

META-ANÁLISIS: UNA ALTERNATIVA METODOLÓGICA A LAS REVISIONES TRADICIONALES DE LA INVESTIGACIÓN

Julio Sánchez Meca
Manuel Ato García
Universidad de Murcia

1. INTRODUCCIÓN

Uno de los principios fundamentales de la investigación científica es que debe ser «repliable». El propósito de la sección «Método» de un informe de investigación es, precisamente, proporcionar una descripción del procedimiento experimental lo suficientemente detallada como para que otro investigador pueda repetir el estudio. Los resultados de una investigación aislada no se pueden elevar a la categoría de hechos científicos hasta que no hayan sido replicados por otros investigadores. En efecto, el considerable trabajo invertido en la realización de una investigación aporta tan sólo una pequeña pieza de un enorme rompecabezas; es preciso retomar investigaciones anteriores introduciendo variaciones en el procedimiento para poner a prueba, no sólo la existencia, sino también la consistencia o robustez de un determinado hallazgo empírico (Evans, 1985).

Este carácter acumulativo del conocimiento científico requiere que entre la investigación pasada y la futura se incluya una etapa intermedia imprescindible para su progreso: la revisión de la literatura y la resolución de contradicciones empíricas. Así pues, las revisiones de investigación cumplen una función primordial en la identificación de hallazgos suficientemente contrastados como

para ser considerados relevantes. Dado su importante papel, las revisiones de investigación han suscitado un creciente interés por la metodología de la revisión. Uno de los detonantes de este interés ha sido la enorme proliferación de investigación que durante las últimas décadas han sufrido las ciencias del comportamiento, junto con una multiplicación de las fuentes y medios de acceso a la información, entre las que cabe destacar la búsqueda informatizada de la literatura (Cooper, 1984). Pero el exponente esencial de esta transformación metodológica proviene de la introducción de métodos estadísticos en el proceso de revisión y, en general, de la aplicación de los mismos criterios de rigor empírico exigidos en investigaciones originales (Hedges, 1984b).

Con anterioridad a esta situación, la tarea de un revisor de la literatura era relativamente sencilla, ya que sobre un determinado tópico de investigación no solía haber más que unos pocos estudios. Actualmente, la capacidad de cualquier revisor se ve con demasiada frecuencia sobrecargada. En la mayoría de las ocasiones resulta incluso impracticable una simple descripción de los resultados de estudios existentes. Como consecuencia, ha surgido una propuesta alternativa para la integración de la investigación, que permite realizar un análisis sistemático de las investigaciones sobre un tema (Glass, 1976,

1977; Glass, McGaw y Smith, 1981). Su nombre más común es el de *meta-análisis* (Glass, 1976) pero también suele denominarse *revisión cuantitativa* (Green y Hall, 1984), *integración de investigaciones* (Walberg y Haertel, 1980) o *análisis cuantitativo de dominios de investigación* (Cooper y Rosenthal, 1980; Rosenthal, 1980). Básicamente, consiste en aplicar métodos cuantitativos para resumir los resultados de muchos estudios y comprobar su relación con las características de los mismos (Glass, McGaw y Smith, 1981; Hattie y Hansford, 1984; Hunter, Schmidt y Jackson, 1982; Smith, 1982).

Aunque los primeros intentos por desarrollar métodos cuantitativos para integrar resultados de investigación se remontan a los años 30 con Tippet (1931), Fisher (1932) y Pearson (1933) como pioneros, esta metodología no había despertado el interés de los revisores en ciencias sociales y del comportamiento. Sin duda alguna, han sido los novedosos e influyentes trabajos de G. V. Glass y colaboradores (Glass y Smith, 1979; Smith y Glass, 1977, Smith, Glass y Miller, 1980) los que han impulsado la práctica, y con ella el desarrollo de técnicas cuantitativas de integración cada vez más sofisticadas. Fruto de este interés han sido las publicaciones de varios textos de meta-análisis en los últimos cinco años (Cooper, 1984; Glass, McGaw y Smith, 1981; Hedges y Olkin, 1985; Hunter, Schmidt y Jackson, 1982; Light y Pillemer, 1984; Mullen y Rosenthal, 1985; Rosenthal, 1980, 1984; Yeaton y Wortman, 1984).

El capítulo que aquí presentamos tiene como objetivo mostrar una panorámica general de los procedimientos al uso en la revisión cuantitativa de las ciencias sociales y del comportamiento. Describiremos, en primer lugar, las debilidades que presentan los métodos tradicionales de revisión. A continuación ofreceremos una definición de meta-análisis integradora y un detenido tratamiento de las etapas que se requieren para su correcta aplicación. Para terminar, expondremos una valoración personal de las ventajas e inconvenientes de este procedimiento metodológico innovador.

2. LA REVISIÓN CUALITATIVA

El término «revisión de la literatura» abarca una amplia gama de significados. Básicamente, pueden distinguirse tres tipos de revisión (Cooper, 1984; Jackson, 1980). En primer lugar, las *revisiones de investigación* integran la investigación dedicada a un tópico y extraen conclusiones globales a partir de un conjunto de trabajos que tratan hipótesis similares. En segundo lugar, las *revisiones teóricas* recogen las teorías formuladas para la explicación de un fenómeno y contrastan su grado de generalidad, su consistencia interna y la naturaleza de sus predicciones. Por último, las *revisiones metodológicas* tienen como objetivo examinar la adecuación de los métodos de investigación y las definiciones operativas aplicados a un tema de investigación. Obviamente, no es difícil encontrar revisiones de literatura que combinen dos de los tipos de revisión expuestos, o incluso los tres. En cualquier caso, el tipo de revisión más común es el que se ocupa de integrar la investigación sobre un área, combinado o no con cualquiera de los otros tipos. En consecuencia, nuestra exposición se centra en las revisiones de investigación.

Por *revisión cualitativa* se entiende la que incluye cualquier síntesis de investigaciones en la que el método de integración no incorpora el uso de procedimientos estadísticos formales (Hedges, 1984b; Smith, 1982). Las revisiones cualitativas también se denominan *narrativas*, *tradicionales*, *literarias* o *verbales*. Generalmente, un revisor cualitativo toma cada estudio por su valor aparente e intenta encontrar una teoría que reconcilie los resultados contradictorios (Hunter *et al.*, 1982). Este modo de proceder resulta provechoso cuando el número de estudios es escaso. Cuando el área de investigación a revisar abarca un número amplio de trabajos, la tarea de procesar toda la información existente se hace materialmente imposible. Ante esta situación, un revisor cualitativo puede poner en práctica una de tres estrategias. En primer lugar, puede intentar abar-

car toda la información; el resultado suele ser «... una revisión donde las sinopsis verbales de los estudios se ensartan en listas zigzagantes» (Glass, 1976, p. 4). Una segunda estrategia consiste en desechar una buena parte de los estudios en base a criterios de calidad de la investigación y aceptar uno o dos como esenciales para explicar las contradicciones empíricas. Esta estrategia tampoco es correcta porque elimina arbitrariamente información que puede ser crucial. Por último, se puede intentar la tarea de integrar mentalmente los resultados de los estudios. Esta alternativa es altamente propensa a adoptar conclusiones conservadoras y pesimistas acerca de la relación objeto de estudio. Es precisamente el caso de las innumerables revisiones de literatura que terminan con el tópico «... se necesita más investigación...» (Cooper y Rosenthal, 1980; Rosenthal, 1983c, 1984).

2.1. Cuantificación en las revisiones tradicionales

El principal inconveniente de la revisión cualitativa es su limitada capacidad para tratar muchos resultados a un mismo tiempo. Para simplificar la tarea de integración, una práctica muy común consiste en resumir los resultados de los estudios clasificándolos según su significación estadística. Es decir, dados k estudios que contrastan una hipótesis común, cada uno de ellos se incluye dentro de tres categorías: *a*) resultado estadísticamente significativo en favor de la hipótesis; *b*) resultado estadísticamente significativo en contra de la hipótesis, y *c*) resultado estadísticamente no significativo. En concreto, si la mayoría de los estudios muestra resultados estadísticamente significativos se toma como evidencia de que el efecto es distinto de cero. Light y Smith denominaron a este procedimiento *voting method* y lo describieron así:

«... Se examinan todos los estudios que aporten datos sobre una variable independien-

te específica de interés. Quedan definidos tres resultados posibles. La relación entre variables dependiente e independiente puede ser significativamente positiva, significativamente negativa o no significativa. Se tabula el número de estudios que se incluyen dentro de cada una de las tres categorías. Si dentro de una de ellas cae un número considerable de estudios y muy pocos en las otras dos, la categoría modal predominante proporciona la mejor estimación de la dirección de la verdadera relación» (1971, p. 433).

Este procedimiento, que aquí denominaremos «método de recuento», se ha utilizado tradicionalmente en las revisiones para comprobar la existencia y consistencia de efectos experimentales. Pero pese a lo intuitivo y sugerente que pueda parecer, es sencillo comprobar que es inadecuado y, en ocasiones, conduce incluso a conclusiones erróneas (Glass, 1977; Hedges y Olkin, 1980).

Un ejemplo contribuirá a comprender esta falacia. Supóngase que dos estudios ponen a prueba la existencia de una relación entre dos variables. La tabla 1 recoge los cuatro resultados posibles. Un revisor típico consideraría que las configuraciones 1 y 2 manifiestan una «consistencia de resultados». Del mismo modo, las configuraciones 3 y 4 las interpretaría como «fallos de replicación». Sin embargo, ninguna de las dos conclusio-

TABLA 1
Resultado

Configuración	Estudio 1	Estudio 2
1	Significativo	Significativo
2	No significativo	No significativo
3	Significativo	No significativo
4	No significativo	Significativo

nes es correcta. Con respecto a la configuración 1, es perfectamente posible que dos relaciones sean estadísticamente significativas, pero también significativamente distin-

tas entre sí. Lo mismo puede ocurrir en la configuración 2: dos relaciones pueden no ser estadísticamente significativas aun cuando sean significativamente diferentes entre sí. Por último, las configuraciones 3 y 4 no tienen por qué interpretarse como inconsistentes. En efecto, el hecho de que una relación no sea estadísticamente significativa mientras que otra sí lo sea no garantiza que esas dos relaciones sean significativamente diferentes entre sí (Humphreys, 1980).

No es posible, por tanto, determinar la consistencia de resultados de un conjunto de estudios a partir únicamente de la significación estadística. El método de recuento, al basarse sólo en este dato, conduce a conclusiones erróneas (Hedges, 1984). Téngase en cuenta que la significación estadística de un resultado se halla directamente relacionada con el tamaño muestral (Cooper, 1979, 1981) y en consecuencia las conclusiones se sesgan en favor de los estudios con tamaños muestrales grandes. Y aunque el problema podría resolverse a posteriori si se comprueba que el tamaño muestral no está relacionado con la magnitud de los resultados, el método de recuento en sí no permite determinar la magnitud de los efectos experimentales. Desafortunadamente, y pese a la proliferación de críticas recibidas (Carver, 1978; Cowger, 1984; Craig, Eison y Metze, 1976; Croll, 1984; Lykken, 1968; Posavac y Sinacore, 1984), los investigadores continúan concediendo a la significación estadística un valor mucho mayor del que realmente le corresponde, en detrimento de la escasa atención prestada a la magnitud del efecto (Glass *et al.*, 1981).

Pero la limitación más grave del método de recuento es que, bajo ciertas circunstancias, la probabilidad de adoptar una conclusión falsa *aumenta* conforme se incrementa el número de estudios, como han mostrado Hedges y Olkin (1980). Dados *k* estudios con tamaño muestral *n* que contrastan la misma hipótesis, si existe una magnitud del efecto diferente de cero, entonces la probabilidad de no detectar el efecto aumenta con el número de investigaciones, en particular cuando el tamaño del efecto no es muy grande ($\delta <$

$< 0,50$), con el consiguiente aumento del error tipo II y la disminución de la potencia. Todo ello hace inaceptable este procedimiento de revisión (Glass *et al.*, 1981; Hedges y Olkin, 1980, 1985; Hunter *et al.*, 1982; Rosenthal, 1984).

2.2. Rigor científico en las revisiones de la literatura

Desde la década de los setenta, numerosos autores han denunciado la carencia de una estructura sistemática y objetiva en la revisión de dominios de investigación. Ya en 1971 Feldman argüía que «... existe poco análisis formal o sistemático en la metodología y en la importancia de... la revisión y la integración... de la literatura» (p. 85). Asimismo, tras un análisis crítico de las prácticas de revisión tradicionales, Light y Smith (1971) encontraron severas deficiencias y concluían recomendando una mayor sistematización en las revisiones de investigación: «Poco progreso puede conseguirse agrupando *palabras en las conclusiones* de un conjunto de estudios...; por el contrario, (el progreso) sólo será posible cuando seamos capaces de agrupar, de manera sistemática, los datos originales de los estudios» (p. 443; el subrayado es nuestro). Pero quizá el trabajo más concluyente sobre los métodos de integración utilizados tradicionalmente por los revisores haya sido el realizado por Jackson (1980). En tal estudio, Jackson seleccionó una muestra aleatoria de 36 revisiones de investigaciones pertenecientes a revistas de psicología, sociología y educación e ideó una minuciosa clasificación de la metodología de la revisión obteniendo básicamente los resultados siguientes:

a) Aunque el 75 por 100 de los revisores citó las revisiones anteriores sobre el mismo tema de investigación, tan sólo el 6 por 100 examinó críticamente la evidencia, los métodos y las conclusiones pertinentes de aquéllas.

b) Muy a menudo, los revisores centraron su atención en el análisis y discusión de algunos trabajos, desechando parte de la información disponible. Además, en ningún caso se especificó el modo de selección del subconjunto de estudios.

c) Los revisores se valieron frecuentemente de descripciones crudas y erróneas de los resultados de los estudios. En efecto, el 15 por 100 de los revisores utilizaron métodos de recuento para clasificar la evidencia empírica, una práctica poco conveniente, como tratamos más arriba.

d) En muchas ocasiones, los revisores fueron incapaces de reconocer que las contradicciones y conflictos empíricos entre los resultados de los estudios podían deberse a errores de muestreo aleatorio.

e) En general, los revisores no solían valorar de modo sistemático las posibles relaciones entre los resultados de los trabajos y las características de los mismos, tales como el diseño de investigación, el tipo de sujetos, el contexto de investigación o la naturaleza de la intervención experimental.

f) Finalmente, trataron tan superficialmente la metodología de la revisión que difícilmente podía juzgarse la validez de sus conclusiones.

El análisis crítico de Jackson (1980) deja clara la casi total carencia de rigor en las prácticas tradicionales de revisión de investigaciones y apunta hacia la necesidad de dedicar más atención a esta importante etapa en la acumulación del conocimiento científico.

2. HACIA UNA DEFINICIÓN DE META-ANÁLISIS

Es preciso, en principio, distinguir tres niveles de análisis sobre los que pueden aplicarse métodos estadísticos en trabajos de investigación (Glass, 1976). La *investigación primaria* es el análisis original de los datos en un estudio de investigación; es la aplicación más común de métodos estadísticos. La

investigación secundaria puede perseguir uno de los 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 (Judd y Kenny, 1981). Aunque la investigación secundaria posee un gran atractivo, tiene una importante limitación: requiere disponer de los datos originales de los trabajos de investigación para ser adecuadamente aplicada. Por último, el *meta-análisis*, que Glass (1976) define como «el análisis de los análisis», o más exactamente, «... el análisis estadístico de una gran colección de resultados de análisis de trabajos individuales con el propósito de integrar los hallazgos obtenidos» (p. 3).

El meta-análisis ha surgido como alternativa a las revisiones tradicionales ante las dificultades planteadas para extraer el conocimiento de un conjunto de estudios sobre un tema común (Cooper, 1982; Green y Hall, 1984; Hedges, 1984b; Strube y Hartman, 1983). Más que un mero conjunto de técnicas estadísticas, representa una nueva perspectiva en la acumulación del conocimiento. Constituye una aproximación sistemática a la integración de dominios de investigación. En palabras de Cooper (1982): «La integración de proyectos diferentes de investigación implica inferencias científicas tan centrales a la validez del conocimiento como las inferencias adoptadas en la interpretación de datos primarios» (p. 291). En efecto, a lo largo del desarrollo de una revisión de investigación, el científico tendrá que tomar numerosas decisiones que afectarán directamente a los resultados y conclusiones de la misma. Cada una de las elecciones adoptadas por el revisor puede crear amenazas contra la validez y la fiabilidad de la revisión. Por tanto, para que el conocimiento transmitido en ciencias sociales y del comportamiento a través de las revisiones de investigación sea objetivo y digno de crédito, los revisores de investigación deben utilizar las mismas normas de rigor científico exigidas a los

investigadores primarios (Anguera y Sánchez, 1982; Cooper, 1984; Curlette y Cannella, 1985; Hattie y Hansford, 1984; Jackson, 1980; Rosenthal, 1984). Sólo mediante la sistematización de todo el proceso de síntesis de la investigación es posible conseguir que distintos revisores sobre un tema común obtengan resultados comparables, aunque sus interpretaciones puedan diferir (Light y Pillemer, 1982; Pillemer, 1984; Pillemer y Light, 1980a, b).

