

La revisión cuantitativa: una alternativa a las revisiones tradicionales

JULIO SANCHEZ MECA

Area de Metodología de las Ciencias del Comportamiento

RESUMEN

En este trabajo se critica el carácter asistemático de las revisiones cualitativas y se insiste en la imposibilidad de proporcionar una síntesis integradora y significativa de los estudios de investigación sobre un tópico, fundamentalmente cuando el número de estudios existentes relacionados con el tópico de investigación es considerablemente grande. Alternativamente, se propone la aplicación de los métodos estadísticos para integrar los resultados de una serie de estudios que tratan un tema común. Por último, se evalúa la eficacia de las revisiones cuantitativas.

ABSTRACT

The asystematic character of qualitative revisions is criticized in this work, where is emphasized the failing to give an integrative synthesis on a topic, basically when the number of studies related with the research topic is considerably large. Alternatively, the application of statistical methods to accumulate findings of a study set dealing with a common subject is outlined. Finally, the effectiveness of quantitative reviews is evaluated.

INTRODUCCION

El progreso del conocimiento en cualquier área de investigación no es posible sino mediante la integración de muchos hallazgos convergentes. Un solo estudio de investigación nunca es definitivo, «... a pesar de lo memorable y sensacional que pueda ser» (Green & Hall, 1984, p. 38). Tradicionalmente, la integración de los resultados de los estudios sobre un tópico de investigación se ha expuesto en artículos y trabajos de revisión, en los que el autor examinaba cada estudio e intentaba encontrar una teoría que reconciliara los diferentes resultados. Este tipo de integración se conoce con el nombre de *revisión cualitativa*¹ y suele prestar un gran servicio al estado del conocimiento sólo cuando existen pocos estudios sobre el problema de investigación. Sin embargo, en las últimas décadas ha sido tal el cúmulo de investigaciones que salen a la luz en Ciencias Sociales y del Comportamiento que, sobre un mismo tópico de investigación pueden encontrarse varios cientos de estudios; estudios que, además, varían entre sí en cuanto al diseño, las variables manipuladas y/o medidas y los hallazgos obtenidos. Como resultado, la tarea de procesar tan enorme cantidad de información que se desprende de los numerosos estudios exige un esfuerzo excesivo para las posibilidades de cualquier revisor. Ante esta situación, el típico revisor cualitativo intentará abarcar toda la información, obteniendo como resultado «... una revisión donde las sinopsis verbales de los estudios se ensartan en listas zig-zagueantes» (Glass, 1976, p. 4); o bien, tratará de simplificar la tarea de integración desechando una parte considerable de los estudios y «... proponiendo uno o dos estudios aceptables como la clave del problema» (Glass, 1976, p. 4). Esta última estrategia elimina injustificadamente mucha información en base a criterios arbitrarios y no empíricos.

Ante las debilidades, inconsistencias y contradicciones generalmente mostradas por las revisiones cualitativas tradicionales, en los últimos años se ha venido desarrollando con gran rapidez y difusión una nueva aproximación a la revisión de los estudios de investigación: se trata del uso de los métodos estadísticos para describir y analizar la literatura de investigación sobre un tema, que tiene sus más remotos antecedentes en los trabajos de Sir Ronald A. Fisher orientados hacia la combinación de niveles de significación de estudios independientes (cfr. Fisher, 1966). Glass (1976) acuñó el término *meta-análisis* para denominar tal enfoque, pero también suele llamarse *integración*

¹ A veces también denominada subjetiva, narrativa o verbal.

de investigación (Walberg & Haertel, 1980) y *análisis cuantitativo de dominios de investigación* (Cooper & Rosenthal, 1980; Rosenthal, 1980). En el presente trabajo, siguiendo los criterios apuntados por Green y Hall (1984), nos referiremos a esta aproximación como *revisión cuantitativa*.

Resulta conveniente distinguir entre las revisiones cuantitativas y los análisis primario y secundario. El *análisis primario* es el análisis original de los datos en un estudio de investigación; es lo que típicamente se considera como la aplicación de los métodos estadísticos. El *análisis secundario* puede perseguir uno de tres objetivos: (a) el reanálisis *exacto* de los datos tal como los examinó el investigador primario; (b) el uso de métodos estadísticos diferentes para contrastar las mismas hipótesis del investigador primario; o (c) la comprobación de nuevas cuestiones con los datos primarios (cf. Judd & Kenny, 1981, p. 212-214). En cambio, la *revisión cuantitativa* es el «análisis de los análisis» o, más concretamente, «... el análisis estadístico de una gran colección de resultados de los análisis de los estudios individuales con el propósito de integrar hallazgos» (Glass, 1976, p. 3). Más que una técnica, la revisión cuantitativa es una perspectiva que utiliza diversas técnicas de medida y de análisis estadístico.

Glas y sus colaboradores (1981, pp. 21-24) exponen tres características de las revisiones cuantitativas:

- (1) Son cuantitativas porque se valen de una variedad de métodos estadísticos para organizar y extraer información de los estudios.
- (2) No prejuzgan los hallazgos de investigación en términos de la calidad del estudio. Al contrario, su posible influencia es comprobada empíricamente.
- (3) Buscan conclusiones generales. Para ello, una revisión cuantitativa debe integrar resultados de estudios con diferencias superficiales, pero en una fase posterior del análisis se pueden examinar las características diferenciales de los estudios.

En esencia, una revisión cuantitativa consiste en la aplicación de los procedimientos de investigación familiares —conceptualización del problema, formulación de hipótesis, definición y medida de las variables, muestreo y análisis de datos— a la integración de estudios de investigación sobre un tópico determinado.

Los apartados que siguen a continuación describen el curso temporal de los pasos o etapas que cualquier revisor cuantitativo debe ejecutar para inte-

gar la literatura de investigación, para finalizar con una evaluación de esta nueva perspectiva metodológica.

DEFINICION Y BUSQUEDA DE LA LITERATURA

El primer paso para realizar una revisión cuantitativa consiste en seleccionar y definir un tópico de investigación que haya suscitado un cuerpo de evidencia empírica considerable. El tópico puede implicar la búsqueda de descripciones de variables, de relaciones entre variables o de supuestos causales. A continuación, deberá definirse la población de los estudios de investigación y localizarla. El objeto ideal de esta primera etapa de la revisión cuantitativa es recuperar todos los estudios de investigación sobre el tema seleccionado. Sin embargo, resulta obvio que, por muy sofisticadas que sean las técnicas de búsqueda utilizadas por el revisor, será prácticamente imposible localizar todos los estudios existentes (Green & Hall, 1984). De cualquier modo, los revisores cuantitativos deberían dedicar más atención a la búsqueda de los estudios de la que hasta ahora suelen prestar (Cooper, 1982, 1984).

1. Fuentes primarias y secundarias

La literatura empírica puede ser obtenida por fuentes primarias o secundarias (véase Ortega y Fernández Dols, 1980). Entendemos por 'fuentes primarias' las englobadas por la literatura periódica de archivo (revistas, disertaciones, trabajos de congresos, etc.). Las 'fuentes secundarias' citan, revisan y organizan el material de las fuentes primarias; tales son las revistas de revisión (por ejemplo, *Psychological Bulletin*, *Psychological Review*, *Annual Review of Psychology*) y *Abstracts* y archivos de citación (por ejemplo, *Psychological Abstracts*, *Current Contents*, *Dissertation Abstracts International*). Así, las fuentes secundarias pueden facilitar la identificación de las fuentes primarias, que constituirán la materia de la revisión (Glass et al., 1981).

2. Cómo realizar una búsqueda de literatura

Hunter y sus colaboradores (1982) recogen tres aproximaciones fundamentales para localizar estudios de investigación sobre un tópico: (a) el examen de los *Index*, (b) la búsqueda de las bibliografías existentes, y (c) las consultas a otros estudiosos familiarizados con el tema de interés.

