

Cómo realizar una revisión sistemática y un meta-análisis*

Julio Sánchez-Meca
Universidad de Murcia

Las revisiones sistemáticas y, en especial, los meta-análisis son un tipo de investigación científica que tiene como propósito integrar de forma objetiva y sistemática los resultados de los estudios empíricos sobre un determinado problema de investigación, con objeto de determinar el 'estado del arte' en ese campo de estudio. Para alcanzar este objetivo, la realización de un meta-análisis requiere desarrollar una serie de etapas similares a las de cualquier investigación empírica: (1) formulación del problema, (2) definición de los criterios de inclusión y búsqueda de los estudios, (3) codificación de las características de los estudios que puedan moderarlos resultados; (4) cálculo del tamaño del efecto; (5) técnicas de análisis estadístico e interpretación y (6) publicación del meta-análisis. Dada la relevancia que está adquiriendo esta metodología, el propósito de este artículo es acercar las revisiones sistemáticas y los meta-análisis a los profesionales de cualquier disciplina empírica. Con este propósito, en este artículo se describen las fases en que se lleva a cabo un meta-análisis y se ilustran con un ejemplo real del ámbito de las ciencias de la educación.

Palabras clave: revisión sistemática, meta-análisis, síntesis de la investigación, tamaño del efecto, calidad de la investigación.

How to carry out a systematic review and a meta-analysis. Systematic reviews and, in particular, meta-analyses are a kind of scientific research aimed to objectively and systematically integrate the results of a set of empirical studies about a given research problem, with the purpose of determining the 'state of the art' in that research field. To accomplish that objective, carrying out a meta-analysis requires to develop a series of phases very similar to those of any empirical research: (1) formulating the problem, (2) defining the inclusion criteria and searching for the studies; (3) coding the study characteristics that can moderate the results, (4) calculating an effect size index, (5) defining the statistical analysis techniques and interpreting their results, and (6) publishing the meta-analysis. Due to the relevance that this methodology is reaching, the purpose of this article was to approximate systematic reviews and meta-analyses to professionals in every empirical discipline. With this vein, in this article the phases of a meta-analysis are described and they are also illustrated with a real example from the education sciences.

Keywords: systematic review, meta-analysis, research synthesis, effect size, research quality.

Introducción

Hasta hace relativamente poco tiempo, la revisión de la literatura científica sobre un

tópico en el ámbito de las ciencias sociales y de la salud era un proceso que descansaba en la subjetividad del revisor y en el que no existían unas normas que garantizaran su objetividad. Esta práctica llevó a la conclusión de que en las ciencias sociales y de la

Correspondencia: Julio Sánchez-Meca
Departamento de Psicología Básica y Metodología
Facultad de Psicología
Universidad de Murcia
Campus de Espinardo, 30100 Murcia
E-mail: jsmecca@um.es

* Este trabajo ha sido financiado por un Proyecto de Investigación del Ministerio de Ciencia e Innovación y los Fondos FEDER (Proyecto n.º: PSI2009/12172).

salud el conocimiento científico se acumulaba muy pobremente, resultando difícil alcanzar conclusiones sólidas sobre las evidencias científicas obtenidas en las investigaciones (Rosenthal, 1991). Las dificultades de las revisiones subjetivas se acrecentaron conforme se produjo un incremento exponencial del volumen de estudios empíricos que se publican sobre cualquier ámbito. Esta explosión de la literatura científica que se produjo a partir de la década de 1970 hizo prácticamente imposible la tarea de revisar la literatura científica sobre un tópico si no es con la ayuda de estrategias objetivas y sistemáticas.

Como una alternativa a las revisiones subjetivas, también denominadas narrativas, surgió el meta-análisis, y las revisiones sistemáticas, como una metodología objetiva y rigurosa para llevar a cabo el proceso de revisión de la investigación en un campo concreto de conocimiento y lograr de esta forma una eficiente acumulación de las evidencias (Hunt, 1997). Esta revolución en el proceso de revisión de la literatura tuvo lugar a principios de la década de 1980, con los trabajos, entre otros, de Gene V. Glass sobre la eficacia de la psicoterapia (Smith y Glass, 1977) y sobre el efecto del tamaño de los grupos sobre el rendimiento de los alumnos (Smith y Glass, 1983), así como los de John E. Hunter y Frank L. Schmidt sobre la validez de los tests de selección de personal (Schmidt y Hunter, 1977), los de Janet S. Hyde sobre las diferencias sexuales en diversas habilidades cognitivas (Hyde y Linn, 1986), los de Robert Rosenthal sobre los efectos de las expectativas del experimentador sobre los resultados de las investigaciones (Rosenthal y Rubin, 1978), o los de Stephen W. Raudenbush sobre los efectos de las expectativas del maestro sobre el rendimiento escolar esperado de sus alumnos (Raudenbush, 1984).