Partiendo de esta concepción, el meta-análisis puede caracterizarse con tres propiedades (Glass *et al.*, 1981):

a) Es una aproximación *cuantitativa* porque se vale de una variedad de métodos estadísticos para organizar y extraer información de trabajos de investigación primaria.

b) *No prejuzga* los hallazgos de investigación en términos de calidad del estudio; por el contrario, su posible influencia puede ser empíricamente corroborada.

c) Busca *conclusiones generales*; para ello, debe integrar resultados de trabajos con diferencias superficiales, pero en una fase posterior del análisis se pueden examinar las características diferenciales de tales trabajos.

Pese a su juventud, las técnicas meta-analíticas han ganado una gran difusión. Aunque es fácil encontrar multitud de aplicaciones en cualquier área de investigación, han logrado su mayor aceptación en psicología educativa y en psicología social y de la personalidad. Sirvan como ejemplo las revisiones cuantitativas sobre la efectividad de los métodos de enseñanza individualizada (Kulik, Kulik y Cohen, 1979) o de los métodos de enseñanza de la ciencia (Bredderman, 1980, 1985); la relación entre rendimiento académico y constructos tales como el estatus socioeconómico (White, 1980), las actitudes (Willson, 1983), el carácter directivo del maestro (Gage, 1978), el locus de control (Findley y Cooper, 1983) o la atribución causal (Whitley y Frieze, 1985); la efectividad de programas de entrenamiento en la eje-

cución de tests de rendimiento y de aptitudes (Bangert-Drowns, Kulik y Kulik, 1983a, b; Kulik, Bangert-Drowns y Kulik, 1984); diferencias sexuales en la decodificación de señales no verbales (Hall, 1978), en habilidades cognitivas (Hyde, 1981), en agresividad (Hyde, 1984), en ejecución motora (Thomas y French, 1985), en el desarrollo del pensamiento moral (Walker, 1984) y en conformidad (Cooper, 1979; Eagli y Carli, 1981); la efectividad de la psicoterapia (Miller y Berman, 1983; Nicholson y Berman, 1983; Shapiro y Shapiro, 1982a, b, 1983; Smith y Glass, 1977, Smith, Glass y Miller, 1980; Steinbrueck; Maxwell y Howard, 1983); los efectos de las expectativas del experimentador (Harris y Rosenthal, 1985; Raudenbush, 1984; Rosenthal, 1976); la relación del bienestar subjetivo con la situación profesional (Haring, Okun y Stock, 1984) con el sexo y la clase social (Haring, Stock y Okun, 1984) o con la orientación en los papeles sexuales (Whitley, 1985).

4. ETAPAS DE UN ESTUDIO META-ANALÍTICO

Hasta hace poco, los textos de metodología al uso (Anderson y Borkowsky, 1978; Arnau, 1978; Christensen, 1980; Cozby, 1981; Dominowsky, 1980; Kerlinger, 1973, 1979; McGuigan, 1978; Plutchik, 1983) habían prestado una atención escasa o nula al desarrollo de procedimientos formales para la síntesis de la investigación. De hecho, la estricta rigurosidad exigida en los informes de investigación primaria contrasta con el carácter personal, subjetivo e idiosincrático de las revisiones de investigación. Afortunadamente, con la llegada del meta-análisis se ha puesto de manifiesto un gran interés por la metodología de la revisión de la investigación, que también comienza a reflejarse en textos más recientes de metodología (vid. p. ej., Rosenthal y Rosnow, 1984). Del mismo modo que a los investigadores primarios se les obliga a detallar todos los aspectos me-

metodológicos en sus informes de investigación —conceptualización del problema, formulación de hipótesis, definición y medida de variables, muestreo y análisis de datos—, los procedimientos meta-analíticos también exigen describir de modo riguroso y formal todas las decisiones tomadas en el proceso de revisión.

Aunque no existe un acuerdo total respecto a las fases o etapas dentro de un estudio meta-analítico (Guskin, 1984; Hunter *et al.*, 1982; Rosenthal, 1984), el paralelismo entre éste y la investigación primaria es marcado. En este trabajo distinguiremos seis etapas básicas por las que debe discurrir cualquier síntesis de investigación que aplique métodos formales de revisión, a saber:

1. Formulación del problema.
2. Búsqueda de la literatura.
3. Codificación de los estudios.
4. Medida de los resultados.
5. Análisis e interpretación de datos.
6. Publicación del estudio.

Cada una de estas etapas obliga al revisor a adoptar decisiones que afectan directamente a los resultados de la revisión y, en consecuencia, son propensas a violaciones de la validez de las conclusiones. Describimos a continuación las seis etapas junto con algunas indicaciones para proteger la validez de los resultados.

4.1. Formulación del problema

Siempre que se realice una revisión de investigación —sea o no cuantitativa—, el primer paso consiste en una consideración detenida del problema de investigación. En un estudio meta-analítico, el problema objeto de interés puede abarcar desde la contrastación de las predicciones de una teoría respecto de una determinada asociación entre variables (como en investigación confirmatoria), hasta el descubrimiento de una posible relación entre variables que pudiera

ser importante (como en investigación exploratoria).

En su forma más simple, la formulación del problema de investigación implica la definición de variables y el razonamiento subyacente en su relación. Del mismo modo que en investigación primaria, las variables tienen que ser definidas de dos formas (Cooper, 1982, 1984). En primer lugar, las variables deben asumir definiciones conceptuales, que describen las cualidades de la variable independientemente del espacio y del tiempo y que permiten determinar los hechos que son relevantes al concepto. De este modo, las definiciones conceptuales tienen como objetivo especificar los estudios primarios que son admisibles para la revisión (Curllette y Cannella, 1985). En segundo lugar, las variables deben ser definidas operativamente. Una definición operativa relaciona un concepto abstracto con hechos observables y consiste en un conjunto de instrucciones que describen los fenómenos observables permitiendo determinar si un concepto está o no presente en una situación particular.

A diferencia de las investigaciones primarias, los estudios meta-analíticos suelen incluir múltiples definiciones operativas de un mismo concepto, abrigados por la idea de que en ciencias sociales el operacionalismo múltiple es beneficioso para la confirmación de hechos científicos (Cook y Campbell, 1979; Kenny, 1979; Plutchik, 1983). No obstante, esto es cierto sólo cuando la mayoría de las operaciones definidas son válidas, es decir, tienen alguna correspondencia con el concepto subyacente (Cooper, 1984).

Es preciso, pues, distinguir dos sentidos en los que la diversidad operacional puede afectar a los resultados de un meta-análisis (Cooper, 1982):

a) Dos revisiones sobre un mismo concepto pueden seleccionar definiciones operativas diferentes. Una revisión puede contener algunas operaciones comunes y algunas no comunes a otra revisión. O también, las operaciones incluidas en una revisión pueden estar completamente contenidas en la otra.

b) Dos revisiones sobre un mismo concepto con las mismas definiciones operativas pueden diferir en el detalle operacional. Un revisor puede decidir identificar las más ínfimas diferencias entre las operaciones mientras que otro puede obviar esos aspectos. Como consecuencia, el primer revisor tendrá mayor probabilidad de detectar, en un análisis ulterior, los efectos de variables moderadoras.

4.1.1. Amenazas contra la validez

En la formulación del problema de investigación, el revisor cuantitativo debe tomar dos decisiones que pueden afectar a la validez de los resultados: la amplitud de la definición conceptual y la selección de las definiciones operativas.

Respecto de la primera, el revisor ha de decidirse por el grado de abstracción de la definición conceptual. Un concepto definido en términos muy restringidos puede afectar a su generalidad y, por ende, al valor de las inferencias adoptadas. Por el contrario, un concepto definido ampliamente puede no tener una representación completa de la población de estudios, dando lugar a una generalización exagerada de los resultados.

Respecto de la segunda, en una revisión tiene que darse una correspondencia entre el concepto y las definiciones operativas. Un revisor que utiliza pocas definiciones operativas pretende asegurar la máxima adecuación de éstas con los conceptos, pero esta estrategia limita el grado de robustez de un resultado. La postura opuesta, es decir, la definición de múltiples operaciones de un mismo concepto, como argumentan Cook y Campbell (1979), tiene la ventaja de que permite desechar hipótesis rivales siempre y cuando las operaciones múltiples produzcan resultados similares.

4.1.2. Protección de la validez

Según Cooper (1984), los resultados de un meta-análisis pueden protegerse de las ame-

nazas contra la validez arriba esbozadas adoptando las siguientes precauciones:

a) Toda revisión cuantitativa debería comenzar con la definición conceptual del problema lo más amplia y abierta posible. Si, una vez efectuada la búsqueda de la literatura, la población de estudios se corresponde con una definición conceptual más limitada, entonces debe restringirse el concepto originalmente propuesto a las características de los estudios. Este modo de proceder resulta ser el más conveniente, ya que, como acertadamente argumenta Cooper: «Una vez que los estudios han sido recuperados y catalogados resulta descorazonador percatarse de que se han olvidado piezas disponibles del rompecabezas y que la búsqueda debe reconstruirse» (1982, p. 26).

b) Correspondiendo con un concepto, debe formularse inicialmente algunas definiciones operativas centrales, pero permaneciendo completamente abiertos a la posibilidad de incluir otras operaciones relevantes descubiertas en la literatura. En aquellos casos en los que se detecten definiciones operativas cuya correspondencia con el concepto sea cuestionable, es preferible incluirlas también en la revisión y comprobar empíricamente su adecuación en una etapa posterior del meta-análisis.

4.2. Búsqueda de la literatura

Una vez definido el problema de investigación, el siguiente paso en la realización de un trabajo meta-analítico consiste en identificar todos los estudios que han tratado dicho tópico. En última instancia, la población de estudios está directamente delimitada por las definiciones conceptuales y operativas asumidas en la etapa anterior. Por tanto, la principal decisión que debe tomar un revisor cuantitativo durante la búsqueda de la literatura implica elegir la población de elementos sobre la que se aplicarán las conclusiones. Desde esta perspectiva, la *población objeti-*

vo de un meta-análisis está compuesta por todos los individuos, grupos u otras unidades que el revisor espera representar en el estudio. Pero en la práctica, resulta materialmente imposible encontrar todos los elementos de una población objetivo, bien sea porque su adquisición resultaría demasiado costosa, o bien por la dificultad de localizar algunos elementos. En consecuencia, la *población accesible* incluirá solamente aquellos individuos, grupos u otras unidades que el investigador es capaz de localizar. Es un requisito fundamental definir en el informe meta-analítico las poblaciones objetivo y accesible así como determinar en qué medida pueden diferir.

Hasta hace poco, los revisores de investigación solían prestar muy poca atención a la localización de los estudios, pese a la crucial importancia de esta etapa respecto a la generalización de los resultados de una revisión. Como acertadamente apuntan Glass *et al.* (1981): «El modo en que se busca determina lo que se encuentra; y lo que se encuentra es la base de las conclusiones de la integración de estudios» (p. 61). Afortunadamente, gracias a la enorme explosión de investigaciones que estamos viviendo en ciencias sociales y del comportamiento, existen en la actualidad sistemas de recuperación de información cada vez más sofisticados y completos, lo que ha suscitado un mayor interés por los procedimientos de localización de estudios.

4.2.1. *Técnicas de localización de estudios*

Dada su capital importancia, un investigador interesado en realizar un meta-análisis ha de localizar el mayor número posible de estudios primarios, por muy tediosa que le pueda resultar la tarea. A grandes rasgos, pueden distinguirse tres fuentes o canales de información que *deben* ser utilizados para asegurar la adecuación de la población accesible a la población objetivo (Cooper, 1984; Glass *et al.*, 1981; Green y Hall, 1984; Hun-

ter *et al.*, 1982; vid. también Ortega y Fernández Dols, 1981).

1. *Fuentes informales*, que incluyen investigaciones primarias realizadas por el propio revisor, los manuscritos localizados a través de «colegios invisibles» y los trabajos presentados en congresos profesionales.

2. *Fuentes primarias*, que básicamente recogen las consultas a bibliografías personales y a las revistas de investigación primaria. Asimismo, cabe incluir aquí el procedimiento de búsqueda ascendente (*ancestry approach*). Por regla general, antes de comenzar un trabajo meta-analítico, el investigador ya conoce algunos estudios sobre el tema. El procedimiento de búsqueda ascendente consiste en examinar las referencias bibliográficas de esos trabajos para localizar otros pertinentes al tema seleccionado y repetir esa misma operación hasta que desaparezca la línea principal del tema o los estudios queden obsoletos respecto al momento de su realización.

3. *Fuentes secundarias*, que constituyen la espina dorsal de toda búsqueda de literatura, dada su naturaleza sistemática y exhaustiva. Pueden incluirse aquí las consultas a bibliografías editadas por autores particulares, empresas privadas o instituciones, el examen de los servicios de *abstracts* y archivos de citación (*index*) y las búsquedas bibliográficas por computador. Estas últimas tan sólo requieren determinar un conjunto de palabras-clave o descriptores del tema de investigación, y poseen la ventaja de un enorme ahorro de tiempo respecto de las búsquedas normales.

El problema de qué fuentes de información utilizar en una búsqueda de literatura no es simple. Para empezar, las distintas fuentes no son mutuamente excluyentes entre sí. Existe un cierto grado de solapamiento en la información que proporcionan. Además, las características de la información en las tres fuentes son distintas. Las fuentes informales suelen incluir trabajos de investigación no publicados, mientras que las fuentes primarias

y secundarias se centran más en estudios publicados. Los trabajos recuperados a través de las fuentes primarias e informales tienden a manifestar una excesiva homogeneidad entre sus resultados y las teorías e hipótesis dominantes dentro del círculo científico del que provienen. Por el contrario, las fuentes secundarias presentan una mayor imparcialidad, ya que su registro en estas fuentes requiere el menor número de requisitos, pero tiene la limitación de que la información que publican no es reciente (suelen pasar uno o dos años desde la publicación de un estudio hasta su aparición en una determinada fuente secundaria). Por último, dado el carácter no publicado de los estudios adquiridos mediante las fuentes informales, éstos suelen ser más propensos a sufrir algunas deficiencias metodológicas, en comparación con el rigor exigido a los estudios publicados.

Obviamente, no existe una respuesta general al tipo y número de canales de información que deben utilizarse en un meta-análisis. Ello dependerá, en parte, del tema de investigación y, en parte, de los recursos del investigador. No obstante, la norma más adecuada sería utilizar el mayor número de fuentes distintas (Cooper, 1982, 1984).

4.2.2. Amenazas contra la validez

Como apuntamos más arriba, en la práctica resulta imposible localizar todos los estudios sobre un problema de investigación. Como consecuencia, la población de estudios accesible es siempre menor que la población objetivo. Esta característica común a las búsquedas de literatura en cualquier meta-análisis crea amenazas contra la validez de los resultados. En esta etapa, un revisor cuantitativo debe preguntarse por la representatividad de los estudios localizados y de los individuos o elementos a que se tiene acceso respecto de la población objetivo. Si todos los estudios pertenecientes a una determinada población objetivo tuvieran la misma probabilidad de ser localizados por un investigador potencial, la representatividad

estaría asegurada. Sin embargo, es un hecho conocido que: *a)* los trabajos no publicados son más difíciles de conseguir que los publicados (Glass *et al.*, 1981); *b)* los estudios que van contra la interpretación teórica más común a un círculo científico tienen menos posibilidades de ser publicados (Cooper, 1984), y *c)* los trabajos de investigación que informan de resultados nulos no suelen superar la selección editorial para su publicación, lo que dificulta extremadamente su localización (Coursol y Wagner, 1986; Lane y Dunlap, 1978; Smith, 1980).