- (a) *Index*. Los diversos *Index* pueden variar entre sí en cuanto a los temas tratados, la minuciosidad del área cubierta, la forma adoptada (impresa o disponible por computador), el tipo de información proporcionada (citación de la fuente completa, palabras-clave, *abstract*, etc.), así como la disponibilidad de los mismos.
- (b) Bibliografías. Un procedimiento muy común de las clásicas revisiones cualitativas —perfectamente aplicable a las revisiones cuantitativas—, consiste en la consulta de las bibliografías contenidas en las fuentes primarias y en los artículos de revisión sobre el tema.
- (c) Otro procedimiento ampliamente extendido es el de requerir de otros estudiosos familiarizados con un tópico particular de investigación, su colaboración en la identificación de estudios. Es aconsejable proporcionarles una lista con los estudios ya localizados por el interesado.

Las diversas técnicas de búsqueda bibliográfica no son exclusivas entre sí. De hecho, en Ciencias Sociales y del Comportamiento todas ellas guardan un cierto solapamiento entre los temas tratados. Por ello, Green y Hall (1984) recomiendan el uso combinado de varias fuentes: «Ningún método simple para localizar literatura es perfecto. Muchos revisores cuantitativos probablemente usan una combinación de métodos. Los principales métodos utilizados, en orden de menor a mayor adecuación, son las bibliografías existentes, las búsquedas por computador y los servicios de *abstracts* y la búsqueda fuente por fuente» (p. 46).

DESCRIPCION, CLASIFICACION Y CODIFICACION DE LOS ESTUDIOS

Una vez localizados y adquiridos los estudios, es necesario describir, clasificar y codificar sus características. Para ello, suelen confeccionarse unos cuestionarios con una lista de las características de los estudios que pudieran afectar a los resultados.

Glass et al. (1981, p. 77) clasifican las características de los estudios en sustantivas y metodológicas. Son características *sustantivas* las específicas del tópico de investigación estudiado; por ejemplo, en una revisión cuantitativa sobre las primeras etapas del procesamiento de información visual en buenos y malos lectores (Sánchez, 1985) se codificaron como características sustantivas el sexo, la edad, el CI, la tarea del sujeto, el procedimiento experimental, el tipo de estímulos y algunas otras más. Las características *meto-*

dológicas se refieren a aspectos más generales de los estudios, tales como la fecha y la fuente de publicación, las características de los sujetos, el tamaño muestral, el diseño de la investigación, el tipo de medida, la mortalidad experimental, etc.

El propósito de esta etapa es permitir un análisis posterior de las posibles relaciones existentes entre estas características de los estudios y sus resultados (Smith, 1982). Precisamente, la búsqueda de variables moderadoras que pudieran estar afectando a la magnitud y dirección de los efectos es de interés primordial en las revisiones cuantitativas (Wise, Peters & O'Connor, 1984).

MEDICION DE LOS HALLAZGOS DE LOS ESTUDIOS

El meta-análisis tiene como objetivo la integración estadística de los estudios empíricos de un fenómeno común. Para que esta integración sea posible, los hallazgos de los estudios deben ir expresados en una escala común; de lo contrario, resulta imposible compararlos e integrarlos. Por eso, cuando los estudios que pretenden revisarse varían en cuanto a las escalas de medida y a los métodos de informe de sus resultados, surgen dificultades.

Los revisores cuantitativos se han enfrentado de diversas formas a este problema, utilizando desde el simple recuento de los resultados estadísticamente significativos frente a los no significativos, pasando por la acumulación de niveles de significación informados en los estudios, hasta el cálculo de los tamaños del efecto a partir de los datos proporcionados en los estudios. De todos éstos, la medida más informativa para la comparación e integración de los estudios es el 'tamaño del efecto'. Lo trataremos a continuación con algún detalle.

1. El tamaño del efecto

Las revisiones cuantitativas que se basan en el recuento de resultados significativos o en la acumulación de niveles de significación no proporcionan información alguna acerca del tamaño, magnitud o fuerza de los efectos. La medida del tamaño del efecto esta adquiriendo una considerable popularidad en las revisiones cuantitativas, ya que, además de permitir la significación estadística de los resultados, también es susceptible de una descripción de su dirección y magnitud (Green & Hall, 1984). Se han desarrollado varios índices del tamaño del efecto aplicables según las características y los objetivos de la

investigación. Para una mayor claridad de exposición, describiremos a continuación los índices más utilizados para los estudios experimentales y correlacionales por separado (cf. Glass et al., 1981; Hedges & Olkin, 1985; Rosenthal, 1984).

1.1. Diseños experimentales

El caso más general suele implicar diseños de dos grupos, tradicionalmente denominados grupos experimental y control. El estadístico comúnmente utilizado en esta situación para el cálculo del tamaño del efecto es la *d* de Cohen (1977, p. 8), que consiste en la *diferencia media tipificada* entre las medias de los grupos experimental y control. Glass et al. (1981) la definieron como la diferencia entre las medias de los grupos experimental y control dividida por la desviación típica intra-grupo:

$$\hat{\Delta}_{E-C} = \frac{\bar{y}_i^E - \bar{y}_i^C}{S_i}, \quad (1)$$

siendo \bar{y}_i^E y \bar{y}_i^C las medias de los grupos experimental y control; y s_i la desviación típica intra-grupo, que se obtiene por

$$S_i = \left[\frac{(n_i^E - 1)(S_i^E)^2 + (n_i^C - 1)(S_i^C)^2}{n_i^E + n_i^C - 2} \right]^{1/2} \quad (2)$$

donde n_i^E y n_i^C son los tamaños muestrales de los grupos experimental y control, respectivamente, del *i*-ésimo estudio; y S_i^E y S_i^C son las desviaciones típicas de los grupos experimental y control.

En realidad, el estadístico 'diferencia media tipificada' definido en (1) constituye un estimador de la 'diferencia media tipificada de la población' del *i*-ésimo estudio, Δ_{E-C} :

$$\Delta_{E-C} = \frac{\mu_i^E - \mu_i^C}{\sigma_i}, \quad (3)$$

donde μ_i^E y μ_i^C representan las medias de las poblaciones de los grupos experimental y control del *i*-ésimo estudio; y σ_i la desviación típica de ambas poblaciones. Suponiendo poblaciones normales y varianzas común

$$[N(\mu_i^E, \sigma_i^2) \text{ y } N(\mu_i^C, \sigma_i^2)],$$

Hedges (1981) demostró que el estimador de Δ_{E-C} dado en (1) está sesgado, constituyendo una estimación positivamente sesgada. Por ello, propuso un estimador insesgado, $\hat{\Delta}_{E-C}^u$, añadiendo a (1) un factor de corrección:

$$\hat{\Delta}_{E-C}^u = c(m) \hat{\Delta}_{E-C}, \quad (4)$$

siendo $c(m)$ el factor de corrección, cuyo valor aproximado puede obtenerse por

$$c(m) \approx 1 - \frac{3}{4m - 1}, \quad (5)$$

con $m = n_i^E + n_i^C - 2$. Es obvio que conforme m aumenta, $\hat{\Delta}_{E-C}$ tiende a $\hat{\Delta}_{E-C}^u$, de modo que el sesgo de $\hat{\Delta}_{E-C}$ es prácticamente nulo para muestras grandes (Hedges, 1982, p. 492).

Aunque la diferencia media tipificada es un estadístico claro y simple, su cálculo puede tener problemas debidos a deficiencias y omisiones de ciertos datos en los informes de los estudios. Cuando el diseño experimental es de dos grupos, el estadístico de contraste más común es t (o F , ya que en este contexto $F = t^2$). En estos casos, se puede calcular el tamaño del efecto si al menos el informe del estudio informa del valor de t y los tamaños muestrales. Para muestras independientes, el tamaño del efecto viene dado directamente por,

$$\hat{\Delta}_{E-C} = t [(1/n_i^E) + (1/n_i^C)]^{1/2}, \quad (6)$$

cuando $n_i^E \neq n_i^C$; y por

$$\hat{\Delta}_{E-C} = t(2/n_i)^{1/2}, \quad (7)$$

cuando $n_i^E = n_i^C$. Para muestras relacionadas, suponiendo que conocemos r_{xy} ,

$$\hat{\Delta}_{E-C} = t_d [2/n_i(1 - r_{xy}^2)]^{1/2}, \quad (8)$$

donde t_d representa el valor de la prueba t para muestras dependientes.

