

PEDRO MORALES (\*)

## 1. INTRODUCCIÓN

### 1.1. *Qué es el meta-análisis*

El meta-análisis es un método (o un enfoque metodológico, pues hay variantes) para hacer *síntesis cuantitativas* de lo ya investigado, establecer el *estado de la cuestión* y llegar, o intentar llegar, a una *generalización* sobre una hipótesis determinada. Dicho en términos más simples, el meta-análisis es una *investigación sobre investigaciones*. Es un procedimiento muy estructurado para *integrar* los resultados de numerosos estudios previos. No se trata, por lo tanto, de un nuevo método de análisis para aplicarlo a los datos de una nueva muestra, con su diseño propio, etc., sino de un método para integrar en un *resultado* único los resultados de todas las investigaciones previas relacionadas con una determinada hipótesis.

La metodología del meta-análisis es fundamentalmente *cuantitativa* (aunque no *exclusivamente* cuantitativa) y se ha popularizado a partir de las obras de Glass (1976) y de sus colaboradores (Glass, MacGaw y Smith, 1981). El término meta-análisis es de Glass, aunque también se emplean otros términos, como el de *revisión cuantitativa* y otros semejantes.

Glass (1976) distingue entre el análisis *primario*, que es el análisis original que hace cada autor con los datos de su estudio particular; el análisis *secundario*, que es el re-análisis de los mismos datos, para utilizar mejores métodos estadísticos o para responder a nuevas preguntas con los mismos datos; y el *meta-análisis*, que es un *análisis de toda una colección de análisis* para poder integrar todos los resultados. La *unidad de análisis* es el estudio o la investigación primaria. La única condición es que todos los estudios se refieran a la misma hipótesis, entendida en un sentido amplio, y que presenten los datos necesarios para poder llevar a cabo un meta-análisis.

Los objetivos de un meta-análisis no se reducen a una simple y única cuantificación de muchos resultados de diversos estudios. Los objetivos son: a) ciertamente y en

---

(\*) Universidad de Comillas.

primer lugar, resumir o describir el *grado* en el que dos constructos están relacionados (expresado de una u otra manera); *b*) y además identificar factores o características de los estudios analizados que explican la variabilidad de los resultados; *c*) sugerir orientaciones para futuras investigaciones; *d*) ofrecer nuevas perspectivas teóricas, nuevas hipótesis que a su vez pueden ser objeto de otros estudios, resolver o explicar resultados conflictivos; y *e*) sugerir posibles aplicaciones de los resultados (Abrami, Cohen y d'Apollonia, 1988).

En estos últimos años han ido apareciendo otros muchos títulos sobre la metodología del meta-análisis, sobre todo a partir del año 1984 (Rosenthal, 1984; Hedges y Olkin, 1985; Wolf, 1986; Cooper, 1989; Hunter y Schmidt, 1990; y en castellano, Gómez, 1987 —de todos estos autores, tienen especial importancia Glass, como iniciador del meta-análisis, y Hedges, quien ha desarrollado la metodología—. Se trata de una metodología y de un enfoque claramente en expansión.

Estas revisiones cuantitativas tienen también sus críticos; se han propuesto variantes e indudablemente tienen, como veremos, sus limitaciones. Aun así, estas revisiones han tenido una gran aceptación y en las buenas revistas se publican constantemente (se pueden mencionar, por ejemplo, las revistas americanas *Psychological Bulletin* y *Review of Educational Research*). La aceptación que tiene el meta-análisis por parte de autores y fuentes serios y de prestigio no deja lugar a dudas sobre su importancia.

Sobre la importancia del meta-análisis podemos al menos citar a Asher (1990, en la inauguración de la conferencia anual de la *American Psychological Association*), para quien la introducción de esta metodología constituye una innovación crucial (*major turning point*) para estructurar y dar sentido a nuestros datos, comparable a lo que en su momento supusieron el análisis de varianza, los conceptos clásicos de los diseños experimentales (validez interna, validez externa) y los coeficientes de fiabilidad Kuder-Richardson 20 y  $\alpha$  de Cronbach.

El meta-análisis es importante además porque, como vamos a ver, está muy relacionado con aspectos *débiles* de los métodos cuantitativos que utilizamos habitualmente y con las limitaciones de las revisiones de estudios de carácter cualitativo.

El meta-análisis, de alguna manera, intenta llenar vacíos importantes en la práctica de la investigación, no surge *porque sí*, no parece una moda pasajera (todavía se *está haciendo* el método), y es útil conocer los *porqués* y los *cómos* de esta metodología.

## 1.2. Novedad del meta-análisis

¿En qué sentido podemos decir que esta orientación y esta metodología son nuevas?

*a*) En Estados Unidos, al menos, el meta-análisis no es una novedad; este tipo de estudios es muy frecuente (aunque en 1984 todavía merecía el calificativo de novedad —Green y Hall, 1984—). En los años 80, sobre todo en los años finales de esta década, ha habido una verdadera explosión de este tipo de estudios. En las buenas revistas aparecen con frecuencia (y son incidentalmente muy útiles, aunque sólo sea por la información bibliográfica que contienen). En España, sin embargo, no son

frecuentes; sí se ven citados ocasionalmente, y últimamente ha aparecido algún buen artículo sobre este método (Vázquez, 1990).

El meta-análisis es lo suficientemente reciente e importante como para situarlo en un lugar destacado entre las orientaciones actuales en el campo de la investigación psicológica y educativa. Y con mayor razón si pensamos que no se trata de una metodología estadística más para llevar a cabo un estudio particular, sino para sintetizar el *estado de la cuestión* a partir de muchos estudios particulares.

b) La metodología propia del meta-análisis incluye el cálculo del *tamaño* o de la *magnitud del efecto* (*effect size*); un punto sobre el que habrá que volver, pues su aplicación es útil, e incluso necesaria, en investigaciones primarias y sencillas como complemento de las pruebas estadísticas más usuales (como el simple contraste de medias). Sin embargo, el *tamaño del efecto* todavía no es recogido por la mayoría de los textos, ni españoles ni extranjeros. Consecuentemente, se utiliza muy poco. Y de hecho, aquí y ahora, todavía es novedad.

No se trata ya de utilizar el *tamaño del efecto* para hacer una síntesis de otros estudios (meta-análisis), sino de incorporarlo a la investigación primaria. Son métodos cuantitativos complementarios, plenamente establecidos, pero de hecho desconocidos por muchos de los que empiezan a investigar. No es una novedad para especialistas o lectores asiduos de la literatura experimental, pero sí es una novedad para muchos otros interesados en la investigación. Llama la atención, por ejemplo, el poco (o nulo) uso de estos cálculos en tesis doctorales recientes, en las que las hipótesis y los métodos de análisis utilizados lo están pidiendo.

Y ahora, una pequeña digresión propia de un profesor de estadística. Esta ausencia (la no aplicación de métodos, sencillos por otra parte, donde cabría esperarlos) no deja de constituir un toque de atención para los profesores de estas materias (de estadística y materias afines, y también de otras asignaturas con base experimental).

El hecho de que estos métodos se hayan desarrollado o, al menos, popularizado muy recientemente puede explicar el que muchos profesores, con una preparación universitaria de hace bastantes años, no los hayan incorporado a su docencia habitual. Modelos y métodos estadísticos para analizar datos hay muchos, pero, con frecuencia, la docencia de los mismos se queda corta. Cuando se imparten clases de estadística y materias afines, a veces se tiende a simplificar o a acortar la asignatura para hacerla más asequible a los alumnos. Sin embargo, el dar algunos temas de manera incompleta (como los diseños experimentales relacionados con el contraste de medias, por mencionar algo muy sencillo) puede inducir a error en las conclusiones (al menos, por la limitación que supone quedarse en la mera significación estadística) o a no aprovechar bien toda la información disponible. En este campo sí puede hablarse en cierta medida de *novedad*, al menos en nuestro contexto. El mismo meta-análisis, que es relativamente simple, debería incluirse en los programas habituales de estadística o de métodos de investigación *experimental* (Asher, 1990).

### 1.3. Planteamiento de esta exposición

En esta exposición, más que entrar en fórmulas y procedimientos estadísticos concretos (algo sí se puede indicar), interesa insistir sobre todo en dos puntos. En

primer lugar, hay que comentar las causas y circunstancias, que son varias, que explican el porqué del éxito actual de los métodos cuantitativos de integración de resultados de otras investigaciones (*integración* y, además, *cuantitativa*); como ya se ha indicado, entre las razones para insistir en estudios integradores está la debilidad o las limitaciones de los métodos cuantitativos habituales.

En segundo lugar, expondremos las características más importantes de esta metodología, cuál es el *proceso* del meta-análisis. En este proceso hay aspectos relevantes para cualquier investigación, aunque no se trate precisamente de un meta-análisis. Además, añadiremos un resumen de sus limitaciones, junto con algunas observaciones generales.

## 2. POR QUÉ EL META-ANÁLISIS HA TENIDO TANTA ACEPTACIÓN

Es de especial interés, a mi juicio, no limitarnos a una exposición de la metodología del meta-análisis. Es importante examinar por qué podemos decir que el terreno estaba *abonado* para una innovación metodológica como ésta. Habrá que hablar de la necesidad de disponer de resúmenes integradores de lo ya investigado (establecer el *estado de la cuestión*), de la insuficiencia de las revisiones meramente cualitativas o narrativas y de limitaciones importantes de los métodos cuantitativos tradicionales. Estas limitaciones no constituyen una novedad, pero a veces sí parecen novedad, a juzgar por la poca atención que se les presta.

### 2.1. Necesidad, en general, de síntesis y estudios integradores que presenten el estado de la cuestión

Una tarea fundamental en el campo de la investigación consiste en organizar, resumir y sintetizar los datos, la información o las conclusiones de otras investigaciones parciales o primarias.

La investigación, múltiple y dispersa en distintas fuentes de información, es útil en la medida en que se puede sintetizar, integrar y comunicar. Muchas notas de campo, en cualquier tema, valiosas en sí mismas, hay que organizarlas para que se conviertan en información útil. Preguntas importantes, a las que se ha respondido en múltiples investigaciones aisladas, se quedan sin la respuesta final, que pone al día el estado de la cuestión, si no hay una integración de los diversos estudios particulares.

El establecer el *estado de la cuestión*, el determinar cuál es la situación de la *respuesta* a una pregunta determinada tienen un interés obvio. Si asociamos la investigación a la idea de descubrir algo *nuevo*, ciertamente el sintetizar para *dar sentido* a información previamente acumulada es investigación. Es algo nuevo en la medida en que permite ver y evaluar lo que era imposible detectar en información dispersa y, con frecuencia, contradictoria. La necesidad de métodos *estadísticos* para integrar resultados de experiencias distintas en las que se intenta responder a las mismas preguntas se planteó ya en los primeros años de este siglo (años 30) respecto a problemas relacionados con la agricultura (Bangert-Drowns, 1986, quien hace una pequeña historia del meta-análisis).

## 2.2. Necesidad específica de estudios integradores de lo ya investigado en un tema, en vez de hacer una nueva investigación sobre lo mismo

Por lo general, las revisiones de lo ya investigado sobre un tema han sido el *punto de partida* para una nueva investigación. Es lo que vemos en las tesis doctorales, que suelen comenzar con una revisión de estudios. Citar y resumir una serie de estudios sirve para situar el propio trabajo en el contexto apropiado. Ahora tratamos las revisiones no como punto de partida, sino como *finalidad* propia de una investigación, que, de momento, termina ahí. La finalidad no es propiamente la mera revisión y la síntesis, sino la generalización a la que se pueda llegar.

Normalmente, el que inicia una investigación revisa las conclusiones de investigaciones precedentes antes de iniciar el propio estudio. Según las conclusiones, limitaciones, etc., de lo encontrado por otros investigadores, plantea su nueva investigación, que supuestamente va a dar la respuesta *definitiva*. La revisión de otras investigaciones con frecuencia no es concluyente (en parte debido a que no se tienen en cuenta diferencias importantes entre los diversos estudios), y del panorama que presentan las anteriores investigaciones (limitaciones, áreas no exploradas, etc.) nace el diseño de la nueva investigación.