Se pueden agrupar en tres clases las principales amenazas a la validez que se producen en esta etapa de la revisión:

a) La población de estudios a que se tiene acceso por el revisor puede no ser representativa de la población objetivo. Un investigador que utiliza pocas fuentes de información para la búsqueda de estudios puede estar sesgando la población accesible. Por tanto, no es correcto generalizar los resultados del meta-análisis a toda la población objetivo.

b) La inclusión en un meta-análisis sólo de estudios publicados puede sobrestimar la magnitud del efecto experimental. Hace tiempo que la evidencia experimental demostró que los estudios publicados en ciencias sociales y del comportamiento constituyen una muestra sesgada de toda la población de estudios (McNemar, 1960; Rosenthal, 1979). Más concretamente, se supone que las revistas contienen un 5 por 100 más de estudios con error tipo I, mientras que los archivos y ficheros de las editoriales recogen el 95 por 100 de los estudios que exhiben resultados no significativos (por ejemplo, $p > 0,05$). Como consecuencia, el tamaño del efecto medio estimado mediante estudios publicados constituye una sobrestimación del tamaño del efecto de la población objetivo de los trabajos de investigación. Este sesgo no sólo puede ser de magnitud, sino también de dirección.

c) La población de individuos, grupos o unidades localizada por el revisor puede no

ser representativa de la población objetivo. En efecto, si los estudios recuperados no son representativos de toda la población, cabe la posibilidad de que ciertas subpoblaciones de elementos no queden representadas fielmente en la población accesible, lo que restringiría las inferencias de los resultados del meta-análisis.

4.2.3. Protección de la validez

La preocupación por los problemas que conlleva la correcta búsqueda de la literatura ha ocasionado una fuerte polémica (Atkinson, Furlong y Wampold, 1982; Dawes, Landman y Williams, 1984; Eaton, 1984; Fagley, 1984, 1985; Fagley y McKinney, 1983; Kurosawa, 1984; Landman y Dawes, 1984; Mahoney, 1977; Orwin y Cordray, 1984). Todavía no hay una respuesta concluyente a estos problemas, pero pueden apuntarse las recomendaciones siguientes:

a) La medida más adecuada para proteger los resultados de un meta-análisis contra la falta de representatividad de los estudios localizados es llevar a cabo una búsqueda de la literatura lo más amplia y exhaustiva posible. Es aconsejable recabar información desde las tres fuentes arriba señaladas para que los sesgos que puedan sufrir cada una de ellas se compensen en alguna medida. Este es el único modo de asegurar que dos investigadores que desarrollen sendos meta-análisis sobre un mismo problema obtengan resultados similares, aun cuando no hayan utilizado exactamente las mismas fuentes.

b) El problema del sesgo de publicación (*file-drawer problem*) ha sido abordado matemáticamente por Rosenthal (1979) y por Orwin (1983). Y aunque la solución que proponen no es definitiva, al menos permite estimar los prejuicios que el sesgo de publicación puede ejercer sobre las conclusiones. Este procedimiento permite calcular matemáticamente la tolerancia de un estudio meta-analítico a resultados nulos. Más concretamente, Rosenthal (1979) ideó una fórmula por

la que es posible determinar el número de resultados nulos que deben existir archivados en los ficheros de editoriales para que la probabilidad del grupo completo de resultados (localizados y no recuperables) exceda de un cierto nivel de significación (p. ej., $p=0,05$). De este modo, si el número de resultados nulos necesarios para contrarrestar la significación media de los resultados localizados es muy elevado, entonces podremos desechar el sesgo de publicación como una amenaza seria a la validez de los resultados. Su computación es muy simple (cf. Rosenthal, 1979, p. 639): dado un grupo de k contrastes independientes de una misma hipótesis sobre los que se ha calculado la puntuación típica z media representativa de la significación del colectivo, z , y suponiendo un nivel de significación del 5 por 100, la ecuación

$$N = \frac{k (k \bar{z}^2 - 2,706)}{2,706} \quad [4.1]$$

proporciona una estimación del número N de resultados nulos no recuperables que deben existir para anular la significación media de los k estudios, siendo 2,706 el cuadrado de la puntuación típica correspondiente a un nivel de significación del 5 por 100 ($z=1,645$). Por ejemplo, Rosenthal y Rubin (1978) integraron los niveles de significación de 345 estudios sobre los efectos de las expectativas obteniendo una puntuación media de 1,22. Así, el número de resultados nulos no recuperados necesarios para que la significación global pudiera deberse al sesgo de publicación tendría que ascender, aplicando la ecuación [4.1], a

$$N = \frac{345 (345 \times 1,22^2 - 2,706)}{2,706} = 65\ 123$$

es decir, debería haber más de 65 000 resultados no recuperados con significación nula ($z=0$) para poder explicar los resultados del meta-análisis por el sesgo de publicación. Lógicamente, un número tan elevado de estudios nulos es muy poco plausible.

El índice de tolerancia a resultados nulos (*fail-safe N*), desarrollado por Rosenthal (1979), tiene la limitación de que sólo es aplicable cuando en un meta-análisis se integran niveles de probabilidad. Sin embargo, es más común encontrar revisiones cuantitativas que acumulan índices del tamaño del efecto experimental, en particular la «diferencia media tipificada» o *d* de Cohen (1977). Con objeto de extender la lógica de Rosenthal (1979), Orwin (1983) ha introducido un estadístico análogo aplicable al tamaño del efecto (TE). Para su computación, es preciso conocer el número de estudios *k*, el TE medio, *d*, y un TE criterio, *d_c*, que puede variar según el tema de investigación. Siguiendo a Cohen (1977), pueden interpretarse los TE criterio de 0,20 como «pequeños», de 0,50 como «medios» y de 0,80 como «grandes». Además, una ventaja del procedimiento desarrollado por Orwin respecto del de Rosenthal es que permite plantear TE medios, *d_{fs}*, distintos de cero para los resultados no recuperables¹. El índice de tolerancia a resultados nulos puede calcularse mediante

$$N = \frac{k(\bar{d} - d_c)}{d_c - d_{fs}} \quad [4.2]$$

Si hacemos *d_{fs}* igual a cero, la ecuación se convierte en el análogo exacto de la formulación de Rosenthal (1979). Por ejemplo, apliquemos esta lógica a los 1766 TEs obtenidos por Smith, Glass y Miller (1980) sobre la efectividad de la psicoterapia. El TE medio obtenido fue de 0,85. Suponiendo que los estudios no recuperables tuvieran un TE medio nulo (*d_{fs}*=0), el número de estudios necesarios para poder explicar los resultados del meta-análisis por el sesgo de publicación ascendería a 1236 para un TE criterio medio (*d_c*=0,50) y a 5740 para un TE criterio pequeño (*d_c*=0,20).

c) Además del cálculo del índice de tolerancia a resultados nulos, el problema del sesgo de publicación puede ser abordado desde

otro frente. Si en un meta-análisis se ha logrado localizar estudios publicados y no publicados, es posible comparar los resultados de cada fuente de información entre sí para determinar la magnitud y el sentido del sesgo, si lo hubiera.

d) El problema de la correspondencia entre los elementos de la población localizados por el revisor y los de la población objetivo está más en función de las características de los individuos seleccionados por los investigadores primarios que por el propio revisor. Para controlar esta fuente del sesgo, el revisor puede resumir las características muestrales de los individuos empleados en los estudios primarios y limitar las inferencias de sus resultados sólo a los tipos de individuos implicados.

e) Por último, en el informe de un meta-análisis debe especificarse con la mayor exactitud todos los detalles del proceso de búsqueda de estudios, incluyendo las fuentes de información utilizadas, los años cubiertos en la búsqueda y las palabras-clave seleccionadas. Sólo de esta forma el lector podrá juzgar la fiabilidad de la búsqueda y la validez de los resultados.

4.3. Codificación de los estudios

El propósito de cualquier meta-análisis es extraer el estado de conocimiento que se posee sobre algún ámbito de investigación. Para ello, es preciso un estudio sistemático y exhaustivo de todas las investigaciones seleccionadas en la fase anterior. Cuando un determinado tema de investigación está plagado de resultados contradictorios, situación muy común en Ciencias sociales y del comportamiento, el revisor debe codificar las características de los estudios que supuestamente afectan a los resultados, pudiendo explicar tales contradicciones. Este proceso de codificación suele formalizarse elaborando

¹ Strube (1985b) ha extendido recientemente la formulación de Rosenthal (1979) para cualquier puntuación típica media de los estudios no recuperados.

un cuestionario que incluya todas las características relevantes.

Glass *et al.* (1981) distinguen entre características metodológicas y características sustantivas. Las características *metodológicas* se refieren a aspectos generales de la investigación, tales como la fecha y la fuente de publicación, el tipo de diseño de investigación, el tamaño muestral, las características de los sujetos (sexo, edad, CI, estatus socioeconómico, etc.), la unidad de análisis, las características del autor (sexo, equipo de investigación al que pertenece, etc.) o la calidad de la investigación. El análisis de las características metodológicas puede decir algo acerca de la coherencia del proceso científico en una determinada área de investigación. Las características *sustantivas* son específicas del tópico de investigación estudiado. Por ejemplo, en un meta-análisis sobre memoria icónica e informe parcial (Sánchez, 1985) se incluyeron características tales como la tarea del sujeto, el procedimiento experimental, el tiempo de exposición del estímulo, el tipo de material estimular, etc.

El proceso de codificación de las características de los estudios constituye un problema de medida y, como tal, está sujeto a posibles deficiencias en su fiabilidad y validez. Algunas características serán muy fáciles de codificar (p. ej., el año y la fuente de publicación). Sin embargo, otras requerirán la formulación de reglas y definiciones detalladas acerca del aspecto que se pretende codificar (p. ej., la calidad de la investigación). Como muy bien han apuntado Bullock y Svyantek (1985), un esquema de codificación «...es un intento de convertir una realidad compleja en un conjunto de números mutuamente excluyentes y exhaustivos» (p. 113). Por tanto, los resultados de cualquier meta-análisis están en función de la coherencia de la codificación.

4.3.1. Amenazas contra la validez

Muchas de las consideraciones acerca de la medición o codificación de las caracterís-

ticas de los estudios tiene que ver con cuestiones de validez. Consideraciones tales como la claridad de las definiciones, la adecuación de la información recogida o el grado de inferencia que tiene que asumir el codificador en la extracción de las características de los estudios. Básicamente, pueden distinguirse dos amenazas a la validez:

a) Las deficiencias en los informes de los estudios afectarán a la calidad de la codificación. Según Orwin y Cordray (1985), estas deficiencias se articulan en dos niveles: «macro» y «micro». La falta de calidad de los informes a nivel «macro» hace referencia a las acostumbradas prácticas en los informes de investigación dentro de un ámbito de investigación determinado. Por ejemplo, ciertas variables relevantes para su codificación pueden quedar excluidas porque no se acostumbra a detallarlas en los informes. Por otra parte, la falta de calidad a nivel «micro» se refiere al acabado, exactitud y claridad con que cada investigador particular escribe sus informes de investigación. Así, el registro inadecuado de las características de un estudio conduce a la incertidumbre del revisor y, a su vez, conduce a error.

b) La falta de fiabilidad en la codificación de las características puede deberse no sólo al problema de la calidad de los informes, sino también a errores del mismo codificador. Esto suele ocurrir especialmente con características complejas que requieren hacer juicios o tomar decisiones por parte del codificador. Si las reglas de codificación no son suficientemente explícitas, distintos codificadores de un mismo estudio pueden hacer juicios diferentes (Glass *et al.*, 1981).

4.3.2. Protección de la validez

Algunos problemas de validez pueden resolverse sin muchas dificultades definiendo las reglas de codificación de cada característica con el máximo detenimiento o bien dividiendo conceptos amplios en otros más restringidos. Por el momento, no existe un

remedio infalible para proteger la fiabilidad de la codificación, pero en cambio parece haber cierto acuerdo en recomendar las precauciones siguientes (Bullock y Svyantek, 1985; Glass *et al.*, 1981; Orwin y Cordray, 1985; Rosenthal, 1984; Stock *et al.*, 1982):

a) Elaborar un *libro de codificación* que incluya lo más exhaustivamente posible todas las definiciones de las características y las normas y reglas para su codificación.

b) Utilizar en el proceso de codificación un equipo de codificadores que deberá recibir unas sesiones de entrenamiento para la correcta interpretación y utilización del libro de codificación. Dicho equipo evaluará a continuación una muestra aleatoria de todos los estudios incluidos en el meta-análisis.

c) Realizar un estudio sistemático de la fiabilidad inter-codificadores. Las características codificadas pueden ser de muy distinta naturaleza desde características cualitativas (como la fuente de publicaciones) hasta cuantitativas (como la edad media o el CI medio de los sujetos de la muestra). Cada característica puede ser evaluada al menos con un índice de fiabilidad. Para llevar a cabo el estudio de la fiabilidad, los métodos estadísticos ofrecen una amplia gama de posibilidades. Si, por ejemplo, se han utilizado dos codificadores y la característica es cuantitativa, es aconsejable calcular el coeficiente de correlación de Pearson, r . Sin embargo, este índice sólo proporciona la fiabilidad «media» de los dos codificadores, no la fiabilidad «efectiva». Téngase en cuenta que cuantos más codificadores utilicemos más fiabilidad ganamos, ya que los errores aleatorios de cada codificador se contrarrestan con los errores de los otros codificadores. Por tanto, cuanto más grande sea el número de codificadores utilizados mayor será la fiabilidad «efectiva», que puede estimarse por la conocida fórmula de Spearman-Brown (Rosenthal, 1984):

$$R = \frac{n r}{1 + (n - 1) r} \quad [4.3] \quad 20$$

siendo R la fiabilidad «efectiva»; n , el número de codificadores y r la fiabilidad media entre los n codificadores (es decir, la media de $n(n - 1)/2$ coeficientes de correlación). Esta fórmula es muy sencilla de utilizar cuando se dispone de un equipo reducido de codificadores. Con un número elevado, empero, el cálculo de la correlación «media» se convierte en una tediosa tarea. Un procedimiento alternativo consiste en aplicar un análisis de varianza de medidas repetidas, donde los estudios corresponden a las observaciones y los codificadores constituyen el factor. Por este procedimiento es posible estimar el coeficiente de correlación «medio» mediante

$$r = \frac{MC \text{ inter-est.} - MC \text{ error}}{MC \text{ inter-est.} + (n-1) MC \text{ error}} \quad [4.4]$$

siendo n el número de codificadores. Y la fiabilidad «efectiva» vendría dada por

$$r = \frac{MC \text{ inter-estudios} - MC \text{ error}}{MC \text{ inter-estudios}} \quad [4.5]$$

Las ecuaciones [4.3] y [4.5] proporcionan resultados similares, salvando errores de redondeo².

Téngase en cuenta que la ecuación [4.3] no sólo puede aplicarse al coeficiente de correlación de Pearson, sino también al de Spearman, a la correlación biserial-puntual o al coeficiente phi. Pero cuando la característica codificada es cualitativa, en lugar de calcular un coeficiente de correlación es posible aplicar índices tales como la proporción de acuerdo o el coeficiente kappa de Cohen (1960). Aunque algunos autores recomiendan

² Los resultados de un estudio de fiabilidad deben documentarse en el informe del meta-análisis, para que el lector pueda evaluar la adecuación del sistema de codificación empleado.

que no se utilice porque pueden producir resultados erróneos (vid., p. ej., Rosenthal, 1984), el coeficiente kappa posee ciertas ventajas, en particular, que puede generalizarse su uso a más de dos codificadores (Conger, 1980; Uebersax, 1982). En cualquier caso, es recomendable la combinación de varios índices de fiabilidad (Orwin y Cordray, 1985).

d) Por último, para proteger la validez de la codificación respecto de la falta de calidad en los informes, es recomendable también adoptar otras dos medidas de precaución: en primer lugar, definir la calidad de los datos obtenidos como una característica más para analizar su posible relación con los resultados y, en segundo lugar, buscar otras fuentes de información externas a los informes para completarlos (p. ej., contactando directamente con los autores).

4.4. Medida de los resultados

Puesto que el propósito final reside en integrar resultados de un conjunto de estudios para interpretar el estado de conocimiento sobre un problema concreto de investigación, es preciso también que el revisor defina dos aspectos de los datos, a saber, determinar cuál es la *unidad de análisis* y encontrar una *medida común* de los resultados de los estudios.