En algunos informes de los estudios no se cita el resultado de la prueba estadística aplicada, sino sólo el nivel de significación alcanzado por dicho valor. En estos casos, si el nivel de significación citado es exacto y se conocen los tamaños muestrales, se puede calcular el tamaño del efecto transformado mediante las tablas del estadístico en cuestión el nivel de significación con los grados de libertad apropiados. Si el estadístico de contraste es t , entonces una

vez obtenido es posible determinar el tamaño del efecto mediante (6), (7) u (8), según las circunstancias. Si el nivel de significación citado en el estudio es sólo aproximativo, entonces todavía se puede calcular una aproximación conservadora al tamaño del efecto haciendo el valor t igual al nivel de significación citado.

Cuando el diseño experimental implica más de dos grupos, el estadístico de contraste más común es F . En estos casos, podemos estimar el tamaño del efecto mediante la diferencia media tipificada comparando los grupos de dos en dos, o cada uno de ellos con un grupo de control común. En cualquier caso, necesitaremos las medias de los grupos y el valor de F . La desviación típica intra-grupo puede ser obtenida mediante,

$$S_i = (\text{MC Error})^{1/2}. \quad (9)$$

Si desconocemos MC Error, podemos obviamente determinarla mediante

$$\text{MC Error} = \text{MC Inter}/F, \quad (10)$$

donde

$$\text{MC Inter} = n_j \sum_{j=1}^r (\bar{y}_{.j} - \bar{y}_{..})^2 / (r - 1), \quad (11)$$

siendo n_j el tamaño muestral, r el número de grupos, $\bar{y}_{.j}$ la media del grupo j y $\bar{y}_{..}$ la media total.

Mayores complicaciones surgen cuando los estudios utilizan diseños factoriales en los que, además del factor principal objeto de la comparación, se incluye(n) otro(s) factor(es) que puede(n) afectar al principal. En estos casos, es imprescindible disponer de las medias de los grupos y de los valores F exactos como mínimo; aunque la información más idónea sería la tabla-resumen del análisis de varianza aplicado. La mayor dificultad estriba en el cálculo de la desviación típica intra-grupo. Glass et al. (1981, p. 119) tratan con algún detenimiento estas dificultades.

Hasta aquí, nos hemos centrado en una de las diversas medidas —quizá la más popular— del tamaño del efecto: la diferencia media tipificada. Cuando disponemos de dos grupos, otra estimación del tamaño del efecto es la *correlación biserial-puntual*, r_{bp} , entre el tratamiento y el efecto, que puede encontrarse en cualquier texto de estadística descriptiva (por ejemplo, Ato et al., 1981):

$$r_{bp} = \frac{\bar{y}_i^E - \bar{y}_i^C}{S_T} (pq)^{1/2}, \quad (12)$$

donde p representa la proporción de personas del grupo experimental ($p = n_i^E/N_i$); q es la proporción de personas del grupo de control ($q = n_i^C/N_i$) y S_T es la desviación típica total de los dos grupos (siendo $N_i = n_i^E + n_i^C$). Si los tamaños muestrales son desiguales, $n_i^E \neq n_i^C$, entonces r_{bp} debe ser corregido mediante,

$$r_C = \frac{r_{bp}}{[4pq(1 - r_{bp}^2) + r_{bp}^2]^{1/2}}, \quad (13)$$

siendo,

$$pq = n_i^E n_i^C / N_i^2.$$

Hunter et al. (1982, p. 95) recomiendan el uso de la correlación biserial-puntual en lugar de la diferencia media tipificada, ya que posee la ventaja de prestarse a técnicas multivariadas tales como la correlación parcial, regresión múltiple y path-análisis. Sin embargo, la interpretación de la diferencia media tipificada es más atractiva para muchos investigadores. De cualquier modo, Glass (1978) derivó una transformación de los índices 'diferencia media tipificada', dada en (1), y 'correlación biserial-puntual', dada en (12), según la expresión,

$$r_{bp} = \frac{\hat{\Delta}_{E-C}}{[(\hat{\Delta}_{E-C})^2 + 4(N_i - 2)/N_i]^{1/2}}. \quad (14)$$

Alternativamente, se han desarrollado indicadores del tamaño del efecto no paramétricos o de distribución libre, aplicables cuando la prueba estadística utilizada no es paramétrica, o cuando se dispone de las medidas originales del estudio (Hedges & Olkin, 1984; Kraemer, 1984; Kraemer & Andrews, 1982).

Otros índices del tamaño del efecto muy comunes son aquéllos que lo estiman mediante la 'proporción de varianza explicada', o la 'fuerza de la asociación entre las variables dependiente e independiente'. Este es el caso, por ejemplo, del coeficiente *omega-cuadrada* de Hays, que informa de la proporción de varianza de la variable dependiente que es explicada por la varianza de la variable independiente y posee propósitos inferenciales. Para diseños de dos muestras donde se aplica la prueba t , *omega-cuadrada* viene dada por (cf. Hays, 1980, p. 295),

$$\omega^2 = \frac{t^2 - 1}{t^2 + n_i^E + n_i^C - 1}, \quad (15)$$

cuando $n_i^E \neq n_i^C$; y por

$$\omega^2 = \frac{t^2 - 1}{t^2 + 2n_i - 1}, \quad (16)$$

para $n_i^E = n_i^C = n_i$. Cuando disponemos de más de dos grupos y se ha aplicado F , una estimación de ω^2 viene dada por (cf. Hays, 1980, p. 349):

$$\omega^2 = \frac{SC \text{ Inter} - (r - 1) MC \text{ Error}}{SC \text{ Total} + MC \text{ Error}}. \quad (17)$$

Aunque el empleo de índices tanto descriptivos como inferenciales del tamaño del efecto posee indudables ventajas, sin embargo, Glass et al. (1981) no aconsejan su aplicación porque «... rara vez resultará satisfactorio expresar los hallazgos experimentales como una medida de la asociación entre varios niveles de una variable independiente y una variable dependiente. Tales medidas de asociación (por ejemplo, ω^2) son descriptivas de una serie completa de condiciones experimentales algo arbitraria que un investigador escoge para investigar en un único estudio» (p. 102).

1.2. Diseños correlacionales

La medida del tamaño del efecto en los estudios correlacionales es más directa que en los experimentales, ya que la fuerza de la relación entre dos variables suele ser medida por el *coeficiente de correlación de Pearson*, r_{xy} , un estadístico descriptivo muy adecuado para reflejar el tamaño del efecto (Rosenthal, 1984; Smith, 1982). De hecho, suele identificarse el tamaño del efecto con la fuerza o magnitud de la relación entre una variable dependiente y una independiente. No obstante, una correlación depende de la heterogeneidad de los casos, por lo que no puede considerarse como *el* índice ideal (Green & Hall, 1984). Además, las correlaciones están sujetas a diversas fuentes de variación, tres de las cuales es fundamental controlar en una revisión cuantitativa, a saber, el error muestral, el error de medida y la variación en el rango (Hunter et al., 1982).