Una *revisión sistemática* es un tipo de investigación científica mediante la cual se revisa la literatura científica sobre un tópico partiendo de una pregunta formulada de forma clara y objetiva, utilizando métodos sistemáticos y explícitos para localizar, selec-

cionar y valorar críticamente las investigaciones relevantes a dicha pregunta y aplicando protocolos sistemáticos para la recogida de datos e información de dichas investigaciones, con el objetivo de alcanzar conclusiones válidas y objetivas sobre qué es lo que dicen las evidencias sobre dicho tópico. En palabras de Last (2001, pp. 176-177), una revisión sistemática “es la aplicación de estrategias que limitan la comisión de sesgos al integrar, analizar críticamente y sintetizar todos los estudios relevantes sobre un tópico. El meta-análisis puede, aunque no necesariamente, formar parte de este proceso”.

Si en una revisión sistemática se obtiene un índice cuantitativo de la magnitud del efecto que cada estudio está investigando y se aplican técnicas de análisis estadístico para integrar dichos efectos, entonces la revisión sistemática se denomina meta-análisis. Así pues, podemos definir el *meta-análisis* como “el análisis estadístico de una gran colección de resultados de trabajos individuales con el propósito de integrar los hallazgos obtenidos” (Glass, 1976, p. 3), o también como “la síntesis estadística de los datos de estudios diferentes pero similares, es decir, estudios comparables, que proporciona un resumen numérico de los resultados globales” (Chalmers, Hedges y Cooper, 2002, p. 17).

En lo que sigue, se presenta una descripción de cuáles son las fases en que se lleva a cabo un meta-análisis o una revisión sistemática. Para profundizar en esta metodología existen actualmente numerosos textos a los que remitimos al lector interesado en profundizar más allá de los contenidos de este artículo (Borenstein, Hedges, Higgins y Rothstein, 2009; Botella y Gambara, 2002; Cooper, 2010; Cooper, Hedges y Valentine, 2009; Hedges y Olkin, 1985; Higgins y Green, 2008; Lipsey y Wilson, 2001; Martín, Tobías y Seoane, 2006; Torgerson, 2003).

Fases de un meta-análisis

La realización de un meta-análisis o de una revisión sistemática pasa necesariamente por las mismas fases que las que se requieren para llevar a cabo un estudio empí-

rico. No obstante, dado que la unidad de análisis en un meta-análisis es el estudio empírico, mientras que en los estudios empíricos la unidad de análisis suelen ser los participantes seleccionados a partir de una o varias poblaciones, las fases de un meta-análisis presentan ciertas peculiaridades que es preciso clarificar. En términos generales, podemos plantear la realización de un meta-análisis siguiendo estas seis etapas:

1. Formulación del problema
2. Búsqueda de los estudios
3. Codificación de los estudios
4. Cálculo del tamaño del efecto
5. Análisis estadístico e interpretación
6. Publicación del meta-análisis

Formulación del problema

Como en cualquier investigación empírica la primera fase consiste en plantear el problema que se pretende investigar. Un meta-análisis no es distinto de los demás tipos de investigación a este respecto. En primer lugar, debe formularse de forma clara la pregunta que se pretende responder, así como definir los constructos y conceptos implicados en la misma. De la formulación de la pregunta surgen a continuación los objetivos que se pretenden alcanzar con el meta-análisis y, en su casos, las hipótesis que se quieren contrastar (Cooper, 2010).

En el ámbito de las ciencias sociales y de la salud, incluidas las ciencias de la educación, el meta-análisis se está aplicando para responder a preguntas de muy diversa índole, pero la más común es evaluar la eficacia de programas, tratamientos e intervenciones en general para prevenir, resolver o paliar problemas educativos, psicológicos, de salud o sociales. Como ejemplo ilustrativo vamos a describir el meta-análisis realizado por Erion (2006) para evaluar la eficacia de los programas de tutorización parental en casa como un medio de mejorar el rendimiento escolar de los hijos. Los objetivos concretos que perseguían con el meta-análisis eran: (a) ¿cuál es la efectividad general de la tutorización parental?; (b) ¿están los resultados de los programas relacionados

con la duración de los mismos, así como con características tales como la disponibilidad de instrucciones escritas, de práctica supervisada o de realizar consultas al profesor? y (c) ¿están los resultados de los programas relacionados con otras características de los estudios, tales como el curso, el área de conocimiento estudiada, el modo de evaluación de la eficacia o la fuente de publicación del estudio?.

Búsqueda de los estudios

Una vez planteados los objetivos del meta-análisis, el siguiente paso consiste en localizar los estudios empíricos que hayan abordado la pregunta objeto de investigación. Esta fase pasa necesariamente por la definición de los criterios de inclusión y exclusión de los estudios. Estos criterios dependen del objetivo del meta-análisis, pero nunca pueden faltar los siguientes: (a) identificar los diseños de los estudios admisibles para el meta-análisis (e.g., ¿sólo se aceptarán estudios experimentales con asignación aleatoria o también se admitirán estudios cuasi-experimentales?); (b) definir los tipos de programas, tratamientos o intervenciones que se pretenden investigar; (c) definir las características de los participantes en los estudios (e.g., ¿sólo se admitirán estudios que han trabajado con muestras de niños y adolescentes, o con adultos, o con pacientes que tienen un determinado trastorno médico o educativo?); (d) determinar los datos estadísticos que deben aportar los estudios para poder calcular los tamaños del efecto (e.g., medias, desviaciones típicas, proporciones, pruebas *t*, pruebas *F* de ANOVA, etc.) y (e) identificar cómo han de venir medidas las variables de resultado (e.g., escalas psicológicas, pruebas de rendimiento debidamente baremadas, medidas de autoinforme, etc.). Finalmente, otros dos criterios de inclusión que no pueden faltar son el idioma en el que tiene que estar escrito el estudio y el rango temporal que se pretende examinar.