Es importante señalar aquí dos enfoques o finalidades en la investigación. Hay investigaciones que responden a una necesidad específica, como son la de evaluar una metodología en una situación determinada, la de evaluar un proyecto educativo o la de hacer un estudio sobre los valores y actitudes de una determinada población. Son investigaciones de interés, en principio, para el que las hace.

Hay, sin embargo, otras investigaciones que marcan el énfasis en otros aspectos, que tienen como finalidad el aportar unas conclusiones que se van a traducir, quizás, en decisiones políticas, que pueden implicar un gasto público importante o que pueden afectar a muchas personas. Es en estos casos en los que a veces no se pide ya una *nueva* investigación, sino que se requiere que se exponga *cómo están las cosas* en el momento presente.

Concretamente en Estados Unidos (donde ha nacido la metodología del meta-análisis), las agencias gubernamentales que financian muchos de estos estudios piden resultados de *verdad*, que expresen cuál es el estado de la cuestión, de manera más definitiva, en asuntos importantes (en relación, por ejemplo, con la integración racial en las escuelas, el tamaño óptimo de las clases, la educación bilingüe, etc.). Las conclusiones de esta puesta al día de la situación pueden condicionar el apoyo a determinadas líneas políticas, la asignación de recursos, determinadas recomendaciones, etc. Es decir, a veces se da la necesidad de disponer *ya* de una respuesta a ciertas preguntas, aunque esta respuesta no pueda considerarse definitiva.

El primer enfoque (primero resumir algo y luego planificar una nueva investigación) no va dejando satisfechos a los que financian estas investigaciones, quienes, en definitiva, lo que buscan es saber adónde se ha llegado en un momento determinado. Si, por ejemplo, sobre un tema ha habido 30 investigaciones que han supuesto un coste apreciable, pero que no han aportado una respuesta definitiva porque presentan resultados contradictorios o no muy coherentes o limitados, ¿qué se puede esperar de la investigación trigésimo primera?, ¿la posibilidad de incluirla en la introducción

de la investigación trigésimo segunda? Es como decir que ha llegado la hora de hacer una buena síntesis de lo que ya hay (Raudenbush, 1991). De alguna manera no se está pidiendo que se haga una *nueva* investigación en el sentido tradicional del término, puesto que ya hay muchas, sino que se comunique de una vez a qué conclusiones se ha llegado.

El meta-análisis aporta una metodología, con todas las limitaciones que se puedan encontrar, a esta necesidad de obtener una respuesta para tomar una decisión. Vázquez (1990) cita el ejemplo de Australia, donde un meta-análisis, financiado por agencias gubernamentales, ha facilitado determinar cuáles deben ser las directrices terapéuticas para tratar la depresión y la esquizofrenia en los hospitales estatales.

La posibilidad de llevar a cabo con rigor estas síntesis ha elevado el *status* del meta-análisis, porque éste ayuda a establecer el *estado de la cuestión*. Más que un preludio para una nueva investigación se convierten estas síntesis en investigaciones, en sentido propio, que aportan algo realmente nuevo (una generalización bien fundamentada) partiendo de lo que ya se ha hecho. Y esto responde a la necesidad de tener respuestas más definitivas en un momento dado, sin volver permanentemente a investigar sobre lo mismo. La capacidad de los investigadores para dar respuestas a preguntas básicas que pueden condicionar determinadas políticas es algo demasiado serio como para tratar a la ligera el tema de la integración de estudios (Bangert-Drowns, 1986).

### 2.3. Limitaciones de las revisiones tradicionales de carácter narrativo

La revisión de estudios e investigaciones para exponer el estado de la cuestión no es naturalmente algo novedoso. Lo que sí es *más* novedoso, y es propio del meta-análisis, es hacer estas revisiones mediante un método más riguroso en todos sus pasos. En el caso del meta-análisis se llega a una *cuantificación*, primero, de cada estudio (con lo que cada uno no aporta un *sí* o un *no*, sino el *grado* en que dice *sí* o *no* a una conclusión) y, después, una cuantificación-resumen de todos los estudios en su conjunto.

La metodología propia del meta-análisis ha servido además para responder a una cierta crisis en la investigación en las ciencias sociales (Raudenbush, 1991). Las revisiones tradicionales de estudios e investigaciones han mostrado un carácter *narrativo*, y en ellas se ha puesto con frecuencia el énfasis en resultados contradictorios. Es muy normal que unos estudios no confirmen las conclusiones de otros estudios. La conclusión de que *hace falta investigar más* es muy común al término de estas revisiones. Es, asimismo, una conclusión casi rutinaria de muchas tesis doctorales.

La *crisis* de las revisiones más bien narrativas ha coincidido, por una parte, con la puesta al día de los métodos que han permitido síntesis cuantitativas y, por otra, con una crítica más rigurosa a este tipo de revisiones narrativas.

El estudio crítico más importante de las revisiones narrativas, muy citado y que en buena medida ha contribuido a que se ponga un mayor énfasis en las revisiones cuantitativas, es de 1980 (Jackson, 1980); fecha en la que, precisamente, estaban apareciendo las primeras publicaciones sobre el meta-análisis. En este estudio se

analizan sistemáticamente 36 revisiones narrativas, escogidas aleatoriamente y publicadas en revistas de prestigio en el campo de la investigación en las ciencias sociales. Dos investigadores independientes analizaron estas revisiones (de estudios de carácter experimental), codificándolas según 66 características. Sobre el meta-análisis no conocemos una crítica tan exhaustiva y sistemática (Abrami, Cohen y d'Apollonia, 1988).

La conclusión de este estudio (Jackson, 1980) era que, *de hecho*, no existía una metodología clara y sistemática sobre cómo hacer estas revisiones. Las limitaciones de estas revisiones estaban relacionadas con aspectos importantes propios de cualquier revisión de estudios, como son: *a)* la selección de hipótesis de investigación, *b)* la identificación y selección de estudios, *c)* la descripción de las características de estas investigaciones primarias, *d)* el análisis de estos estudios, *e)* la interpretación de los resultados, y *f)* la exposición misma de la revisión de estudios.

Típicamente estas revisiones se dividen en las siguientes categorías:

- a)* las que se limitan a *enumerar* los estudios que muestran un resultado *significativo* en relación con la variable que se analiza,
- b)* las que *excluyen* todos los estudios que no apoyan el punto de vista del autor,
- c)* las que no van más allá de calcular una *media* de un determinado estadístico a partir de diversos estudios,
- d)* las que contabilizan el número de *votos* o resultados a favor de una determinada hipótesis.

Los inconvenientes que suelen presentar estas revisiones los sistematizan muchos autores (Jackson, 1980; Slavin, 1987). Aquí recogemos los aspectos más importantes:

1) En general, se echa de menos en estas revisiones un *método sistemático*, comparado y replicable. Una característica de esta falta de sistematización es que con frecuencia no se mencionan criterios claros para incluir los estudios revisados (criterios de inclusión o de exclusión).

L'Hommedieu, Menges y Brinko (1987), quienes, por otra parte, son partidarios de utilizar el meta-análisis en conexión con las revisiones narrativas, llaman la atención sobre el hecho de que estas revisiones se limitan frecuentemente a reformular el resumen (*abstract*) que suele venir al comienzo de los artículos.

2) Posiblemente la crítica más generalizada es que muy frecuentemente se limitan a especificar el número de estudios que está a favor de una determinada conclusión (*vote-count*); práctica que lleva a conclusiones erróneas cuando los estudios son muchos, y las muestras, pequeñas.

En estos casos, típicamente, se utiliza sólo la *significación estadística* como criterio de eficiencia, de *resultado* digno de tenerse en cuenta. Se pueden detectar dos dificultades en esta práctica. En primer lugar, se pierde mucha información y se da el mismo valor a resultados muy distintos. Como ya se ha indicado, no es lo mismo un *sí* o un *no* que especificar el *grado* del *sí* o el del *no* (sobre este punto, las referencias y críticas a los resúmenes narrativos son muy abundantes; véase, por ejemplo, Green y Hall, 1984; Abrami, Cohen y d'Apollonia, 1988).

Cuando las conclusiones se limitan a un *sí* o a un *no* (un método es eficaz o no lo es), la *mera* conclusión de cada estudio en particular, a favor o en contra de una

determinada hipótesis, depende de muchas variables (por ejemplo, se rechaza la *hipótesis nula* con mayor facilidad si el número de sujetos es grande). Los procedimientos del meta-análisis, con todas sus limitaciones, son mucho más matizados, porque tienen en cuenta el *peso* de cada estudio (magnitud del efecto, error típico).

Con el criterio de la significación estadística se llega a veces a revisiones *mecánicas* que sacrifican información valiosa y se incluyen estudios con metodología cuestionable o poco relevantes para el tema de interés.

Además, cuando se utiliza como criterio (de inclusión, de prueba *a favor*) un resultado estadísticamente significativo, con frecuencia subyace la interpretación de que la *no* significación estadística equivale a la *no* diferencia o a la *no* relación. Esto es literalmente falso; la interpretación adecuada es que *no* se rechaza la *hipótesis nula*, pero que puede haber diferencia o relación, aunque no se haya detectado (se puede aceptar la *hipótesis nula* siendo falsa; lo que se conoce por error Tipo II, que no se tiene en cuenta en estos casos y queda minimizado en estas revisiones; en general, este tipo de error es poco reconocido y considerado —Asher, 1990—). Ampliamos más adelante este aspecto relacionado con el rechazo o la aceptación (o, con más propiedad, no rechazo) de la *hipótesis nula*.

3) Otra crítica que se hace a estas revisiones es que, en ocasiones, se trata de revisiones *sesgadas*, por cuanto incluyen los estudios que confirman el propio punto de vista y eluden otros más comprometedores. Posiblemente esto es menos frecuente en investigaciones *serias* que en otro tipo de trabajos o publicaciones, pero siempre cabe la tentación poco científica de prescindir del dato o del estudio cuyas conclusiones *no interesan*. Como advierten Falchicov y Boud (1989), estas revisiones, sobre todo si son puramente cualitativas, pueden estar sujetas a un doble prejuicio o sesgo (*bias*): el sesgo del que lleva a cabo el experimento o investigación primaria y el sesgo o subjetivismo del que la selecciona o revisa. En el meta-análisis se intenta, al menos, que el *mismo método* controle el sesgo en la selección y en el análisis de cada investigación.

Las revisiones *narrativas* han retardado, de hecho, la acumulación de conocimientos porque con frecuencia se han citado solamente los estudios que favorecen una determinada posición (la del que hace la revisión). Como los *prejuicios* (e intereses) de los que hacen estas revisiones de estudios pueden ser muy distintos, se termina presentando un cuadro bastante caótico del estado de la cuestión. Esta crítica, referida a la selección de la información en las revisiones narrativas, la aducen muchos autores (por ejemplo, Slavin, 1984; Hunter y Schmidt, 1990 —según la revisión crítica que de su obra hace Raudenbush, 1991—, y muchos otros).

La búsqueda selectiva de unas determinadas conclusiones es un peligro real; por eso, en la técnica del meta-análisis los *criterios de selección* de los estudios que se van a revisar están previstos y resultan relevantes.

Por mencionar un ejemplo, en un tema que puede ser conflictivo, como es el de la evaluación que hacen los alumnos de sus profesores, cabe siempre una *búsqueda selectiva* o una *divulgación sesgada* del dato que confirme las propias opiniones (Cohen, 1983). Así, el estudio de Rodin y Rodin presenta una correlación alta y negativa (-0,75) entre la calidad del profesor (juzgada por los alumnos) y el rendimiento; y en



el extremo contrario tenemos la investigación de Frey (1976), con una correlación positiva y notablemente alta (0,79), o la de Centra (1977), con una correlación igualmente positiva y muy alta (0,92). Puestos a buscar un tipo concreto de resultados, no es difícil encontrar la cita apropiada.

