



Educación XXI

ISSN: 1139-613X

educacionxx1@edu.uned.es

Universidad Nacional de Educación a  
Distancia  
España

Botella, Juan; Zamora, Ángela  
**EL META-ANÁLISIS: UNA METODOLOGÍA PARA LA INVESTIGACIÓN EN  
EDUCACIÓN**  
Educación XXI, vol. 20, núm. 2, 2017, pp. 17-38  
Universidad Nacional de Educación a Distancia  
Madrid, España

Disponible en: <http://www.redalyc.org/articulo.oa?id=70651145001>

- ▶ Cómo citar el artículo
- ▶ Número completo
- ▶ Más información del artículo
- ▶ Página de la revista en [redalyc.org](http://redalyc.org)

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal  
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

## EL META-ANÁLISIS: UNA METODOLOGÍA PARA LA INVESTIGACIÓN EN EDUCACIÓN

(META-ANALYSIS: A METHODOLOGY FOR RESEARCH IN EDUCATION)

Juan Botella

*Universidad Autónoma de Madrid*

Ángela Zamora

*Universidad Nacional de Educación a Distancia*

DOI: 10.5944/educXX1.18241

### Cómo referenciar este artículo/How to reference this article:

Botella, J. y Zamora, Á. (2017). El meta-análisis: una metodología para la investigación en educación. *Educación XXI*, 20(2), 17-38, doi: 10.5944/educXX1.18241

Botella, J. & Zamora, Á. (2017). El meta-análisis: una metodología para la investigación en educación [Meta-analysis: a methodology for research in education]. *Educación XXI*, 20(2), 17-38, doi: 10.5944/educXX1.18241

## RESUMEN

Se presenta la metodología del meta-análisis como herramienta para la investigación educativa. Un meta-análisis implica una síntesis cuantitativa de la evidencia acumulada sobre una pregunta de investigación previamente definida. La respuesta se basará en la información contenida en los estudios previamente publicados (estudios primarios). Las principales características del meta-análisis son la precisión, la objetividad y la replicabilidad. Su aplicación permite obtener una estimación combinada del tamaño del efecto. También permite evaluar la heterogeneidad observada en un campo de estudio. Ello permite a su vez formular nuevas hipótesis que incorporan el papel de variables que no se habían tenido en cuenta hasta el momento. Un meta-análisis comienza con la formulación del problema. A continuación se lleva a cabo la localización de estudios relacionados con ella. En la fase de codificación se caracterizan los trabajos recogidos en la fase anterior. Finalmente, un análisis estadístico conduce a los resultados, que se presentarán en una publicación que asegure la replicabilidad del estudio. Entre las herramientas gráficas asociadas al meta-análisis destaca el llamado *Forest Plot*, que permite presentar de forma conjunta las estimaciones puntuales y de intervalo de los estudios seleccionados para

el meta-análisis. Aunque el potencial del meta-análisis es evidente, también presenta algunas debilidades derivadas de la falta de acotación de los constructos estudiados o de la existencia de un sesgo en la publicación de resultados, sobre todo de los estadísticamente significativos. La valoración general que se hace del papel del meta-análisis en la investigación en educación es positiva aunque también se señalan áreas de mejora. Se indican algunos recursos, tanto bibliográficos como de *software*, que permiten a los investigadores interesados adentrarse en esta herramienta tan útil en el campo de la investigación educativa.

## PALABRAS CLAVE

Meta-análisis; investigación en educación; revisión sistemática.

## ABSTRACT

The methodology of meta-analysis is presented as a tool for educational research. A meta-analysis implies a quantitative synthesis of the cumulated evidence about a research question previously defined. The answer will be based on the information contained in the studies previously published (primary studies). Precision, objectivity and replicability are the main characteristics of this methodology. Its applications aim at obtaining a combined estimation of the effect size. The meta-analysis is also applied to explore the observed heterogeneity in a body of results. This allows the statement of new hypothesis that incorporate the role of new variables not taking into account before. A meta-analysis starts with the formulation of a problem. Then, a literature search is carried out. In the codification phase all the documents found are characterised. Finally, a statistical analysis leads to the results, which in turn will be published in a way that ensures the replicability of the meta-analysis. Among the available tools in the context of meta-analysis the so-called Forest Plot plays a remarkable role. The use of this type of graphics allows the joined plotting of both point and interval estimates of the studies selected for the meta-analysis. Even though the potential of the meta-analysis is obvious, it also exhibits some weaknesses derived from the lack of clear boundaries in the constructs studied or from the presence of bias in the publication of the results. The general appraisement of meta-analysis in the field of educational research is positive, although areas for improvement are also highlighted. A number of resources both bibliographic and software-related are available for researchers who wish to venture into this fruitful tool for educational research.

## KEYWORDS

Meta-analysis; educational research; systematic review.

## EL META-ANÁLISIS NACIÓ EN EL CAMPO DE LA EDUCACIÓN

Algunas aportaciones metodológicas nacidas en disciplinas de las ciencias sociales han acabado teniendo una gran aceptación en otras. En el caso de la educación su mayor triunfo de exportación es la metodología del *meta-análisis* (MA). Fue en una conferencia impartida por Gene V. Glass como presidente de la *American Educational Research Association* (AERA), publicada como artículo en *Educational Researcher* (Glass, 1976), cuando se empleó este término por primera vez para hacer referencia a la naciente metodología. Sin duda no fue casual que naciese precisamente en una disciplina como la educación, llena de resultados dispersos y a veces contradictorios. La necesidad de un mayor rigor y sistematicidad eran palpables y fueron el caldo de cultivo idóneo. Los problemas asociados a la síntesis de la evidencia acumulada en la investigación era ya una preocupación antigua, pero no se había creado una metodología específica para realizar esta tarea. Curiosamente, en muchas disciplinas (incluida la educación) se oían voces reclamando un mayor rigor metodológico en las investigaciones, pero ese mismo rigor no se reclamaba para las revisiones. El espíritu de los tiempos (el *zeigeist* de los historiadores) pedía a gritos algún cambio en esa dirección.

Los primeros ejemplos de MAs fueron revisiones en los ámbitos de la educación y la psicología. Concretamente, Glass (1976) ilustró su propuesta inicial con sendas síntesis de resultados, aún muy rudimentarias, sobre la eficacia de la psicoterapia (posteriormente publicada: Smith y Glass, 1977) y sobre la relación entre el estatus socioeconómico y el rendimiento académico (resultados de una tesis doctoral presentada ese mismo año). Glass hizo dos aportaciones principales: el propio nombre de la metodología y la propuesta de emplear índices de *tamaño del efecto* (TE) para traducir los resultados de todos los estudios a una métrica común. Por el contrario, sus métodos estadísticos aún no eran los apropiados. Hubo que esperar hasta la publicación del libro de Hedges y Olkin (1985) para contar con modelos estadísticos adecuados para el escenario meta-analítico. El artículo seminal de Glass es ya uno de los artículos metodológicos más citados (4.948 citas según *google académico*; acceso el 1-1-2017). Cuarenta años después de su nacimiento el MA es una metodología utilizada en todas las disciplinas en las que se emplea el análisis estadístico, aunque seguramente es en medicina donde se han dedicado más esfuerzos y recursos a su desarrollo y refinamiento. El propio Glass ha proporcionado una valoración aprovechando el aniversario (Glass, 2016).