A pesar de su simplicidad, cuando se pretende revisar un amplio campo de investigación correlacional, podemos encontrarnos con una variedad desconcertante de estadísticos: coeficientes de correlación de Pearson y biserial-puntual, coeficientes Phi, coeficientes de contingencia, tablas de contingencia con contrastes Chi-Cuadrado, pruebas t , análisis de varianza, etcétera. General-

TABLA I

Estadístico informado	Transformación a r_{xy}	Referencia
(a) correlación biserial-puntual.	<p>(*) $r_{xy} = r_{bp} u N(n_1 n_2)^{1/2}$</p> <p>u: ordenada en el origen de la distribución normal estandarizada.</p> <p>N: tamaño muestral total.</p>	Glas & Stanley (1970, p. 171).
(b) $t = \frac{\bar{X}_1 - \bar{X}_2}{\left[S^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \right]^{1/2}}$	$r_{bp} = \left[\frac{t^2}{t^2 + n_1 + n_2 - 2} \right]$ <p>luego pasar r_{bp} a r_{xy} según (a).</p>	Glass & Stanley (1970, p. 318)
(c) t basada en grupos extremos.	$\rho = \frac{t(2/n)^{1/2}}{\left\{ \frac{4z^2}{p^2} + \left[\frac{z^2}{p^2} - \frac{xz}{p} \right] t^2(2/n) \right\}^{1/2}}$ <p>n: tamaño del grupo. p: proporción de corte en cada extremo. z: ordenada de la curva normal en el punto de corte. x: unidad de desviación típica normal correspondiente a p (valor en la abscisa).</p>	Feldt (1961, p. 315)
(d) $F = MC \text{ Inter} / MC \text{ Error}$ para $J = 2$ grupos.	$F^{1/2} = t $. Luego proceder según (b)	
(e) $F = MC \text{ Inter} / MC \text{ Error}$ para $J > 2$ grupos.	<p>1) juntar los grupos en dos y luego proceder según (d); o</p> <p>2) $r_{xy} = \eta = [SC \text{ Inter} / (SC \text{ Inter} + SC \text{ Error})]^{1/2}$</p>	Hays (1973, pp. 683-684)
(f) χ^2 sólo (es decir, sin disponer de las frecuencias para una tabla de contingencia).	<p>(**) $r_{xy} \approx P = \left[\frac{\chi^2}{\chi^2 + N} \right]^{1/2}$</p> <p>N: tamaño muestral total.</p>	Kendall & Stuart (1967, pp. 557 y ss.)
(g) Tabla de contingencia 2×2 .	Calcular la correlación tetracórica por las tablas.	Glass & Stanley (1970, pp. 165 y ss.)
(h) Tabla de contingencia $F \times C$.	Formar una tabla 2×2 y calcular según (g).	
(i) Correlación ordinal de Spearman, r_s .	$r_{xy} = r_s$, ya que la transformación de r_s a r_{xy} bajo normalidad es casi una línea recta.	Kruskal (1958)
(j) U de Mann-Whitney.	<p>Transformar U a la correlación biserial por rangos mediante:</p> $r_{bp} = 1 - 2U/n_1 n_2$	Willson (1976)

(*) $r_{xy} \approx 1,25r_{bp}$ cuando $p = n_1/N$ está entre 0,2 y 0,8 (Magnusson, 1966, p. 205).

(**) P es el coeficiente de contingencia de Pearson y $P^2 \rightarrow \rho^2$ conforme aumenta el número de categorías de la tabla. Con pocas categorías, esta estimación puede ser demasiado baja (tomado de Glass et al., 1981, pp. 149-150).

mente, existe una transformación algebraica de esos estadísticos al coeficiente de correlación de Pearson, r_{xy} . Glass et al. (1981, pp. 149-150) recogen algunas de estas transformaciones en una tabla similar a la que reproducimos en la tabla I.

1.3. Conclusión

Según se desprende de los dos puntos anteriores, existen diversos índices para medir el tamaño del efecto de un estudio, pero todos van encaminados hacia un mismo fin: encontrar una medida que permita comparar los resultados de los estudios de un modo significativo y revelador. El recuento de resultados significativos y la acumulación de niveles de significación casi nunca resultan ser reveladores; además, las técnicas de recuento son propensas a arrojar conclusiones falsas. Sin embargo, el tamaño del efecto es ampliamente aplicable y a menudo significativo. Por regla general, la diferencia media tipificada y el coeficiente de correlación de Pearson, para los diseños experimentales y correlacionales respectivamente, son las dos medidas del tamaño del efecto más adecuadas. De cualquier modo, no hay que olvidar que tales índices se deben aplicar cuando las variables dependientes de los estudios que se quieren integrar no son directamente comparables; en caso contrario, es más directo comparar los estudios por sus propios resultados (Glass et al., 1981; Green & Hall, 1984; Hedges & Olkin, 1985; Rosenthal, 1984).

INTEGRACION DE LOS ESTUDIOS

La última etapa de las revisiones cuantitativas consiste en el análisis e integración de los hallazgos de los estudios. Téngase en cuenta que toda revisión cuantitativa tiene como unidad básica el 'estudio'. Las diversas características de los estudios —previamente codificadas— se convierten en las variables independientes y la medida de los hallazgos actúa como variable dependiente. No obstante, el modo como se puede llevar a cabo una integración cuantitativa varía según dos criterios básicos: (a) el tipo de medida de los hallazgos utilizado (recuento de resultados significativos, niveles de significación, tamaños del efecto), y (b) el carácter descriptivo o inferencial desde el que el revisor enfoca la integración.

Con el paso del tiempo, las técnicas de análisis e integración se han ido sofisticando cada vez más. Sin pretensiones de exhaustividad, describimos a continuación las técnicas de integración cuantitativa más conocidas en orden creciente de eficacia.

1. Técnicas de recuento (vote-counting methods)

Constituyen unas de las primeras —y rudimentarias— técnicas desarrolladas para integrar estudios de investigación. En su forma más simple, consisten en una sencilla tabulación de los resultados 'significativos' frente a los 'no significativos' acerca de un determinado tópico de investigación. Un análisis posterior mediante Chi-Cuadrado de Pearson permite al revisor determinar la significación global de los resultados de los estudios. Light y Smith (1971) bautizaron esta técnica con el nombre de *voting method* y la describieron así: «Se examinan todos los estudios que aporten datos sobre una variable independiente específica de interés. Quedan definidos tres resultados posibles. La relación entre las variables dependiente e independiente puede ser significativamente positiva, significativamente negativa o no significativa en ninguna dirección. Se tabula el número de estudios que se incluyen dentro de cada una de estas tres categorías. Si un número considerable de estudios cae dentro de una de estas tres categorías y muy pocos en las otras dos, entonces esa categoría modal proporciona la mejor estimación de la dirección de la verdadera relación entre las variables dependiente e independiente» (1971, p. 433).

Aunque los procedimientos de recuento son más estandarizados y replicables que la integración cualitativa, presentan ciertas limitaciones. En primer lugar, el método de recuento no tiene en cuenta el tamaño de la muestra. Como sabemos, las muestras grandes producen más frecuentemente resultados estadísticamente significativos que las muestras pequeñas. No obstante, esta complicación puede eliminarse *post hoc* si se comprueba que el tamaño muestral no está relacionado con la magnitud de los hallazgos de los estudios.

Incluso cuando se comprueba que el tamaño muestral no afecta a la interpretación de los niveles de significación, la cuestión de la magnitud de los efectos queda sin responder. Desafortunadamente, los investigadores suelen asignar a los niveles de significación más valor informativo del que realmente poseen, en detrimento de la atención prestada al tamaño del efecto.

Pero el problema más grave de los procedimientos de recuento es que conducen a conclusiones falsas. En efecto, Hedges y Olkin (1980) han demostrado que, si existe un efecto verdadero, entonces en una serie cualquiera de estudios cuya potencia estadística media sea menor de 0,50, la probabilidad de cometer una conclusión falsa al utilizar este método *aumenta* con el número de estudios.

A pesar de la asiduidad con que estos procedimientos son utilizados en la actualidad, las críticas planteadas dicen mucho en contra de éstos como mé-

todos de integración cuantitativa de estudios (cf. por ej., Glass et al., 1981; Hedges, 1984; Hedges & Olkin, 1980; Hunter et al., 1982).