4) Otra limitación (que no tiene que ver solamente con las revisiones narrativas) es que con frecuencia se analiza una *selección* de estudios. Las revisiones de una serie limitada de estudios representativos, sin la revisión más *exhaustiva* propia del meta-análisis, o hecha con criterios muy definidos, lleva a inconsistencias, a resultados contradictorios y no concluyentes. Esto puede deberse simplemente a diferencias puramente aleatorias entre los resultados de diversas investigaciones.

El que esto suceda (resultados no coherentes en investigaciones distintas) es normal, sobre todo cuando se trabaja con muestras pequeñas (posiblemente, el caso más habitual en la investigación psicológica). Los resultados obtenidos en cada investigación parecen específicos de cada situación; es difícil explicar discrepancias y llegar a leyes o incluso a conclusiones más generales. Las revisiones propias del meta-análisis, hechas con todo su rigor (rigor en la selección de estudios y en los análisis estadísticos complementarios), permiten aprovechar mejor estas *pequeñas* investigaciones y llegar a conclusiones o *tendencias* más claras (Raudenbush, 1990).

A propósito de la selección de los estudios, puede además hablarse de un *sesgo no intencionado*, cuando solamente se analizan estudios publicados (experimentos y análisis que no llegan a un resultado estadísticamente significativo ni siquiera se publican).

5) Podemos señalar otra limitación, de otro orden, de las revisiones más narrativas, y es que resulta muy difícil hacerlas bien (descripción, análisis) cuando el número de estudios es muy grande (Green y Hall, 1984). La metodología del meta-análisis se presta a abarcar, con mayor facilidad, un número grande de estudios.

Muchos meta-análisis no pasan de los 40 ó 50 estudios analizados, pero con frecuencia incluyen muchos más. Por ejemplo, Kulik y Kulik (1988) revisan 53 estudios sobre la eficacia del *feedback* (de resultados de exámenes); Falchicov y Boud (1989) revisan 57 estudios sobre la relación entre la autoevaluación de los alumnos y la evaluación hecha por sus profesores; Kulik, Kulik y Bangert-Drowns (1990) revisan 108 estudios sobre la eficacia de los programas de *mastery testing*; y Linn y Hastings analizan los datos de 154 facultades de derecho (sobre predictores de éxito académico). Buscando números mayores, encontramos los más de 1.500 resultados que revisan y sintetizan Saphiro y Saphiro (1982) y otros tantos (distintos de los anteriores) que revisan Smith, Glass y Miller (1980); en ambos casos analizan la eficacia de la psicoterapia frente a *placebos* (los dos estudios son citados y comentados por Green y Hall, 1984).

Parece claro que una revisión sistemática hecha dentro del estilo narrativo-explicativo resulta muy difícil (si se quiere hacer *bien*) ante tantas fuentes de información, si no se quiere caer en una mera contabilización de resultados a favor o en contra.

Hay que mencionar, finalmente, que la importante crítica de Jackson (1980) a este tipo de revisiones se refiere a las revisiones y síntesis bibliográficas que tienen como objetivo principal el *inferir generalizaciones* sobre un tema determinado de estudio, no, en cambio, a las que sólo pretenden exponer de manera resumida y coherente una información que proviene de diversas fuentes.

Las revisiones de estudios de carácter más cualitativo tampoco hay por qué excluirlas en principio; sí hay que buscar una mayor sistematización y un mayor rigor en el proceso, y no faltan buenas orientaciones para llevarlas a cabo (Cooper, 1982, 1989; Light y Pillemer, 1984).

#### 2.4. *Insuficiencia de las investigaciones primarias* cuantitativas

Vamos a centrarnos en dos puntos muy relacionados entre sí, pero que, vistos de manera independiente, ofrecen dos perspectivas que facilitan el análisis de limitaciones importantes de la investigación cuantitativa. Un punto que vamos a considerar es la insuficiencia de cada investigación primaria, tomada de forma aislada cuando se trata de llegar a generalizaciones (los estudios meramente descriptivos de una situación ofrecen menos problemas). El otro punto está relacionado con el paradigma usual de la *hipótesis nula* y con el modo de interpretarlo. En la respuesta a estas limitaciones se encuentra el fundamento *cuantitativo* de la metodología del meta-análisis.

##### 2.4.1. La dificultad de controlar *otras explicaciones* en los estudios experimentales

Una perspectiva que apoya la necesidad de *síntesis integradoras* es la insuficiencia de cada investigación si se toma aisladamente. Según la teoría estadística, entendida de modo un tanto ingenuo, con un buen diseño experimental (validez interna) y una muestra adecuada (validez externa) es suficiente para llegar a conclusiones generalizables. Esto no es así; y aunque no tratemos ahora este tema con amplitud, sí conviene recordar algunas cosas, porque esta insuficiencia de cada investigación particular está en la base del énfasis actual en los estudios integradores.

La visión clásica de la investigación experimental se basa en confirmar que las variables A, B, C, etc. producen X. El modelo clásico más sencillo es el diseño experimental, con un pre-test, un post-test y un grupo de control. La investigación evaluadora se basa en este paradigma. Pero esta concepción es muy simple. Puede suceder que A, B y C sean condiciones suficientes, pero no necesarias, para que se cumpla X; o que otras variables puedan producir el mismo efecto. El paradigma para concluir una relación de *causalidad* está tomado de las ciencias físicas: las mismas circunstancias producen los mismos efectos. Pero en la realidad social y humana hay *otras cosas* que interfieren; hay muchas interacciones desconocidas e incontrolables. Las *mismas circunstancias* casi nunca se dan. Por esta razón, los resultados de muchos estudios experimentales son discrepantes.

En el caso típico de la evaluación de un método, el profesor no sólo aplica el método, además conoce qué pasa con cada alumno, cómo debe relacionarse con alumnos concretos, etc. No se trata sólo de aplicar un método. Y esto lo saben los padres cuando buscan un profesor determinado, más que un método determinado. Una solución ante la imposibilidad de controlar todas las variables y todas las interacciones es precisamente el meta-análisis, que integra muchos estudios particulares (House, 1991; quien trata de la insuficiencia de los experimentos con el control de variables).

Las críticas a los estudios experimentales —tomados aisladamente, cuando de un único estudio (con un *buen* diseño) se quiere generalizar— son muy frecuentes en autores relevantes (podemos citar a Meelh, 1978, coautor con Cronbach de la teoría más establecida sobre la validez de constructo; y la revisión crítica de los experimentos que hacen Rorer y Widiger, 1983). En este contexto cabe mencionar la opinión de Cronbach (1975), quien afirma que cuando se trata de establecer generalizaciones, los grandes estudios correlacionales pueden ser mejores que la investigación experimental con el control artificial de variables (de *alguna manera*, el meta-análisis equivale a un gran estudio correlacional; aunque esto no es verdad literalmente). No sobra añadir que las doce fuentes de falta de validez *clásicas*, que aparecen rutinariamente en muchos textos (a partir de la conocida obra sobre diseños de Campbell y Stanley, 1966), han pasado a ser 33 (Cook y Campbell, 1976); demasiadas para controlarlas de modo eficaz en un experimento *definitivo*.

Frente a estas limitaciones, las conclusiones del meta-análisis no dependen de un único experimento o de un único estudio (con frecuencia, válido y de interés para una situación dada), sino del análisis y la cuantificación de los resultados obtenidos en circunstancias distintas; de esta manera, es posible descubrir con mayor seguridad tendencias generales. Los días en los que se consideraba que dos o tres estudios podían ser una buena guía u orientación para establecer una teoría están pasando rápidamente (Asher, 1990). Críticas al énfasis que se pone en un único estudio o experimento pueden encontrarse en muchas fuentes (el *mito del estudio único y decisivo* —Light y Pillemer, 1984; Rosnow y Rosenthal, 1989—). Estos autores, y tantos otros en la actualidad, expresan su preferencia por una visión *cumulativa* de la ciencia, en la que cada estudio no da una respuesta definitiva, aunque sí aporta una modesta y valiosa contribución.

Las interacciones, que pueden pasar desapercibidas en un estudio particular, pueden examinarse en el meta-análisis, comprobando si determinadas características de los estudios primarios están relacionadas con el resultado final. Una ventaja adicional está en que el meta-análisis puede incorporar nuevos datos en el análisis (como el sexo del autor: ¿tiene que ver el sexo del autor con los resultados de los estudios sobre diferencias entre los sexos?; los investigadores varones tienden a encontrar a las mujeres más *influenciables* que las investigadoras —Eagly y Carli, 1981; comentado por Green y Hall, 1984—). Este tipo de análisis sólo es posible hacerlo cuando la unidad del estudio es el análisis, la investigación primaria, que entra con todas sus características (en este caso, el sexo del autor).

Otro ejemplo que muestra que, analizando muchos estudios, se pueden descubrir interacciones que no aparecen en un único estudio es el meta-análisis de Kulik y Kulik (1982) sobre una cuestión polémica y que puede tener implicaciones prácticas muy importantes, como es la de agrupar a los alumnos según su capacidad. En este meta-análisis, llevado a cabo con 52 estudios, se llega a la conclusión de que el agrupamiento según la capacidad beneficia a los alumnos más capaces, pero el efecto es muy pequeño. Muestra, por otro lado, una gran variedad en los estudios individuales: en unos, el resultado es positivo y, en otros, el resultado es negativo. Esta diversidad invita a otros análisis, y la conclusión final es que los alumnos se aprovechan más de este agrupamiento cuando los más capaces reciben una instrucción mucho más enriquecida y adaptada. En los demás, las consecuencias son irrelevantes;

no hay diferencias importantes (y no hay efecto negativo en los menos capaces). No hay tampoco diferencia cuando la división se hace entre alumnos que podemos llamar *normales* y alumnos *menos aptos*. Analizando este tema en un *único* buen estudio, o no se llega a ninguna parte o se puede encontrar el resultado que más interese.

#### 2.4.2. Limitaciones del paradigma de la *hipótesis nula*

Me voy a referir de manera más explícita al contraste de medias, como ejemplo frecuente y sencillo; aunque todo esto es generalizable a otros tests estadísticos.

Cuando comparamos dos medias, intentamos responder a tres preguntas: 1) ¿Se puede afirmar que *existe* una diferencia (que las muestras proceden de poblaciones distintas), que la diferencia *no es cero*?; 2) ¿es *grande* la diferencia (de *magnitud* apreciable)? y 3) ¿es *relevante* la diferencia?

Con el contraste estadístico se responde solamente a la primera pregunta. Si la respuesta es afirmativa, diremos que sí hay diferencia, que la diferencia es estadísticamente significativa. A mayor valor de *t* tenemos mayor seguridad para afirmar que sí hay diferencia (por encima de lo esperable por efecto del azar), pero eso no implica que podamos concluir que la diferencia sea precisamente grande o relevante.

Cabe hacer las siguientes observaciones:

- a) A una idéntica diferencia entre medias (utilizando el mismo instrumento) pueden corresponder valores de *t* muy distintos, porque en esos valores de *t* influyen las desviaciones típicas y el número de sujetos. Basta con ver la fórmula para caer en la cuenta de que con muestras grandes es más fácil obtener diferencias estadísticamente significativas (disminuye el denominador y aumenta el cociente).
- b) La *mera* diferencia entre dos medias en valores absolutos, en puntuaciones directas, no es fácil de interpretar, porque depende de la escala o instrumento utilizado (¿qué magnitud debe tener una diferencia para poder decir que es *grande*?).
- c) Una diferencia no es comparable con otras diferencias, en la misma variable dependiente o rasgo, entendido a un nivel más genérico, si se han obtenido con otra escala métrica.
- d) Una diferencia estadísticamente significativa puede ser pequeña e irrelevante a efectos prácticos.
- e) Una diferencia *no significativa* puede ser *grande* e importante en una situación dada.