4.4.1. La unidad de análisis

Si en una investigación primaria la unidad de análisis de datos la constituyen las observaciones empíricas, en una revisión cuantitativa la unidad de análisis procede de los estadísticos ya analizados en los estudios primarios. No obstante, existe cierta polémica en torno a esta cuestión (vid. p. ej., Bangert-Drowns, 1986). Para algunos autores, los datos de un meta-análisis deben ser los mismos estudios (Bangert-Drowns, Kulik y Kulik, 1983a; Landman y Dawes, 1982; Wortman y Bryant, 1985); es decir, cada estudio debe

estar representado por un solo dato en el meta-análisis para evitar problemas de dependencia entre los datos. Una postura más flexible consiste en considerar que un mismo estudio puede aportar varios resultados interesantes y en consecuencia deberían ser introducidos por separado en el meta-análisis (Bredderman, 1983; Hansford y Hattie, 1982; Smith, Glass y Miller, 1980). Por ejemplo, en un meta-análisis acerca de la efectividad de la psicoterapia sería muy común encontrar estudios que comparasen la efectividad de varias técnicas psicoterapéuticas con un grupo de control. La información relevante de este estudio implicaría incluir una medida de la efectividad de cada técnica psicoterapéutica en el meta-análisis, ya que el resultado global carecería de sentido en caso contrario. Esta postura tiene la ventaja de que no pierde información intra-estudio, permitiendo la comprobación de posibles efectos moderadores de algunas características, pero de ningún modo asegura el supuesto de independencia entre los resultados integrados.

La postura de considerar el estudio como unidad de análisis es también propensa a cierta ambigüedad. Muy frecuentemente, un solo informe de investigación suele contener múltiples contrastes de una misma hipótesis. Ello puede deberse a que: 1) se utilizaron diferentes muestras de personas y sus datos se analizaron por separado; 2) se emplearon medidas múltiples de un mismo constructo y cada medida fue analizada individualmente, o 3) un mismo trabajo de investigación informa de más de un proyecto de investigación.

Cooper (1984) propone una solución de compromiso a este problema, consistente en utilizar una unidad de análisis variable. Concretamente, Cooper propone que se codifique cada prueba estadística por separado como si fueran independientes entre sí. Cada resultado debe estar vinculado con sus características peculiares codificadas. Cuando se pretende obtener un resultado global en base al meta-análisis, los resultados de las pruebas estadísticas son ponderados de tal modo que cada estudio contribuye por igual al resultado general. Sin embargo, cuando se

examinan los efectos de características potencialmente moderadoras, los resultados de un mismo estudio se agrupan únicamente dentro de las categorías individuales de las características objeto de estudio. Esta estrategia permite retener de cada estudio la máxima información y asumiendo una mínima violación del supuesto de independencia de los resultados.

4.4.2. El problema de la medida

Para poder integrar los resultados de un conjunto de estudios, es preciso que se expresen en una escala de medida común. Por regla general, las variables dependientes empleadas en los estudios difieren entre sí, razón por la cual no son directamente comparables. Ésta ha sido precisamente una de las dificultades que impidió un desarrollo más acelerado de las técnicas cuantitativas a la revisión de investigación (Bangert-Drowns, 1986; Hedges y Olkin, 1985). Este problema se ha intentado solucionar mediante la aplicación de dos estrategias diferentes, a saber: 1) el uso de los niveles de significación y 2) la obtención de medidas de la magnitud del efecto (Glass *et al.*, 1981; Hunter *et al.*, 1982; Rosenthal, 1984).

4.4.2.1. El nivel de significación

Esta aproximación al problema de la medida de los resultados de múltiples estudios consiste en obtener el nivel de significación exacto, p , que corresponde a cada contraste de significación realizado acerca de una hipótesis común. El conjunto de niveles p unilaterales así obtenido es susceptible de ser estadísticamente analizado, ya que convierte los resultados de los estudios a una misma medida. Tal es el caso de los trabajos ya clásicos de Rosenthal acerca de los efectos de las expectativas del experimentador (Rosenthal, 1966).

Pero por atractiva que pueda parecer esta estrategia, presenta varios inconvenientes. El

principal de ellos es que tal solución se basa en la significación estadística y no nos dice nada acerca de la magnitud del efecto experimental. Dada la estrecha dependencia de las pruebas estadísticas con el tamaño muestral, su aplicación en meta-análisis no es aconsejable si no va acompañada de alguna medida del tamaño del efecto (Hunter *et al.*, 1982; Rosenthal, 1978, 1984).

20

4.4.2.2. El tamaño del efecto

La medida más informativa del resultado de una investigación es el *tamaño o magnitud del efecto* (Green y Hall, 1984). Glass (1976) ha sido el principal responsable de la introducción de las medidas del tamaño del efecto de (TE) a la integración cuantitativa de resultados de investigación. La definición de TE propuesta por Cohen (1977) pone de manifiesto sus ventajas frente a la significación estadística: «Resulta conveniente utilizar el término “tamaño del efecto” para indicar el grado en que el fenómeno está presente en la población o el grado en que la hipótesis nula es falsa. ... Cuando la hipótesis nula es falsa, lo es en algún grado específico, es decir, *el tamaño del efecto es un valor concreto, distinto de cero, de la población*. Cuanto mayor es este valor, tanto mayor es el *grado* en que se manifiesta el fenómeno bajo estudio» (pp. 9-10; el subrayado es nuestro).

El nivel de significación estadística sólo permite determinar si un efecto experimental es distinto de cero. Por el contrario, el TE es capaz de responder a la pregunta: ¿en qué medida es distinto de cero? (Cooper, 1984) y, por ende, la información que proporciona es más cabal para los propósitos perseguidos.

Del mismo modo que los métodos estadísticos han desarrollado múltiples pruebas para cada diseño de investigación particular, se han desarrollado paralelamente una amplia gama de indicadores del TE, tanto paramétricos (Carroll y Nordholm, 1975; Charter, 1982; Cohen, 1973; Dodd y Schultz, 1973;

Dwyer, 1974; Friedman, 1968, 1982; Fleiss, 1969; Gaebelin *et al.*, 1976; Haase, 1983; Kennedy, 1970; Keren y Lewis, 1979; Lane y Dunlap, 1978; Maxwell *et al.*, 1981; Ronis, 1981; Serlin, 1982; Smith, 1982; Vaughan y Corballis, 1969; Wang, 1982) como no paramétricos (Preece, 1983; Serlin *et al.*, 1982; Strahan, 1982). Con todo, aquí centraremos nuestra atención en los dos tipos de índices del TE más utilizados en meta-análisis, a saber, los basados en la diferencia media entre dos condiciones experimentales y los basados en el coeficiente de correlación, sin olvidar tampoco los estimadores no paramétricos introducidos por Kraemer y Andrews (1982).

Un breve comentario acerca de las medidas de TE basadas en la proporción de varianza explicada es necesario antes de iniciar nuestra exposición. A pesar de su todavía no muy extendida utilización en investigación experimental, las medidas de la proporción de varianza explicada, tales como el coeficiente de determinación (r^2), omega cuadrada de Hays (ω^2) o eta cuadrada (η^2) no son aconsejables para la integración de resultados, entre otras cuestiones planteadas por O'Grady (1982), por estar en función de los niveles de la variable independiente que un investigador particular decide incluir en un estudio y de la relación subyacente entre las variables dependiente e independientes (Glass *et al.*, 1981; Green y Hall, 1984; Fowler, 1985; Rosenthal, 1984). Además, estos índices no especifican la dirección de las diferencias. Por ejemplo, dos estudios en los que la media del grupo experimental difiere de la media del grupo de control en una desviación típica, pero en sentido opuesto, obtendrían la misma proporción de varianza explicada (Hedges y Olkin, 1985).

Los inconvenientes citados, junto con el escaso carácter intuitivo de estas medidas (Rosenthal y Rubin, 1979a; Yeaton y Sechrest, 1981), desaconsejan su uso en la integración cuantitativa de resultados de investigación.

Íntimamente vinculada con el anterior se encuentra la conveniencia de que los revisores de investigación no utilicen estimadores

del TE para contrastes de significación con más de un grado de libertad. Es decir, todos los TEs deberían ir expresados como comparaciones entre dos grupos, como medidas de correlación entre dos variables o como cualquier otro contraste con un solo grado de libertad. Cooper (1984) explica la razón de ser de esta medida: «...En un análisis de varianza, la razón F con más de un grado de libertad nos dice si existe una variabilidad significativa entre las medias de grupo, pero no nos informa acerca de en qué grado difieren significativamente entre sí. Por tanto, en cualquier caso es esencial que las pruebas de significación con más de un grado de libertad vayan seguidas de comparaciones con un solo grado de libertad. Esto mismo debería hacerse con los tamaños del efecto asociados con estos contrastes» (p. 97).

4.4.2.2.1. La diferencia media tipificada

Uno de los índices del TE más divulgados en meta-análisis es la *diferencia media tipificada*, un índice desarrollado por Cohen (1969, 1977) que permite determinar el efecto experimental resultante de la comparación entre dos grupos, generalmente un grupo experimental y un grupo de control. La denominada d de Cohen consiste en calcular la diferencia entre las medias de los grupos experimental y control, \bar{y}^E e \bar{y}^C , y dividirla por la desviación típica conjunta de ambas poblaciones, σ .

$$d = \frac{\bar{y}^E - \bar{y}^C}{\sigma} \quad [4.6]$$

Glass (1976) retomó este índice para aplicarlo en su meta-análisis sobre la efectividad de la psicoterapia (Smith, Glass y Miller, 1980), definiéndolo como la diferencia entre las medias de los grupos experimental y control dividida por la desviación típica del grupo de control.

$$\Delta_{E-C} = \frac{\bar{y}^E - \bar{y}^C}{S_C} \quad [4.7]$$

siendo S^C la desviación típica del grupo de control,

$$S^C = \left[\frac{\sum \overset{y^c}{\underset{d_i}{(y^E - \bar{y}^C)^2}}}{n^C - 1} \right]^{1/2} \quad [4.8]$$

El argumento que propone Glass para justificar el uso de S^C en lugar de un promedio de las desviaciones típicas de los dos grupos es que los tratamientos experimentales pueden inflar o disminuir desproporcionadamente las desviaciones típicas correspondientes en comparación con la desviación típica del grupo de control no tratado. Por ejemplo, un conjunto de tratamientos comparados con un mismo grupo de control darían lugar a TEs diferentes, aun cuando sus diferencias medias fueran idénticas. Cuando ocurra esta situación, Glass *et al.* (1981) recomiendan el uso de la desviación típica del grupo de control para tipificar las diferencias medias ya que sólo de este modo las estimaciones de los TEs de los distintos tratamientos serían idénticas.

Sin embargo, hay otros muchos casos en los que es plausible el supuesto de homogeneidad de las varianzas (homoscedasticidad). Hedges (1981) ha derivado un estimador del TE de la población suponiendo que las distribuciones de los grupos experimental y control son normales y con varianza común $[N(\mu^E, \sigma^2), N(\mu^C, \sigma^2)]$. El TE paramétrico, δ , vendrá dado por

$$\delta = \frac{\mu^E - \mu^C}{\sigma} \quad [4.9]$$

El estimador de Hedges consiste en dividir la diferencia media por la desviación típica intragrupos, S ,

$$g = \frac{\bar{y}^E - \bar{y}^C}{S} \quad [4.10]$$

donde

$$S = \left[\frac{(n^E - 1)(S^E)^2 + (n^C - 1)(S^C)^2}{n^E + n^C - 2} \right]^{1/2} \quad [4.11] \quad 20$$

siendo n^E, n^C, S^E y S^C los tamaños muestrales y las desviaciones típicas respectivas de los grupos experimental y control.

Los estimadores Δ_{E-C} y g , por ser estadísticos muestrales, tienen su correspondiente distribución muestral. Concretamente, siendo

$$\tilde{n} = \frac{n^E n^C}{n^E + n^C} \quad [4.12]$$

entonces $\sqrt{n} \Delta_{E-C}$ y $\sqrt{\tilde{n}} g$ siguen sendas distribuciones t no centrales con parámetro de no centralidad $\sqrt{\tilde{n}} \delta$ y grados de libertad $(n^C - 1)$ y $(n^E + n^C - 2)$, respectivamente. Hedges (1981; vid. también Hedges y Olkin, 1985) ha demostrado que g es mejor estimador que Δ_{E-C} , ya que está menos sesgado y tiene menor varianza. El sesgo de la g de Hedges como estimación de δ puede corregirse mediante el factor de corrección $c(m)$,

$$g^u = c(m) g \quad [4.13]$$

donde g^u es un estimador insesgado de δ y $c(m)$ es un factor de corrección que puede obtenerse aproximadamente mediante

$$c(m) \approx 1 - \frac{3}{4m - 1} \quad [4.14]$$

siendo $m = n^E + n^C - 2$. Obviamente, para muestras grandes el sesgo de g es mínimo, por lo que se aconseja el uso del factor de corrección sólo para muestras pequeñas (p. ej., menores de 20).

La diferencia entre la d de Cohen, dada en la ecuación [4.6] y la g de Hedges de la ecuación [4.10] radica en que la primera utiliza N para calcular la desviación típica, σ (siendo $N = n^E + n^C$) mientras que g utiliza los grados de libertad de la distribución correspondiente (es decir, $n^E + n^C - 2$).

Cualquiera de los índices de la diferencia media tipificada descritos se caracteriza por su facilidad de cálculo y su interpretación intuitiva. De hecho, asumiendo distribuciones normales, es posible expresar cualquier diferencia media tipificada mediante una figura que represente dos distribuciones normales —una para el grupo experimental y otra para el grupo de control—, cuyo grado de solapamiento está en función del TE obtenido. Por ejemplo, en un estudio meta-analítico sobre la ejecución de niños buenos y malos lectores en tareas experimentales que reflejan la actividad de las etapas iniciales de procesamiento de la información visual (Sánchez, 1984), se observó que la diferencia media tipificada global resultó ser de $-0,43$. Este resultado puede expresarse tal y como aparece en la figura 1.

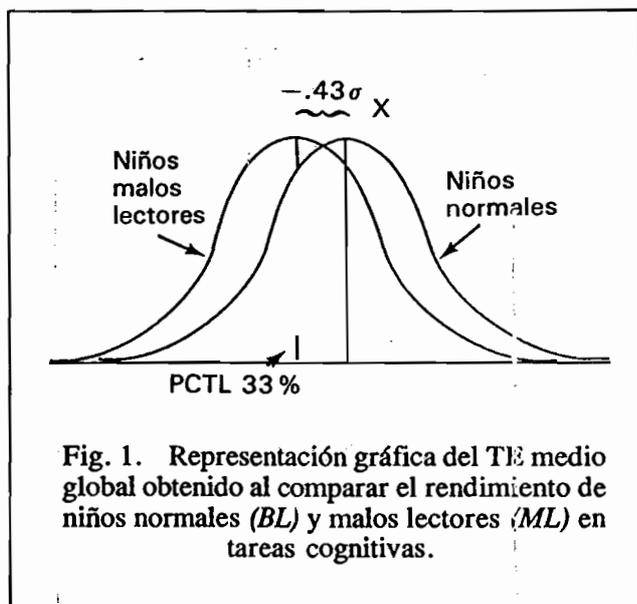


Fig. 1. Representación gráfica del TE medio global obtenido al comparar el rendimiento de niños normales (BL) y malos lectores (ML) en tareas cognitivas.

Asimismo, este resultado puede explicarse en términos del porcentaje de la distribución del grupo de control que supera la media del grupo experimental. Para ello, basta con consultar en las tablas de la distribución normal tipificada el percentil correspondiente al TE obtenido. En nuestro ejemplo, una diferencia media tipificada de $-0,43$ se corresponde con el percentil 33. Por tanto, puede decirse que la media de los niños malos lec-

tores (grupo experimental) sólo superó al 33 por 100 de la distribución de los niños buenos lectores (grupo control).

Los índices d de Cohen, Δ_{E-C} de Glass y g de Hedges son los más utilizados en meta-análisis. Pero hay otros índices del TE basados en las diferencias medias para aquellos casos en los que no se comparan medias, sino otros estadísticos. Así, por ejemplo, el índice d' permite comparar diferencias entre dos proporciones independientes P_1 y P_2 (Cohen, 1962).

$$d' = P_1 - P_2 \quad [4.15]$$

Posteriormente, Cohen (1977) propuso otro índice del TE para la comparación de dos proporciones independientes, basado en la transformación arcoseno para estabilizar las varianzas y normalizar las distribuciones muestrales de las proporciones:

$$h = \phi_1 - \phi_2 \quad [4.16]$$

siendo

$$\phi = 2 \arcsen \sqrt{P}, \quad [4.17]$$

Del mismo modo, Cohen (1962) propuso inicialmente como índice del TE para la comparación entre dos coeficientes de correlación de Pearson independientes la diferencia entre ambos, r_1 y r_2 , pero debido a que diferentes coeficientes de correlación tienen distribuciones muestrales distintas, Cohen (1977) aconsejó posteriormente el uso del estadístico q consistente en la transformación previa de r a la distribución Z de Fisher, para normalizar la distribución y estabilizar la varianza.

$$q = Z_{r_1} - Z_{r_2}, \quad [4.18]$$

siendo

$$Z_r = 1/2 \log_e \left[\frac{1+r}{1-r} \right] \quad [4.19]$$

Los índices q y h son sencillos de calcular mediante el uso de tablas específicas (vid. Rossi, 1985).

