaje.
Hay que dotar de tiempo al personal docente para planificar la evaluación como parte del proceso de aprendizaje.

- El impulso de la evaluación permanente debe permitir ampliar los objetivos más allá del rendimiento, hacia otras esferas del aprendizaje. En este sentido, el trabajo de autoevaluación o de evaluación entre iguales en la educación primaria parece tener efectos importantes sobre la capacidad de autorregulación del alumnado y puede comportar mejoras en la transición a la secundaria, etapa donde el control del docente disminuye y se requieren más capacidades de autogestión. La autoevaluación o la evaluación entre iguales requieren de una planificación especial, de recursos para su desarrollo y de la formación de los docentes para su puesta en marcha.
- En otro orden de cosas, ante el peso explicativo que tienen los conocimientos previos sobre el rendimiento de los alumnos y la certeza de que estos conocimientos condicionan el aprendizaje futuro, parece oportuno dotar a los centros escolares de herramientas para la evaluación diagnóstica que permitan afinar el diseño del proceso de aprendizaje. En este sentido, se necesitan recursos y formación para identificar adecuadamente los conocimientos previos. Asimismo, sería conveniente que la Administración pública utilizara esta evaluación inicial como mecanismo de control en su análisis de la evaluación de docentes y centros educativos.
Parece oportuno dotar a los centros escolares de herramientas para la evaluación diagnóstica que permitan afinar el diseño del proceso de aprendizaje.

- La Administración pública precisa de datos y de información para diseñar mejor los procesos educativos. A tal efecto, las evaluaciones cobran gran importancia. Es necesario, sin embargo, que estas evaluaciones sean completas y permitan captar el mayor número de variables que intervienen en el proceso (capital instructivo familiar, motivación, calidad docente...). Además, estas evaluaciones deben estar a disposición del personal docente y del alumnado para poder utilizarlas con finalidad formativa.
- Por último, a la luz de la ambigüedad de algunos resultados, hay que impulsar aquella investigación que evalúe más profundamente el impacto de los distintos instrumentos evaluadores, la intermediación de otras variables, como la calidad del profesorado, y, muy especialmente, los efectos respecto al alumnado de bajo rendimiento.

Bibliografía

- [1] OCDE (2013). *Synergies for Better Learning*, OECD Publishing.
- [2] Elliot, V.; Baird, J.-A.; Hopfenbeck, T.; Ingram, J.; Thompson, I.; Usher, N.; Zantout, M.; Richardson, J. y Coleman, R. (2016). "A marked improvement?".
- [3] Dunn, K. E. y Mulvenon, S. W. (2009). "A Critical Review of Research on Formative Assessment: The Limited Scientific Evidence of the Impact of Formative Assessment in Education". *Pract. Assessment, Res. Eval.*, vol. 14, núm. 7, pp. 1–11.
- [4] Bell, B. y Cowie, B. (2000). "The Characteristics of Formative Assessment in Science Education". *Sci. Educ.*, vol. 85, pp. 536–553.
- [5] Shute, V. J. (2007). "Focus on Formative Feedback", núm. marzo.
- [6] Hattie, J. y Timperley, H. (2007). "The Power of Feedback". *Rev. Educ. Res.*, vol. 77, núm. 1, pp. 81–112.
- [7] Klute, M.; Apthorp, H.; Harlacher, J. y Reale, M. (2017). "Formative assessment and elementary school student academic achievement: A review of the evidence".
- [8] Ross, J. A. (2006). "The Reliability, Validity, and Utility of Self-Assessment". *Pract. Assessment, Res. Eval.*, vol. 11, núm. 10, pp. 1–11.
- [9] Dochy, F.; Segers, M. y Buehl, M. M. (1999). "The Relation Between Assessment Practices and Outcomes of Studies: The Case of Research on Prior Knowledge". *Rev.*, vol. 69, núm. 2, pp. 145–186.
- [10] Black, P. y Wiliam, D. (1998). "Assessment and Classroom Learning". *Assess. Educ. Princ. Policy Pract.*, vol. 5, núm. 1, pp. 7–74.
- [11] Kluger, A. N. y DeNisi, A. (1996). "The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory". *Psychol. Bull.*, vol. 119, núm. 2, pp. 254–284.
- [12] Meusen-Beekman, K. D.; Brinke, D. J. y Boshuizen, H. P. A. (2016). "Studies in Educational Evaluation Effects of formative assessments to develop self-regulation among sixth grade students: Results from a randomized controlled intervention". *Stud. Educ. Eval.*, vol. 51, pp. 126–136.
- [13] De Marcos, L.; Hilera, J. R.; Barchino, R.; Jiménez, L.; Martínez, J. J.; Gutiérrez, J. A.; Gutiérrez, J. M. y Otón, S. (2010). "An experiment for improving students performance in secondary and tertiary education by means of m-learning auto-assessment". *Comput. Educ.*, vol. 55, núm. 3, pp. 1069–1079.
- [14] Wiliam, D.; Lee, C.; Harrison, C. y Black, P. (2004). "Teachers developing assessment for learning: impact on student achievement". *Assess. Educ. Princ. Policy Pract.*, vol. 11, núm. 1, pp. 49–65.
- [15] Mcdonald, B. y Boud, D. (2003). "The Impact of Self-assessment on Achievement: The effects of self-assessment training on performance in external examinations". *Assess. Educ. Princ. Policy Pract.*, vol. 10, núm. 2, pp. 209–220.
- [16] Schunk, D. H. (1996). "Goal and Self-Evaluative Influences During Children's Cognitive Skill Learning". *Am. Educ. Res. J.*, vol. 33, núm. 2, pp. 359–382.
- [17] Education Endowment Foundation (2017). "Teaching & Learning Toolkit: Feedback".

Primera edició: mayo de 2018
© Fundació Jaume Bofill, Ivàlua, 2018
fbofill@fbofill.cat, info@ivalua.cat
www.ivalua.cat
www.fbofill.cat

Autora: Sheila González Motos
Edició: Bonalletra Alcompàs
Coordinació editorial: Anna Sadurní
Responsable de projectos internacionales: Valtencir
Mendes
Diseño y maquetación: Enric Jardí
ISBN: 978-84-947888-4-0

Esta obra está sujeta a la
licencia Creative Commons de
**Reconocimiento-No Comercial-
SinObraDerivada (by-nc-nd)**. Se
permite la reproducción, distribución y
comunicación pública de la obra siempre
que se reconozca su autoría. No se
permite el uso comercial de la obra ni la
generación de obras derivadas.