Hace falta, por lo tanto, disponer de algún tipo de cuantificación más clara de la diferencia, que sea a su vez importante para responder a la tercera pregunta sobre la relevancia. En este juicio entrarán también criterios no estadísticos y un conocimiento sustantivo de la situación. Lo que sí parece claro es que la dependencia de la decisión sobre la *hipótesis nula* (rechazarla o no rechazarla) constituye una base muy limitada para la interpretación.

Sin entrar en otras consideraciones, puede ser suficiente aducir algunas citas sobre el mal uso, o el uso limitado, que se hace del modelo de la *hipótesis nula*.

Meelh (1978) afirma que construir la ciencia «rechazando hipótesis nulas es un terrible error, un procedimiento básicamente inadecuado, una pobre estrategia científica y una de las peores cosas que han sucedido en la historia de la psicología». Asimismo, afirma que son tantos los factores que afectan las variables que medimos, y es tan difícil controlarlos, que prácticamente «nunca hay dos poblaciones con idéntica media; la hipótesis nula es siempre falsa».

Nunnally ya señalaba en 1960 que, cuando no se rechaza la *hipótesis nula*, ello suele ser por falta de sujetos. Con muestras grandes y tomando como criterio los niveles de significación, *se puede probar casi cualquier cosa*. En su opinión, «el modelo de comprobación de hipótesis peor utilizado es el de la *hipótesis nula*, el énfasis en la *hipótesis nula* es poco informativo; en la vida real, la *hipótesis nula* casi nunca es verdadera». Para Nunnally, citando un viejo adagio latino (*natura abhorret vacuum*), de la misma manera que la naturaleza aborrece el vacío, la probabilidad aborrece la correlación cero (o la diferencia cero). Con cierto humor, el mismo autor termina su artículo diciendo que cuando un psicólogo sonríe y dice que la correlación es significativa a un nivel de .01, quizás es lo más que puede decir, «pero no tiene ninguna razón para sonreír» (una solución más informativa es añadir los intervalos de confianza, entre qué límites se encuentra una correlación o una diferencia).

Otros autores expresan que la práctica de confiar en la significación estadística como si fuera un *índice de certeza* es ridícula; y el que el nivel de significación ( $\alpha$ ) se haya convertido en el *criterio* para aceptar o rechazar los resultados de una investigación es «uno de los ejemplos más impresionantes de ignorancia generalizada (*mass ignorance*) en la historia de la ciencia» (Cohen y Hyman, 1979, quienes analizan con rigor el uso y la interpretación de la significación estadística).

Este tipo de críticas bien razonadas es frecuente a partir de los años 70, e incluso en los anteriores (Bakan, 1966; Derrick, 1976, que aduce abundantes citas críticas de otros autores; Carver, 1978; Schwartz y Dalglish, 1982; entre muchos otros). No se trata de una novedad, aunque la práctica habitual, incluidos los libros de texto y los *trabajos de los alumnos*, no siempre se hace eco de estas críticas.

En buena parte debido a estas críticas, suele recomendarse (Nunnally, 1960; y otros autores) utilizar intervalos de confianza (son más informativos que una mera diferencia o correlación). Igualmente por razones de *información*, los valores de probabilidad ( $F$ ,  $t$ ) no deben exponerse sin los valores de las medias, y es preferible dar los valores exactos de  $p$  (unilaterales, sin limitarse al clásico  $p < .05$ ). Otra línea crítica está relacionada con el límite tradicional de la probabilidad de .05 («seguramente Dios ama el .06 casi tanto como el .05» —Rosnow y Rosenthal, 1989—) y de las pruebas unilaterales y bilaterales (pueden verse comentadas estas cuestiones en Pillemer, 1991). Todas estas limitaciones ciertamente tienen menos peso en la medida en que se aporten todos los datos informativos (intervalos de confianza, valores exactos de  $p$ , etc.) y se tienda a la acumulación e integración de estudios sin depositar toda la confianza, para generalizar, en «el único» estudio.

Sin entrar en un análisis más profundo y matizado de estas críticas, recordemos al menos dos cosas: 1) la frecuente confusión entre significación estadística y *magnitud*

y relevancia, y 2) la frecuente ausencia del cálculo de la *magnitud* o del *tamaño del efecto* que, al menos en parte, obvia estos problemas y en el que se basan las técnicas cuantitativas del meta-análisis. Este punto lo tratamos a continuación y, de hecho, es una respuesta a las limitaciones interpretativas del paradigma de la *hipótesis nula*. La respuesta para utilizar bien este modelo iría en esta línea: *a)* calcular el tamaño del efecto o medidas equivalentes (incluso cuando se va más a *describir* que a *generalizar*) y *b)* replicar los experimentos o tener en cuenta muchos estudios, que es la línea del meta-análisis (sobre todo, para generalizar conclusiones), ya que las pruebas estadísticas no pueden sustituir la replicación (Schwartz y Dalglish, 1982).

## 2.5. La disponibilidad de una metodología para hacer síntesis integradoras de carácter cuantitativo: El tamaño del efecto

La puesta a punto de un procedimiento cuantitativo para integrar resultados de diversos estudios ha facilitado este tipo de síntesis. Todo tipo de investigación necesita disponer de un método. Incluso la investigación *cualitativa* necesita procedimientos, compartidos y suficientemente establecidos, para organizar la información de manera que se pueda llegar a conclusiones claras y verificables (Miles y Humberman, 1984; Quinn, 1987; Marshall y Rossman, 1989).

Cuando los datos son numéricos, como los que proceden de la investigación experimental en las ciencias sociales, existen ya procedimientos explícitos, con criterios también muy explícitos, para hacer estas síntesis. Cuando Glass (1976) habla por primera vez de meta-análisis no se refiere a una nueva técnica, sino al uso de los procedimientos ya conocidos para integrar resultados.

Esta cuantificación se puede y se debe hacer a dos niveles: 1) en la investigación primaria, en cada investigación independiente, y 2) cuando se trata de integrar los resultados de muchas investigaciones independientes.

Estos cálculos evitan algunas limitaciones de las revisiones meramente narrativas, sobre todo cuando casi se limitan a especificar el número de estudios que están a favor de una determinada conclusión, sin cuantificar la magnitud de la relación.

El cálculo básico que incorpora y hace posible el meta-análisis es el de la *magnitud* o *tamaño del efecto*, hecho en cada estudio particular y que va a servir para realizar una síntesis final integradora en la que cada estudio tiene su propio peso. Este punto es aplicable no ya al meta-análisis, sino a cada investigación particular. Es oportuno exponerlo brevemente antes de comentar de manera más detallada el procedimiento del meta-análisis.

La expresión *efecto* se refiere, obviamente, al resultado de un tratamiento experimental o a la consecuencia asociada a una determinada variable independiente. Este término (*efecto*) debe entenderse en un sentido genérico, no necesariamente en un planteamiento rígido de diseños experimentales; aunque es en los diseños experimentales, con un grupo experimental y un grupo de control, en los que tiene su aplicación más obvia.

Cuando decimos *tamaño* o *magnitud* estamos ya señalando que hay *grados*, que el efecto en cuestión puede ser mayor o menor. Se trata de la *cuantificación* de una

diferencia (hablando del contraste de medias, pero aplicable asimismo a otros estadísticos) independientemente de la escala métrica utilizada en las puntuaciones originales. Esta cuantificación se hace en cada experimento o estudio particular y es el dato que permite integrar diversos estudios en el meta-análisis (medias, errores típicos).

Hay dos procedimientos para hacer esta cuantificación: uno se basa en la transformación de los valores de  $t$  (y de otros estadísticos) en un coeficiente de correlación (que necesariamente tendrá como valores extremos  $-1$  y  $+1$ ), y el otro consiste en calcular una *diferencia tipificada*, como aclararemos enseguida.

Existe una cierta ambivalencia en el uso de los términos. Por lo general, se denomina *tamaño o magnitud del efecto* (*effect size*) a ambos cálculos; aunque no hemos visto consistencia en el uso de esta denominación. Algunos autores (Abrami, Cohen y d'Apollonia, 1988) denominan *tamaño del efecto* (*effect size*) a la diferencia tipificada y *magnitud del efecto* (*effect magnitude*) al coeficiente de correlación. Es frecuente denominar tamaño del efecto (*effect size*) a ambos cálculos, aunque con más frecuencia se reserva este término para la diferencia tipificada simbolizada con la letra  $d$  (Cohen, 1981, propone el símbolo *delta*).

1) Cualquier valor de  $t$  o  $z$  (refiriéndonos al contraste de medias) se puede transformar en un coeficiente de correlación biserial puntual: una variable es la variable dependiente y la otra es la pertenencia a un grupo u otro ( $1$  ó  $0$ ). Entre una diferencia y una relación existe una relación conceptual; se puede formular el mismo problema en forma de relación (¿existe relación entre el sexo y la actitud A?) y en forma de diferencia (¿difieren varones y mujeres en su actitud hacia A?). A esta relación conceptual le corresponde una relación empírica o metodológica. La transformación puede hacerse mediante esta fórmula:

$$r_{bp} = \sqrt{\frac{t^2}{t^2 + \text{grados de libertad}}}$$

Un mismo valor de  $t$  va a dar valores más bajos de relación en la medida en que aumente el número de sujetos. Es conocido que los valores de *ji cuadrado* y del análisis de varianza se pueden convertir también en coeficientes de correlación. Los valores de los coeficientes de correlación oscilan entre  $0$  y  $\pm 1$ , cualesquiera que sean las puntuaciones directas originales, y permiten comparar y resumir con más claridad las distintas magnitudes (en diferencias entre medias, en valores de  $t$ , etc.) procedentes de estudios distintos (es una manera de reducir todo a una *métrica común*). Otros resultados estadísticos (*ji cuadrado*,  $F$ , etc.) admiten igualmente transformaciones a coeficientes de correlación.

2) Las fórmulas que convierten diversos estadísticos en coeficientes de correlación son más conocidas; menos lo es (al menos, juzgando por los textos) la fórmula específica del *tamaño del efecto* entendido como *diferencia tipificada*. Ha sido el método más común de expresar la magnitud del efecto, popularizado por Cohen (1977; en la bibliografía citamos la segunda edición), poco antes de que el meta-análisis alcanzara la difusión que hoy tiene. El símbolo usual es  $d$ , aunque también se utilizan los símbolos *delta*,  $ES$  (*effect size*, en inglés),  $ME$  o  $TE$  (Magnitud o Tamaño del Efecto):

$$d = \frac{\bar{X}_e - \bar{X}_c}{\sigma}$$

$\bar{X}_e$  y  $\bar{X}_c$  son las medias de los grupos experimental y control (o de los dos grupos comparados); la desviación típica ( $\sigma$ ) suele ser la del grupo de control, aunque también se utiliza, y posiblemente sea preferible, la desviación típica de los grupos combinados en uno solo. Cuando se comparan las medias de un pre-test y de un post-test (muestras relacionadas), se utiliza la desviación típica del post-test.

a) Lo que obtenemos mediante estas fórmulas es una diferencia *tipificada*: nos dice a cuántas desviaciones típicas equivale la diferencia entre dos medias.

b) El valor de esta magnitud es *independiente* de las puntuaciones originales; por lo tanto, estos valores son comparables aun cuando vengan de estudios distintos e incluso aunque se hayan utilizado instrumentos distintos (por ejemplo, distintas medidas de autoconcepto con distinto número de items, etc.).

El uso de instrumentos distintos que miden el mismo rasgo o constructo es habitual. No hay un único instrumento para medir un rasgo. Esta comparación (y con mayor razón, el calcular las medias de *los tamaños del efecto* de distintos estudios que utilizan distintos instrumentos) puede ser cuestionable, y hace falta una valoración racional previa de los instrumentos que asegure su comparación. En esta dirección apunta una de las críticas al meta-análisis que comentaremos más adelante.