2. Técnicas de acumulación de niveles de significación

Aunque pueden catalogarse dentro de las técnicas de recuento, los métodos de acumulación de niveles de significación están más elaborados. Consisten en la integración de los niveles de significación de una serie de estudios a los que el estadístico de contraste alcanzó (o no) la significación. A continuación, se obtiene el nivel de significación promedio, p , que representa a todos los estudios. Si este valor p es suficientemente bajo, el revisor podrá concluir que el efecto probado en dichos estudios es significativo (Rosenthal, 1978; Rosenthal & Rosnow, 1984; Rosenthal & Rubin, 1979).

Existen diversos modos de acumular niveles de significación (Hedges & Olkin, 1985; Rosenthal, 1984; Strube, 1985). Rosenthal (1978) ha evaluado siete de ellos.

2.1. Suma de unidades logarítmicas

Este método, también llamado método de Fisher, supone que si tenemos K estudios independientes con niveles de significación p_1, p_2, \dots, p_k , entonces partiendo de una hipótesis nula común a todos los estudios,

$$\sum_{i=1}^k (-2 \log_e p_i) \sim \chi^2_{2K} \quad (18)$$

Una vez calculado el sumatorio, sólo resta consultar el resultado en las tablas de Chi-Cuadrado con $2K$ grados de libertad, para determinar el nivel de significación representativo de la serie de estudios.

Este método, aunque muy utilizado, sufre la desventaja de que produce el rechazo de la hipótesis nula de forma rutinaria cuando el número de estudios es grande ($K > 5$). Además, es sensible a la no independencia de los estudios. A pesar de todo, esta técnica continúa siendo la más conocida y discutida de todas las que combinan probabilidades independientes.

2.2. Suma de probabilidades

Esta técnica, propuesta por Edgington (1972a), permite calcular el nivel de

significación, p , de una serie de estudios mediante la razón entre la suma de las probabilidades, p_i , de los K estudios elevada a la k -ésima potencia y $K!$:

$$P = \left(\sum_{i=1}^k p_i \right)^k / K! \quad (19)$$

Este método es aconsejable para series pequeñas de estudios, ya que requiere que la suma de los niveles p no exceda mucho la unidad; de lo contrario, el nivel P global obtenido tiende a ser bastante conservador, a no ser que se apliquen correcciones especiales.

2.3. Suma de puntuaciones t

Consiste en sumar las puntuaciones t de cada estudio y dividir por la raíz cuadrada de la suma de las varianzas de las distribuciones t de cada estudio. Así pues, siendo la varianza de la distribución t de Student para el i -ésimo estudio,

$$V(tv_i) = v_i / (v_i - 2),$$

entonces el estadístico

$$z = \frac{\sum_{i=1}^k t_i}{\left\{ \sum_{i=1}^k [v_i / (v_i - 2)] \right\}^{1/2}}, \quad (20)$$

sigue una ley normal. Por tanto, el valor z obtenido es consultado en las tablas de la distribución normal tipificada para obtener el valor de significación, p , representativo de la serie de estudios integrados.

La limitación de esta técnica ideada por Winer (1971) es que no puede ser aplicada cuando el tamaño de las muestras es inferior a tres, ya que en este caso la varianza sería cero o un valor negativo. Además, este método no proporciona buenas aproximaciones a la distribución normal para $v_i < 10$.

2.4. Suma de puntuaciones z

Esta técnica denominada 'método Stouffer', asume que si tenemos K estudios independientes, la razón entre la suma de las K puntuaciones típicas

(obtenidas de cada nivel de significación) y $K^{1/2}$, constituye una unidad de desviación normal, z , y, como tal, adopta la forma de una distribución normal,

$$z = \frac{\sum_{i=1}^k z_i / k^{1/2}}{\sum_{i=1}^k z_i / k^{1/2}} \quad (21)$$

Consultando las tablas de la distribución normal tipificada, obtenemos el nivel de significación global, p , de los estudios integrados.

2.5. Suma de puntuaciones z ponderadas

Mosteller y Bush (1954) sugirieron una técnica que consiste en ponderar la puntuación típica de cada estudio por su tamaño muestral, o por los grados de libertad, o por cualquier otro criterio, tales como la elegancia, la validez interna o la validez ecológica del estudio individual. Así pues, suponiendo una ponderación por los grados de libertad, la razón entre la suma de los productos de cada puntuación típica por sus grados de libertad y la raíz cuadrada de la suma de los grados de libertad, se distribuye según una ley normal:

$$z = \left(\frac{\sum_{i=1}^k z_i v_i}{\left(\sum_{i=1}^k v_i^2 \right)^{1/2}} \right) \quad (22)$$

A continuación, procedemos como en (21).

2.6. Contraste del nivel P medio

Edginton (1972b) ha propuesto una técnica por la que la razón dada por la diferencia entre 0,50 y la probabilidad media, \bar{P} , de los K niveles de significación y la raíz cuadrada de la razón entre la varianza de los p_i niveles y K , sigue una ley normal:

$$z = (0,50 - \bar{P}) / (\sigma_p / K)^{1/2} \quad (23)$$

siendo

$$\sigma_p = \left[\frac{\sum_{i=1}^k (p_i - \bar{P})^2}{K - 1} \right]^{1/2}$$

A continuación, procedemos como en (21).

La ventaja de esta técnica es su simplicidad de cálculo, pero tiene la limitación de que el número de estudios no debe ser inferior a cuatro.

2.7. Contraste de la puntuación z media

Mosteller y Bush (1954) modificaron el método Stouffer transformando los niveles de significación en puntuaciones típicas, y computando una prueba t (o F) sobre la puntuación típica media, con $K - 1$ grados de libertad para t ; y con 1 y $K - 1$ grados de libertad para F :

$$t = \left(\frac{\sum_{i=1}^k z_i / K}{(\sigma_z / K)^{1/2}} \right) \quad (24)$$

siendo σ_z ,

$$\sigma_z = \left[\frac{\sum_{i=1}^k (z_i - \bar{z})^2}{K - 1} \right]^{1/2}$$

Esta técnica presenta una baja potencia cuando el número de estudios es pequeño, concretamente, cuando $K < 5$.

Resumen, aunque no existe una técnica perfectamente válida para todas las condiciones, la que se ajusta a un mayor rango de situaciones es la de sumar puntuaciones típicas, ponderadas o no (Rosenthal, 1978, p. 191). El problema básico de los métodos de acumulación de niveles de significación es que no aportan información alguna acerca del tamaño del efecto (Hunter et al. 1982). El mismo Rosenthal (1978, p. 192; ver también Rosenthal & Rubin, 1979) ha admitido la necesidad de completar las técnicas de acumulación de probabilidades con un análisis del tamaño del efecto. De hecho, Hedges y Olkin (1980, pp. 363-366) han desarrollado tres procedimientos de recuento alternativos que permiten estimar el tamaño del efecto, pero sólo son aplicables cuando los tamaños muestrales se mantienen constantes para todos los estudios integrados.

3. Técnicas de meta-análisis

Los procedimientos de recuento y de acumulación de niveles de significación arriba expuestos coinciden en dos aspectos básicos: (a) no miden el tamaño del efecto (excepto las técnicas desarrolladas por Hedges y Olkin, 1980) y (b) su enfoque es eminentemente inferencial. A diferencia de éstas, el meta-análisis pone el énfasis en el cálculo del tamaño del efecto, pero su enfoque puede ser descriptivo y/o inferencial. De hecho, de las dos técnicas de meta-análisis más conocidas actualmente, una es básicamente descriptiva y la otra inferencial. Nos estamos refiriendo a los meta-análisis de Glass (1976, 1977; Glass et al., 1981) y de Hunter-Schmidt (Hunter et al., 1982). No obstante, las dos técnicas no son

incompatibles, ya que el meta-análisis de Hunter-Schmidt constituye una extensión del meta-análisis de Glass.