En las páginas siguientes hacemos una presentación, necesariamente sucinta, de las claves de esta metodología, de sus fortalezas y debilidades y de sus perspectivas futuras. Además de divulgar sus potencialidades, nues-

tra pretensión es invitar a los investigadores del ámbito de la educación que no la conozcan a asomarse a la misma.

## DEFINICIÓN Y CARACTERÍSTICAS PRINCIPALES

Como definición no exhaustiva podemos decir que el MA es una metodología para el análisis cuantitativo de revisiones de la literatura científica sobre una pregunta específica (Botella y Sánchez-Meca, 2015). Nació con el impulso de tratar de mejorar la forma poco rigurosa como se hacían las revisiones clásicas, que desde entonces pasaron a llamarse revisiones narrativas. Se decía que las revisiones narrativas eran imprecisas y subjetivas, lo que conllevaba que fueran poco replicables. Las características que pretende tener un buen meta-análisis se establecen precisamente a partir de esta constatación y de la voluntad de combatirlas. Se dice que las características del MA son la precisión, la objetividad y la replicabilidad (Botella y Sánchez-Meca, 2015). Por tanto, una forma de valorar la calidad de un meta-análisis es fijarse en la medida en que ha logrado ser preciso, objetivo y replicable. La *precisión* se obtiene dando respuestas numéricas en términos de estadísticos con propiedades conocidas. La *objetividad* se refiere a una operacionalización explícita y clara de los conceptos involucrados. La *replicabilidad* se traduce en que las decisiones adoptadas tengan la suficiente transparencia como para que una repetición independiente con los mismos criterios de decisión conduzca a los mismos resultados.

En muchas investigaciones primarias el objetivo es responder a la pregunta de si dos variables están relacionadas o no. La significación estadística de las diferencias de medias o de una correlación son ejemplos de herramientas estadísticas adecuadas para responder a esta pregunta. Para ello se aprovecha la información contenida en los datos recogidos en esa investigación. El meta-análisis se puede emplear para responder a esa misma pregunta, pero la información que se analiza está contenida en un conjunto de estudios primarios en lugar de estar contenida en uno solo y se utiliza para proporcionar una *estimación combinada* del TE. La estimación se suele expresar mediante un intervalo de confianza. Que dicho intervalo incluya el valor que refleja la ausencia de efecto (ya sea 0 para diferencias de medias o de correlaciones, o 1 para razones de ventajas o riesgos relativos) permite responder a la pregunta. La estimación combinada no será una simple media aritmética, sino una combinación ponderada. De esta forma los estudios realizados con muestras más grandes tienen mayor peso en la estimación combinada que los estudios realizados con muestras pequeñas.

Por ejemplo, Sirin (2005) sintetizó en su MA 58 investigaciones primarias sobre la relación entre el nivel socioeconómico de los estudiantes

y su rendimiento académico (actualización del trabajo esbozado en el artículo de Glass). Dado que los estudios, seleccionados con unos criterios bien establecidos, examinan las relaciones entre las variables indicadas, se utilizó el coeficiente de correlación de Pearson como índice de TE. La síntesis proporcionó una estimación combinada de  $r = .27$ , con un intervalo de confianza del 95 % entre .23 y .30. Como dicho intervalo no incluye el 0 se concluye que el nivel socioeconómico y el rendimiento académico no son independientes. Cada estudio primario es un intento de responder a esa pregunta, pero el MA proporciona una respuesta basada en los 58 estudios y su intervalo de confianza es más preciso. Otro ejemplo se puede encontrar en el MA de Patall, Cooper y Robinson (2008). En este se analiza la influencia de la implicación de los padres en la realización de los deberes escolares de los hijos. Para ello incluyen en su trabajo 14 investigaciones primarias que utilizan diseños experimentales de dos grupos (el índice de TE es la diferencia media tipificada). En uno de ellos se realizan intervenciones para preparar a los padres para ayudar a sus hijos con los deberes. El otro actúa como grupo de control. La estimación combinada del TE es  $d = 0.09$  con un intervalo de confianza del 95 % entre -0.16 y 0.34. Como en este caso el intervalo de confianza contiene el 0 hay que concluir que no hay evidencia de que el entrenamiento de los padres para ayudar a sus hijos tenga relación con el rendimiento de estos.

Pero además de hacer estimaciones más precisas e informativas de los índices de TE que reflejan los fenómenos de interés, el MA permite analizar los matices con los que se presenta dicho efecto. Por un lado, mediante el ajuste de modelos de efectos aleatorios podemos valorar si el fenómeno se presenta de forma homogénea o si lo hace en magnitudes diferentes (Borenstein, Hedges, Higgins y Rothstein, 2010). Por otro, el análisis de variables moderadoras también permite analizar dichos matices. En muchos estudios primarios se estudian las interacciones con las que se presentan estos efectos. El MA no solo permite sintetizar las estimaciones de dichas interacciones. También permite analizar moderadoras que son muy difíciles (si no imposibles) de analizar en los estudios primarios. Veamos también algunos ejemplos. En el MA mencionado anteriormente (Sirin, 2005) el autor examina el papel de diversas variables moderadoras, entre ellas la etapa escolar del estudiante. En este caso los resultados indican que la etapa escolar es un moderador efectivo de las relaciones entre las variables. Concretamente, el TE medio de los estudios realizados con alumnos de edades equivalentes a la ESO fue significativamente mayor que los del resto de niveles académicos. Esta es una moderadora que podría haber sido valorada en algún estudio primario, incluyendo muestras de diferentes etapas escolares. Por su parte, en el MA de Patall, Cooper y Robinson (2008) se analiza como moderadora una variable metodológica casi imposible de valorar en un estudio primario. En concreto, se valoraron las diferencias entre las esti-

maciones del TE obtenidas con estudios manipulativos aleatorizados frente a estudios aleatorizados por grupos escolares y frente a los realizados con metodología cuasi-experimental. Los resultados muestran que las diferentes metodologías proporcionan estimaciones distintas, aunque el pequeño número de estudios en alguna de las condiciones aconsejan prudencia en su interpretación.