La falta de homogeneidad conceptual entre los instrumentos no constituye un peligro cuando se utiliza el mismo instrumento. En este caso, ciertamente, el tamaño del efecto abre la puerta, al margen del meta-análisis, a la posibilidad de integrar estudios pequeños en uno solo, dentro de una misma investigación. Con frecuencia, un investigador realiza el mismo experimento, o ensaya el mismo método, con muestras pequeñas y sucesivas. Cabe siempre una integración final, dando mayor calidad al estudio, utilizando la magnitud del efecto (en vez de limitarse a señalar *cuántas veces* ha habido un resultado significativo).

c) Como todos los valores de la *magnitud del efecto* son comparables entre sí, de estos valores se puede calcular la media y el error típico procedentes de estudios diferentes, para *resumir* todos los resultados en un único dato. En esta posibilidad se basa la metodología del meta-análisis.

d) El valor de  $d$  es una diferencia tipificada (una puntuación típica), que, si observamos la fórmula, nos dice en cuántas desviaciones típicas se aparta la media del grupo experimental de la media del grupo de control. Si traducimos este valor a las probabilidades asociadas en la distribución normal, nos da una información adicional. Si  $d = 1.0$ , como sabemos que por debajo de  $z = 1$  se encuentra el 84 por 100 de los casos, tenemos que el percentil 50 (mediana) del grupo experimental corresponde al percentil 84 en el grupo de control. Podemos decir también que el sujeto medio del grupo experimental supera el 84 por 100 del grupo de control (estamos suponiendo que la distribución es normal y que existe homogeneidad de varianzas). Este valor se interpreta, por estas razones, como una medida de coincidencia o solapamiento (*overlap*) de ambas distribuciones.

e) Respecto a la cuestión de cuándo se puede considerar *grande* o *pequeño* un determinado valor, suelen aceptarse estas orientaciones (Cohen, 1988): si  $d = .20$ , el



valor es pequeño; si  $d = .50$ , el valor es moderado, y si  $d = .80$ , el valor es grande. Estas orientaciones resultan un tanto arbitrarias, aunque son muy aceptadas.

Posiblemente sea más informativo comparar unos valores con otros o buscar, como referencia, qué suele obtenerse en estudios similares. Convencionalmente tiende a considerarse un valor no inferior a  $.50$  como de *significación práctica* (importante) cuando se trata de resultados sobre el rendimiento escolar (Cohen y Hyman, 1979); aunque a veces se aceptan valores menores (en torno a  $.30$ , referencias tomadas de Wolf, 1986, p. 27).

En cualquier caso, es un dato sobre la *magnitud* de la diferencia y no sobre la *seguridad* con la que afirmamos la diferencia (como  $t$  o  $p$ ), y resulta importante, al menos, para *cuantificar cambios*, hacer comparaciones o resumir resultados (calculando la media de la magnitud del efecto procedente de estudios, experimentos o pares de grupos distintos).

Una magnitud del efecto grande (en torno a  $.80$  ó  $1$ ) puede corresponder a un valor no significativo de  $t$ . Podemos decir que la diferencia es *significativa* (en el sentido de *importante*) en la muestra, aunque no sea extrapolable a la población.

f) Esta consideración complementa la interpretación de la *mera significación estadística* que nos dan los valores de  $t$  y otros semejantes. La significación estadística depende en gran medida del número de sujetos y, en realidad, es una afirmación (como  $p < .01$ ) que se refiere a la *población*, no a la *muestra* con la que trabajamos. Si obtenemos un valor de  $t$  cuya probabilidad, según las tablas, es inferior a  $.01$ , esto quiere decir que si la *hipótesis nula* fuera cierta (diferencia = 0) obtendríamos este valor de  $t$  (en la misma población) menos de una vez de cada 100. Pero no nos dice nada sobre nuestra muestra concreta. Aquí, en la investigación educativa o psicológica, la magnitud del efecto nos puede decir más sobre la *significación práctica* o la relevancia de los resultados. Puede que no sea cierto en el universo, pero es importante *aquí* (Cohen y Hyman, 1979).

g) En estudios experimentales (que pretenden detectar cambios en función de un tratamiento o diferencias en función de una variable independiente), la magnitud del efecto que se puede detectar depende en buena parte del número de sujetos de las muestras. Un efecto grande se detecta obviamente con mayor facilidad, por lo que es fácil detectarlo en muestras pequeñas. Los efectos *grandes*, si los hay, se pueden detectar en muestras relativamente pequeñas (en torno a los 50 sujetos); los efectos *moderados* (en torno a  $.50$  o menos) pueden detectarse en muestras de entre 100 y 200 sujetos; y para detectar efectos *pequeños* (en torno a  $.20$ ) hacen falta muestras mayores (cercasas a los 1.000 sujetos —pueden verse tablas específicas en Light, Singer y Willet, 1984, p. 197—).

h) Hemos mencionado las fórmulas para convertir un valor de  $t$  en un coeficiente de correlación y para calcular el tamaño del efecto (diferencia tipificada,  $d$ ). Existen otras muchas fórmulas para traducir resultados, expresados en estadísticos distintos, a una métrica común. En el caso del análisis de varianza para dos muestras independientes, ya sabemos que  $t^2 = F$ .

A título indicativo añadimos las fórmulas de conversión de  $d$  en  $r$ , de  $t$  en  $d$  y de  $r$  en  $d$ :

$$t \text{ en } d, d = \frac{2t}{\sqrt{\text{grados de libertad}}}$$

$$r \text{ en } d, d = \frac{2r}{\sqrt{1-r^2}}$$

$$d \text{ en } r, r = \frac{d}{\sqrt{d^2 + 4}}$$

Cuando se dispone de los valores exactos de  $p$  (y no simplemente de  $p < .05$  ó  $p < .01$ ), también se pueden combinar en un único valor mediante procedimientos sencillos (Rosenthal, 1982, p. 64 y ss.). Los valores exactos de  $p$  se pueden consultar en tablas mayores que las convencionales de los textos, y hoy día es fácil encontrarlas en calculadoras ya programadas.

Estas fórmulas las hemos puesto como ilustración; en las obras citadas sobre el meta-análisis pueden encontrarse otras muchas. Lo que importa es ver la posibilidad, en primer lugar, de cuantificar mejor una diferencia (con lo que esto añade a la apreciación de la *relevancia práctica* y no meramente estadística) y, en segundo lugar, de comparar mejor e incluso *integrar* resultados (medias) que provienen de estudios distintos.

El cálculo del *tamaño del efecto* ( $d$ ) no es el único método para obtener más información, una vez rechazada la *hipótesis nula* (y aunque no se rechace, se puede y se debe calcular el tamaño del efecto), ni para combinar resultados, independientemente de que se esté realizando, o no, un meta-análisis. Los métodos para combinar resultados (probabilidades) no son de ahora. Rosenthal (1978) cita estudios de los años 30 y presenta y evalúa una serie de métodos sencillos para combinar valores de  $p$ ,  $t$  y  $z$  (también pueden encontrarse en Glenn, 1983, y, en general, en los textos sobre meta-análisis).

### 3. PROCESO DEL META-ANÁLISIS

Enumeramos a continuación los pasos del meta-análisis. Precisamente, una de las características del meta-análisis es la *sistematización* clara de un proceso, de manera que autores independientes puedan llegar a los mismos resultados (aunque esto no sucede fácilmente). Ciertamente, uno de los méritos de los autores del meta-análisis (Glass, Hedges) es precisamente su claridad al especificar las *reglas* de la revisión; aunque después, de hecho, veamos que no todos las aceptan de idéntica manera.

Al hablar del proceso del meta-análisis, nos referimos a sus líneas básicas; en realidad hay varios enfoques, y quizás ninguno sea todavía el definitivo (una tipología clara de diversos tipos o enfoques del meta-análisis puede verse en Bangert-Drowns, 1986).

#### 3.1. Amplitud de la revisión de estudios

Una vez seleccionado el tema de interés (la *hipótesis* de interés, hablando con mayor propiedad), los proponentes del meta-análisis (Glass y otros, 1981) abogan por

la inclusión de *todos* los estudios localizados que cumplan los requisitos que previamente se hayan establecido y justificado.

La razón para incluir *todo*, sin distinguir según la calidad de los diversos estudios, es que el investigador puede dejarse llevar por sus propios prejuicios al determinar cuáles son los *buenos* y cuáles los *malos* estudios metodológicamente hablando. De hecho, el meta-análisis se ha impuesto por la frecuencia con la que aparecen revisiones cualitativas sin un sistema claro en la selección de los estudios y sin justificar los criterios de inclusión.

Una advertencia importante consiste en incluir igualmente estudios *no publicados*. Una razón para incluir estudios no publicados es que con frecuencia no se publican los estudios con resultados no significativos o que resulten en contra de la hipótesis del investigador. Un ejemplo de meta-análisis en el que se analizan, sobre todo, tesis doctorales es el de Schaeffli, Rest y Thoma (1985), referido a la eficacia de programas de *educación moral* (utilizando exclusivamente el *Defining Issues Test*).

El llevar a la práctica una inclusión exhaustiva, tal como se propone en la técnica del meta-análisis, no es fácil. Según algunos autores, como Slavin (1986, 1987), tampoco es conveniente dicha inclusión exhaustiva, por lo que proponen escoger solamente los mejores estudios; esto supone establecer unos criterios de calidad. Slavin denomina a su estrategia *best-evidence synthesis*, con la que pretende aunar las ventajas del análisis cuantitativo del meta-análisis y de las revisiones tradicionales. Los criterios de calidad que propone son fundamentalmente la validez interna (control de variables extrañas) y la validez externa (excluyendo, por ejemplo, estudios con muestras muy pequeñas y resultados de difícil extrapolación).

Ciertamente, el problema de pretender incluir todo viene de la abundancia de la literatura experimental en algunos campos, como reconocen otros autores (Falchikov y Boud, 1989; comentando a Slavin). Abrami, Cohen y d'Apollonia (1988) adoptan una postura intermedia y advierten las ventajas de ser generosos a la hora de incluir muchos estudios aun cuando su calidad sea cuestionable.

De la inclusión exhaustiva propuesta por Glass (1976) le han venido al meta-análisis muchas críticas. Una crítica citada con frecuencia es la de Eysenck (1978): El meta-análisis es un ejercicio en *mega-silliness* (mega-tontería), porque la calidad de los resultados no va a ser mejor que la calidad de los estudios incluidos (*garbage in-garbage out*). Realmente, esta cita es muy mencionada y la vemos en muchos artículos (Green y Hall, 1984; Bangert-Drowns, 1986; Vázquez, 1990), pero se emplea para matizarla después, y no porque se esté de acuerdo literalmente con la cita. La calidad metodológica de los estudios puede también ser objeto de análisis, y se pueden analizar por separado los estudios según posean, o no, una determinada característica. De hecho, y como responde Glass (Glass y otros, 1981), el meta-análisis se ocupa, asimismo, de la calidad de los estudios (los criterios de calidad se pueden codificar y analizar).

Slavin insiste en dos puntos, que es oportuno tener en cuenta. Uno es la posibilidad de analizar por separado los buenos y los no tan buenos estudios, y otro, especificar con claridad los criterios de *no inclusión* (por qué se prescinde de algún tipo de estudios).

Obviamente, los meta-análisis ya publicados son una buena fuente bibliográfica, específica y seleccionada por cada autor.

### 3.2. Criterios para la inclusión de estudios

Ya hemos mencionado la calidad como un posible criterio, a juicio de algunos autores. Está claro que una condición es la pertinencia del tema (que se estudie la misma hipótesis). Otra condición es que en cada estudio particular se haya calculado la magnitud del efecto o estén presentes los datos que permitan calcularla.

En la metodología del meta-análisis se da especial importancia a la determinación de criterios específicos para incluir estudios particulares en una determinada revisión de estudios. Este énfasis es válido para cualquier revisión de estudios que se haga con otras finalidades; por ejemplo, cuando se resume una serie de estudios en una tesis doctoral para situar el contexto o justificar el propio trabajo.

Estos criterios deben estar suficientemente justificados. A veces son muy obvios; por ejemplo, Kulik y Kulik (1984), en un meta-análisis sobre los efectos de la *enseñanza acelerada*, ponen como condiciones de inclusión, entre otras: *a*) que los resultados estén expuestos en términos cuantitativos (no en forma de impresiones, apreciaciones, etc.), *b*) que haya siempre un grupo de control, y *c*) que los grupos de control y experimental estén igualados de alguna manera en la aptitud.