Partiendo de los criterios de inclusión y exclusión de los estudios, se procede a realizar una búsqueda bibliográfica lo más amplia posible para identificar los estudios que

pueden cumplir con los criterios de selección. Se recomienda en esta fase combinar procedimientos formales e informales de búsqueda de estudios. De entre las fuentes formales, no puede faltar la consulta de bases bibliográficas electrónicas, como ERIC, PsycInfo, MedLine, CINAHL, etc., dependiendo del ámbito de estudio del meta-análisis. Otras fuentes formales se basan en la consulta directa de revistas especialmente sensibles al problema investigado, así como la revisión de las referencias de los estudios que se vayan localizando. Con objeto de paliar los posibles efectos nocivos del fenómeno del sesgo de publicación, se recomienda complementar el uso de fuentes formales con fuentes informales de búsqueda, tales como contactar con expertos de reconocido prestigio en el campo para solicitarles estudios no publicados, acceder a los 'colegios invisibles' o consultar libros de actas de congresos, tesis doctorales y otras estrategias que permitan acceder, en la medida de lo posible, a la 'literatura fugitiva' (Rothstein y Hopewell, 2010). Con objeto de analizar la fiabilidad del proceso de selección de los estudios, al menos dos codificadores deben realizar dicho proceso de forma independiente y comprobar el grado de acuerdo.

En el meta-análisis de Erion (2006) se definieron los siguientes criterios de inclusión de los estudios: (a) el estudio tenía que incluir dos grupos de estudiantes, uno de ellos actuando como grupo de control; (b) el programa de intervención tenía que implicar tareas de tutorización realizadas por los padres, u otros miembros de la familia, en el hogar; (c) los participantes tenían que encontrarse en edad escolar; (d) las variables dependientes tenían que referirse a habilidades escolares básicas (e.g., lectura, pronunciación, matemáticas y expresión escrita); (e) debía aportar los datos estadísticos pertinentes para poder calcular el tamaño del efecto, y (f) tenía que estar escrito en inglés. Respecto de los procedimientos de búsqueda de estudios, el meta-análisis de Erion (2006) realizó búsquedas electrónicas en las bases ERIC, PsycInfo y Dissertation Abstracts Online entre los años 1970 y 2004. Además,

se revisaron las referencias de los estudios empíricos que se fueron seleccionando.

Codificación de los estudios

Una vez seleccionados los estudios que cumplen con los criterios establecidos en el meta-análisis, el paso siguiente es elaborar un *Manual de Codificación* en el que se hagan explícitos los criterios mediante los cuales se van a codificar las características de los estudios. La razón de examinar dichas características no es otra que comprobar qué características de los estudios pueden estar moderando o afectando a los resultados. La codificación de las características de los estudios es, pues, una tarea imprescindible si queremos explicar por qué los estudios sobre un mismo tema alcanzan resultados diferentes, e incluso en ocasiones contradictorios (Lipsey y Wilson, 2001).

Con este propósito se elabora un *Protocolo de Registro* de las variables moderadoras de los estudios y se aplica a todos ellos. Atendiendo a su procedencia conceptual, es habitual distinguir entre características metodológicas, sustantivas y extrínsecas. Las *características metodológicas* son aquellas que tienen que ver con la metodología y el diseño del estudio como, por ejemplo, el tipo de diseño (experimental vs. cuasi-experimental) el tipo de grupo de control (placebo vs. control puro), la inclusión o no de medidas pretest, la mortalidad experimental, el tamaño muestral o el uso de evaluadores enmascarados, o ciegos, al tratamiento recibido por el sujeto que está siendo evaluado. En el meta-análisis de Erion (2006) una característica metodológica registrada fue si el instrumento de evaluación del resultado del alumno estaba o no debidamente baremado o referenciado.