## POR QUÉ Y PARA QUÉ HACER UN META-ANÁLISIS

La motivación más evidente para realizar un MA es el deseo de obtener una *estimación* de un TE basado en toda la evidencia disponible. Esa fue la primera función reconocida para esta metodología. Hoy en día podemos identificar algunas otras. Es importante que el autor deje bien explícito cuál es la motivación y el objetivo de su MA. De esa forma se comprenderán mejor las decisiones que adopta. Por ejemplo, cada vez es más difícil encontrar una pregunta de investigación para la que no se haya publicado un MA. Sin embargo, al cabo de algunos años la evidencia disponible puede haber aumentado considerablemente. A veces el objetivo de un MA no es otro que el de una *actualización* de los resultados de otro anterior, incorporando los estudios publicados desde entonces.

En otras ocasiones hay además (o en lugar de) el deseo de modificar algunos de los criterios de inclusión y exclusión de estudios, de los criterios empleados para la codificación de los estudios o de algunas decisiones adoptadas en su ejecución. Son ocasiones en las que se mantiene una posición crítica respecto a cómo se hizo el MA anterior. El objetivo es ofrecer un MA alternativo que a juicio de los autores será más adecuado, con una *reformulación* de sus conceptos y criterios. En otras ocasiones se trata de aclarar el papel de variables moderadoras no evaluadas hasta el momento, porque nunca se había creído que fueran relevantes o porque no había suficientes estudios con datos relevantes. La formulación de nuevas teorías o modelos puede llevar a valorar el papel de esas nuevas variables que hasta el momento no habían sido consideradas. El MA se emplea entonces para una *evaluación alternativa* de la heterogeneidad observada en los estudios.

El análisis de la heterogeneidad es seguramente el aspecto más atractivo de un MA para muchos investigadores. Es en este terreno en el que puede realmente hacer aportaciones más originales, proporcionando evidencia específica propia no contenida en los estudios primarios. Por supuesto, la mera estimación de un TE sigue teniendo varias utilidades. Por un lado, dichas estimaciones permiten establecer objetivos cuantitativos más realistas al introducir intervenciones educativas. Al diseñar programas de intervención y de las actividades de valoración de dichos programas se

deben establecer objetivos, preferiblemente en términos cuantitativos. Un MA puede ser una fuente de información para esta tarea. Es difícil anticipar el impacto cuantitativo sin una base empírica adecuada. Por otro lado, aportan una información crucial para establecer el tamaño muestral de futuros estudios primarios, de forma que sus contrastes tengan la potencia deseada (Faul, Erdfelder, Lang y Buchner, 2007). Una de las funciones de dichas estimaciones es aportar este dato, necesario para establecer el tamaño de las muestras para investigaciones futuras. Para dicha tarea se suele fijar una potencia deseable manteniendo una probabilidad de error tipo I (alfa). El procedimiento exige involucrar alguna estimación del TE.

Una vez decidido el efecto que se pretende estudiar y establecido el objetivo del MA hay que formar un equipo de trabajo. Es recomendable que en dicho equipo se integren al menos un experto en esta metodología y un experto en el tema al que se refiere. El primero ayudará a dar con seguridad los pasos necesarios. El segundo será imprescindible para adoptar múltiples decisiones que solo alguien con conocimiento y experiencia en el campo puede adoptar.

El MA es una poderosa herramienta para hacernos una idea de lo que a veces se llama el «estado del arte». Permite hacer esta valoración, lo que implica obtener estimaciones del TE, pero también la detección de lagunas.

El análisis de las moderadoras permite identificar asociaciones entre variables. No debemos olvidar que dichas asociaciones solo representan meras covariaciones. Incluso aunque en los estudios primarios la metodología sea manipulativa y permita realizar inferencias causales, en el MA no podemos hacerlo. Sin embargo, identificar variables con un papel de moderadoras de los efectos es una rica fuente de hipótesis para nuevas investigaciones primarias.

## FASES DE UN META-ANÁLISIS

Como ya hemos explicado con detalle en otros sitios cuáles son las fases en la realización de un MA aquí nos limitaremos a mencionarlas por encima (Botella y Sánchez-Meca, 2015; Sánchez-Meca y Botella, 2010). En ellas se aprecia el parecido y equivalencia con las fases en otros tipos de investigaciones.

Las fases que vamos a señalar no tienen que por qué realizarse de una forma estrictamente secuencial, de forma que cada una solo comience cuando haya terminado la anterior. Más bien, la experiencia nos dice que hay que estar dispuesto a volver a las fases anteriores a la vista de lo que se

va obteniendo en cada una. Hay que tomarlas de una forma flexible. Aun así, estas fases se refieren a lo que podríamos llamar un MA típico, que se realiza con el objetivo de sintetizar la evidencia sobre una determinada pregunta. Pero hay otras motivaciones que pueden propiciar otras secuencias de fases. Volveremos sobre ellas en el próximo apartado. Las fases en un MA son las siguientes.

- a) *Formulación del problema.* La pregunta habitual en un MA típico se refiere a la asociación entre dos variables. Se trata de responder a la pregunta de si dichas variables están relacionadas y, en caso de estarlo, cuál es la magnitud de dicha asociación. Las formulaciones demasiado vagas no son prácticas. En esta fase es necesario traducir esas formulaciones a definiciones operativas útiles. Por ejemplo, en uno de los primeros MA en el campo de la educación se trataba de responder a la pregunta de si el tamaño de los grupos docentes en las escuelas moderaba el éxito en los aprendizajes (Glass y Smith, 1979). La idea de que con grupos pequeños la enseñanza es más personalizada y por tanto más ajustada a los individuos, estaba (y está) en el ambiente de la educación. Se trataba de localizar la evidencia relevante y de dar una respuesta a la pregunta basada en toda esa evidencia. Tanto la variable supuestamente explicativa (el tamaño de los grupos) como la de resultado (el aprendizaje) deben ser traducidas de forma operativa (por ejemplo, la primera como el tamaño de los grupos docentes para al menos el 80 % de las horas en la escuela y la segunda como las calificaciones medias de los grupos en las asignaturas de mayor relevancia, como la lengua y las matemáticas).
- b) *Búsqueda de los estudios.* Una vez planteada la pregunta se puede proceder a localizar y reunir las fuentes de evidencia. Esta suele ser la fase más tediosa, sobre todo si hay abundante investigación sobre la cuestión planteada. Se suelen especificar unos criterios de inclusión y exclusión de los estudios en el MA, cuyo objetivo no es otro que conseguir una base de datos suficientemente homogénea como para permitir una generalización razonable. La búsqueda se debe hacer por todos los canales posibles, incluyendo tanto estudios publicados como no publicados. No hay que olvidar que muchas veces los estudios aportan evidencia relevante aunque sus objetivos fueran muy diferentes de los del propio meta-análisis. A veces es incluso mejor que sea así. Veamos de nuevo un ejemplo con el tamaño de los grupos escolares. Hay muchos estudios en los que se exploran variables asociadas al rendimiento académico, pero el tamaño del grupo no figura entre ellas. Sin embargo, si en los diferentes estudios hay grupos de control en los que se em-

plean métodos habituales, pero estos son de diferentes tamaños, entonces se pueden obtener estimaciones de la diferencia en el rendimiento para diferentes tamaños grupales. Esta perspectiva tiene la ventaja de que los autores que realizaron esos estudios no tenían preferencias ni ideas previas conocidas, ya que no era este el objetivo del estudio. Sin embargo, sus estimaciones pueden ser muy útiles para responder a la pregunta planteada.