3.1. *El meta-análisis de Glass*

Básicamente, Glass aborda la integración de los estudios desde la perspectiva del análisis multivariado, donde el tamaño del efecto constituye la medida o variable dependiente y las características codificadas de los estudios son variables independientes. Para ello, se vale de la amplia gama de los métodos estadísticos, desde los estadísticos descriptivos univariados, tablas de frecuencia y correlaciones, hasta la estimación de modelos lineales, análisis de regresión, análisis factorial y, en definitiva, «... cualquiera de los métodos de análisis estadístico que hayan demostrado ser útiles para la extracción del significado de los datos» (Glass et al., 1981, p. 153).

Las propiedades más importantes de esta técnica son las siguientes:

(1) Un fuerte énfasis en el tamaño del efecto en lugar del nivel de significación. Según Glass, el propósito de la investigación es más descriptivo que inferencial, y considera que los estadísticos descriptivos más importantes son los que indican más claramente la magnitud de los efectos. Su meta-análisis utiliza generalmente estimaciones de la diferencia media tipificada, para los estudios experimentales, y el coeficiente de correlación de Pearson, para los correlacionales. El primer paso de la integración es, pues, obtener los estadísticos descriptivos más informativos, es decir, el tamaño del efecto medio y su desviación típica, $\hat{\sigma}_{ES}$. A continuación, en base a alguna(s) característica(s) de los estudios se fragmenta la serie y se computan las medias de esos fragmentos. Así, se observa la posible influencia de determinada(s) característica(s) sobre los tamaños del efecto.

(2) Aceptación de la varianza de los tamaños del efecto, $\hat{\sigma}_{ES}^2$, en su valor aparente. Es decir, se supone que la variación entre los tamaños del efecto de los estudios es real y, como tal, debería tener alguna explicación sustantiva. Estas explicaciones se buscan en las características variables de los estudios calculando las correlaciones entre las características de los estudios y los tamaños del efecto para determinar su potencia explicatoria.

(3) Las características de los estudios que deben ser codificadas y contrastadas para su posible asociación con los resultados de los estudios son seleccionadas en base a criterios eminentemente empíricos. Glass (1976, 1977) cree que tales cuestiones son 'empíricas' y parece restar importancia al papel de la teoría

o la lógica en la determinación de las variables que deben ser contrastadas como moderadores potenciales de los resultados de los estudios.

3.2. *El meta-análisis de Hunter-Schmidt*

Hunter et al. (1982, pp. 140-142) han desarrollado un enfoque inferencial del meta-análisis de Glass, así como técnicas para corregir los tamaños del efecto de los errores de medida y de muestreo y de la restricción del rango. Este procedimiento nació de los estudios de la validez de los tests de empleo (cf. por ejemplo, Schmidt & Hunter, 1977) y, por tanto, su aplicación está más difundida en los estudios correlacionales que en los experimentales. A diferencia del meta-análisis de Glass, estos procedimientos recomiendan como medida del tamaño del efecto el coeficiente de correlación en lugar de la diferencia media tipificada.

Veamos las propiedades más importantes:

(1) Atención especial al tamaño del efecto en lugar del nivel de significación. Los tamaños del efecto deben ser corregidos de los artefactos estadísticos y de medida, que los desvían de sus verdaderos valores.

(2) A diferencia del meta-análisis de Glass, el de Hunter-Schmidt no toma la varianza del tamaño del efecto en su valor aparente. Al contrario, el primer paso después de obtener una estimación del verdadero tamaño del efecto medio es contrastar la hipótesis de que $\hat{\sigma}_{ES}^2$ se debe completamente a diversos artefactos estadísticos. Estos artefactos incluyen: (a) el error muestral; (b) diferencias entre los estudios en la fiabilidad de las variables dependiente e independiente; (c) diferencias en la restricción del rango; (d) diferencias en la validez del instrumento de medida; y (e) los errores de transcripción, tipográficos y computacionales. Hunter et al. (1982, pp. 35 y ss.) exponen métodos para estimar la varianza debida a los tres primeros artefactos citados. Por regla general, si estos tres artefactos explican el 75% o más de $\hat{\sigma}_{ES}^2$, se puede concluir que la varianza residual (el 25% o menos) es probablemente debida a los otros dos artefactos y, por tanto, que $\hat{\sigma}_{ES}^2 = 0$. Este resultado indicaría que el tamaño del efecto es constante a través de los estudios, a pesar de las diversas características que varían entre los estudios revisados. En este caso, el resultado final del meta-análisis es el tamaño del efecto medio estimado.

(3) Si la hipótesis de que $\hat{\sigma}_{ES}^2 = 0$ no puede ser aceptada, entonces se correlacionan las propiedades de los estudios con los tamaños del efecto, tal como sugiere Glass, excepto que el método de Hunter-Schmidt se fia de consideraciones teóricas, lógicas, estadísticas y psicométricas cuando es posible decidir qué características de los estudios codificar y cómo hacerlo.

(4) Estas correlaciones entre las características de los estudios y los tamaños del efecto son corregidas de los errores de muestreo. Si todas las correlaciones no son significativas, pasamos al punto (8).

(5) Las mismas correlaciones son corregidas de la falta de fiabilidad de las características de los estudios.

(6) Se obtienen mediante análisis de regresión múltiple las puntuaciones pronosticadas del tamaño del efecto sobre las características de los estudios. Las ponderaciones beta resultantes deben interpretarse como índices de los posibles efectos causales de las verdaderas características de los estudios sobre los verdaderos tamaños del efecto.

(7) La verdadera correlación múltiple resultante, R , debe ser corregida mediante la fórmula de corrección apropiada. El valor R^2 corregido indica la proporción de la varianza del tamaño del efecto explicada por la variación en las características de los estudios.

(8) A continuación, pueden derivarse tres tipos diferentes de distribución. Estas constituyen los productos finales del meta-análisis. Las dos primeras tienen la misma media: el tamaño del efecto medio; además, la desviación típica de las tres distribuciones es corregida de los artefactos estadísticos:

(a) Una distribución presenta una desviación típica corregida para los efectos de los artefactos estadísticos sólo. Esta distribución describe los verdaderos tamaños del efecto que se separan cuando las características de los estudios se permite que varíen. Se puede obtener intervalos confidenciales para esta distribución.

(b) Otra distribución presenta una desviación típica corregida para los artefactos estadísticos y 'para los efectos de las desviaciones de las características de los estudios de sus valores promedio'. Esta distribución describe el verdadero tamaño del efecto que se espera cuando las características de los estudios son mantenidas constantes en sus valores medios. Se pueden obtener intervalos confidenciales también.

(c) Por último, un grupo de distribuciones presenta desviaciones típicas corregidas para los artefactos estadísticos y manteniendo constantes las características de los estudios en valores diferentes a los de sus medias. Suponiendo homoscedasticidad, estas desviaciones típicas deberían ser iguales a la de la distribución (b). Pero, obviamente, las medias de los tamaños del efecto pueden diferir. Así, definiendo previamente los valores de las características es posible estimar el tamaño del efecto medio esperado bajo tales condiciones.

Lógicamente, cuando las correlaciones entre las características de los estu-

dios y los tamaños del efecto son nulas, las distribuciones (a), (b) y (c) serán idénticas.

3.3 Conclusión

Las técnicas de meta-análisis constituyen en la actualidad los métodos de revisión cuantitativa más perfeccionados e informativos. No obstante, aún están en período de formación y, por esta razón, sufren de limitaciones. Dos problemas graves son la capitalización sobre el azar y la baja potencia estadística que se produce cuando se aplica el análisis de regresión múltiple de los tamaños del efecto sobre las características de los estudios. No existe una solución puramente estadística a estos problemas, pero pueden ser mitigados basándonos en la selección de las características de los estudios y de las conclusiones finales, no sólo en los estadísticos a mano, sino también en otros hallazgos teóricamente relevantes (por ejemplo, los resultados de otros meta-análisis) y en consideraciones teóricas (Hunter et al., 1982).