De especial interés en esta fase es la valoración de la calidad metodológica de los estudios empíricos, ya que ésta puede estar estadísticamente relacionada con los resultados de los estudios. Básicamente, existen dos enfoques para la medición de la calidad, el que se centra en el análisis de componentes individuales de calidad y el que aplica escalas de calidad (Littell, Corcoran y Pillai, 2008). La medición de la calidad a partir de

componentes individuales consiste en identificar dimensiones específicas que hagan referencia a algún aspecto de la calidad metodológica de un estudio evaluativo, operacionalizados en ítems individuales, tales como el tipo de asignación de los sujetos a los tratamientos o el porcentaje de mortalidad o pérdida de sujetos. En esta línea, en el manual de revisiones sistemáticas de la Colaboración Cochrane, Higgins y Green (2008) identifican los siguientes aspectos que nunca deberían faltar en la valoración de la calidad metodológica de los estudios evaluativos:

- (a) Sesgos de selección: diferencias sistemáticas en la composición inicial de los grupos pueden comprometer la validez interna del diseño.
- (b) Sesgos de ejecución: diferencias sistemáticas en los cuidados proporcionados a los grupos aparte de las intervenciones objeto de estudio (e.g., contaminación del tratamiento).
- (c) Sesgos por mortalidad diferencial: diferencias sistemáticas en las características de los participantes que abandonan los grupos de tratamiento.
- (d) Sesgos de detección: diferencias sistemáticas en la evaluación de las variables de resultado (e.g., efectos de las expectativas debidos al no enmascaramiento de los evaluadores).

Por su parte, las escalas de calidad reúnen múltiples ítems, definen un sistema de puntuación para cada ítem y posibilitan atribuir a cada estudio una puntuación global de calidad que considere conjuntamente todos los aspectos que inciden en ella. Un buen sistema de valoración de la calidad para la posterior comprobación de su relación con los resultados.

Las *características sustantivas* son las que tienen que ver con el objeto propio de estudio del meta-análisis. Como tales, dependen de dicho objeto, pero los meta-análisis sobre la eficacia de programas e intervenciones suelen clasificar, a su vez, las ca-

racterísticas sustantivas en características de tratamiento, de los participantes y del contexto. Por *características de tratamiento* se entienden aquellas que tienen que ver con el modo en que se ha definido e implementado el programa. En el meta-análisis de Erion (2006) son ejemplos de esta categoría la duración del programa de tutorización, la disponibilidad de instrucciones escritas, la disponibilidad de práctica supervisada o la posibilidad de realizar consultas al profesor. Las *características de los participantes*, como su nombre indica, tienen que ver con éstos, tales como la edad media de la muestra, el porcentaje de varones en la muestra, el tipo de trastorno padecido, la gravedad del trastorno, la extracción social de la muestra, etc. Así, en el meta-análisis de Erion (2006) se registró el curso en el que se encontraban estudiando los alumnos. Y las características de contexto tienen que ver con el lugar, o con el contexto temporal, en el que se aplicó el programa, tal como el país, si el programa se aplicó en un centro (escolar, sanitario, social) o en la casa del participante.

Por último, las *características extrínsecas* son aquellas que, en principio, no tendrían por qué afectar a los resultados de los estudios, al no tener nada que ver con la aplicación del método científico, y sin embargo en ocasiones se observan asociaciones entre ellas. Son ejemplos de tales variables el año de realización del estudio, la fuente de publicación del estudio (publicado vs. no publicado), la formación del investigador principal del estudio (e.g., educador, psicólogo, trabajador social, etc.) o incluso el sexo del investigador principal. El meta-análisis de Erion (2006) codificó, por ejemplo, el año del estudio y la fuente de publicación.

Una tarea muy importante que no puede faltar en un meta-análisis es examinar la fiabilidad del proceso de codificación de las características de los estudios, ya que algunas de éstas serán complejas de registrar. Con este propósito, dos codificadores, al menos, deberían codificar de forma independiente los estudios (o una muestra aleatoria de éstos) y comprobar el grado de acuerdo (Orwin y Vevea, 2010).

Cálculo del tamaño del efecto

Al mismo tiempo que se codifican las características de los estudios, se hace preciso calcular un índice estadístico que sea capaz de reflejar la magnitud del efecto obtenido en cada estudio. Ese índice estadístico tiene que ser tal que pueda calcularse de forma homogénea en todos los estudios, de forma que sea capaz de poner en la misma métrica los resultados de los estudios. Esta homogeneización de la unidad de medida de los resultados es necesaria porque los estudios analizan sus resultados de formas muy diversas, tales como a través de medias y desviaciones típicas, proporciones de éxito (o de fracaso), mediante la aplicación de pruebas estadísticas diferentes, tales como pruebas *t* de diferencias entre medias, pruebas *F* de ANOVA, pruebas Chi-cuadrado de Pearson, etc.

La mejor forma de unificar la unidad métrica del resultado de los estudios es mediante el cálculo de un índice del tamaño del efecto. El *tamaño del efecto* es un índice estadístico que mide el grado en que existe el fenómeno que se está investigando en cada estudio y no se deja afectar por el tamaño muestral, como ocurre con las pruebas de significación estadística (Cohen, 1988). Se han propuesto en la literatura diferentes familias de índices del tamaño del efecto cuya adecuada aplicación depende del objetivo del estudio, del diseño empleado y de la naturaleza de las variables medidas (Borenstein et al., 2009; Hedges y Olkin, 1985; Sánchez-Meca, 2008). Los índices del tamaño del efecto más frecuentemente utilizados en los meta-análisis son: (a) la familia *d*, especialmente indicada para estudios que aplican diseños con dos grupos y el resultado se mide de forma continua, y que consiste en calcular la diferencia entre las medias de los dos grupos y, opcionalmente, dividirla por la desviación típica conjunta de ambos grupos; (b) la familia *r*, especialmente indicada para estudios correlacionales, que se basa en la obtención de un coeficiente de correlación, y (c) la familia de índices de riesgo, que está indicada para estudios con dos grupos en los que la variable de respuesta es dicotómi-

ca (e.g., aprueba vs. suspende, mejora vs. no mejora).