- c) *Codificación de los estudios.* Es la fase de codificación de las características de los estudios primarios y su registro en la base de datos. Este proceso comienza con frecuencia con un listado de características y termina con otro, ya que al avanzar en este proceso se advierte la importancia de características no consideradas al principio o se decide redefinir las categorías o las formas de valorarlas.

Una forma de clasificar las características de los estudios que se codifican y registran es seguir el esquema de Lipsey (1994), que incluye: (1) variables de tratamiento, (2) variables de los participantes, (3) características del contexto, (4) características metodológicas, y (5) características extrínsecas al proceso científico. La base de datos organizada de esta forma será el material con el que se trabajará a partir de esta fase.

- d) *Análisis estadístico e interpretación.* Para empezar, se calcula la estimación o estimaciones independientes que aporta cada estudio del TE que refleja el efecto estudiado. Como ya hemos visto anteriormente en algunos ejemplos, con estos valores se obtendrá una estimación combinada y un intervalo de confianza. Con la estimación combinada, basada en todos los estudios, más su intervalo de confianza, ya se puede responder a las dos preguntas mencionadas más arriba. Si el intervalo incluye el valor de no efecto del índice de TE que se esté empleando (ya sea 0 o 1, dependiendo del índice elegido) se debe concluir que las variables no están asociadas. De lo contrario se concluye que sí lo están y se interpreta el intervalo como una conjectura inferencial de su magnitud. Si los indicadores de heterogeneidad (Huedo-Medina et al, 2006) y el modelo estadístico asumido así lo aconsejan, será el momento de analizar las fuentes de las variaciones observadas en los valores de TE. Para ello se ajustarán modelos categoriales y de meta-regresión (Botella y Sánchez-Meca, 2015).
- e) *Publicación.* El informe meta-analítico tiene algunas peculiaridades, propias de sus características especiales. Existen guías y reco-

mendaciones (Botella y Gambara, 2006; Clarke, 2009; Rosenthal, 1995). Como norma general, un buen informe meta-analítico será el que permita una más fiel replicabilidad del estudio.

## UNA HERRAMIENTA GRÁFICA: EL *FOREST PLOT*

En el contexto del MA se han desarrollado también algunas herramientas gráficas específicas. Si se comprenden bien pueden ser muy útil para ayudarnos a tener una visión de conjunto de lo que se recoge en una base de datos meta-analítica. Entre ellas destaca la figura conocida como *forest plot*.

El *forest plot* es una figura en la que se exponen las estimaciones puntuales y de intervalo de un determinado TE, aportados por los estudios integrados en un MA (figura 1). Aunque tiene diversos diseños alternativos y puede incorporar muchos elementos opcionales, lo más característico de esta figura es una rejilla en la que aparece, habitualmente en la parte inferior, una escala métrica del índice de TE empleado. Por encima de dicha escala se representan las estimaciones, una en cada línea horizontal. Cada estudio o estimación independiente se identifica al comienzo de cada línea. La estimación puntual suele ser un círculo o un cuadrado que se sitúa a la altura correspondiente de la escala. El cuadrado o círculo suele tener un tamaño proporcional al tamaño de la muestra empleada en el estudio. Sobre dicha representación se dibuja un segmento horizontal cuyos extremos coinciden en la escala con los límites del intervalo de confianza. En la parte inferior se dibuja un rombo horizontal cuyo centro coincide con la estimación combinada y sus extremos con los de su intervalo de confianza. Se suele añadir una línea vertical sobre el valor de no efecto. Algunos autores recomiendan incluir también una línea intermitente vertical sobre el valor de la estimación combinada. Los estudios suelen aparecer en orden alfabético (el mismo orden que en las referencias), pero esta ordenación tiene poca utilidad. Es mejor aprovechar el gráfico como una oportunidad para valorar visualmente la relación entre el TE y alguna variable moderadora o circunstancial. En el ejemplo de la figura 1 superior se muestra un *forest plot* con 20 estudios con la diferencia de medias tipificada como índice de TE. Así, el estudio de Anterich et al aportó una estimación de 0,30 con un intervalo de confianza al 95 % de 0,05 – 0,55. La estimación combinada resultó ser 0,38 con un intervalo de confianza al 95 % entre 0,33 y 0,43. Mientras en la figura 1 superior los estudios aparecen en orden alfabético, en la figura 1 inferior aparecen ordenados por el sexo (porcentaje de mujeres en la muestra). De esta forma se aprecia la relación entre el TE y dicha moderadora (mayor TE cuanto mayor es la presencia proporcional de mujeres).

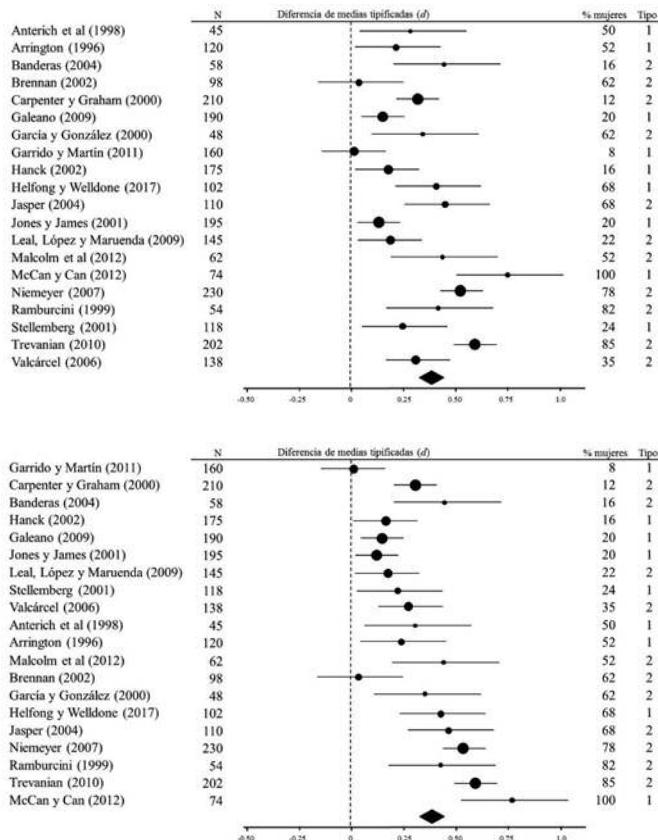


Figura 1. Dos figuras *forest plot*, con los estudios ordenados alfabéticamente (figura superior) y según el porcentaje de mujeres en la muestra (figura inferior)

En educación contamos con múltiples ejemplos de ordenación de los estudios. Aparte del orden alfabético tradicional (e.g., Klassen y Tze, 2014; Steffgen, Recchia, y Viechtbauer, 2013) se han ordenado por el valor del TE (e.g., Schwinger, Wirthwein, Lemmer y Steinmayr, 2014) o por la fecha de publicación (e.g., Zheng, Warschauer, Lin y Chang, 2016).