4.4.2.2.2. El coeficiente de correlación

Un índice alternativo a la diferencia media tipificada para estimar el TE de un estudio consiste en calcular un coeficiente de correlación. Este índice está especialmente indicado para estudios correlacionales, ya que el mismo informe de la investigación suele especificar el coeficiente de correlación de Pearson si las variables son continuas. Pero incluso con diseños de dos grupos es posible calcular el *coeficiente de correlación biserial-puntual* r_{bp} mediante

$$r_{bp} = \frac{\bar{y}^E - \bar{y}^C}{S_T} (pq)^{1/2} \quad [4.20]$$

siendo S_T la desviación típica total de los dos grupos; $p=n^E/N$ y $q=n^C/N$, y $N=n^E+n^C$.

Rosenthal (1984) y Hunter *et al.* (1982) recomiendan el uso en meta-análisis del coeficiente de correlación en lugar de la diferencia media tipificada por su mayor facilidad de cálculo, ya que la diferencia media tipificada no suele ser incluida en los informes de investigación. Pero la elección de un índice del TE debe hacerse según las características del área de investigación que se revisa. En cualquier caso, es siempre posible transformar la diferencia media tipificada, d , a coeficiente de correlación biserial puntual mediante

$$r_{bp} = \frac{d}{\sqrt{d^2+1/pq}} \quad [4.21]$$

Si $p=q$, entonces la ecuación [4.21] se hace

$$r_{bp} = \frac{d}{\sqrt{d^2+4}} \quad [4.22]$$

Estimadores

4.4.2.2.3. ~~Estimadores~~ no paramétricos

Los índices del TE descritos en las dos secciones anteriores son convenientes cuando los datos del estudio cumplen con los supuestos de la prueba t de Student. Sin embargo, este enfoque no es siempre realista. Kraemer y Andrews (1982) han sugerido que, en muchas ocasiones, los datos no son simétricos y requieren alguna transformación para aproximarse a la normalidad. Si el supuesto de normalidad no se cumple, los TEs calculados antes y después de la transformación pueden diferir considerablemente. Una solución consiste en estimar δ calculando el TE sobre los datos transformados, ya que éstos serán razonablemente normales.

Kraemer y Andrews (1982) proponen como solución alternativa un estimador del TE que no se vea afectado por transformaciones monotónicas de los datos. Este estimador no paramétrico del TE permite interpretar la diferencia media tipificada resultante en términos de los percentiles de la distribución normal tipificada, tal y como sucede con los estimadores paramétricos. La aproximación de Kraemer y Andrews (1982) ha sido extendida por Hedges y Olkin (1984); véase también Hedges y Olkin, 1985, pp. 93-100, quienes han sugerido siete estimadores no paramétricos del TE. De éstos, quizá el de aplicación más generalizada sea el estadístico \hat{v}_1^* , que implica calcular la proporción \hat{q}_1^* de puntuaciones del grupo de control menores que la mediana de las puntuaciones del grupo experimental y transformar dicha proporción a puntuación típica mediante la función de distribución normal tipificada:

$$\hat{v}_1^* = \Phi^{-1}(\hat{q}_1^*). \quad [4.23]$$

Los supuestos de la prueba t tampoco se cumplirán ante la presencia de puntuaciones desviadas (*outliers*). La presencia de puntuaciones desviadas distorsiona los valores de los estadísticos de contraste y las estimaciones del TE. Puede calcularse una estimación del TE después de haber eliminado tales pun-

tuciones desviadas, pero también puede aplicarse una estimación del TE robusta ante la presencia de puntuaciones desviadas. Hedges y Olkin (1984) han propuesto un índice de este tipo, que implica estimar las medias de las poblaciones experimental y control a través de las medianas de los grupos respectivos, $Mdn(Y^E)$ y $Mdn(Y^C)$ y dividir la diferencia por el rango:

$$\tilde{\delta} = \frac{Mdn(Y^E) - Mdn(Y^C)}{\tilde{\sigma}} \quad [4.24]$$

siendo

$$\tilde{\sigma} = a_2 y_2^C + \dots + a_{n-1} y_{n-1}^C \quad [4.25]$$

donde $y_1^C \leq y_2^C \leq \dots \leq y_{n-1}^C \leq y_n^C$ son los valores ordenados del grupo de control.

Los coeficientes óptimos a_2, \dots, a_{n-1} se eligen de modo que minimicen la varianza de $\tilde{\delta}$. El estadístico $\tilde{\delta}$ es un ejemplo de estimador robusto de δ , ya que elimina las dos observaciones desviadas.

Téngase en cuenta que los estimadores no paramétricos que acabamos de citar lo son en el sentido de que no dependen explícitamente del modelo estructural paramétrico de las observaciones, pero no son estadísticos de distribución libre (Krauth, 1983). Poco se sabe acerca de las propiedades muestrales de los estimadores no paramétricos del TE, pero probablemente son menos eficientes que sus alternativas paramétricas cuando se cumplen los supuestos de éstas. Su uso debe restringirse únicamente a aquellos casos en que los supuestos paramétricos se hayan violado. El gran inconveniente de los estimadores no paramétricos del TE citados es que requieren disponer de las observaciones empíricas del estudio para su obtención (Hedges y Olkin, 1985).

4.4.2.3. Obtención del tamaño del efecto

Aunque la diferencia media tipificada es un estadístico claro y simple, su cálculo suele

tener problemas debidos a deficiencias de ciertos datos en los informes de los estudios. Cuando el diseño de la investigación incluye sólo dos grupos (experimental *versus* control) el estadístico de contraste más común es t (o F , ya que en este contexto $F = t^2$). Es posible estimar d si al menos el estudio proporciona el valor de t y los tamaños muestrales n^E y n^C . Para muestras independientes, el TE viene dado directamente por

$$d = t \left[\frac{1}{n^E} + \frac{1}{n^C} \right]^{1/2} \quad [4.26]$$

Si $n^E = n^C$, entonces

$$d = t (2/n)^{1/2} \quad [4.27]$$

Para muestras relacionadas, suponiendo conocido r_{xy} ,

$$d = t \left[\frac{2}{n(1 - r_{xy}^2)} \right]^{1/2} \quad [4.28]$$

En algunos informes de investigación no se cita el resultado de la prueba t , sino sólo el nivel de significación alcanzado y los grados de libertad. Si el nivel de significación es exacto, puede obtenerse el valor del estadístico de contraste consultando las tablas de la distribución correspondiente y calcular d mediante la ecuación [4.26], [4.27] o [4.28], según los casos. Si el nivel de significación especificado es sólo aproximado, puede obtenerse una estimación conservadora de d igualando el valor de t con el nivel de significación y proceder en consecuencia.

Cuando el diseño experimental incluye más de dos grupos, suele aplicarse como estadístico de contraste la razón F . Un diseño de este tipo es susceptible de asimilar tantas estimaciones de d como comparaciones dos a dos admita el diseño; o bien, puede obtenerse un valor d para la comparación de cada grupo experimental con un grupo de control común. En cualquier caso, la información mínima necesaria para su cálculo incluye las medias de

cada grupo, los tamaños muestrales y el resultado de la razón F global para poder obtener la desviación típica intragrupos, S (McGaw y Glass, 1980). Si los grupos son independientes, S se obtiene por

$$S = (MC \text{ error})^{1/2} \quad [4.29]$$

La MC error puede también obtenerse mediante

$$MC \text{ error} = MC \text{ tratamientos} / F \quad [4.30]$$

siendo

$$MC \text{ tratamientos} = \frac{\sum_j n_j (\bar{y}_{.j} - \bar{y}_{..})^2}{(r-1)}, \quad [4.31]$$

donde n_j es el tamaño muestral del grupo j , r es el número de grupos, $\bar{y}_{.j}$ es la media del grupo j e $\bar{y}_{..}$ es la media total (Ato *et al.*, 1981).

La obtención de la desviación típica intra-grupo cuando disponemos de un diseño de un factor (A) con medidas repetidas requiere que se conozca la tabla de las fuentes de variación del ANOVA correspondiente (Glass *et al.*, 1981). Una estimación de S consiste en combinar la variación debida a los sujetos con la variación debida a la interacción del factor con los sujetos mediante

$$S = \left[\frac{SC \text{ intersujetos} + SC \text{ } A \times \text{sujetos}}{g.l. \text{ intersujetos} + g.l. \text{ } A \times \text{suj.}} \right]^{1/2} \quad [4.32]$$

siendo $SC \text{ } A \times \text{sujetos} / (r-1) (n-1)$ el término de error para contrastar la razón F global.

Mayores complicaciones surgen cuando los estudios emplean diseños factoriales en los que, además del factor principal objeto de la comparación, se incluyen otros factores secundarios que pueden afectar al principal. Por ejemplo, un estudio puede comparar diversos tratamientos con un grupo de control (factor A) cruzándolo con otro factor B se-

cundario. La desviación típica más aproximada para estimar los TEs de los tratamientos debe incluir una combinación de la variación debida al factor B , de la debida a la interacción $A \times B$ y del error:

$$S = \left[\frac{SC \text{ } B + SC \text{ } A \times B + SC \text{ error}}{g.l. \text{ } B + g.l. \text{ } A \times B + g.l. \text{ error}} \right]^{1/2} \quad [4.33]$$

Si el diseño es de medidas repetidas y el factor principal es A , entonces

$$S = \left[\frac{SC \text{ inters.} + SC \text{ } A \times \text{suj.} + SC \text{ } B + SC \text{ } B \times \text{suj.} + SC \text{ } A \times B + SC \text{ } A \times B \times \text{suj.}}{g.l. \text{ inters.} + g.l. \text{ } A \times \text{suj.} + g.l. \text{ } B + g.l. \text{ } B \times \text{suj.} + g.l. \text{ } A \times B + g.l. \text{ } A \times B \times \text{suj.}} \right]^{1/2} \quad [4.34]$$

Por último, si el diseño es mixto la obtención de la desviación típica dependerá de la naturaleza del factor principal objeto de la estimación de los TEs. Si el factor principal (A) es inter, la desviación típica coincide con su término de error,

$$S = \left[\frac{SC \text{ suj.} / A}{g.l. \text{ suj.} / A} \right]^{1/2} \quad [4.35]$$

Pero si el factor principal (B) es intra, entonces

$$S = \left[\frac{SC \text{ } A + SC \text{ suj.} / A + SC \text{ } A \times B + SC \text{ } B \times \text{suj.} / A}{g.l. \text{ } A + g.l. \text{ suj.} / A + g.l. \text{ } A \times B + g.l. \text{ } B \times \text{suj.} / A} \right]^{1/2} \quad [4.36]$$

Cuando los estudios aplican pruebas estadísticas no paramétricas, la obtención de la diferencia media tipificada presenta problemas adicionales. Al no contrastar hipótesis simples acerca de las medias de población, las pruebas no paramétricas no son suscep-

tibles de una transformación sencilla a sus contrapartidas paramétricas. No obstante, Glass *et al.* (1981) proponen como solución conservadora convertir el nivel de significación alcanzado por la prueba no paramétrica a su alternativa paramétrica. Si, por ejemplo, se ha aplicado la *U* de Mann-Whitney, su nivel de significación se convierte al correspondiente valor *t* y a continuación se prosigue como en la ecuación [4.26] ó [4.27].

Cabe también otra solución cuando en el informe de la investigación se recogen los datos empíricos de los grupos. En este caso, se pueden aplicar los estimadores no paramétricos del TE desarrollados por Kraemer y Andrews (1982), Krauth (1983) y Hedges y Olkin (1984).

Por regla general, la estimación del TE basada en el coeficiente de correlación de Pearson es la más fácil de obtener, ya que los informes de los estudios (correlacionales) suelen especificarlo. No obstante, en muchas ocasiones es posible encontrar estadísticos tan dispares como la correlación biserial-puntual, el coeficiente phi, tablas de contingencia con contrastes chi-cuadrado, pruebas *t*, análisis de varianza y otras alternativas. Para estos casos resulta de interés consultar la tabla 2 de la página siguiente, reproducida de Glass *et al.* (1981), que permite transformar algebraicamente al coeficiente de correlación de Pearson los estadísticos citados.

4.4.3. Amenazas contra la validez

La elección de una medida significativa e informativa de los resultados de un estudio constituye un aspecto fundamental de cualquier meta-análisis. El revisor debe poner especial cuidado en la estimación de los TEs (o de los niveles de significación), ya que dichos valores van a representar a cada estudio en la integración global de resultados. Citamos a continuación tres posibles problemas que pueden aparecer en esta etapa y que, en mayor o menor medida, pueden afectar a la validez de los resultados.

a) Algunos estudios tendrán que ser excluidos del meta-análisis por falta de datos suficientes para estimar los TEs correspondientes. La eliminación de estudios por datos incompletos puede sesgar la muestra de estudios incluidos.

b) Contrariamente, en otras ocasiones pueden salvarse estudios con datos incompletos pero que proporcionan información suficiente para estimar de un modo aproximado los TEs. La calidad de la estimación puede afectar a los resultados finales del meta-análisis.

c) Finalmente, hay una serie de frecuentes situaciones que pueden influir notablemente en la estimación del TE (Cooper, 1984). En primer lugar, la fiabilidad de las medidas implicadas en el estudio. A menor fiabilidad, menor sensibilidad en la detección de los efectos. En segundo lugar, la elección de la desviación típica para estimar la diferencia media tipificada puede modificar considerablemente las estimaciones. Y por último, no debemos olvidar el sesgo muestral que manifiestan los estimadores del TE. En efecto, TEs obtenidos en muestras pequeñas (p. ej., menores de $n = 10$) pueden «inflar» la estimación global del TE.

4.4.4. Protección de la validez

Las recomendaciones que debe seguir un revisor cuantitativo dependerán en gran medida de las características peculiares del área de investigación seleccionada para revisión. Las que citamos a continuación son de naturaleza general:

a) Para evitar la exclusión de estudios por falta de datos, el revisor debe ponerse en contacto con el autor del estudio y solicitarle la información pertinente (Eaton, 1984).

b) Las estimaciones de los TEs pueden diferir de un codificador a otro si muchos estudios presentan datos incompletos. Es, por tanto, aconsejable un estudio de la fiabilidad de las estimaciones que permita evaluar la exactitud de las reglas incluidas en el libro de codificación (Stock *et al.*, 1982).

TABLA 2

Estadístico informado	Transformación a r_{xy}	Referencia
a) Correlación biserial puntual, r_{bp}	(*) $r_{xy} = r_{bp} \sqrt{(n_1 n_2) / u N}$ u : ordenada en el origen de la distribución normal tipificada. N : tamaño muestral total.	Glass y Stanley (1970, p. 171)
b) $t = \frac{\bar{X}_1 - \bar{X}_2}{\left[S^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \right]^{1/2}}$	$r_{bp} = \left[\frac{t^2}{t^2 + n_1 + n_2 - 2} \right]^{1/2}$ luego pasar r_{bp} a r_{xy} según a).	Glass y Stanley (1970, p. 318)
c) t basada en grupos extremos	$q = \frac{t(2/n)^{1/2}}{\left\{ \frac{4z^2}{p^2} + \left[\frac{z^2}{p^2} - \frac{xz}{p} \right] t^2(2/n) \right\}^{1/2}}$ 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 abcisa).	Feldt (1961, p. 315)
d) $F=MC$ inter/ MC error para $J=2$ grupos.	$F^{1/2} = t $. Luego proceder según b).	
e) $F=MC$ inter/ MC error para $J > 2$ grupos.	1) Juntar los grupos en dos y luego proceder según d); o 2) $r_{xy} = \eta = [SC \text{ inter} / (SC \text{ inter} + SC \text{ error})]^{1/2}$	Hays (1973, pp. 683-684)
f) X^2 solo (es decir, sin disponer de las frecuencias para una tabla de contingencia).	(**) $r_{xy} = P = \left[\frac{X^2}{X^2 + N} \right]^{1/2}$ N : tamaño muestral total.	Kendall y Stuart (1967, pp. 557 y ss)
g) Tabla de contingencia 2×2 .	Calcular la correlación tetracórica por las tablas.	Glass y 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.	Transformar U a la correlación biserial por rangos mediante: $r_{bp} = 1 - 2U/n_1 n_2$.	Willson (1976)

(*) $r_{xy} = 1,25r_{bp}$ 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 $P^2 \rightarrow q^2$ conforme aumenta el número de categorías de la tabla. Con pocas categorías, esta estimación puede ser demasiado baja.

c) Con objeto de comprobar si la calidad de las estimaciones afecta a la magnitud de los TEs, debe codificarse la calidad de las estimaciones como una característica más del meta-análisis.

d) Con respecto a la comprobación de la posible relación entre la fiabilidad de las medidas y los TEs, éste es un problema de difícil solución. La única posibilidad sería recuperar las fiabilidades de estudios individuales, pero estos datos no suelen especificarse (Cooper, 1984).

e) La elección de la desviación típica para el cálculo de la diferencia media tipificada, por regla general no será muy problemática porque en los estudios no suelen darse las desviaciones de cada grupo sino los resultados globales de las pruebas estadísticas aplicadas, que permiten estimar únicamente la desviación típica intragrupos o global. Pero si disponemos de tal información, es aconsejable utilizar en el denominador del estadístico d la misma desviación típica en todos los casos (Cooper, 1984).

f) Para evitar en lo posible los efectos del sesgo sobre las estimaciones de los TEs, se deben aplicar los factores de corrección adecuados. En el caso de la diferencia media tipificada, las ecuaciones [4.13] y [4.14] describen el modo de corregir las estimaciones obtenidas sobre muestras pequeñas. Si se está utilizando el coeficiente de correlación de Pearson es conveniente su transformación a Z de Fisher mediante la ecuación [4.19] (Rosenthal, 1984).