Un ejemplo típico es el de los meta-análisis hechos para analizar la validez de la evaluación de los profesores realizada por los alumnos. En este caso se entiende por validez la relación entre la evaluación del profesor y el rendimiento académico de los alumnos (naturalmente, se puede hablar de otros criterios de validez).

Los criterios de inclusión (claros y bien justificados) los exponen Cohen (1981, 1987; autor de varios meta-análisis muy conocidos sobre este tema) y otros autores (Marsh, 1987; Abrami, d'Apollonia y Cohen, 1990).

*a*) Se deben llevar a cabo en cursos divididos en varias secciones en las que distintos profesores enseñen la misma asignatura, con los mismos textos, objetivos y un mismo examen final. De esta manera se mantienen constantes numerosas condiciones y, además, se neutralizan características personales de los profesores y de la situación que no se podrían considerar indicadores de la *calidad docente*. Esta condición impone ya un límite, y en la práctica, estos análisis sólo son posibles en grandes universidades.

*b*) Los alumnos deben ser asignados a las diversas secciones aleatoriamente. En este caso, las diferencias entre secciones, al evaluar a los mismos profesores, se pueden atribuir con mayor propiedad a diferencias entre los profesores, no a diferencias iniciales entre los alumnos. En las grandes universidades americanas, en las que el alumno tiene libertad para escoger una sección entre varias, diversos estudios muestran (por ejemplo, Leventhal y otros, 1975) que los alumnos que escogen una u otra sección pueden diferir bastante en los datos biográficos y en las razones para escoger una sección u otra.

Como esta condición es difícil de cumplir, el *minimum* requerido es que los alumnos no sepan quién será su profesor.

*c*) La *unidad de análisis* debe ser la media de la clase en cada ítem de evaluación (o en cada *factor*, según el tipo de análisis que se haga), no las respuestas individuales

de cada alumno (un resumen de las razones para utilizar medias de clases en vez de respuestas individuales puede verse en Marsh, 1987). Se trata de un punto importante, y más en este tipo de análisis. Al tema de la *unidad de análisis* no siempre se le presta la atención debida.

d) Los profesores deben ser los principales responsables de la enseñanza, y no sus ayudantes, profesores auxiliares, etc. (es una de las limitaciones más serias del estudio de Rodin y Rodin, 1972, con resultados muy negativos hacia estas evaluaciones y muy difundido a pesar de su mala calidad).

e) El criterio de rendimiento de los alumnos no debe ser las meras calificaciones, sino datos más objetivos.

No siempre son tantos los criterios ni tan estrictos, pero es útil ver un ejemplo riguroso como el que acabamos de mencionar. Naturalmente son muchos los estudios sobre este tema (evaluación del profesorado) que no cumplen estas condiciones, pero es en estos meta-análisis en los que, en última instancia, se puede probar la validez de estas evaluaciones (los resultados finales son favorables a este tipo de evaluación).

Cuando no todos los estudios son de la misma calidad (calidad determinada según criterios objetivos y previamente codificados), caben análisis separados para distintos grupos de estudios; es lo que hace Cohen (1987) reanalizando por separado los estudios de mayor calidad. En este caso, la *mejor calidad* viene dada por lo siguiente: a) existe un mayor rigor en los datos sobre el rendimiento del alumno, y b) entre las diferentes secciones del mismo curso no hay diferencias iniciales o éstas están bien controladas (en los estudios que considera de *calidad óptima* hay por lo menos diez secciones por curso y los cuestionarios de evaluación son más claramente comparables).

### 3.3. Codificación y análisis de los estudios

Los estudios seleccionados se analizan y codifican según ciertas características importantes. Típicamente se procede de este modo:

- 1) Se determina qué datos, características, información, etc. nos interesa *buscar* en cada estudio.
- 2) Se construyen tablas o cuadros que sinteticen la información más importante para facilitar: a) el cálculo de la magnitud del efecto de cada estudio, si no viene ya calculada; b) los análisis posteriores, como la relación entre determinadas características y el efecto o la eficacia; c) la presentación de la información; y d) los comentarios y síntesis posteriores.

Light, Singer y Willett (1990), sin referirse explícitamente al meta-análisis, dan orientaciones prácticas sobre cómo analizar estudios. En una investigación experimental, el objeto de estudio o la unidad de análisis suelen ser los sujetos o los grupos. Se recogen datos, se sistematizan y se analizan. La presentación típica de estos datos consiste en un cuadro de doble entrada en el que las *filas* son los sujetos y las *columnas* son los ítems, las variables o las preguntas.

Para hacer estas revisiones se puede pensar en algo análogo. En estas revisiones, el *sujeto* o la unidad de análisis son los estudios o investigaciones ya realizados, a *los que hacemos las mismas preguntas*: recogemos la información pertinente de cada estudio, la organizamos y la analizamos. Es como pasar el mismo cuestionario a los distintos estudios. Hacemos, con los estudios que revisamos, algo parecido a lo que hacemos con los sujetos de nuestra investigación.

*Qué preguntas* podemos hacer a cada estudio o investigación que vamos a revisar es algo importante, y hay que pensarlo bien. El *determinar de antemano* qué queremos ver en cada estudio facilita esta revisión. No se trata de una lectura sin sistema; vamos buscando determinada información (que son las preguntas que hacemos a cada estudio).

Las *preguntas* que podemos hacer a los estudios que revisamos (o la información que debemos buscar y sistematizar) van a variar según el tipo de investigaciones seleccionadas, pero algunas *preguntas* serán siempre las mismas. Además de anotar el autor o los autores, la fecha, las características de la fuente (publicada o no, informe interno, tesis, etc.), habrá que tener en cuenta el número de sujetos, tipo de muestra, las situaciones o características peculiares (si se trata de un curso normal o de un seminario, cuál es su duración, cuál es la comparabilidad de los grupos experimental y control, etc.), la variable dependiente (normalmente será la misma, pero a veces puede haber subconstructos, matizaciones, etc.), el diseño o método de análisis (*t* de Student, ANOVA, etc.) y los valores estadísticos finales. Las características que se codifican suelen ser muchas.

También se puede sistematizar las conclusiones, o determinar en qué circunstancias se ha llegado a esas conclusiones, etc. En un meta-análisis interesan, sobre todo, los valores de la magnitud del efecto o los datos que permitirán calcularlo.

La información se organiza en tablas (una o varias) que facilitan la presentación de los datos básicos.

Si la revisión se ha hecho no como parte del proceso del meta-análisis, sino como paso previo para situar una nueva investigación o un estudio experimental (como sería el caso de una tesis doctoral), cabría ahora hacer una síntesis y un comentario de carácter más cualitativo. Podríamos plantearnos, por ejemplo, a qué *preguntas de interés* (de interés, al menos, para el que presenta una nueva investigación) no se responde en las investigaciones revisadas; cómo se pueden explicar las inconsistencias en las conclusiones de varios autores, dónde ponen el *énfasis*; qué áreas se han investigado más y qué áreas han sido menos investigadas. Se puede, finalmente, intentar llegar a una generalización, extraer unas conclusiones generales que centren el estado de la cuestión.

#### 3.4. Cálculo del tamaño del efecto en cada estudio y de la media de todos

No todos los estudios particulares vienen con su *tamaño del efecto* calculado, pero sí suelen venir los datos necesarios para calcularlo. Los procedimientos y fórmulas concretos pueden verse en la bibliografía específica del meta-análisis (una buena síntesis metodológica se encuentra en Wolf, 1986). Pueden incluso calcularse a partir de los estadísticos propios de los métodos no paramétricos. Las fórmulas usuales son

ya muy conocidas, pero existen variantes que habrá que considerar en cada caso (por ejemplo, para ponderar, o no, el tamaño del efecto según el tamaño de la muestra).

Cada estudio particular suele presentar sus resultados con un valor de  $t$ ,  $F$ ,  $r$ ,  $X^2$ , etc. Como ya hemos indicado, con la magnitud del efecto se reducen todos los resultados a una métrica común, y esto permite no sólo una comparación más clara entre los mismos, sino también integrar todos los tamaños del efecto individuales en una media común. Éste es el paso final del meta-análisis, al que hay que añadir cálculos complementarios, como es el del error típico.

El cálculo del tamaño del efecto presenta problemas especiales; por ejemplo, cuando las desviaciones típicas no son homogéneas o cuando en un estudio particular no hay grupo de control. No es difícil encontrar en la literatura del meta-análisis soluciones metodológicas para estas situaciones (McGaw y Glass, 1980). No deja de ser una dificultad en el meta-análisis la enorme variedad de fórmulas y estadísticos con los que se presentan los resultados. Los tamaños del efecto hay que *rescatarlos* a partir de los datos y no hay un método único de conversión. Sin embargo, en muchos casos es sencillo y la bibliografía sobre el meta-análisis presenta métodos asequibles y aplicables ya a la investigación primaria (que es por donde se debería empezar).

Cuando algunos tamaños del efecto difieren mucho de la norma, puede cuestionarse el que estos estudios estén probando la misma hipótesis. Varios autores (Rosenthal, 1984; Hedges y Olkin, 1984) presentan métodos para comprobar la homogeneidad de los estudios. En el caso de una heterogeneidad apreciable, el efecto medio final puede calcularse en subconjuntos de estudios más homogéneos. En todo el procedimiento existe, pues, una cierta indeterminación que hay que tener en cuenta.

#### 4. LIMITACIONES DEL META-ANÁLISIS

La práctica del meta-análisis tiene sus limitaciones y así lo han expuesto y criticado autores que reconocen a la vez la importante contribución de estos análisis a la investigación en las ciencias sociales (Jackson, 1980; Slavin, 1984; Wolf, 1986, quien hace un buen resumen de las críticas más comunes del meta-análisis). Además de las limitaciones que podamos encontrar en el meta-análisis como método, cualquier meta-análisis puede estar tan mal hecho como cualquier investigación primaria (Slavin, 1984).

Las críticas más frecuentes las exponemos a continuación. Unos puntos están muy relacionados con otros, por lo que estas críticas se podrían sistematizar de otra manera.

##### 4.1. *Criterios de inclusión*

Uno de los puntos más cuestionados son los criterios de inclusión. Los autores que proponen el método inicialmente (Glass y otros, 1981) son partidarios de incluir *todos* los estudios pertinentes. Ya hemos visto que Slavin (1986, 1987) es más restrictivo (*best-evidence synthesis*) y aboga por criterios de calidad más claros y estrictos (más precisión en la definición de lo que se investiga, criterios de calidad metodológica, como la validez interna y la externa). A pesar de esta crítica, que parece muy obvia

(como es el tener en cuenta la calidad), son muchos los autores que defienden incluir todos los estudios; siendo la calidad una de las características codificadas y consideradas. Examinando los efectos de la calidad en los resultados, el meta-análisis es probablemente más objetivo que las revisiones más literarias y no cuantitativas (Wolf, 1986).

El meta-análisis da a la investigación una mayor apariencia de ciencia (ya sabemos que estamos en una ciencia *blanda*, no fácilmente comparable con las ciencias físicas), pero no se cae siempre en la cuenta de que los estudios analizados pueden ser muy distintos en cuanto al rigor metodológico y a las condiciones experimentales, por lo que el tamaño del efecto final no puede tomarse como la última palabra.

#### 4.2. *Inconsistencia en la aplicación del método: Resultados discrepantes en el meta-análisis sobre un mismo tema*

De hecho, distintos meta-análisis sobre el mismo tema llevan a conclusiones distintas. Un ejemplo lo presentan Abrami, Cohen y d'Apollonia (1988), quienes analizan diversos meta-análisis sobre la validez de la evaluación del profesor en la universidad (relación entre evaluación y rendimiento de los alumnos). Las conclusiones del meta-análisis de Cohen (1981 —tiene además otros meta-análisis posteriores más rigurosos—) apoyan la validez de estas evaluaciones, mientras que el meta-análisis de Dowell y Neal (1982) llega a una conclusión mucho más cautelosa.