Para finalizar, recomendamos como una estimación del tamaño del efecto medio un procedimiento desarrollado simultáneamente por Hedges (1982, pp. 493-495) y por Rosenthal y Rubin (1982, pp. 500-501). No obstante, sólo será aplicable cuando los tamaños del efecto de los estudios cumplan con los supuestos de homogeneidad e independencia.

VENTAJAS E INCONVENIENTES DE LA REVISIÓN CUANTITATIVA

No cabe duda de que los métodos formales para sintetizar hallazgos de investigación de muchos estudios representan una innovación metodológica significativa. De las diversas técnicas cuantitativas, el meta-análisis parece la más prometedora. No obstante, los métodos de integración cuantitativa se encuentran aún en esa primera etapa de 'aprendizaje por ensayo-error', a pesar de su extendida difusión. Muchos investigadores se han preocupado por la valoración de estas técnicas innovadoras (por ejemplo, Bullock & Svyantek, 1985; Cook & Levinton, 1980; Cooper, 1984; Eysenck, 1978; Glass et al., 1981; Green & Hall, 1984; Hunter et al., 1982; Pillemer, 1984; Wortman, 1983). Expondremos a continuación las ventajas que ofrecen estas técnicas de revisión cuantitativa sobre las revisiones cualitativas, así como sus debilidades más comentadas.

1. Ventajas

Un buen número de autores coinciden en apuntar las siguientes ventajas de las revisiones cuantitativas:

(a) La revisión cuantitativa es un método eficiente para resumir literaturas extensas. Ante un tópico de investigación muy documentado, el revisor cualitativo es incapaz de describir y abarcar todos los resultados y diseños de los estudios.

(b) Probablemente, la revisión cuantitativa es más objetiva que la tradicional. Debido a la meticulosidad en la búsqueda de la literatura, el revisor cuantitativo difícilmente podrá ponderar desmerecidamente aquellos estudios que sean coherentes con sus prejuicios.

(c) El análisis estadístico puede detectar relaciones y tendencias demasiado sutiles al revisor cualitativo. El revisor cuantitativo puede estudiar las interacciones entre los estudios; interacciones que pasan desapercibidas al método tradicional.

(d) Como consecuencia del punto anterior, el revisor cuantitativo puede contrastar hipótesis que no pueden comprobarse en los estudios individuales.

(e) Debido a la mayor seguridad de sus conclusiones, la revisión cuantitativa permite llenar vacíos en la literatura y proponer convenientemente nuevas direcciones a la investigación.

2. Inconvenientes

A pesar de su gran popularidad, los métodos de revisión cuantitativa están sufriendo duras críticas (Eysenck, 1978, 1983; Gallo, 1978; Presby, 1978; Searles, 1985). Básicamente, pueden resumirse del modo siguiente:

(a) Aunque cualquier revisor puede equivocarse al establecer los resultados de un estudio, el revisor cuantitativo es más propenso a cometer errores. La razón estriba en que el revisor tendrá que realizar cálculos para obtener los tamaños del efecto. No obstante, parece estar garantizada una alta fiabilidad entre distintos revisores de una misma serie de estudios.

(b) Muchos críticos del meta-análisis (por ejemplo, Eysenck, 1978) afirman que no tiene sentido mezclar resultados de estudios que son *diferentes*, es decir, con técnicas de medida diferentes, tipos de personas distintos, etcétera ('el problema de las manzanas y las naranjas'). Contra esta crítica, Glass et al. (1981)

arguyen que tales comparaciones son las únicas que tienen sentido, ya que los estudios *iguales* no vale la pena compararlos. Además, en el meta-análisis las manzanas son codificadas como manzanas y las naranjas como naranjas.

(c) El meta-análisis descarta la calidad de los estudios como criterio para seleccionar los estudios. Así, se dice que el meta-análisis descansa sobre cantidades indiscriminadas de datos más que sobre la calidad de los estudios. Sin embargo, Glass et al. (1981) y Hunter et al. (1982) aseguran que eliminar estudios *a priori* por inadecuaciones metodológicas es una estrategia errónea, ya que el posible efecto de la calidad del diseño puede ser valorado *a posteriori* sin necesidad de excluir de antemano algunos estudios.

(d) La revisión cuantitativa depende de los estudios localizados. Generalmente, suelen localizarse sólo los estudios publicados y, como se sabe, los hallazgos de éstos presentan diferencias sistemáticas respecto de los no publicados. En este sentido, Lane y Dunlap (1978) han argumentado que los tamaños del efecto de los estudios publicados son probablemente sobre-estimaciones del valor paramétrico, dado que los estudios con resultados no significativos no suelen superar la selección editorial. No obstante, Rosenthal (1979) se enfrentó a este problema desarrollando una fórmula matemática por la que cualquier revisor cuantitativo puede estimar el número de resultados nulos no localizados que deberían existir en los ficheros para que la probabilidad conjunta del grupo completo de resultados —los localizados y los no recuperables— exceda un cierto nivel de significación. De este modo, si el número de resultados nulos necesarios para contrarrestar los resultados localizados es muy elevado, entonces podremos valorar con cierta seguridad los resultados obtenidos. En definitiva, las conclusiones alcanzadas en cualquier meta-análisis deben tener en cuenta esta posible amenaza a la validez de los resultados. En palabras de Rosenthal y Rosnow (1984): «Resultaría muy útil a los lectores si para cada valor *p* combinado que presentaran los revisores, también proporcionaran su tolerancia a futuros resultados nulos asociada con su nivel de significación» (p. 382).

(e) Las revisiones cuantitativas son realizadas sobre grandes bloques de datos en los que de un mismo estudio se pueden obtener varias estimaciones del tamaño del efecto. Esto resta independencia a los datos y provoca una impresión equivocada de la fiabilidad de los resultados. Glass et al. (1981, p. 229) reconocen que ésta es la crítica más sólida de las planteadas al meta-análisis. La principal consecuencia de la no independencia es una reducción de la fiabilidad de la estimación de los promedios o de las ecuaciones de regresión. Hunter et al. (1982) manifiestan que no existe una respuesta única a este problema. El modo de integración de los resultados dependerá de la naturaleza del proceso de replicación utilizado en el estudio. Pero, en cualquier caso, la inclusión de

resultados múltiples de un mismo estudio debe estar en función de si las variables manipuladas en el experimento afectarán o no a la integración de los estudios.

3. Conclusión

De los inconvenientes arriba expuestos se desprende la idea de que los revisores cuantitativos deben desarrollar nuevos métodos y soluciones imaginativas a la integración de la investigación. No obstante, las ventajas de las técnicas descritas sobre las revisiones tradicionales son manifiestas. Por supuesto, no existe un único modo de realizar algo tan complejo como una revisión de literatura. Está claro que en aquellas áreas donde se acumulan pocos estudios la revisión tradicional a menudo será muy beneficiosa; pero en otras muchas áreas, los métodos estadísticos pueden ayudar e incluso convertirse en imprescindibles. Como exponen Green y Hall (1984): «Los métodos estadísticos, para que sean útiles, deben utilizarse con consideración. El análisis de los datos es una ayuda para el pensamiento, no un sustituto. Ninguna fórmula mecánica funcionará. Pero las revisiones cuantitativas meticulosas es probable que desempeñen un papel importante en los próximos avances de la psicología» (p. 52).

A pesar de su carácter novedoso, las técnicas de meta-análisis se están aplicando con gran asiduidad actualmente. Es, pues, de esperar que, conforme se desarrollen procedimientos más sofisticados dentro de este campo, el papel de las revisiones cuantitativas dentro del progreso científico del conocimiento irá incrementándose.