## LAS PRÁCTICAS META-ANALÍTICAS EN EDUCACIÓN

Todo este bagaje se ha traducido en unas prácticas muy consolidadas en educación. Para hacerse una idea del papel que esta metodología desempeña hoy en educación recomendamos dar un repaso al libro *Visible Learning* (Hattie, 2009). En él se aborda la cuestión del rendimiento académico empleando como evidencia la que proporcionan más de 800 MAs relevantes para su objetivo. Un simple hojeo de esta obra nos da una idea del impo-

nente papel que el MA tiene hoy en educación, teniendo en cuenta que se restringe exclusivamente al rendimiento académico.

Pero también es muy recomendable una lectura del trabajo de Ahn, Ames y Myers (2012), en el que se ofrece una revisión crítica de la forma en que se están realizando los MAs en educación. Seleccionaron 56 MAs publicados en algunas de las principales revistas de la disciplina y analizaron sus prácticas y la forma en que han incorporado los desarrollos más recientes. Sus resultados muestran que las temáticas de estos MAs son muy variadas. Igualmente, los métodos son también variados y dependen mucho del problema que se esté abordando.

Su valoración es en general positiva, especialmente en cuanto a las primeras fases que hemos explicado anteriormente. Sin embargo, señalan como áreas de mejora la evaluación y análisis de los datos. Por ejemplo, con frecuencia no se especifica el modelo estadístico asumido (efecto fijo o efectos aleatorios) ni su justificación. Lo mismo ocurre con ciertos aspectos del análisis, como la cuestión de las comparaciones múltiples a posteriori o el análisis de la potencia de los contrastes. También falta a veces un análisis de la calidad de los estudios primarios y la valoración del sesgo de publicación como amenaza.

Hay que distinguir también entre las deficiencias de los MAs y las de sus informes. A veces las prácticas son correctas pero se informan de manera deficiente. En general, los defectos del informe son los que limitan la replicabilidad. Muchos defectos no se pueden resolver en el MA, ya que aparecen en origen en el estudio primario. Por otro lado, es importante incluir todo lo que pueda ayudar al lector a valorar las limitaciones del MA, especialmente el grado en que las amenazas a su validez están presentes. Es recomendable emplear alguna guía de referencia para adoptar las decisiones en la elaboración del informe. Por ejemplo, se ha desarrollado para las ciencias sociales, especialmente la psicología, la guía MARS (*meta-analysis reporting standards*; American Psychological Association, 2008), que sirve como lista de comprobación.

## **EXTENSIONES Y LÍNEAS ACTUALES DE DESARROLLO**

Ya hemos visto cómo originalmente el MA nació en contextos en los que se buscaba estimar la magnitud de un determinado TE y analizar sus fuentes de variabilidad. Posteriormente se han ido generando formas específicas de MA cuyos objetivos son diferentes pero que se basan en la misma filosofía.

Como ejemplo de ello vamos a destacar los estudios de generalización de la fiabilidad (*reliability generalization*). Son estudios de MA en los que el foco

de interés es la fiabilidad de las medidas (Vacha-Haase, 1998). Contra lo que muchas personas creen, la fiabilidad no es una propiedad de los instrumentos, sino de las puntuaciones. Un mismo test puede proporcionar puntuaciones con niveles de fiabilidad muy distintos cuando se aplican a personas pertenecientes a subpoblaciones con características específicas o cuando se aplican en contextos especiales. En muchas investigaciones se justifica la elección de un test para medir una característica invocando las propiedades psicométricas que recoge el manual y que fueron estimadas en el proceso de baremación. Sin embargo, ese mismo test puede proporcionar puntuaciones que no son homogéneas en su fiabilidad. Es posible que en determinadas subpoblaciones los niveles de fiabilidad sean sistemáticamente superiores o inferiores. Conociendo estas matizaciones se pueden adoptar mejores decisiones a la hora de seleccionar un instrumento de medida. Los estudios de generalización de la fiabilidad sintetizan la información relativa a las estimaciones de la fiabilidad que ofrecen los estudios primarios (Botella, Suero y Gambara, 2010). La mayoría de los estudios en los que se emplea un test no ofrecen el valor del coeficiente de fiabilidad obtenido en su muestra, a pesar de que es una práctica insistentemente recomendada. Con los estudios que sí lo hacen se puede proceder con la misma filosofía que en cualquier MA. Por un lado, obtener estimaciones combinadas de la fiabilidad, con su intervalo de confianza. Por otro, estudiar las fuentes de variabilidad que pudieran explicar las variaciones observadas en los coeficientes localizados. Naturalmente, en la mayoría de las ocasiones serán coeficientes de consistencia interna, mientras que la fiabilidad test-retest y la de formas paralelas se evalúan e informan con mucha menos frecuencia.

Los estudios de generalización de la fiabilidad pueden referirse a instrumentos que miden características propias del aprendizaje, como por ejemplo serían las puntuaciones que proporcionan instrumentos como la Escala de Eficacia para el Profesor (*TES*; Henson, Kogan y Vacha-Haase, 2001) o el cuestionario *SPQ* para valorar la dimensión superficial/profunda del enfoque del aprendizaje (Biggs, Kember y Leung, 2001). Pero también puede referirse a características de otro tipo cuya relación con el aprendizaje se quiere explorar, como por ejemplo algunas características psicológicas o el test *How I Think (HIT)* (Gini y Pozzoli, 2013) para evaluar distorsiones cognitivas. Otro ejemplo podría ser la escala de identidad étnica, un constructo que tiene repercusiones relevantes en el rendimiento académico. En el estudio de generalización de la fiabilidad de Herrington et al (2016) sobre esta escala se concluye que en general es un test que proporciona puntuaciones fiables y bastante uniformes. Sin embargo, dicha fiabilidad disminuye cuando se aplica a personas con niveles educativos bajos.

Pero además los instrumentos pueden referirse tanto a personas como a centros. Así, Anafarta e Irmak (2014) realizan un estudio de generalización de la fiabilidad del *SERVQUAL*, un instrumento para valorar de forma

multidimensional la calidad de un servicio. En su trabajo seleccionan aplicaciones de este instrumento a centros de educación superior por todo el mundo. Las estimaciones combinadas fueron adecuadas en su magnitud, pero mostraban altos niveles de heterogeneidad. Los autores exploran la posible asociación entre los valores y algunas variables moderadoras, como el sexo, el país, el idioma, el número de puntos y el número de ítems (había una versión acortada), pero ninguna de ellas mostró una asociación significativa.