4.5. Análisis e interpretación de datos

El carácter innovador de los procedimientos meta-analíticos alcanza su máxima dimensión en la etapa de análisis de los datos e interpretación de los resultados. Como apuntamos más arriba, esta aproximación implica una nueva perspectiva al desarrollo del conocimiento mediante la integración de estudios individuales. Una de sus principales aportaciones ha consistido en la aplicación

de los métodos estadísticos típicos de las investigaciones primarias a la revisión de estudios. Del mismo modo que el investigador primario es obligado a utilizar técnicas de análisis estadístico y de interpretación en sus informes, el meta-análisis exige al revisor que adopte las mismas normas de rigor científico en el análisis e interpretación de un conjunto de resultados de estudios diferentes (Cooper, 1982, 1984; Hedges, 1984b).

En la etapa de análisis e interpretación, los resultados de los estudios, ya cuantificados mediante niveles de significación o estimaciones del TE, son sintetizados para obtener un índice global y representativo del conjunto de estudios, acompañado de otros índices descriptivos tales como la desviación típica, los cuartiles, el rango, etc. (Smith, 1982). Pero los datos integrados en un meta-análisis rara vez serán lo suficientemente homogéneos entre sí como para suponer que la medida global es representativa de todos los estudios; en particular, cuando los estudios integrados presentan resultados contradictorios. Es precisamente en estas situaciones en las que el enfoque meta-analítico ofrece su máxima utilidad. En efecto, la variabilidad entre los resultados de los estudios puede ser tratada eficazmente analizando los efectos que determinadas características (codificadas previamente) ejercen sobre los TEs (Light, 1979; Pillemer y Light, 1980a, 1980b). Es decir, los TEs actúan como variable dependiente, mientras que las características de los estudios son las variables independientes.

Según los objetivos de una revisión de la literatura y según las características del área de investigación, el enfoque meta-analítico ofrece distintos procedimientos (vid. Bangert-Drowns, 1986 para una descripción detallada), pero todos ellos coinciden en apuntar la importancia de los métodos estadísticos para resumir el estado de conocimiento en ese área. Los primeros trabajos meta-analíticos de Glass y colaboradores proponían el uso único y exclusivo de estadísticos descriptivos para sintetizar los resultados globales del meta-análisis (Glass *et al.*, 1981). A este respecto, el enfoque «exploratorio» preconiza-

do por Tukey (1977) resulta especialmente indicado para describir los resultados, desde la aplicación de estadísticos robustos hasta el uso de tablas del tipo *stem-and-leaf* y gráficos *schematic boxplot*. Pero el uso exclusivo de estadísticos descriptivos en meta-análisis ha sido criticado porque limita seriamente sus posibilidades. La razón que argumentaba Glass es que los datos procedentes de un meta-análisis no suelen cumplir con los supuestos de las pruebas inferenciales tradicionales. Como trataremos a continuación, el argumento es incuestionable, pero precisamente por esta razón se ha desarrollado un intenso interés por el desarrollo de nuevas pruebas estadísticas de los datos al uso en los trabajos meta-analíticos (Cooper, 1984; Hedges y Olkin, 1985; Hunter *et al.*, 1982; Rosenthal, 1984).

4.5.1. Las pruebas inferenciales convencionales

Es común encontrar estudios meta-analíticos que aplican las pruebas estadísticas convencionales (análisis de varianza, análisis de regresión, etc.) tal y como si se tratara de investigaciones primarias. Sin embargo, su uso está sujeto a serias violaciones de sus supuestos. En primer lugar, los datos de un meta-análisis no siempre cumplirán con el supuesto de independencia, ya que en muchas ocasiones un solo estudio suele aportar más de un TE al conjunto de datos. En segundo lugar, las pruebas inferenciales tradicionales no permiten determinar si la variabilidad entre los TEs se debe únicamente a error de muestreo. Finalmente, el supuesto de homogeneidad de las varianzas (homocedasticidad) difícilmente se cumple en un meta-análisis, ya que las varianzas de los TEs son inversamente proporcionales a sus tamaños muestrales (Cooper, 1984; Glass *et al.*, 1981; Hedges y Olkin, 1985).

Todos estos argumentos suscitan serias dudas acerca de la validez de los procedimientos estadísticos convencionales en meta-análisis. En palabras de Hedges y Olkin

(1985): «No parece haber ningún argumento rigurosamente defendible en favor del uso de las pruebas *t* convencionales, el análisis de varianza o el análisis de regresión para analizar TEs o correlaciones» (p. 12).

Como consecuencia de las debilidades que sufren los métodos estadísticos convencionales en meta-análisis, se están dedicando actualmente grandes esfuerzos por encontrar nuevas técnicas inferenciales adecuadas al tratamiento estadístico de los TE. Aunque nos encontramos en sus inicios, esta línea de investigación metodológica ya ha dado algunos frutos. Presentamos a continuación un breve exponente de las nuevas técnicas de análisis estadístico de los TEs, junto con los métodos de acumulación de niveles de significación, que son técnicas más tradicionales.

4.5.2. Técnicas de acumulación de niveles de probabilidad

Posiblemente, los primeros procedimientos estadísticos desarrollados para sintetizar cuantitativamente los resultados de un conjunto de estudios se deban a L. H. C. Tippett (1931), R. A. Fisher (1932) y K. Pearson (1933), quienes desarrollaron varias técnicas de acumulación de niveles de significación. No obstante, hasta la década de los 70 no se comenzaron a aplicar con cierta asiduidad en las ciencias sociales y del comportamiento. R. Rosenthal (1978) ha ayudado a popularizar estos procedimientos. Una característica común a todas las técnicas de acumulación que aquí describimos es que los niveles *p* proceden de la comparación de los grupos únicamente (p. ej., *t* o *F* con un grado de libertad en el numerador); por lo general, un grupo experimental y un grupo de control.

Las técnicas de acumulación de niveles de probabilidad son pruebas globales (*omnibus*) o *no paramétricas* porque no dependen de la forma de los datos subyacentes, sino de los niveles de significación, *p*, exactos. Dado un conjunto de *k* estudios independientes con parámetros $\theta_1, \theta_2, \dots, \theta_i, \dots, \theta_k$, que se

contrastan con un estadístico T_i , es posible obtener el nivel de significación exacto y unilateral de cada una, $p_1, p_2, \dots, p_i, \dots, p_k$. Con estos datos se pueden hacer inferencias acerca de la hipótesis nula

$$H_0: \theta_1 = \theta_2 = \dots = \theta_i = \dots = \theta_k = 0$$

Esta hipótesis supone que *todos* los parámetros son cero. Es decir, basta con que uno solo de ellos no sea nulo para que la H_0 sea rechazada. Si la hipótesis se rechaza, estos procedimientos permiten concluir que el nivel de probabilidad medio manifestado por un conjunto de estudios que contrastan un mismo efecto experimental es significativo; si la hipótesis se acepta, ello implica que el tratamiento experimental ejerce un efecto nulo (Rosenthal, 1978; Rosenthal y Rosnow, 1984). Rosenthal (1978) y Hedges y Olkin (1985, cap. 3) han descrito varios procedimientos de acumulación de niveles de probabilidad. Presentamos a continuación los más interesantes.

4.5.2.1. Suma de logaritmos

Este método, también llamado chi-cuadrado inverso o método de Fisher (1932), supone que si disponemos de k estudios independientes sobre una hipótesis común con niveles de probabilidad unilaterales $p_1, p_2, \dots, p_i, \dots, p_k$, el estadístico

$$-2 \sum_i^k \log_e p_i \quad [4.37]$$

se distribuye según chi-cuadrado con $2k$ grados de libertad. Para obtener el nivel de probabilidad global representativo de la serie de estudios, se consultan las tablas de chi-cuadrado.

Aunque muy utilizado, este método tiene el inconveniente de que provoca el rechazo de la hipótesis nula de forma rutinaria cuando el número de estudios es elevado (p. ej., $k > 5$). Y además, resulta muy sensible a la no independencia de los niveles p .

4.5.2.2. Suma de probabilidades

Propuesta inicialmente por Edginton (1972a), esta técnica permite calcular el nivel de significación medio o global de una serie de estudios mediante la razón entre la suma de los k niveles p_i elevada a la k -ésima potencia y $k!$ (k factorial):

$$\bar{p} = \left(\sum_i^k p_i \right)^k / k! \quad [4.38]$$

La técnica es aconsejable con pocos estudios, ya que requiere que la suma de los niveles p_i no exceda mucho la unidad; de lo contrario, el nivel \bar{p} global tiende a ser bastante conservador, a no ser que se apliquen correcciones especiales.

4.5.2.3. Suma de puntuaciones t

Winer (1971) ha desarrollado un método que consiste en dividir la suma de puntuaciones t_i por la raíz cuadrada de la suma de las varianzas respectivas. Siendo la varianza de la distribución t de Student

$$V(t_i) = g.l._i / (g.l._i - 2)$$

el estadístico

$$z = \frac{\sum_i^k t_i}{\left\{ \sum_i^k \left[g.l._i / (g.l._i - 2) \right] \right\}^{1/2}} \quad [4.39]$$

sigue una ley normal tipificada. El valor z obtenido puede consultarse en las tablas correspondientes para obtener el nivel de probabilidad global \bar{p} .

La limitación fundamental de esta técnica radica en que no puede aplicarse con tamaños muestrales menores de 3. De hecho, este método no proporciona buenas aproximaciones a la distribución normal para $g.l. < 10$.

4.5.2.4. Suma de puntuaciones z

Esta técnica, denominada *método Stouffer* (Stouffer *et al.*, 1949), asume que dados k estudios independientes, la razón entre la suma de las k puntuaciones típicas z (correspondientes a cada nivel de probabilidad) y $k^{1/2}$ constituye una unidad de desviación normal tipificada

$$z = \frac{\sum_i^k z_i}{k^{1/2}} \quad [4.40]$$

Consultando las tablas de la distribución normal tipificada, se obtiene el nivel de probabilidad global, de \bar{p} , de los estudios integrados.

Este método es el más simple y, por lo general, el más aplicado. Su limitación esencial se encuentra en que asume varianzas iguales cuando en algunas condiciones los errores tipo I y tipo II pueden haberse incrementado.

4.5.2.5. Suma de puntuaciones z ponderadas.

Mosteller y Bush (1954) sugirieron una modificación del método de Stouffer, consistente en ponderar cada puntuación z_i por su tamaño muestral, o por sus grados de libertad, o por cualquier otro criterio tales como la elegancia, la validez interna o la validez ecológica del estudio:

$$z = \left(\frac{\sum_i^k w_i z_i}{\sum_i^k w_i^2} \right)^{1/2} \quad [4.41]$$

siendo w_i la ponderación aplicada. A continuación, se procede como en 4.5.2.4.

4.5.2.6. Contraste del nivel P medio

Edgington (1972b) propuso 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 probabilidad y la raíz cuadrada de la razón entre la varianzas de los p_i niveles y k sigue una ley normal tipificada. Abreviadamente,

$$z = \frac{(0,50 - \bar{p})}{\left(\frac{\sum_i^k p_i(1-p_i)}{k} \right)^{1/2}} \quad [4.42]$$

y a continuación se procede como en 4.5.2.4. La ventaja de esta técnica es su simplicidad de cálculo, pero tiene como limitación que el número mínimo de estudios no puede ser inferior a cuatro.

4.5.2.7. Contraste de la puntuación z media

Mosteller y Bush (1954) modifican el método de Stouffer transformando los niveles de probabilidad en puntuaciones z y computando la prueba t sobre la puntuación típica media, \bar{z} , con grados de libertad $k-1$:

$$t = \frac{\frac{\sum_i^k z_i}{k}}{\left(\frac{\sum_i^k (z_i - \bar{z})^2}{k-1} \right)^{1/2}} \quad [4.43]$$

siendo

$$\sigma_z^2 = \frac{\sum_i^k (z_i - \bar{z})^2}{k-1}$$

Esta técnica presenta una baja potencia estadística cuando el número de estudios es pequeño ($k < 5$).

4.5.2.8. El método de Tippett

Tippett (1931) desarrolló el primer procedimiento de acumulación de niveles de probabilidad. Este método supone que tenemos $p_1, p_2, \dots, p_i, \dots, p_k$ niveles de probabilidad independientes y permite contrastar la hipótesis nula de que ningún nivel p es significativo con la hipótesis alternativa de que

al menos uno sí lo es. El estadístico de contraste es $P_{[i]}$, definido como el nivel de probabilidad más pequeño de todo el conjunto de k niveles,

$$P_{[i]} = \text{mín} [p_1, p_2, \dots, p_i, \dots, p_k] \quad [4.44]$$

Una vez fijado un nivel de significación, α , la hipótesis nula queda rechazada si $P_{[i]} < 1 - (1 - \alpha)^{1/k}$.

4.5.2.9. Método logit

Otro método para combinar k niveles de probabilidad independientes ha sido desarrollado por George (1977) y consiste en transformar cada nivel p en unidad logit y combinarlos mediante el estadístico L :

$$L = \sum_i^k \log_e (p_i / 1 - p_i) \quad [4.45]$$

A continuación se calcula el estadístico L^* mediante

$$L^* = |L| [0,3 (5k+4) / k(5k+2)]^{1/2}, \quad [4.46]$$

que se distribuye según una ley t de Student con grados de libertad $(5k + 4)$.

4.5.2.10. Suma de puntuaciones z dependientes

Los nueve procedimientos para combinar niveles de probabilidad hasta aquí descritos tienen una característica en común: sólo son aplicables cuando es posible suponer independencia entre los niveles p . Este es un supuesto que suele violarse no pocas veces en meta-análisis, ya que es muy común que un mismo estudio aporte varios datos al conjunto. Strube (1985a) ha tratado este problema aplicando el «método de Stouffer» de suma de puntuaciones z_i al caso de niveles p_i dependientes.

Supóngase, por ejemplo, que tenemos dos

estudios que comparan un grupo tratado con una técnica psicoterapéutica y un grupo de control, y que ambos estudios han medido la efectividad del tratamiento con dos variables dependientes, a saber, autoinforme e informe del terapeuta. En la tabla 3 se exponen los niveles p y sus correspondientes puntuaciones típicas z .