Al igual que el proceso de codificación de las características de los estudios, el cálculo de los tamaños del efecto en los estudios debe someterse a un análisis de su fiabilidad, mediante la realización de dichos cálculos por, al menos, dos codificadores de forma independiente, con objeto de comprobar el acuerdo alcanzado entre ellos.

En el meta-análisis de Erion (2006), al incluir estudios con dos grupos de estudiantes, uno que recibía el programa de tutorización parental y el otro que actuaba como grupo de control, y medir los resultados de forma continua mediante pruebas de rendimiento escolar, se utilizó como índice del tamaño del efecto la diferencia media tipificada, un índice de la familia *d*, que se obtiene mediante:

$$d = c(m) \frac{\bar{y}_T - \bar{y}_C}{S},$$

siendo n_T y las medias de los grupos tratados y de control una vez finalizado el programa, S es la desviación típica conjunta de los dos grupos, que se obtiene mediante:

$$S = \sqrt{\frac{(n_T - 1)S_T^2 + (n_C - 1)S_C^2}{n_T + n_C - 2}},$$

Siendo \bar{y}_T y \bar{y}_C los tamaños muestrales de los dos grupos, y $c(m)$ es un factor de corrección para muestras pequeñas, que se calcula mediante:

$$c(m) = 1 - \frac{3}{4N - 9},$$

Siendo $N = n_T + n_C$. En el meta-análisis de Erion (2006), valores *d* positivos reflejaron un mejor rendimiento en el grupo que recibió el programa de tutorización parental, mientras que los valores negativos indicaron lo contrario.

Análisis estadístico e interpretación

Una vez codificadas las características de los estudios y calculados los tamaños del efecto, toda esta información se informatiza

creando una base de datos en la que las filas son los estudios y las columnas son las variables potencialmente moderadoras de los resultados de los estudios, así como el tamaño del efecto obtenido en cada estudio. Un paso previo en el análisis estadístico consiste en describir las características de los estudios que se han codificado, con objeto de obtener una imagen prototípica de los estudios. Una vez hecho esto, el análisis estadístico típico de un meta-análisis pasa por tres fases (Lipsey y Wilson, 2001; Marín-Martínez, Sánchez-Meca y López-López, 2009): (1) cálculo del tamaño del efecto medio con su intervalo de confianza y valoración de su significación estadística; (2) análisis de la heterogeneidad de los tamaños del efecto, y (c) si los tamaños del efecto son heterogéneos, búsqueda de variables moderadoras de tal variabilidad.

En este punto es importante hacer notar que los estudios individuales de un meta-análisis habrán utilizado tamaños muestrales diferentes, lo cual afecta a la precisión de las estimaciones del tamaño del efecto obtenidas. Cuanto mayor es el tamaño muestral de un estudio, tanto mayor es su precisión. Los análisis estadísticos que se realizan en un meta-análisis ponderan el tamaño del efecto obtenido en cada estudio en función de su precisión, de forma que el factor de ponderación de cada estudio es la inversa de su varianza. Pero para llevar a cabo estos análisis el meta-analista tiene primero que plantear el modelo estadístico desde el cual va a ponderar los tamaños del efecto. Son dos los modelos estadísticos típicos en meta-análisis: los modelos de efectos fijos y de efectos aleatorios (Sánchez-Meca, Marín-Martínez y Huedo, 2006).

Bajo el modelo de efectos fijos se asume que todos los estudios están estimando a un mismo tamaño del efecto poblacional, de forma que la única variabilidad entre los tamaños del efecto se deberá al error de muestreo. Desde el modelo de efectos aleatorios se asume que cada estudio estima a su propio tamaño del efecto poblacional, entendiendo que existe una distribución de tamaños del efecto poblacionales y no uno sólo

común a todos los estudios. La decisión de qué modelo estadístico asumir debería guiarse en función de qué grado de heterogeneidad espera encontrar el meta-analista entre los tamaños del efecto o, si no es capaz de hacer una predicción a este respecto, puede tomar la decisión en función del resultado de alguna prueba de heterogeneidad. Una descripción más detallada de estos dos modelos puede consultarse en Konstantopoulos y Hedges (2010), Raudenbush (2010), en Sánchez-Meca et al. (2006) o en Sánchez-Meca y Marín-Martínez (en prensa).