En esta misma línea hay que considerar a los estudios que valoran los instrumentos sencillos de cribaje (*screening*). Con frecuencia se buscan indicadores simples que permitan anticipar situaciones indeseadas en el desarrollo del aprendizaje, con objeto de intervenir preventivamente. Por ejemplo, muchas instituciones de educación superior tratan de identificar cuanto antes a los estudiantes de nuevo ingreso que probablemente tendrán dificultades para lograr un desempeño adecuado. El objetivo es poder intervenir y así impedir que se llegue a esas situaciones. Adviértase que el objetivo no es otro que hacer una simple clasificación dicotómica que permita identificar a los estudiantes «en riesgo», normalmente para derivarlos. Un ejemplo son los llamados *Curriculum-Based Measurements of Oral Reading*, que se aplican en los primeros cuatrimestres de una carrera. Se trata de valorar la capacidad de estos tests para una clasificación efectiva, lo que supone obtener una alta *sensibilidad* (probabilidad de un resultado positivo en el test en personas que realmente fracasarán) y una alta *especificidad* (probabilidad de un resultado negativo en el test en personas que no fracasarán). Si ambas medidas se combinan con la prevalencia del suceso (fracaso académico) se obtienen otros índices con mucha utilidad práctica: el *valor predictivo positivo* y el *valor predictivo negativo*. Cuando un instrumento se hace suficientemente popular comienzan a aparecer estudios en los que se estima su capacidad diagnóstica. De nuevo entra en juego el MA ofreciendo técnicas para sintetizar la información y para explorar las fuentes de variabilidad en los índices correspondientes (Botella y Huang, 2012; véase un MA alternativo en Reschly, Busch, Betss, Deno y Long, 2009).

Hoy en día el MA sigue su desarrollo, avanzando en múltiples líneas y afrontando nuevas preguntas. Destacaremos aquí algunos de los que nos parecen más importantes. En primer lugar, hay que destacar el papel que el MA está desempeñando en el cuestionamiento que los científicos están haciendo de su propio trabajo. En el último decenio se ha producido lo que se ha venido en llamar una «crisis de confianza» en lo que los científicos hemos venido haciendo (Baker, 2016). La metodología de investigación se ha aplicado a veces de forma poco rigurosa y asumiendo lo que se han llamado «prácticas cuestionables» (e.g., véase Pashler y Wagenmakers, 2012 y el número especial que presentan). Dichas prácticas han producido mucho ruido y una pérdida de confianza en sus resultados y conclusiones. Ejemplo

de ello es el sesgo de publicación, que revela una pretensión de publicar los trabajos más informativos y diferenciados, en lugar de aquellos que suman a los desarrollos previos. Aparte de los defectos específicos que la forma de aplicar el método en cada disciplina pueda haber desarrollado, también hay factores externos a la ciencia que sin duda han tenido un efecto colateral indeseado de fomento de dichas prácticas. En particular, se ha producido una fuerte presión a publicar y a producir tomando como referencia los indicadores de impacto que han elaborado las agencias evaluadoras. Sin duda que esta presión, en un movimiento generalizado a nivel internacional y en todas las disciplinas, ha tenido el efecto beneficioso de que la investigación ha crecido de forma espectacular. Sin embargo, la presión para publicar ha tenido también el efecto de que se han publicado muchos trabajos intrascendentes o cerrados con demasiada prisa. El MA ha aportado herramientas para valorar los efectos de dichas prácticas y así poder evaluar la solidez del cuerpo de evidencias proporcionadas en un campo particular.

Respecto a los desarrollos más técnicos, de procedimientos cada vez más sofisticados, podemos citar el MA multivariante o el MA en red. El primero se refiere a la síntesis de resultados que no se refieren a la asociación entre dos variables, sino a la interrelación entre varias variables. En términos teóricos se había planteado hace tiempo (e.g., Becker y Schram, 1994) pero recientemente el trabajo de Cheung ha relanzado esta línea con la síntesis de evidencia en términos de modelos de ecuaciones estructurales (Cheung, 2015). El segundo se refiere a una situación particular en la que muchas intervenciones son comparadas en diferentes estudios con un mismo grupo de control (no tratado, placebo o con intervención «habitual»), pero nunca o casi nunca se comparan directamente entre sí. El MA en red (e.g., Higgins y Welton, 2015) aporta un modelo y las herramientas para poder tratar la información sin perder la capacidad de inferencia causal que los diseños originales permitían.

Igualmente, se sigue investigando en procedimientos para afrontar el problema del sesgo de publicación, tanto para su detección como para su valoración y, si es el caso, su corrección. Uno de los métodos que están recibiendo más atención es el que se auto-restringe directamente a solo los estudios significativos, mediante el método conocido como *p-curve* o *p-uniform* (van Assen, van Aert y Wicherts, 2015; véase en Blázquez, Botella y Suero, 2017, su papel en un debate metodológico).

## **CRÍTICAS Y DEBILIDADES DEL MA**

Pero el meta-análisis no es tampoco una metodología perfecta. También sufre amenazas que es necesario conocer para poder emplearla de

forma juiciosa. Algunas de sus debilidades han sido señaladas casi desde su nacimiento. Vamos a mencionar dos: el problema de la heterogeneidad y el sesgo de publicación.

A veces se dice que en meta-análisis se mezclan peras con manzanas y que esta heterogeneidad resta solidez a sus conclusiones. Esta crítica puede ser muy acertada, pero no debe referirse a la metodología como tal sino a la aplicación inadecuada que a veces se ha hecho de ella. Si se mezclan estudios que se refieren a constructos muy heterogéneos o mal definidos, al final las conclusiones no son claras. Por eso es importante que el meta-analista defina con claridad los límites de los constructos involucrados en sus síntesis. Cuando estos no están bien definidos o no se respetan en alguna fase del MA, se estará haciendo una aplicación defectuosa del MA, tanto en sus operaciones como en la filosofía que la inspira. Al valorar la calidad de un MA se debe prestar atención a la cuestión de la heterogeneidad, pues es una amenaza real a la validez de sus conclusiones.

El sesgo de publicación es un fenómeno bien conocido desde hace mucho tiempo. Con este término se hace referencia a que no todos los estudios realizados acaban siendo ‘visibles’ (esencialmente, publicados) y a que los que sí lo son no suponen una muestra aleatoria de los estudios realizados. La existencia de algún criterio que facilite la publicación haría que lo visible fuera una muestra sesgada de lo realizado. El criterio más conocido es el de la significación estadística. Si se publican sobre todo los estudios con resultados estadísticamente significativos al final la conclusión está sesgada hacia estos y parece que los efectos sean más grandes y sólidos de lo que realmente son. Cuando se habla de sesgo de publicación, sin apellidos, se sobreentiende que se refiere a una mayor probabilidad de publicación para los estudios significativos.