TABLA 3

Estudios	P_i	z_i	r_{ii}
<i>Estudio 1</i>			
— Autoinforme	0,0125	2,24	0,70
— Informe del terapeuta	0,0033	2,72	
<i>Estudio 2</i>			
— Autoinforme	0,5517	-0,13	0,60
— Informe del terapeuta	0,7454	-0,66	
	Σ	4,17	

Si obviamos la falta de independencia que muestran los datos y aplicamos el método Stouffer definido por la ecuación [4.40] obtenemos,

$$z = \frac{\sum_i^k z_i / k^{1/2}}{2,085} = 4,17 / (4)^{1/2} = 2,085 \rightarrow p = 0,0188 \text{ (unilateral).}$$

Sin embargo, este modo de proceder aumenta desproporcionadamente el valor de z y, en consecuencia, la significación del nivel de probabilidad global, \bar{p} . El procedimiento alternativo sugerido por Strube (1985a) requiere disponer de las correlaciones, r_{ii} , entre las pruebas estadísticas pertenecientes al mismo estudio. En nuestro ejemplo, suponemos que la correlación entre el autoinforme y el informe del terapeuta es de 0,70 y 0,60 en los dos estudios, respectivamente. Esto supuesto, el estadístico

$$z = \frac{\sum_i^k z_i}{(k+2 \sum_i^k r_{ii})^{1/2}} \quad [4.47]$$

se distribuye según una ley normal tipificada. Téngase en cuenta que r_{ii} será distinto de cero sólo para las correlaciones entre pruebas estadísticas pertenecientes al mismo estudio, ya que se supone independiente entre los estudios. En el ejemplo de la tabla 3,

$$z = 4,17/[4+2(0,70+0,60)]^{1/2} \\ = 1,623 \rightarrow p = 0,0526 \text{ (unilateral).}$$

Obsérvese la drástica reducción de la significación del nivel de probabilidad global al tener en cuenta la dependencia de los niveles p_i intra-estudio. Obviamente, esta reducción tendrá lugar siempre que las correlaciones intra-estudio sean positivas.

La gran ventaja de este método de combinación de niveles de probabilidad frente a los anteriores es que permite abordar el problema de la independencia, pero su aplicación está limitada a aquellas situaciones en las que es posible obtener (o estimar) las correlaciones de las medidas intra-estudio (Strube, 1985a).

4.5.2.11. Usos e interpretaciones

No parece haber una técnica de acumulación especialmente válida para cualquier situación (Rosenthal, 1978). De hecho, Strube y Muller (1986) han comparado la potencia estadística de un grupo de técnicas sobre un amplio rango de tamaños muestrales individuales, de estudios acumulados y para diferentes TEs, sin encontrar diferencias apreciables, sobre todo cuando el número de estudios es elevado. No obstante, los distintos métodos de acumulación difieren en su simplicidad de cálculo, haciendo preferibles a ciertos métodos bajo determinadas condiciones. Concretamente, el método Stouffer de la suma de puntuaciones z_i ponderadas o no, e independientes o no, parece ser el que mejor se ajusta al mayor número de situaciones (Rosenthal, 1978).

Los métodos de acumulación o de combinación de niveles p proporcionan un índice de probabilidad promedio, \bar{p} , que nos da una

idea de la significación global de los contrastes efectuados sobre una hipótesis común. Sin embargo, si los niveles p combinados son muy heterogéneos entre sí, la interpretación del nivel medio \bar{p} resulta ambigua y susceptible de error. Probablemente, ciertas variables moderadoras pueden estar afectando a los resultados. Es por ello imprescindible comprobar el grado de heterogeneidad manifestado por los niveles p en torno a la probabilidad media global. Rosenthal y Rubin (1979) han desarrollado una prueba estadística que consiste en sumar las diferencias cuadráticas de cada puntuación z_i (transformada para cada nivel p) respecto de su media, \bar{z} . Es decir,

$$\chi^2 = \sum_i^k (z_i - \bar{z})^2, \quad [4.48]$$

que se distribuye según chi-cuadrado con grados de libertad $(k-1)$.

Si la prueba de homogeneidad dada en la ecuación [4.48] resulta significativa, es posible efectuar contrastes lineales para comprobar los efectos de variables moderadoras mediante una modificación del método Stouffer,

$$z = \frac{\sum_i^k \lambda_i z_i}{(\sum_i^k \lambda_i^2)^{1/2}} \quad [4.49]$$

siendo λ_i el peso o ponderación de contraste asignado a cada estudio, teniéndose que cumplir la igualdad $\sum_i^k \lambda_i = 0$ (Rosenthal y Rubin, 1979).

Las pruebas definidas en las ecuaciones [4.48] y [4.49] sólo son válidas para niveles p independientes. Si no puede suponerse independencia, la prueba chi-cuadrado de homogeneidad global (ec. [4.48]) no puede aplicarse. No obstante, puede modificarse el método Stouffer para permitir una comparación o combinación lineal entre niveles de probabilidad. Strube (1985a) ha realizado esta modificación proponiendo el estadístico

$$z = \frac{\sum_i^k \lambda_i z_i}{\left(\sum_i^k \lambda_i^2 + 2 \sum_i^k \lambda_i \lambda_{ii} r_{ii} \right)^{1/2}} \quad [4.50]$$

que se distribuye según una ley normal tipificada.

A pesar del carácter intuitivo de estas técnicas de acumulación, su interpretación es un tanto problemática (Hedges y Olkin, 1985). Estos métodos proporcionan un índice de probabilidad global representativo del conjunto de estudios, que informa de la significación global de los efectos de un determinado tratamiento o hipótesis conceptual. Recuérdese que la hipótesis nula común a todos estos métodos era que *todos* los efectos son nulos, $H_0: \theta_1 = \theta_2 = \dots = \theta_i = \dots = \theta_k = 0$. Para rechazar la hipótesis nula basta que un solo efecto θ_i sea distinto de cero. En consecuencia, el rechazo de la hipótesis nula no dice gran cosa acerca de la eficacia de un tratamiento.

Las dificultades en la interpretación de las técnicas de acumulación de niveles de probabilidad radica en su naturaleza no paramétrica. El rechazo de la hipótesis nula no permite concluir que los efectos del tratamiento son mayores de cero, ni que existe un efecto consistente a través de los estudios. Preguntas tan interesantes como éstas sólo pueden responderse mediante el tratamiento estadístico de medidas del TE. Estos análisis son preferibles para integrar resultados de investigación en ciencias sociales y del comportamiento (Hedges y Olkin, 1985). Siempre que se utilicen métodos de acumulación de niveles de probabilidad es, pues, aconsejable acompañarlos con tratamientos estadísticos de los TEs correspondientes (Rosenthal, 1978, 1984; Rosenthal y Rubin, 1979).

4.5.3. Técnicas de integración de tamaños del efecto

Los métodos de combinación de niveles de significación y los basados en la integración de TEs constituyen los dos bloques básicos de análisis estadístico para realizar una sín-

tesis cuantitativa de los resultados sobre un tema de investigación. Las técnicas basadas en la combinación de TEs son más informativas que las basadas en los niveles de significación, ya que permiten formular inferencias acerca de la dirección, la magnitud y la consistencia de los efectos (Green y Hall, 1984; Hedges y Olkin, 1985).

Hay distintas aproximaciones al análisis estadístico de resultados de investigación (Banbert-Drowns, 1986). Las más conocidas son las de Glass (1976, 1977; Glass *et al.*, 1981) y las de Hunter-Schmidt (Hunter *et al.*, 1982). Glass propone un análisis estadístico eminentemente descriptivo y toma la varianza de los TEs integrados en su valor aparente, es decir, supone que la variabilidad entre los TEs es real y, en consecuencia, debe ser explicada correlacionando los TEs con las características de los estudios. Por el contrario, la aproximación de Hunter-Schmidt se caracteriza por su naturaleza inferencial y porque no toma la varianza en su valor aparente, sino que supone que parte de la misma se explica por los efectos de diversos artefactos estadísticos, tales como el error muestral, la restricción del rango y la falta de fiabilidad de las medidas (Sánchez, 1986). Pero, cualquiera que sea el enfoque adoptado, no deben olvidarse los tres pasos fundamentales que han de guiar el uso de los métodos estadísticos (Hedges, 1984b): *a*) la búsqueda de modelos para los resultados de un conjunto de estudios; *b*) la comprobación de los supuestos del modelo, y *c*) la estimación de los parámetros del modelo para que se ajuste óptimamente a los datos empíricos.

Aunque hay diferentes enfoques para realizar un estudio meta-analítico, presentaremos aquí un seleccionado conjunto de técnicas analíticas que han resultado útiles para una amplia gama de objetivos. Se deben principalmente a Hedges y colaboradores (Hedges, 1981, 1982a, 1982b, 1982c, 1983; Hedges y Olkin, 1983b, 1985) y a Rosenthal y colaboradores (Rosenthal, 1982, 1983a, 1983b, 1983c; Rosenthal y Rubin, 1982b, 1982c, 1986). Centraremos nuestra atención en la diferencia media tipificada y

en el coeficiente de correlación de Pearson como los índices de TE más populares.

4.5.3.1. Ajuste al modelo de la media global

Dado un conjunto de k estudios independientes, cada uno de los cuales proporciona un índice de la magnitud del efecto (d o r) acerca de una hipótesis conceptual común, el primer modelo que razonablemente debe probarse es el de que todos los estudios estiman a un mismo TE paramétrico.

4.5.3.1.1. Diferencia media tipificada

Sean d_1, d_2, \dots, d_k las estimaciones del TE de k estudios independientes que contrastan un grupo experimental con un grupo de control. Supóngase que, para el mismo estudio, las poblaciones experimental y control se distribuyen según una ley normal con varianza común, $N(\mu_i^E, \sigma_i^E), N(\mu_i^C, \sigma_i^C)$. En consecuencia, el TE paramétrico del i -ésimo estudio será

$$\delta_i = (\mu_i^E - \mu_i^C) / \sigma_i$$

Esto supuesto, el modelo de la media global postula que los TEs paramétricos de los k estudios no difieren entre sí:

$$H_0 : \delta_1 = \delta_2 = \dots = \delta_i = \dots = \delta_k = \delta^*$$

Si se cumple esta hipótesis, puede representarse el conjunto de estudios mediante la media global. Hedges (1981, 1982a) ha propuesto un estimador del TE medio global, $d.$, que se obtiene mediante una combinación lineal ponderada de los k TEs. Las ponderaciones asignan mayores pesos específicos a las estimaciones d_i , de los estudios con los tamaños muestrales más grandes, ya que éstas serán más precisas, y viceversa. Las ponde-

raciones asignan pesos inversamente proporcionales a la varianza de cada estudio

$$w_i = \frac{1 / V(d_i)}{\sum_i 1 / V(d_i)} \quad [4.51]$$

donde $V(d_i)$ representa la varianza del i -ésimo estudio y se obtiene por

$$V(d_i) = \frac{n_i^E + n_i^C}{n_i^E n_i^C} + \frac{d_i^2}{2(n_i^E + n_i^C)} \quad [4.52]$$

Por tanto, una estimación del TE medio global viene dada por

$$d. = \sum_i w_i d_i = \frac{\sum_i d_i / V(d_i)}{\sum_i 1 / V(d_i)} \quad [4.53]$$

Si los tamaños muestrales de los estudios son muy pequeños ($n_i \approx 10$), es conveniente sustituir d_i por el estimador insesgado del TE definido por Hedges (1981) y que formulamos en la ecuación [4.13].

Para comprobar la bondad de ajuste del modelo, Hedges (1982a) y Rosenthal y Rubin (1982b) proponen una prueba de homogeneidad de las estimaciones del TE en torno al TE medio, $d.$, mediante

$$Q = \sum_i \frac{(d_i - d.)^2}{V(d_i)} \quad [4.54]$$

donde $d.$ se obtiene mediante la ecuación [4.53] y $V(d_i)$ mediante la ecuación [4.52]. El estadístico Q se aproxima a una distribución chi-cuadrado de Pearson con $k - 1$ grados de libertad. Si a un determinado nivel α Q supera el punto crítico, la hipótesis nula queda rechazada y, en consecuencia, el modelo de la media global no se ajusta a los datos; es decir, no es adecuado representar todo el conjunto de estimaciones del TE me-

diente el TE medio global, $d.$, ya que éstos no estiman el mismo parámetro δ^* . Si, por el contrario, la hipótesis nula es aceptada, el modelo se ajusta a los datos y entonces el TE medio global es un índice representativo de todos los estudios. Es aconsejable entonces estimar δ^* mediante un intervalo confidencial del tipo

$$p [d - c_{\alpha/2} V (d.) < \delta^* < d. + c_{\alpha/2} V (d.)] = 1 - \alpha, \quad [4.55]$$

siendo $c_{\alpha/2}$ la puntuación de la distribución normal tipificada y $V (d.)$ la varianza del TE medio $d.$, que se obtiene por

$$V (d.) = \left[\sum_i^k 1 / V (d_i) \right]^{-1} \quad [4.56]$$

La estimación de δ^* mediante $d.$, puede estar sesgada si de un conjunto de estudios sólo se han podido incluir aquellos que obtuvieron resultados significativos, a causa de la práctica común de no dar información suficiente cuando se acepta la hipótesis nula. En este caso, $d.$ sobrestimaré a δ^* . Para tratar este problema, Hedges (1984a) ha propuesto un estimador de δ^* que tiene en cuenta la exclusión de estudios con resultados no significativos.

4.5.3.1.2. Coeficiente de correlación

Lo que acabamos de presentar para la diferencia media tipificada puede generalizarse al caso en que tenemos r_1, r_2, \dots, r_k coeficientes de correlación obtenidos de k estudios independientes sobre una hipótesis común. El modelo del coeficiente de correlación global supone que los k estudios estiman un coeficiente de correlación paramétrico común, ρ :

$$H_0 = \rho_1 = \rho_2 = \dots = \rho_i = \dots = \rho_k \cong \rho$$

Puede obtenerse una estimación del coefi-

ciente de correlación global ponderando las transformaciones a la z de Fisher de cada coeficiente por la inversa de sus respectivas varianzas. Es decir, siendo $Z_{r_1}, Z_{r_2}, \dots, Z_{r_k}$ las transformaciones a z de Fisher mediante la expresión [4.19], la z de Fisher ponderada media, $z.$, se puede obtener por

$$z. = \frac{\sum_i^k w_i z_{r_i}}{\sum_i^k (n_i - 3)} \quad [4.57]$$

ya que la varianza de la z de Fisher es $1/(n - 3)$. Para obtener el coeficiente de correlación medio global, $r.$, operamos a la inversa en la expresión [4.19].

La bondad de ajuste del modelo puede comprobarse mediante el estadístico

$$Q = \sum_i^k (n_i - 3) (z_i - z.)^2 \quad [4.58]$$

que se distribuye según chi-cuadrado con $k - 1$ grados de libertad (Hedges y Olkin, 1985; Rosenthal, 1984). Q indica la homogeneidad de los coeficientes de correlación en torno a la media global y se interpreta del mismo modo que la diferencia media tipificada. La estimación por intervalo del coeficiente de correlación de la población, ρ , se obtiene mediante.

$$p [z. - c_{\alpha/2} V (z.) < \zeta < z. + c_{\alpha/2} V (z.)] = 1 - \alpha \quad [4.59]$$

siendo ζ la z de Fisher correspondiente a ρ y $V (z.)$ la varianza de $z.$,

$$V (z.) = \left(\sum_i^k n_i - 3k \right)^{-1} \quad [4.60]$$

y para obtener límites confidenciales del coeficiente de correlación se opera a la inversa en [4.19].

4.5.3.2. Ajuste de modelos categóricos

Por regla general, los TEs estimados de un conjunto de estudios suelen ser heterogéneos, resultando significativo el estadístico Q . Una estrategia a seguir entonces consiste en probar los efectos de las características codificadas de los estudios sobre los TEs. Esto puede hacerse ajustando distintos modelos a los datos. Hedges (1982b) ha propuesto un procedimiento para el ajuste de modelos categóricos a los TEs similar al análisis de varianza.

Este modelo supone que un conjunto de estimaciones del TE puede ser clasificado en p clases mutuamente excluyentes. Siendo $m_1, m_2, \dots, m_j, \dots, m_p$ el número de TEs incluidos en cada clase. Sea δ_{ij} el TE paramétrico del i -ésimo estudio correspondiente a la clase j .

$$\delta_{ij} = (\mu_{ij}^E - \mu_{ij}^C) / \sigma_{ij} \quad [4.61]$$

Esto supuesto, cabe formular tres hipótesis: La primera, hipótesis nula, afirma que todos los TEs subyacentes a cada estudio son homogéneos y estiman un parámetro común, δ^* , independientemente de las clases. Es decir,

$$\begin{array}{l} H_0: \text{Clase 1: } \sigma_{11} = \delta_{21} = \dots = \delta_{m_p 1} = \delta^* \\ \quad \cdot \quad \quad \cdot \quad \quad \quad \cdot \\ \quad \cdot \quad \quad \cdot \quad \quad \quad \cdot \\ \quad \cdot \quad \quad \cdot \quad \quad \quad \cdot \\ \text{Clase } p: \delta_{1p} = \delta_{2p} = \dots = \delta_{m_p p} = \delta^* \end{array}$$

La segunda, H_1 , supone que los TEs dentro de cada clase son homogéneos, pero que existen diferencias significativas entre las clases; es decir, cada clase estima un parámetro diferente δ_j^* .

$$\begin{array}{l} H_1: \text{Clase 1: } \sigma_{11} = \delta_{21} = \dots = \delta_{m_p 1} = \delta_1^* \\ \quad \cdot \quad \quad \cdot \quad \quad \quad \cdot \\ \quad \cdot \quad \quad \cdot \quad \quad \quad \cdot \\ \quad \cdot \quad \quad \cdot \quad \quad \quad \cdot \\ \text{Clase } p: \delta_{1p} = \delta_{2p} = \dots = \delta_{m_p p} = \delta_p^* \end{array}$$

y, finalmente, H_2 supone que cada TE subyacente a cada estudio, δ_{ij} , es diferente de los demás,

H_2 : Todos los δ_{ij} son diferentes.