Para ilustrar los resultados de un meta-análisis, siguiendo con el de Erion (2006), el tamaño del efecto medio obtenido para el conjunto total de 32 estudios fue una diferencia media tipificada $d_+ = 0,54$ (límites confidenciales: 0,43 y 0,65) asumiendo un modelo de efectos fijos y $d_+ = 0,71$ (límites confidenciales: 0,48 y 0,92), asumiendo un modelo de efectos aleatorios. Ambos efectos medios fueron estadísticamente significativos, al no contener sus respectivos intervalos de confianza el efecto nulo ($d = 0$). Además, siguiendo la guía orientativa de Cohen (1988), según la cual índices d en torno a 0,2, 0,5 y 0,8 se pueden interpretar como reflejando una magnitud del efecto baja, media y alta, respectivamente, podemos considerar que el efecto medio de los programas de tutorización parental sobre el rendimiento escolar es de magnitud media-alta y, por tanto, representa una significación educativa relevante.

Un tipo de representación gráfica muy útil en meta-análisis es el *forest plot*, que presenta para cada estudio individual su tamaño del efecto junto con los límites confidenciales, tanto de forma numérica como gráfica, y en la parte inferior del mismo figura el tamaño del efecto medio. Como ejemplo, en la Figura 1 se presenta el forest plot con los resultados del meta-análisis de Erion (2006). Su inspección permite obtener una visión global de los resultados y el grado de heterogeneidad existente entre los tamaños del efecto individuales.

Pero el modo más apropiado de alcanzar el segundo objetivo analítico de un meta-

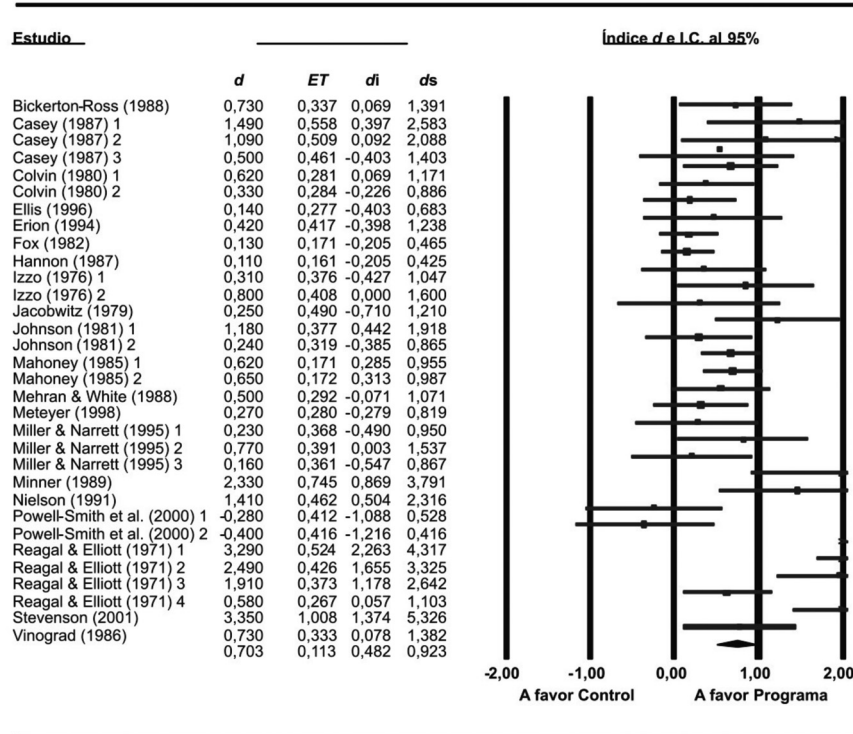


Figura 1. Forest plot del meta-análisis de Erion (2006) sobre la eficacia de la tutorización parental. *d*: diferencia media tipificada de cada estudio. *ET*: error típico de cada índice *d*. *di* y *ds*: límites confidenciales inferior y superior al 95% en torno al índice *d*.

análisis, el de comprobar si los tamaños del efecto son heterogéneos, consiste en aplicar la prueba de heterogeneidad, Q , que si alcanza la significación estadística significa que los tamaños del efecto son heterogéneos entre si y, en consecuencia, el tamaño del efecto medio no los representa bien. El resultado del estadístico Q se suele complementar con el índice I^2 , un estadístico que cuantifica, en porcentaje, la cantidad de heterogeneidad existente entre los estudios, de forma que índices I^2 en torno al 25%, 50% y 75% se pueden interpretar como reflejando heterogeneidad baja, media y alta, respectivamente (Borenstein et al., 2009). En el meta-análisis de Erion (2006), la prueba Q de heterogeneidad fue $Q(31) = 117,17$, $p < 0,001$, estadísticamente significativa, y el índice $I^2 = 73,5\%$, reflejando pues, ambos

resultados una alta heterogeneidad entre los tamaños del efecto de los estudios.

Si existe heterogeneidad entre los tamaños del efecto, entonces se hace preciso examinar el influjo de características de los estudios previamente codificadas, que actúan como variables predictoras, sobre los tamaños del efecto, que actúan en el análisis como la variable dependiente. Así, mediante técnicas de ANOVA y de análisis de regresión, ambas ponderadas, es posible analizar el influjo de variables moderadoras. Como ejemplo, la Tabla 1 presenta los resultados del ANOVA calculado con los datos del meta-análisis de Erion (2006) para comprobar si la duración del programa de tutorización (clasificado como corta, 1-9 semanas, media, 10-29 semanas, y larga, 30 ó más semanas) afectó a los tamaños del efecto. Como

Tabla 1. Resultados del ANOVA ponderado aplicado sobre los tamaños del efecto tomando la duración del programa como variable independiente.