REFERENCIAS BIBLIOGRAFICAS

- ATO, M.; LOPEZ, J. A. & SERRANO, J. M. (1981). *Fundamentos de Estadística Inferencial*. Murcia: Yerba.
- BULLOCK, R. J. & SVYANTEK, D. J. (1985). Analyzing meta-analysis: Potencial problems, an unsuccessful replication, and evaluation criteria. *Journal of Applied Psychology*, 70(1), 108-115.
- COHEN, J. (1977): *Statistical power analysis for the Behavioral Sciences*. New York: Academic Press.
- COOK, T. D. & LEVINTON, L. C. (1980). Reviewing the literature: A comparison of traditional methods with meta-analysis. *Journal of Personality*, 48(4), 449-472.
- COOPER, H. M. (1982). Scientific guidelines for conducting integrative research reviews. *Review of Educational Research*, 52(2), 291-302.
- COOPER, H. M. (1984): *The integrative research review: A systematic approach*. Beverly Hills: Sage.
- COOPER, H. M. & ROSENTHAL, R. (1980). Statistical versus traditional procedures for summarizing research findings. *Psychological Bulletin*, 87, 442-449.
- EDGINTON, E. S. (1972a). An additive method for combining probability values from independent experiments. *Journal of Psychology*, 80, 351-363.
- EDGINTON, E. S. (1972b). A normal curve method for combining probability values from independent experiments. *Journal of Psychology*, 82, 85-89.
- EYSENCK, H. (1978). An exercise in mega-silliness. *American Psychologist*, 33, 517.
- EYSENCK, H. (1983). Special Review. *Behaviour Research & Therapy*, 21(3), 315-320.
- FELDT, L. S. (1961). The use of extreme groups for the presence of a relationship. *Psychometrika*, 26, 307-316.
- FISHER, R. A. (1966). *The design of experiments*. Oliver & Boyd, 8th ed.
- GALLO, P. (1978). Meta-analysis a mixed meta-phor? *American Psychologist*, 33, 515-517.
- GLASS, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5, 3-8.
- GLASS, G. V. (1977). Integrating findings: The meta-analysis of research. *Review of Research in Education*, 5, 351-379.
- GLASS, G. V. (1978). Reply to Mansfield and Bussey. *Educational Researcher*, 7, 3.
- GLASS, G. V. & STANLEY, J. C. (1970). *Statistical methods in education and psychology*. Englewood Cliffs: Prentice-Hall (Edición en castellano: Prentice-Hall, 1974).
- GLASS, G. V.; MCGAW, B. & SMITH, M. L. (1981). *Meta-analysis in Social Research*. Beverly Hills: Sage.
- GREEN, B. F. & HALL, J. A. (1984). Quantitative methods for literature reviews. *Annual Review of Psychology*, 35, 37-53.
- HAYS, W. L. (1980). *Statistics*. New York: Holt, Rinehart & Winston, 8th ed.
- HEDGES, L. V. (1981). Distribution theory for Glass's estimation of effect size and related estimators. *Journal of Educational Statistics*, 6, 107-128.
- HEDGES, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin*, 92(2), 490-499.

- HEDGES, L. V. (1984). Research synthesis: The state of the art. *International Journal Aging & Human Development*, 19(2), 85-93.
- HEDGES, L. V. & OLKIN, I. (1980). Vote-counting methods in research synthesis. *Psychological Bulletin*, 88(2), 359-369.
- HEDGES, L. V. & OLKIN, I. (1984). Nonparametric estimators of effect size in meta-analysis. *Psychological Bulletin*, 96(3), 573-580.
- HEDGES, L. V. & OLKIN, I. (1985). *Statistical Methods for Meta-analysis*. Orlando: Academic Press.
- HUNTER, J. E.; SCHMIDT, F. L. & JACKSON, G. B. (1982). *Meta-analysis: Cumulating findings across Studies*. Beverly Hills: Sage.
- JUDD, C. M. & KENNY, D. A. (1981). *Estimating the effects of Social Interventions*. Cambridge University Press.
- KENDALL, M. G. & STUART, A. (1967). *The advanced theory of statistics vol. 2*. London: Griffin.
- KRAEMER, H. C. (1984). Nonparametric effect size estimation: A reply. *Psychological Bulletin*, 96(3), 569-572.
- KRAEMER, H. C. & ANDREWS, G. (1982). A nonparametric technique for meta-analysis effect size calculation. *Psychological Bulletin*, 91(2), 404-412.
- KRUSKAL, W. H. (1958). Ordinal measures of association. *Journal of the American Statistical Association*, 53, 814-861.
- LANE, D. M. & DUNLAP, W. P. (1978). Estimating effect size: Bias resulting from the significance criterion in editorial decisions. *British Journal of Mathematical & Statistical Psychology*, 31, 107-112.
- LIGHT, R. J. & SMITH, P. V. (1971). Accumulating evidence: Procedures for resolving contradictions among different research studies. *Harvard Educational Review*, 41, 429-471.
- MAGNUSSON, D. (1966). *Test theory*. Reading: Addison-Wesley (edición en castellano: Trillas, 1977).
- MOSTELLER, F. M. & BUSH, R. R. (1954). Selected quantitative techniques. En G. Lindzey (ed.) *Handbook of Social Psychology. Vol. I: Theory and Method*. Cambridge: Addison-Wesley.
- ORTEGA, J. E. & FERNANDEZ DOLS, J. M. (1980): *Fuentes documentales en psicología*. Debate.
- PILLEMER, D. B. (1984). Conceptual issues in research synthesis. *Journal of Special Education*, 18(1), 27-40.
- PRESBY, S. (1978). Overly broad categories obscure important differences between therapies. *American Psychologist*, 33, 514-515.
- ROSENTHAL, R. (1978). Combining results of independent studies. *Psychological Bulletin* 85(1), 185-193.
- ROSENTHAL, R. (1979). The 'file drawer' problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638-641.
- ROSENTHAL, R. (1980). Quantitative assessment for research domains. En R. Rosenthal (ed.) *New Directions for Methodology of Social and Behavioral Science*, n.º 5. San Francisco: Jossey-Bass.
- ROSENTHAL, R. (1984). *Meta-analytic Procedures for Social Research*. Beverly Hill: Sage.
- ROSENTHAL, R. & ROSNOW, R. L. (1984). *Essentials of Behavioral Research: Methods and Data Analysis*. McGraw Hill.
- ROSENTHAL, R. & RUBIN, D. B. (1979). Comparing significance levels of independent studies. *Psychological Bulletin*, 86(5), 1.165-1.168.
- ROSENTHAL, R. & RUBIN, D. B. (1980). Comparing within and between-subject studies. *Sociological Methods & Research*, 9, 127-136.

- ROSENTHAL, R. & RUBIN, D. B. (1982). Comparing effect sizes of independent studies. *Psychological Bulletin*, 92(2), 500-504.
- SANCHEZ, J. (1985). La hipótesis del déficit perceptivo del retraso específico en lectura: Un estudio meta-analítico. *Anales de Psicología*, 2, 75-91.
- SCHMIDT, S. L. & HUNTER, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, 62, 529-540.
- SEARLES, J. S. (1985). A methodological and empirical critique of psychotherapy outcome meta-analysis. *Behavioral Research Therapy*, 23(4), 453-463.
- SMITH, M. L. (1982). Research integration. En H. E. Mitzel (ed.) *Enciclopedia of Educational Research*, vol. III (pp. 1.613-1.618). The Free Press.
- STRUBE, M. J. (1985). Combining and comparing significance levels from nonindependent hypothesis tests. *Psychological Bulletin*, 97(2), 334-341.
- WALBERG, H. J. & HAERTEL, E. H. (1980). Research integration: The state of the art. Special issue of *Evaluation in Education: An International Review Service*, 4, 1-135.
- WILLSON, V. L. (1976). Critical values of the rank-biserial correlation coefficient. *Educational & Psychological Measurement*, 36, 297-300.
- WINER, B. J. (1971). *Statistical Principles in Experimental Design*. New York: McGraw Hill.
- WISE, S. L.; PETERS, L. H. & O'Connor, E. J. (1984). Identifying moderator variables using multiple regression: A reply to Darrow and Kahl. *Journal of Management*, 10(2), 227-236.
- WORTMAN, P. M. (1983). Evaluation research: A methodological perspective. *Annual Review of Psychology*, 34, 223-260.