En parte, estas discrepancias no pueden resolverse mediante análisis cuantitativos, porque entran también consideraciones conceptuales distintas. Pero, en general, sobre estas discrepancias se puede afirmar:

1) El meta-análisis incluye una serie de pasos metodológicos en los que *puede no haber consistencia* entre los diversos autores. Ya hemos mencionado el paso inicial, los criterios de inclusión en un meta-análisis, pero los pasos o fases son más: *a)* primero, especificar los criterios de inclusión (que pueden no estar claros en cada estudio y pueden necesitar un juicio personal del que hace la revisión); *b)* y además, localizar los estudios (esta localización puede ser más o menos exhaustiva, se pueden buscar, o no, estudios no publicados, etc.); *c)* codificar las características de cada estudio (se pueden codificar más o menos características, y las no codificadas no entrarán en otros análisis posteriores); *d)* calcular los resultados (tamaño del efecto) de cada estudio individual (con frecuencia hay varias alternativas); y *e)* realizar el análisis final de los datos.

2) Aunque en cada paso puede haber diferencias entre investigadores, *estas discrepancias se pueden aclarar* precisamente por lo *sistemático* del método (como advierten y analizan Abrami, Cohen y d'Apollonia, 1988). Las discrepancias se perciben con mayor facilidad por el esfuerzo de sistematización que supone hacer un meta-análisis. Y éste es un resultado positivo.

#### 4.3. *Excesiva simplificación*

El dato básico lo constituye (y el énfasis está puesto en) el resultado final de cada estudio (el tamaño del efecto), por eso permanece el problema de las interacciones, del influjo de unas variables en otras. Todo esto puede ser muy distinto en cada



estudio. También es algo que se puede codificar y analizar, pero no es tan fácil. El meta-análisis tiende a centrarse en el resultado final, y pueden pasar más desapercibidas o resultar menos enfatizadas las características de los estudios, más o menos relacionadas con determinados efectos. El *resumen final* puede quedar muy claro y convincente, pero falta de matización.

Por otra parte, el tamaño del efecto, que es sin duda un estadístico descriptivo útil, enmascara muchas decisiones subjetivas realizadas al calcularlo en cada estudio particular, ya que no hay un único procedimiento (Green y Hall, 1984; Wolf, 1986).

Muchas de estas limitaciones se compensan si se analiza un número grande de estudios. Con pocos estudios, y más si son bastante distintos, los resultados del meta-análisis son mucho más discutibles. Esto nos lleva otra vez a la *amplitud* de la revisión. La validez del meta-análisis depende de que la revisión de estudios sea completa o, al menos, de que se seleccione una muestra representativa (L'Hommedieu, Menges y Brinko, 1987).

#### 4.4. Integración de estudios muy diferentes

En el meta-análisis típicamente se incluyen estudios semejantes (en torno a la misma hipótesis), pero que pueden ser distintos, y lo son de hecho, en muchas variables: el instrumento utilizado, el tipo de tratamiento, etc. Cuando el *rasgo* (variable dependiente) es el mismo, pero el instrumento es distinto, no siempre es fácil llegar a la conclusión de que se está estudiando el mismo rasgo. En expresión popularizada por Gallo (1978), existe el peligro de mezclar *manzanas con naranjas*, y es éste uno de los aspectos más criticados del meta-análisis. Glass (1978, citado por L'Hommedieu, Menges y Brinko, 1987) responde en defensa de su meta-análisis que lo que pretende es analizar *frutas*, es decir, analizar los constructos y sus relaciones a un nivel más general. Es una respuesta que hay que tener en cuenta, porque los constructos pueden concebirse en diversos niveles de abstracción. En qué nivel de abstracción nos movemos es algo que ha de estar siempre claro. Es una instancia más que nos recuerda la necesidad de juicios *cualitativos* en un proceso de integración cuantitativa.

Un ejemplo típico en el que se analizan variables medidas con instrumentos distintos e incluso con denominaciones distintas lo tenemos en el meta-análisis de Barrick y Mount (1991), hecho a partir de 117 investigaciones primarias, sobre la relación entre rasgos de personalidad y éxito en el trabajo (*job performance*). En los procesos de selección de personal se utilizan numerosos tests que a su vez miden numerosos rasgos. En este meta-análisis quedan resumidos en cinco: 1) extraversión-introversión (que incluye asertividad, sociabilidad, etc.), 2) estabilidad emocional (neuroticismo, inseguridad, etc.), 3) amabilidad (*likeability*, conformidad social, etc.), 4) responsabilidad (*conscientiousness*, fidelidad, voluntad de trabajo, capacidad de organización y esfuerzo, etc.) y 5) lo que podríamos llamar actitud cultural (inteligencia, cultura, sensibilidad artística, creatividad). Los autores aducen suficiente base empírica para hacer esta reducción a cinco rasgos de lo que miden numerosos instrumentos (la presentan como una *taxonomía* aceptada, aunque no deja de ofrecer una cierta imprecisión).

Indudablemente, en un ejemplo como el mencionado es necesario un juicio racional y una decisión sobre el significado de estas variables. En este meta-análisis es la cuarta característica (responsabilidad, que casi equivale a lo que suele denominarse *motivación de logro*) la que está más relacionada, en todos los grupos ocupacionales analizados, con todos los criterios de éxito propuestos (y que son la productividad, la permanencia en la empresa, las promociones salariales y los ascensos dentro de la propia empresa).

Siempre cabe hacer subanálisis estudiando por separado *manzanas y naranjas* (Green y Hall, 1984), y de hecho, se hace así. Estos subanálisis (según ciertas diferencias importantes entre grupos de estudios) requieren una buena codificación previa (*paso fundamental en el meta-análisis*) y la actualización de algunos métodos de análisis para diferenciar estudios con resultados atípicos (Rosenthal, 1984; Hedges y Olkin, 1984). Los análisis complementarios, el tener en cuenta las diferencias entre estudios y el utilizar sus características como nuevas variables independientes constituyen la respuesta (Green y Hall, 1984) para no combinar acríticamente y sin análisis complementarios estudios muy diferentes.

Está claro que la hipótesis que se desea investigar puede ser definida de manera muy amplia o muy restringida; la mayoría de los investigadores opta por definiciones amplias (Green y Hall, 1984).

#### 4.5. *Pretensión de respuestas definitivas*

Una dificultad o un inconveniente que mencionan varios autores (Slavin, 1984; L'Hommedieu, Menges y Brinko, 1987) radica más bien en el estilo de los que presentan los meta-análisis. Suelen presentarse estos estudios con cierto matiz *autoritario*, como si se tratara de la *última palabra* que descansa en la supuesta *exhaustividad* de estas revisiones y en el prestigio de un método muy sistemático, muy ordenado, que, de hecho, está siendo muy utilizado.

La impresión de *estudio definitivo* propia de estas revisiones puede desanimar a otros autores para hacer nuevos experimentos o nuevos re-análisis. El uso creciente del meta-análisis puede desanimar para hacer nuevas investigaciones, dando las conclusiones ya alcanzadas por definitivas y sin tener en cuenta que estos análisis integradores pueden no estar tan bien hechos, que es posible hacerlos de otra manera o que se pueden actualizar.

#### 4.6. *Observaciones sobre estas limitaciones*

Las respuestas metodológicas a muchas de estas limitaciones se encuentran en las mismas obras sobre el meta-análisis. Es obvio, por otra parte, que, a pesar de tratarse de un método muy estructurado, no es tampoco un método exacto ni se pretende que lo sea en la medida en que pueden serlo los métodos de otras ciencias. De hecho, casi en cada paso hay que ir tomando decisiones racionales y, en buena parte, subjetivas. Es cierto además que el meta-análisis, realizado con rigor, permite apreciar sus mismas limitaciones, cosa que no sucede tan fácilmente con las revisiones pura-

mente narrativas de corte más tradicional (como subrayan Abrami, Cohen y d'Apollonia, 1988).

Por otra parte, resulta también claro que el análisis de los datos es una ayuda al pensamiento, pero no su sustituto (Green y Hall, 1984); la estadística no puede sustituir la lógica y el sentido común (Schwartz y Dalglish, 1982, en un buen análisis crítico de la inferencia estadística).

El meta-análisis es una contribución importante a lo que puede llegar a ser una metodología rigurosa para hacer revisiones en estudios, pero no es un sustituto de las revisiones cualitativas. Es importante una revisión *enciclopédica* y es importante la descripción-resumen propia del meta-análisis, pero ambas deben usarse junto con la revisión narrativa. En esto, unos autores ponen más énfasis que otros; después de todo, el auge del meta-análisis se debe en buena parte al cansancio ante tantas revisiones narrativas que han sido objeto de muchas y justificadas críticas.

L'Hommedieu, Menges y Brinko (1987), comentando sus propias dificultades para llevar a cabo un meta-análisis (poner al día el de Cohen, 1981, sobre la evaluación del profesorado), sostienen que el meta-análisis es un buen complemento, más que un sustituto, de otras revisiones de carácter más explicativo y cualitativo. Esta observación no constituye una novedad; que el dato estadístico es una ayuda importante, pero no un sustituto del propio pensamiento, es algo que mencionan los autores que hemos ido citando. El análisis estadístico es lo que es, un *instrumento*.

A pesar de sus limitaciones, el éxito del meta-análisis, su aceptación por buenos especialistas, avalan su utilidad para integrar los resultados de los numerosos estudios que pueden encontrarse sobre casi cualquier tema de interés. La integración es una necesidad, y el meta-análisis aporta una serie de métodos para llevarla a cabo. El meta-análisis no es una *moda* (Bangert-Drowns, 1986); está enraizado en los valores fundamentales de la investigación científica, como son la cuantificación y la replicación. Hay mucha información valiosa, pero dispersa, que necesita una integración. Las limitaciones del método son muchas, y las hemos ido notando, pero no expresan una debilidad que invalide el método; simplemente hay que tenerlas en cuenta, y posiblemente habrá que hacer adaptaciones según determinados propósitos o áreas de investigación. La realidad es que en las ciencias sociales, los *tamaños del efecto* son generalmente modestos, los resultados de los diversos estudios sobre la misma hipótesis ofrecen resultados muy distintos (e incluso contradictorios). Si se trata de generalizar, de buscar tendencias, la solución está en la línea del meta-análisis, en el que, en última instancia, se basarán las decisiones (Pillemer, 1991).

## 5. OBSERVACIONES FINALES

1) En nuestro medio no es tan fácil hacer un meta-análisis con investigaciones propias, desarrolladas en nuestro país. El meta-análisis gana en validez cuando se acumulan muchos estudios sobre el mismo tema; hace falta una investigación primaria muy abundante. Esto sucede en Estados Unidos, y de ahí viene la mayoría de los meta-análisis. Las fuentes son más numerosas y más cercanas.

2) Para cualquier estudio, los meta-análisis ya hechos aportan una gran riqueza bibliográfica, seleccionada con criterios claramente explícitos. Puede ser discutible el estado de la cuestión que establecen si lo aplicamos a un medio distinto, pero la información que acumulan es muy útil para cualquier otro estudio.

3) Sin hacer un meta-análisis en sentido estricto, el método es muy adecuado para integrar estudios propios o ajenos que tienen entre sí gran semejanza. El profesor, el psicólogo, etc. que hacen pequeños experimentos (experiencias, cursos, actividades, etc.) con grupos reducidos, de la misma población general; que utilizan los mismos o parecidos instrumentos, o estudian constructos del mismo ámbito, van acumulando datos que, con ayuda de la metodología del meta-análisis, pueden integrar en un estudio más serio por cuanto permite generalizar con mayor seguridad. No se trataría de un meta-análisis en sentido propio, pero la metodología puede ser muy útil.

4) El rigor que se pretende con el meta-análisis (en la selección de estudios, la codificación, etc.), aunque tenga sus puntos débiles, contiene muchas sugerencias útiles para otras situaciones en las que es necesaria una mayor sistematización que la que se ve habitualmente.