Duración del programa	k	d_+	I. C. al 95%	
			d_i	d_s
Corta	12	0,41	0,05	0,77
Media	10	0,70	0,27	1,14
Larga	10	1,09	0,69	1,45
Resultados del ANOVA:	$Q_B(2) = 6,15, p = 0,046$			
	$Q_W(29) = 111,02, p < 0,001$			

k : número de estudios. d_+ : tamaño del efecto medio de cada categoría de la variable independiente. d_i y d_s : límites confidenciales inferior y superior del I.C. al 95% en torno al tamaño del efecto medio de cada categoría. Q_B : prueba estadística de contraste de diferencias entre los tamaños del efecto medios de las categorías. Q_W : prueba de homogeneidad intra-categoría. p : nivel crítico de probabilidad.

se puede observar en dicha tabla, conforme aumentó la duración del programa de tutorización el tamaño del efecto medio se incrementó de 0,41 para los programas cortos hasta 1,09 para los programas largos. Además, la diferencia entre los tres tamaños del efecto medios fue estadísticamente significativa, según el resultado de la prueba Q_B de comparación inter-categorías: $Q_B(2) = 6,15, p = 0,046$. Además, el ANOVA aporta el estadístico Q_W , que permite comprobar si existe homogeneidad dentro de cada categoría. En este caso, dicha prueba también alcanzó la significación estadística: $Q_W(29) = 111,02, p < 0,001$, por lo que debemos concluir que deben existir otras variables moderadoras relevantes, aparte de la duración del programa, capaces de explicar parte de la heterogeneidad de los tamaños del efecto.

Publicación

La última etapa en la realización de un meta-análisis, como la de cualquier otra investigación, es su publicación. Al tratarse de una investigación empírica, las secciones que debe incluir el informe escrito del meta-análisis son las típicas de un estudio empírico: introducción, método, resultados y discusión y conclusiones (Botella y Gambará, 2006; Clarke, 2010; Sánchez-Meca y Botella, 2010).

La *Introducción* del informe debe incluir una panorámica del problema objeto de

estudio, una definición de los constructos implicados en el planteamiento del problema, así como los objetivos concretos perseguidos y, en su caso, las hipótesis. La sección *Método* debe contener todos los datos y decisiones tomadas en el proceso de revisión para que el meta-análisis pueda ser replicado por otros investigadores. Así, no puede faltar una descripción de los criterios de inclusión y exclusión de los estudios, las estrategias de búsqueda de los estudios utilizadas, el proceso de codificación de las características de los estudios y una descripción del índice del tamaño del efecto junto con las técnicas de análisis estadístico aplicadas. De la precisión y meticulosidad con que se reporte esta sección del informe dependerá el grado en que el lector podrá hacer una lectura crítica del meta-análisis y valorar sus posibles deficiencias. En la sección de *Resultados* se presentan los tamaños del efecto, el efecto medio y los datos estadísticos pertinentes para valorar su significación y, en su caso, los análisis realizados para comprobar el influjo de variables moderadoras de los resultados. Por último, en la sección de *Discusión y Conclusiones* los resultados del meta-análisis se ponen en relación con los objetivos e hipótesis, así como con las evidencias previas. También se debe incluir una descripción de las implicaciones clínicas, educativas o sociales de los resultados, cuáles pueden ser las limitaciones del

estudio meta-analítico y apuntar líneas futuras de investigación en ese campo. Puede consultarse en Sánchez-Meca y Botella (2010) una guía orientativa sobre cómo debe escribirse un meta-análisis.

Reflexiones finales

No cabe duda que las revisiones sistemáticas y los meta-análisis están ejerciendo un importante papel en la acumulación del conocimiento científico y, en consecuencia, se han convertido en un potente instrumento para ayudar en la toma de decisiones que los profesionales de cualquier ámbito profesional de las ciencias sociales y de la salud

tienen que tomar continuamente para optimizar su práctica profesional. Es por ello que las revisiones sistemáticas y los meta-análisis se han imbricado profundamente en el enfoque de la Práctica Basada en la Evidencia como una herramienta metodológica imprescindible para desvelar cuáles son las mejores evidencias, o pruebas científicas, acerca de cualquier problema o pregunta social, educativa, clínica, etc. Al mismo tiempo, la evaluación de la calidad metodológica de los estudios primarios, y de la investigación en general, es una cuestión fundamental que trasciende al meta-análisis y se imbrica en las propias raíces del método científico.