Sabemos que existe sesgo de publicación en muchas disciplinas y la educación no es una excepción. Pigott, Valentine, Polanin, Williams y Canada (2013) nos proporcionan una evidencia directa de este fenómeno en el ámbito de la educación analizando una muestra de tesis publicadas posteriormente como artículos científicos. Comprobaron que al trasladar los resultados a un artículo es más probable que en este último aparezcan los resultados de las tesis si fueron significativos que si no lo fueron.

En un interesante trabajo Banks, Kepes y Banks (2012) abordan el problema del sesgo de publicación, también en el contexto de la educación. Tras revisar los principales procedimientos para afrontarlo los ilustran aplicándolos a las bases de datos de dos meta-análisis relacionados con la alfabetización. El dinamismo en este área de investigación en MA es tal que tras solo cuatro años desde su publicación ya disponemos de nuevos mé-

todos que estos autores no incluyen. Dos buenos ejemplos son los métodos de meta-regresión (Stanley y Doucouliagos, 2014) o el método *p-curve* (van Assen, van Aert y Wicherts, 2015). En cualquier caso, los resultados que obtienen con los métodos más clásicos demuestran que el impacto del sesgo de publicación puede ir desde que sea de magnitud despreciable hasta suponer un impacto relativamente importante en la estimación del TE. Concluyen sugiriendo que la recomendación de que en cada MA se analice la presencia de esta amenaza (*American Psychological Association*, 2010) se convierta en un requisito para su publicación.

El meta-análisis no es responsable de que se produzca un sesgo de publicación ni este fenómeno es intrínseco a él. Muy al contrario, es la única metodología que se ha ocupado de desarrollar procedimientos para detectarlo, para valorar su magnitud e incluso para corregirlo. Aunque parezca paradójico, la cuestión del sesgo de publicación es en realidad una de las fortalezas del MA.

## ALGUNOS RECURSOS PRÁCTICOS

Muchas veces lo que dificulta hacer un largo viaje es que no sabemos muy bien cómo dar el primer paso, cómo empezar. Para facilitar un poco las cosas vamos también a facilitar la identificación de algunos recursos prácticos de varios tipos: bibliográficos, programas informáticos y recursos en *internet*.

Hay mucha bibliografía sobre MA. En castellano debemos recomendar, modestamente, el libro de Botella y Sánchez-Meca (2015) que cubre prácticamente todas las necesidades para realizar la mayoría de los MA que podrían hacerse en el ámbito de la educación. En inglés hay mucho más. El *Handbook of Research Synthesis* (Cooper, Hedges y Valentine, 2009) es el volumen que todo usuario de esta metodología debería tener a su alcance como libro de consulta. Para aprender sobre MA hay otros libros más didácticos, como el de Borenstein, Hedges, Higgins y Rothstein (2009) o el de Lipsey y Wilson (2001). Disponemos de otras fuentes para otros temas más específicos, como el MA multivariado (Cheung, 2015) o el sesgo de publicación (Rothstein, Sutton y Borenstein, 2005).

Respecto a los programas para hacer MA también tenemos muchos recursos. Por un lado, las macros de SPSS o SAS desarrollados por Lipsey y Wilson para acompañar su libro. Se pueden descargar gratuitamente de la página de David Wilson.

Son el recurso más recomendable si se va a aprender MA con su libro. Una alternativa más completa, pero comercial (y no precisamente barata), es el *Comprehensive Meta-analysis* (Borenstein, Hedges, Higgins y Rothstein, 2013). Este programa ha sido desarrollado por Borenstein y su equipo, de forma que si el lector va a aprender MA con su manual esta es la mejor sugerencia informática. Todos sus ejemplos están hechos con su propio programa. Hay también varios paquetes en *R* (R Development Core Team, 2010) y por tanto gratuitos. Uno de los más completos y utilizados es *METAFOR* (Viechtbauer, 2010), pero cada vez hay más alternativas (Polanin et al., 2016). Aunque sean menos sofisticados, hay también alternativas que funcionan directamente en *Excel* (Neyeloff et al., 2012).

Antes de analizar los datos habrá que obtener los valores de las estimaciones del TE. Algunos programas como el *Comprehensive Meta-analysis* contienen módulos específicos para esta tarea. Si no se dispone de este programa se puede recurrir a *METAFOR*. También disponemos de calculadoras on-line que nos permiten hacer los cálculos, como *ES-calculator*, que puede obtenerse en la página de David Wilson que ya hemos indicado. Para realizar MA multivariados disponemos de programas también en *R*, como *MetaSEM* (Cheung, 2015).

## NOTAS<sup>1</sup>

<sup>1</sup> <http://mason.gmu.edu/~dwilsonb/ma.html>

## REFERENCIAS BIBLIOGRÁFICAS

- Ahn, S., Ames, A. & Myers, N. (2012). A Review of Meta-Analyses in Education: Methodological Strengths and Weaknesses. *Review of Educational Research*, 82(4), 436-476.
- American Psychological Association (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist*, 63, 839-851.
- American Psychological Association (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.
- Anafarta A. & Irmak, S. (2014). A Reliability Generalization Study of The Service Quality Scale (Servqual) in *The Clute Institute International Business Conference*, Múnchen, Alemania, 534-541.
- Baker, M. (2016). Is there a reproducibility crisis? *Nature*, 533(7604), 452-454.
- Banks, G., Kepes, S. & Banks, K. (2012). Publication Bias: The Antagonist of Meta-Analytic Reviews and Effective Policymaking. *Educational Evaluation and Policy Analysis*, 34(3), 259-277.
- Becker, B. J. & Schram, C. M. (1994). Examining explanatory models through research synthesis. En H. Cooper & L. V. Hedges (Eds.). *The handbook of research synthesis* (pp. 357-381). Nueva York: Russell Sage Foundation.
- Biggs J., Kember D. & Leung D. Y. P. (2001). The revised two-factor Study Process Questionnaire: R-SPQ-2F. *British Journal of Educational Psychology*, 71(1), 133-149.
- Blázquez, D., Botella, J. & Suero, M. (2017). The debate on the ego-depletion effect: Evidence from meta-analysis with the p-uniform method. *Frontiers in Psychology*, 8, 197.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: John Wiley and sons.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. (2010). A basic introduction to fixed-effects and random-effects models for meta-analysis. *Research Synthesis Methods*, 1, 97-111.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. (2013). *Comprehensive Meta-analysis Version 3.0*. Biostat Inc.
- Botella, J. & Gambara, H. (2006). Doing and reporting a meta-analysis. *International Journal of Clinical and Health Psychology*, 6, 425-440.
- Botella, J. y Huang, H. (2012). Procedimientos para realizar meta-análisis de la precisión de instrumentos de clasificación binaria. *Psicothema*, 24(1), 133-141.
- Botella, J. y Sánchez-Meca, J. (2015). *Meta-análisis en Ciencias Sociales y de la Salud*. Madrid: Editorial Síntesis.
- Botella, J., Suero, M. & Gambara, H. (2010). Psychometric inferences from a meta-analysis of reliability and internal consistency coefficients. *Psychological Methods*, 15, 386-397.
- Cheung, M. W. L. (2015). *Meta-analysis: A structural equation modeling approach*. Chichester, Reino Unido: John Wiley & Sons.
- Clarke, M. (2009). Reporting format. En H. Cooper, L. V. Hedges y J. C. Valentine (Eds.) (2009), *The handbook of research synthesis and meta-analysis* 2.<sup>a</sup> ed. (pp. 521-534). Nueva York: Russell Sage Foundation.