Más concretamente, las revisiones bibliográficas propias de trabajos serios, como son las tesis doctorales, pueden encontrar aquí una metodología más precisa y racional. Con un método inicialmente bien estructurado, aunque se utilice después con flexibilidad, se gana en precisión, se potencia la propia capacidad de análisis e incluso se ahorra tiempo.

Las características metodológicas del meta-análisis son útiles, en general, cuando se trata de hacer una investigación primaria, un estudio sencillo o una tesis doctoral. La aportación metodológica del meta-análisis va más allá de su objeto propio, que es hacer una síntesis de estudios previos. Todo lo relacionado tanto con el tamaño del efecto como con la sistematización (que podría ser otra en otros casos) del proceso se puede aplicar a otros tipos y modelos de investigación.

## BIBLIOGRAFÍA

- Abrami, Ph. C.; Cohen, P. A. y d'Apollonia, S. «Implementation Problems in Meta-Analysis». *Review of Educational Research*, 58 (2), 1988, pp. 151-179.
- Asher, W. «Educational Psychology, Research Methodology and Meta-Analysis». *Educational Psychologist*, 25 (2), 1990, pp. 143-158.
- Bakan, D. «The Tests of Significance in Psychological Research». *Psychological Bulletin*, 66, 1966, pp. 423-437.
- Bangert-Drowns, R. B. «Review of Developments in Meta-Analytic Method». *Psychological Bulletin*, 99 (3), 1986, pp. 388-399.
- Barrick, M. R. y Mount, K. «The Big Five Dimensions and Job Performance: A Meta-analysis». *Personnel Psychology*, 44 (1), 1991, pp. 1-26.
- Campbell, D. T. y Stanley, J. C. *Experimental and Quasi-Experimental Designs for Research*. Chicago, Rand McNally, 1966.

- Carver, R. «The Case Against Statistical Significance Testing». *Harvard Educational Review*, 48, 1978, pp. 378-399.
- Centra, J. A. «Student Ratings of Instruction and Their Relationship to Student Learning». *American Educational Research Journal*, 14, 1977, pp. 17-24.
- Cohen, J. «Statistical Power Analysis for the Behavioral Sciences». Hillsdale, NJ, Erlbaum, 1988 (2.<sup>a</sup> edic.).
- Cohen, P. A. «Student Ratings of Instruction and Student Achievement: A Meta-analysis of Multisection Validity Studies». *Review of Educational Research*, 51, 1981, pp. 281-309.
- «Comment on a Selective Review of the Validity of Student Ratings of Teaching». *Journal of Higher Education*, 54, 1983, pp. 448-458.
- «An Updated and Expanded Meta-analysis of Multisection Student Rating Validity Studies». Ponencia presentada en el Congreso de la *American Educational Research Association*, San Francisco, 1986.
- «A Critical Analysis and Reanalysis of the Multisection Validity Meta-analysis». Ponencia presentada en el Congreso de la *American Educational Research Association*. Washington, D. C., 1987.
- Cohen, S. A. y Hyman, J. S. «How Come So Many Hypotheses in Educational Research Are Supported? (A Modest Proposal)». *Educational Researcher*, 8 (11), 1979, pp. 12-16.
- Cook, T. D. y Campbell, D. T. *Design and Analysis of Quasi-Experiments for Field Settings*. Chicago, Rand McNally, 1979.
- Cooper, H. M. «Guidelines for Conducting Integrative Research Reviews». *Review of Educational Research*, 52 (2), 1982, pp. 291-302.
- *Integrating Research: A guide for literature reviews*. Newbury Park, CA, Sage, 1989 (2.<sup>a</sup> edic.).
- Cronbach, L. J. «Beyond the Two Disciplines of Scientific Psychology». *American Psychologist*, 30, 1975, pp. 116-127.
- Derrick, T. «The Criticism of Inferential Statistics». *Educational Research*, 19, 1976, pp. 35-40.
- Dowell, D. A. y Neal, J. A. «The validity and accuracy of students ratings of instruction: A reply to Peter A. Cohen». *Journal of Higher Education*, 54, 1982, pp. 459-463.
- Eagly, A. H. y Carli, L. L. «Sex of Researchers and Sex-Typical Communications as Determinants of Sex Differences in Influenceability: A Meta-analysis of Social Influence Studies». *Psychological Bulletin*, 90, 1980, pp. 1-20.
- Eysenck, H. J. «An Exercise in Mega-silliness». *American Psychologist*, 33, 1978, pp. 15-17.
- Falchikov, N. y Boud, H. D. «Student Self-assessment in Higher Education: A Meta-analysis». *Review of Educational Research*, 59 (4), 1989, pp. 395-430.
- Frey, P. W. «Validity of Student Instructional Ratings: Does Time Matter?». *Journal of Higher Education*, 47, 1976, pp. 327-336.
- Gallo, P. S. (Jr.) «Meta-analysis. A Mixed Meta-phor?». *American Psychologist*, 33, 1978, pp. 515-517.
- Glass, G. V.; McGaw, B. y Smith, A. B. *Meta-analysis in social research*. Beverly Hills, CA, Sage, 1981.

- Glass, G. V. «Primary, secondary and meta-analysis of research». *Educational Researcher*, 5 (10), 1976, pp. 3-8.
- «Commentary on "Interpersonal expectancy effects: The first 345 studies"». *The Brain and Behavioral Sciences*, 3, 1978, pp. 394-395.
- Glenn, N. D. «Replications, Significance Tests and Confidence in Findings in Survey Research». *Public Opinion Quarterly*, 47, 1983, pp. 261-269.
- Gómez, J. *Meta-análisis*. Barcelona, Promociones y Publicaciones Universitarias, 1987.
- Green, B. y Hall, J. «Quantitative methods for literature review». *Annual Review of Psychology*, 35, 1984, pp. 37-53.
- Hedges, L. V. y Olkin, I. O. *Statistical Methods for Meta-analysis*. Orlando, FL, Academic Press, 1985.
- House, E. R. «Realism in Research». *Educational Researcher*, 20 (6), 1991, pp. 2-9.
- Hunter, J. E. y Schmidt, F. L. *Methods of Meta-analysis: Correcting Error and Bias in Research Findings*. Newbury Park, CA, Sage, 1990.
- Jackson, G. B. «Methods for integrative reviews». *Review of Educational Research*, 50, 1980, pp. 438-460.
- Kulik, Ch. L. y Kulik, J. A. «Effects of Ability Grouping on Secondary School Students: A Meta-analysis of Evaluation Findings». *American Educational Research Journal*, 19 (3), 1982, pp. 415-428.
- «Effects of Accelerated Instruction on Students». *Review of Educational Research*, 54 (3), 1984, pp. 409-425.
- «Timing of Feedback and Verbal Learning». *Review of Educational Research*, 58 (1), 1988, pp. 79-97.
- Kulik, Ch. L.; Kulik, J. A. y Bangert-Drowns, R. L. «Effectiveness of Mastery Learning Programs: A Meta-analysis». *Review of Educational Research*, 60 (2), 1990, pp. 265-299.
- L'Hommedieu, R.; Menges, R. G. y Brinko, K. T. «Putting the "But" Back in Meta-analysis: Issues Affecting the Validity of Quantitative Reviews». Ponencia presentada en el Congreso anual de la *American Educational Research Association*, Washington, DC, 1987.
- Leventhal, L.; Abrami, Ph. C.; Perry, R. P. y Breen, L. J. «Section Selection in Multi-Section Courses: Implications for the Validation and Use of Teacher Rating Forms». *Educational and Psychological Measurement*, 35, 1975, pp. 885-895.
- Light, R. J.; Singer, J. D. y Willet, J. B. *By Design, Planning Research on Higher Education*. Cambridge, Mas., Harvard University Press, 1990.
- Light, R. L. y Pillemer, D. B. «Summing up: The science of reviewing research». Cambridge, MA, Harvard University Press, 1984.
- Linn, R. L. y Hastings, C. «A Meta-analysis of the Validity of Predictors of Performance in Law School». *Journal of Educational Measurement*, 21 (3), 1984, pp. 245-260.
- Marsh, H. W. *Students' Evaluations of University Teaching: Research Findings, Methodological Issues and Directions for Future Research*. Elmsford, New York, Pergamon Press, 1987.
- Marshall, C. y Rossman, G. B. «Designing Qualitative Research». Newbury Park, Sage, 1989.

- McGaw, B. y Glass, G. V. «Choice of Metric for Effect Size in Meta-analysis». *American Educational Research Journal*, 17 (3), 1980, pp. 325-337.
- Meellh, P. E. «Theoretical Risks and Tabular Asterisks: Sir Karl, Sir Ronald and the Slow Progress of Soft-Psychology». *Journal of Consulting and Clinical Psychology*, 46, 1978, pp. 806-874.
- Miles, M. B. y Huberman, A. M. «Drawing valid meaning from qualitative data: Toward a shared craft». *Educational Researcher*, 13 (5), 1984, pp. 12-30.
- Nunnally, J. C. «The Place of Statistics in Psychology». *Educational and Psychological Measurement*, 20, 1960, pp. 641-650.
- Pillemer, D. B. «One- Versus Two-Tailed Hypothesis Test in Contemporary Educational Research». *Educational Researcher*, 20 (9), 1991, pp. 13-17.
- Quinn, P. *How to use qualitative methods in evaluation*. Newbury Park, Sage, 1987.
- Raudenbush, W. «Summarizing Evidence: Crusaders for Simplicity». *Educational Researcher*, 20 (7), 1991, pp. 33-37.
- Rodin, M. y Rodin, B. «Student Evaluations of Teachers». *Science*, 177, 1972, pp. 1164-1166.
- Rorer, L. G. y Widiger, T. A. «Personality Structure and Assessment». *Annual Review of Psychology*, 34, 1983, pp. 431-463.
- Rosenthal, R. «Combining Results of Independent Studies». *Psychological Bulletin*, 85 (1), 1982, pp. 185-193.
- Rosenthal, R. «Valid Interpretation of Quantitative Research Results», en D. Brinberg y L. H. Kidder (Eds.), *Forms of Validity in Research*, San Francisco, Jossey-Bass, 1982, pp. 59-76.
- Rosenthal, R. *Meta-analysis procedures for social research*. Beverly Hills, CA, Sage, 1984.
- Rosnow, R. L. y Rosenthal, R. «Statistical Procedures and Justification of Knowledge in Psychological Science». *American Psychologist*, 44, 1989, pp. 1276-1284.
- Shapiro, D. A. y Saphiro, D. «Meta-analysis of Comparative Therapy Outcomes Studies: A Replication and Refinement». *Psychological Bulletin*, 92, 1982, pp. 581-604.
- Schaeffli, A.; Rest, J. R. y Thoma, S. J. «Does Moral Education Improve Moral Judgment? A Meta-analysis of Intervention Studies Using the Defining Issues Test». *Review of Educational Research*, 55 (3), 1985, pp. 319-352.
- Schwartz, S. y Dalglish, L. «Statistical Inference in Personality Research». *Journal of Research in Personality*, 16, 1982, pp. 290-302.
- Slavin, R. E. «Best-evidence Synthesis: An Alternative to Meta-analysis and Traditional Reviews». *Educational Researcher*, 15 (9), 1986, pp. 5-11.
- «Best-evidence Synthesis: Why Less is More». *Educational Researcher*, 16 (4), 1987, pp. 15-16.
- «Meta-analysis in education: How has it been used?». *Educational Researcher*, 13 (8), 1984, pp. 6-15.
- Smith, M. L.; Glass, G. V. y Miller, T. I. *The Benefits of Psychotherapy*. Baltimore, John Hopkins Univ. Press, 1980.
- Vázquez, C. «Revisión cuantitativa de la literatura: El meta-análisis». *Evaluación Psicológica/Psychological Assessment*, 6 (3), 1990, pp. 261-288.
- Wolf, F. M. *Meta-analysis, Quantitative Methods for Research Synthesis*. Beverly Hills, Sage, 1986.