Referencias

- Borenstein, M., Hedges, L.V., Higgins, J.P.T. y Rothstein, H.R. (2009). *Introduction to meta-analysis*. Chichester, UK: Wiley.
- Botella, J. y Gambara, H. (2002). *Qué es el meta-análisis*. Madrid: Biblioteca Nueva.
- Botella, J. y Gambara, H. (2006). Doing and reporting a meta-analysis. *International Journal of Clinical and Health Psychology*, 6, 425-440.
- Chalmers, I., Hedges, L.V. y Cooper, H. (2002). A brief history of research synthesis. *Evaluation and the Health Professions*, 25, 12-37.
- Clarke, M. (2010). Reporting format. En H. Cooper, L.V. Hedges y J.C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* 2ª ed. (pp. 279-293). Nueva York: Russell Sage Foundation.
- Cooper, H. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (3ª ed.). Thousand Oaks, CA: Sage.
- Cooper, H., Hedges, L.V. y Valentine, J.C. (Eds.) (2009). *The handbook of research synthesis and meta-analysis* (2ª ed.). Nueva York: Russell Sage Foundation.
- Erion, J. (2006). Parent tutoring: A meta-analysis. *Education and Treatment of Children*, 29, 79-106.
- Glass, G.V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 10, 3-8.
- Hedges, L.V. y Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Higgins, J.P.T. y Green, S. (Eds.) (2008). *Cochrane handbook for systematic reviews of interventions*. Chichester, UK: Wiley-Blackwell.
- Hunt, M. (1997). *How science takes stock: The story of meta-analysis*. Nueva York: Russell Sage Foundation.
- Hyde, J.S. y Linn, M.C. (1986). *The psychology of gender: Advances through meta-analysis*. Baltimore, MD: Johns Hopkins University Press.
- Konstantopoulos, S. y Hedges, L.V. (2010). Analyzing effect sizes: Fixed-effects models. En H. Cooper, L.V. Hedges y J.C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* 2ª ed. (pp. 279-293). Nueva York: Russell Sage Foundation.
- Last, J.M. (2001). *A dictionary of epidemiology*. Oxford, UK: Oxford University Press.
- Lipsey, M.W. y Wilson, D.B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- Littell, J. H., Corcoran, J. y Pillai, V. (2008). *Systematic reviews and meta-analysis*. Oxford, UK: Oxford University Press.
- Marín-Martínez, F., Sánchez-Meca, J. y López-López, J.A. (2009). El meta-análisis en el ámbito de las Ciencias de la Salud: Una metodología imprescindible para la eficiente acumulación del conocimiento. *Fisioterapia*, 31, 107-114.
- Martín, J.L.R., Tobías, A. y Seoane, T. (Coords.) (2006). *Revisiones sistemáticas en las ciencias de la vida*. Toledo: FISCAM.
- Orwin, R.G. y Vevea, J.L. (2010). Evaluating coding decisions. En H. Cooper, L.V. Hedges y J.C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* 2ª ed. (pp. 177-203). Nueva York: Russell Sage Foundation.

- Raudenbush, S.W. (1984). Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy induction: A synthesis of findings. *Journal of Educational Psychology*, 76, 85-97.
- Raundebush, S.W. (2010). Analyzing effect sizes: Random-effects models. En H. Cooper, L.V. Hedges y J.C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* 2ª ed. (pp. 295-315). Nueva York: Russell Sage Foundation.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research* (ed. rev.). Newbury Park, CA: Sage.
- Rosenthal, R. y Rubin, D.B. (1978). Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences*, 3, 377-415.
- Rothstein, H.R. y Hopewell, S. (2010). Grey literature. En H. Cooper, L.V. Hedges y J.C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* 2ª ed. (pp. 103-125). Nueva York: Russell Sage Foundation.
- Sánchez-Meca, J. (2008). Meta-análisis de la investigación. En M.A. Verdugo, M. Crespo, M. Badía y B. Arias (Coords.), *Metodología en la investigación sobre discapacidad*. Salamanca: Publicaciones del INICO.
- Sánchez-Meca, J. y Botella, J. (2010). Revisiones sistemáticas y meta-análisis: Herramientas para la práctica profesional. *Papeles del Psicólogo*, 31, 7-17.
- Sánchez-Meca, J. y Marín-Martínez, F. (en prensa). Meta-analysis. En B. McGaw, E. Baker y P.P. Peterson (Eds.), *International encyclopedia of education* (3ª ed.). Oxford: Elsevier.
- Sánchez-Meca, J., Marín-Martínez, F. y Huedo, T. (2006). Modelo de efectos fijos y modelo de efectos aleatorios. En J.L.R. Martín, A. Tobías y T. Seoane (Coords.), *Revisiones sistemáticas en ciencias de la vida* (pp. 189-204). Toledo: FISCAM.
- Schmidt, F.L. y Hunter, J.E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, 62, 529-540.
- Shadish, W.R., Cook, T.D. y Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Smith, M.L. y Glass, G.V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32, 752-760.
- Smith, M.L. y Glass, G.V. (1983). Meta-analysis of research on class size and its relationship to attitude and instruction. En R.J. Light (Ed.), *Evaluation studies annual review*, Vol. 8 (pp. 433-447). Beverly Hills, CA: Sage.
- Torgerson, C. (2003). *Systematic reviews*. Londres: Continuum International Pub. Group.