- Cooper, H., Hedges, L. V. & Valentine, J. C. (2009). *The Handbook of Research Synthesis and Meta-analysis*. Nueva York: Russell Sage Foundation.
- Faul, F., Erdfelder, E., Lang, A. G. y Buchner, A. (2007). G\* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2), 175-191.
- Gini, G. & Pozzoli, T. (2013). Measuring self-serving cognitive distortions: A meta-analysis of the psychometric properties of the How I Think Questionnaire (HIT). *European Journal of Developmental Psychology*, 10(4), 510-517.
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5, 3-8.
- Glass, G. V. (2016). One Hundred Years of Research Prudent Aspirations. *Educational Researcher*, 45(2), 69-72.
- Glass, G. V. y Smith, M. K. (1979). Meta-analysis of research on class size and achievement. *Educational Evaluation and Policy Analysis*, 1, 2-16.
- Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Henson, R. K., Kogan, L. R., & Vacha-Haase, T. (2001). A reliability generalization study of the teacher efficacy scale and related instruments. *Educational and Psychological Measurement*, 61(3), 404-420.
- Herrington, H. M., Smith, T. B., Feinauer, E. & Griner, D. (2016). Reliability generalization of the Multigroup Ethnic Identity Measure-Revised (MEIM-R). *Journal of counseling psychology*, 63(5), 586-593.
- Higgins, J. P. & Welton, N. J. (2015). Network meta-analysis: a norm for comparative effectiveness? *The Lancet*, 386(9994), 628-630.
- Huedo-Medina, T., Sánchez-Meca, J., Marín-Martínez, F. & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistics or I<sup>2</sup> index? *Psychological Methods*, 11, 193-206.
- Klassen, R. M. & Tze, V. M. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review*, 12, 59-76.
- Lipsey, M. W. (1994). Identifying potentially interesting variables and analysis opportunities. En H. M. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 111-123). New York: Sage Foundation.
- Lipsey, M. W. y Wilson, D. B. (2001). *Practical meta-analysis*. Thousand oaks, CA: Sage Pub.
- Neyeloff, J. L., Fuchs, S. C., & Moreira, L. B. (2012). Meta-analyses and Forest plots using a microsoft excel spreadsheet: step-by-step guide focusing on descriptive data analysis. *BMC research notes*, 5(1), 1.
- Pashler, H. y Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7(6), 528-530.
- Patall, E. A., Cooper, H. y Robinson, J. C. (2008). Parent involvement in homework: A research synthesis. *Review of educational research*, 78(4), 1039-1101.
- Pigott, T., Valentine, J., Polanin J., Williams, T. y Canada, D. (2013). Outcome-Reporting Bias in Education Research. *Educational Researcher*, 42(2), 424-432.

- Polanin, J. R., Hennessy, E. A., & Tanner-Smith, E. E. (2016). A Review of Meta-Analysis Packages. *R Journal of Educational and Behavioral Statistics*. Prepublished November 30, 2016. Doi 1076998616674315.
- R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL <http://www.R-project.org/>.
- Reschly, A. L., Busch, T. W., Betts, J., Deno, S. L. & Long, J. D. (2009). Curriculum-based measurement oral reading as an indicator of reading achievement: A meta-analysis of the correlational evidence. *Journal of School Psychology*, 47(6), 427-469.
- Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin*, 118, 183-192.
- Rothstein, H. R., Sutton, A. J. y Borenstein, M. (2005). *Publication Bias in Meta-Analysis*. Chichester: John Wiley & Sons.
- Sánchez-Meca, J. y Botella, J. (2010). Revisiones sistemáticas y meta-análisis: herramientas para la práctica profesional. *Papeles del Psicólogo*, 31(1), 7-17.
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of educational research*, 75(3), 417-453.
- Smith, M. y Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32, 752-760.
- Stanley, T. D. y Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, 5(1), 60-78.
- Steffgen, G., Recchia, S. & Viechtbauer, W. (2013). The link between school climate and violence in school: A meta-analytic review. *Aggression and Violent Behavior*, 18(2), 300-309.
- Schwinger, M., Wirthwein, L., Lemmer, G. & Steinmayr, R. (2014). Academic Self-Handicapping and Achievement: A Meta-Analysis. *Journal of Educational Psychology*, 106(3), 744-761.
- Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement*, 58(1), 6-20.
- Van Assen, M. A., Van Aert, R. & Wieters, J. M. (2015). Meta-analysis using effect size distributions of only statistically significant studies. *Psychological methods*, 20(3), 293.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1-48.
- Zheng, B., Warschauer, M., Lin, C. H. & Chang, C. (2016). Learning in One-to-One Laptop Environments A Meta-Analysis and Research Synthesis. *Review of Educational Research*, 86(4), 1052-1084.

## PERFIL ACADÉMICO Y PROFESIONAL

Juan Botella. Doctor en psicología por la Universidad Autónoma de Madrid, en cuya facultad de psicología es catedrático de metodología de las ciencias del comportamiento. Ha participado en numerosos proyectos de investigación sobre diferentes temáticas propias de la psicología, especialmente la atención y su relación con otros procesos cognitivos. Actualmente dirige o participa en varios proyectos de investigación relacionados con el meta-análisis y la síntesis de evidencia científica.

Ángela Zamora. Doctora en Ciencias de la Educación por la Universidad Nacional de Educación a Distancia (UNED). Actualmente ejerce su labor docente en un centro adscrito a la Universidad de Oviedo y es profesora tutora en el Centro Asociado de la UNED en Asturias. Sus intereses en investigación se centran en el estudio de los errores en el proceso de enseñanza-aprendizaje y su relación con la autorregulación, la motivación y las estrategias de aprendizaje.

Dirección de los autores: Juan Botella

Universidad Autónoma de Madrid  
Facultad de Psicología  
Campus de Cantoblanco  
C/ Ivan Pavlov, 6  
28049 Madrid

Ángela Zamora  
UNED  
CA de Asturias  
Avenida del Botánico 1435  
Gijón - Asturias  
E-mail: angzamora@gijon.uned.es

Fecha Recepción del Artículo: 31. Enero. 2017

Fecha Aceptación del Artículo: 01. Marzo. 2017

Fecha Revisión para Publicación: 06. Marzo. 2017