Puesto que tenemos un conjunto de k estimaciones del TE, d_{ij} , clasificados en p clases, podemos definir un estimador ponderado de la media de cada clase, δ_j^* , mediante

$$d_{.j} = \frac{\sum_i^{m_j} d_{ij} / V(d_{ij})}{\sum_i^{m_j} 1 / V(d_{ij})} \quad [4.62]$$

donde $V(d_{ij})$ representa la varianza de cada estimación del TE y se obtiene mediante la ecuación [4.52]. El TE medio global, δ^* , puede estimarse por

$$d_{..} = \frac{\sum_j^p \sum_i^{m_j} d_{ij} / V(d_{ij})}{\sum_j^p \sum_i^{m_j} 1 / V(d_{ij})} \quad [4.63]$$

El primer paso para comprobar el ajuste de un modelo categórico implica contrastar H_0 versus H_2 . Ello es posible aplicando el estadístico Q_T ,

$$Q_T = \sum_j^p \sum_i^{m_j} \frac{(d_{ij} - d_{.j})^2}{V(d_{ij})} \quad [4.64]$$

que se distribuye según chi-cuadrado con $k-1$ grados de libertad y es similar al estadístico Q definido en la ecuación [4.54]. Si se rechaza H_0 , se supone que los TEs no comparten un mismo TE común. En consecuencia, el segundo paso consistirá en contrastar H_0 versus H_1 , que es similar a la prueba F del análisis de varianza y permite determinar si las clases muestran TEs medios diferentes. Ello se consigue definiendo el estadístico Q_B ,

$$Q_B = \sum_j^p \frac{(d_{.j} - d_{..})^2}{v(d_{.j})}$$

donde $v(d_{.j})$ es la varianza de la clase j y se obtiene por

$$V(d_{.j}) = \left[\sum_i^{m_j} 1/V(d_{ij}) \right]^{-1} \quad [4.66]$$

Q_B se distribuye según chi-cuadrado con $p-1$ grados de libertad y representa un estadístico de ajuste interclases. Si la H_0 vuelve a ser rechazada, es posible asumir que las clases estiman TEs medios diferentes o, lo que es lo mismo, que la característica analizada puede estar actuando como variable moderadora. También puede estimarse el TE medio de cada clase, δ_j^* , aplicando

$$p(d_{.j} - c_{\alpha/2} V(d_{.j}) < \delta_j^* < d_{.j} + c_{\alpha/2} V(d_{.j})) = 1 - \alpha \quad [4.67]$$

obteniendo $V(d_{.j})$ por la ecuación [4.66].

Es preciso también averiguar si existe homogeneidad *dentro* de las clases. Para ello, se contrasta H_1 versus H_2 mediante el estadístico de ajuste interclase Q_W ,

$$Q_W = \sum_j^p \sum_i^{m_j} \frac{(d_{ij} - d_{.j})^2}{V(d_{ij})} \quad [4.68]$$

Q_W sigue también una ley chi-cuadrado, con $k-p$ grados de libertad, cuando H_1 es verdadera, es decir, cuando los TEs son homogéneos dentro de las clases.

Si Q_B resulta significativo, pero Q_W no, el modelo se ajusta correctamente a los datos. Si Q_W es significativo, las clases no son homogéneas y el modelo no se ajusta a los datos. Se hace necesario entonces formular otro modelo introduciendo algún otro principio de clasificación. Para comprobar cuáles son las clases heterogéneas, Q_W se descompone en p estadísticos de homogeneidad intra-clase (uno para cada clase), tales que

$$Q_W = Q_{W_1} + Q_{W_2} + \dots + Q_{W_j} + \dots + Q_{W_p}$$

Cada estadístico Q_{W_j} se distribuye según chi-cuadrado con m_j-1 grados de libertad y se obtiene mediante

$$Q_{W_j} = \sum_i^{m_j} \frac{(d_{ij} - d_{.j})^2}{V(d_{ij})} \quad [4.69]$$

Como el lector habrá sin duda comprobado, existe una estrecha analogía entre el análisis de varianza y el análisis estadístico propuesto por Hedges (1982b). En efecto, el estadístico Q_T se descompone en dos partes independientes, una que determina el ajuste inter-clases, Q_B , y otra el ajuste intra-clase, Q_W , de tal modo que

$$Q_T = Q_B + Q_W,$$

expresión que refleja la partición de la suma de cuadrados total en un análisis de varianza.

El análisis estadístico descrito puede continuar ulteriormente. Si Q_B resulta significativo, se pueden plantear combinaciones o contrastes lineales entre TEs medios, de forma similar a la técnica de las comparaciones post-hoc del análisis de varianza (Arnau, 1981; Hedges, 1982b). Se puede contrastar cualquier comparación del tipo

$$\gamma = c_1 \delta_1^* + c_2 \delta_2^* + \dots + c_p \delta_p^*,$$

donde, c_1, c_2, \dots, c_p son constantes conocidas, mediante la estimación

$$c = c_1 d_{.1} + c_2 d_{.2} + \dots + c_p d_{.p}$$

El estadístico $z = C/S_C$ se distribuye según una ley normal, siendo

$$S_C = \left[\sum_i^p c_j^2 V(d_{.j}) \right]^{1/2} \quad [4.70]$$

y $V(d_{.j})$ se obtiene mediante la ecuación [4.66].

El procedimiento de análisis aquí desarro-

lado se basa en un modelo estructural de efectos fijos, y es perfectamente aplicable a la integración de coeficientes de correlación. López, Sánchez y López (1986a) han informatizado este tipo de ajuste. El ajuste de modelos categóricos aplicados al meta-análisis ha demostrado ser un método de análisis estadístico muy útil para la detección de variables moderadoras capaces de explicar contradicciones empíricas (cf., p. ej., Bangert-Drowns, Kulik y Kulik, 1983a; Hillocks, 1984; Sánchez, 1984, 1985). Hedges (1983) ha propuesto también la aplicación de un modelo de efectos aleatorios a la diferencia media tipificada (vid. asimismo Hedges y Olkin, 1985, para su extensión al coeficiente de correlación). Otros autores han propuesto también la aplicación del modelo de efectos aleatorios, pero desde un enfoque bayesiano (p. ej., Hunter *et al.*, 1982; Raudenbush y Bryk, 1985; Willson, 1981, 1984).

4.5.3.3. Ajuste de modelos continuos

Un procedimiento alternativo para explicar la variabilidad manifestada por las estimaciones del TE de un conjunto de estudios consiste en probar el ajuste de modelos de regresión. Hedges (1982c; vid. también Hedges y Olkin, 1983b, 1985) ha desarrollado un método de análisis de regresión múltiple ponderado en el que las estimaciones del TE se ponderan por la inversa de sus respectivas varianzas.

4.5.3.3.1. Diferencia media tipificada

Dada una serie de k estudios independientes, el vector $d' = (d_1, d_2, \dots, d_k)$ representa las estimaciones de los TEs paramétricos $\delta = (\delta_1, \delta_2, \dots, \delta_k)$; definamos el vector de TEs paramétricos δ en función de p variables fijas, x_i , según el modelo

$$\delta = X\beta, \quad [4.71]$$

donde X representa la matriz del diseño y β es un vector de coeficientes de regresión paramétricos. Dicho modelo asume que las p variables pueden ser discretas o continuas.

El modelo lineal expresado en la ecuación [4.71] puede formularse también como

$$d = X\beta + \epsilon, \quad [4.72]$$

donde $\epsilon = d - \delta$ es un vector de residuales que se distribuye según una ley aproximadamente normal k -dimensional con media cero y cuya matriz de covarianza es una matriz diagonal que se estima mediante

$$\hat{\Sigma} = \text{diag} [V(d_1), V(d_2), \dots, V(d_i), \dots, V(d_k)] \quad [4.73]$$

obteniéndose $V(d_i)$ mediante la ecuación [4.52]. Es decir, los elementos del vector son independientes pero no se distribuyen idénticamente, ya que las estimaciones d_i pueden tener varianzas distintas.

El análisis por mínimos cuadrados ponderados proporciona una estimación del vector de coeficientes de regresión mediante

$$\hat{\beta} = (X' W X)^{-1} X' W d, \quad [4.74]$$

siendo W la matriz de ponderación definida como $W = \hat{\Sigma}_d^{-1}$. Por tanto, cada elemento de la diagonal principal de W se corresponde con la inversa de la varianza del TE,

$$w_i = 1/V(d_i) = \frac{2(n_f^E + n_f^C) n_f^E n_f^C}{2(n_f^E + n_f^C)^2 + n_f^E n_f^C d_i^2} \quad [4.75]$$

El vector de coeficientes de regresión estimado, $\hat{\beta}$, se aproxima a una ley normal para muestras grandes, con media β y matriz de covarianza Σ_β , que se estima por

$$\hat{\Sigma}_\beta = (X' W X)^{-1}. \quad [4.76]$$

Es posible establecer intervalos confidenciales en torno a cada coeficiente de regresión mediante

$$p (\beta_j - c_{\alpha/2} \sqrt{\sigma_{jj}} < \beta_j < \beta_j + c_{\alpha/2} \sqrt{\sigma_{jj}}) = 1 - \alpha \quad [4.77]$$

siendo $c_{\alpha/2}$ el punto crítico de la distribución normal tipificada y σ_{jj} el j -ésimo elemento diagonal de la matriz de covarianza estimada de los coeficientes de regresión $\hat{\Sigma}_\beta$.

Para comprobar si las p características muestran simultáneamente una relación funcional significativa con los TEs, contrastamos la hipótesis nula

$$H_0: \beta = 0$$

mediante el estadístico

$$Q_R = \hat{\beta}' \hat{\Sigma}_\beta^{-1} \hat{\beta}, \quad [4.78]$$

que se distribuye según chi-cuadrado con p grados de libertad. Q_R es la suma de cuadrados (ponderada) debida a la regresión y, en consecuencia, si resulta significativo ello será evidencia de que las características incluídas en el modelo afectan a los TEs.

No obstante, la significación del estadístico Q_R no tiene sentido si no se comprueba la especificación del modelo. Si el modelo de regresión planteado no se ajusta adecuadamente a los datos, los coeficientes de regresión $\hat{\beta}_j$ no son fácilmente interpretables. La adecuación del modelo puede contrastarse mediante el estadístico Q_E , que es la suma de cuadrados (ponderada) residual,

$$Q_E = d' W d - Q_R \quad [4.79]$$

Cuando el modelo está correctamente especificado, es decir, cuando $\delta = X\beta$, Q_E se aproxima a la distribución chi-cuadrado con $k-p$ grados de libertad. Si Q_E excede al punto crítico, el modelo no se ajusta adecuadamente, ya que la varianza residual resulta mayor que la esperada por azar.

4.5.3.3.2. Coeficiente de correlación

Del mismo modo que es posible ajustar modelos de regresión sobre estimadores del TE basados en la diferencia media tipificada, también puede aplicarse la misma estrategia al análisis de un conjunto r_1, r_2, \dots, r_k de coeficientes de correlación independientes (Hedges y Olkin, 1983b, 1985). No obstante, para aplicar análisis de regresión ponderado a coeficientes de correlación es necesario transformarlos previamente a valores correspondientes de la distribución z de Fisher mediante la ecuación [4.19].

Sea $Z' = (Z_1, Z_2, \dots, Z_k)$ el vector de coeficientes de correlación transformados a la Z de Fisher y $\zeta = (\zeta_1, \zeta_2, \dots, \zeta_k)$ el vector de parámetros. El modelo

$$\zeta = X\beta \quad [4.80]$$

pretende relacionar los coeficientes de correlación con un conjunto de p variables predictoras o características conocidas.

El estimador ponderado de los coeficientes de regresión es

$$\hat{\beta} = (X'WX)^{-1} X'WZ, \quad [4.81]$$

donde W es una matriz diagonal de ponderaciones que se corresponden con la matriz de covarianza de z , Σ_z^{-1} , definida por

$$\Sigma_z^{-1} = \text{diag} (n_1 - 3, n_2 - 3, \dots, n_k - 3), \quad [4.82]$$

siendo n_1, n_2, \dots, n_k los tamaños muestrales de los estudios. Nótese que cada elemento de la diagonal es el recíproco de la varianza de la distribución Z de Fisher.

Para muestras grandes, $\hat{\beta}$ se distribuye según una ley normal multivariada con media β y matriz de covarianza $\Sigma_\beta = (X'WX)^{-1}$. Como consecuencia, es posible establecer intervalos de confianza para cada coeficiente de regresión $\hat{\beta}_j$ del mismo tipo que el desarrollado en la ecuación [4.77].

Por la misma razón, la suma de cuadra-

dos (ponderada) debida a la regresión se obtiene por

$$Q_R = \hat{\beta}' \Sigma_{\hat{\beta}}^{-1} \beta, \quad [4.83]$$

y el estadístico para el contraste de la adecuación del modelo a los datos es

$$Q_E = Z'WZ - Q_R \quad [4.84]$$

Resumiendo. La búsqueda de modelos de regresión múltiple ponderada es una estrategia de análisis muy útil para encontrar variables moderadoras que pueden producir resultados empíricos aparentemente contradictorios (p. ej., Bargert-Drowns, Kulik y Kulik, 1983a; Sánchez, 1985; Stock *et al.* 1985). Su computación puede obtenerse de cualquier paquete estadístico al uso que incluya cuadrados mínimos ponderados (BMPD, SPSS^x, SAS o SYSTAT; vid. también López, Sánchez y López, 1986b).

Los métodos de análisis estadístico que hemos presentado en este epígrafe pueden aplicarse a una amplia gama de situaciones, pero son tan sólo una pequeña muestra del gran abanico de posibilidades que un revisor de investigación tiene a su disposición para el tratamiento estadístico de la evidencia empírica existente acerca de cualquier problema de investigación. Otras técnicas de análisis estadístico interesantes, pero de menor aplicabilidad, son, por citar algún ejemplo, las aproximaciones cluster (Hedges y Olkin, 1983a; 1985), el enfoque no paramétrico (Katz, Marascuilo y McSweeney, 1985) o las técnicas de recuento desarrolladas por Hedges y Olkin (1980). El interés por la aplicación habitual de tales técnicas ha alentado también la elaboración de programas estadísticos para facilitar la computación (p. ej., Gorman, Primavera y Karras, 1983; McDaniell, 1986; Mullen y Rosenthal, 1985).

4.5.4. Amenazas contra la validez

La etapa de análisis y, sobre todo, la de interpretación, están sometidas a una buena

20
dosis de subjetividad, tanto en investigación primaria como en meta-análisis (Cooper, 1984). No resultará inútil, por tanto, que se especifiquen qué problemas suelen plantear contra la validez de la investigación los procedimientos citados.

En primer lugar, respecto a la etapa de análisis de datos, los supuestos de las pruebas estadísticas pueden no cumplirse, dando así lugar a reglas de inferencia incorrectas (Shapiro y Shapiro, 1982). Uno de los supuestos que en mayor medida suele pasarse por alto es el de la independencia de los datos, y ello principalmente porque un mismo estudio puede aportar más de un dato a la revisión (Glass *et al.*, 1981; Rosenthal, 1984).

En segundo lugar, y respecto a la interpretación de resultados, es fundamental distinguir entre *evidencia generada por la revisión* y *evidencia generada por el estudio*. Esta última tiene lugar cuando cada estudio individual contiene los resultados que ponen a prueba la relación o el efecto experimental que está siendo considerado. En este caso, la integración de todos los resultados en el meta-análisis permite afirmar alguna relación de causalidad. El otro tipo de evidencia que puede darse en un estudio meta-analítico es aquella que no procede de los estudios individuales, sino de las variaciones en las características y detalles del procedimiento de los estudios. Tal es el caso, por ejemplo, de la calidad del estudio o del tipo de diseño experimental aplicado. Estas características no pueden ser manipuladas en un solo estudio y, en consecuencia, la comparación de los estudios en base a dichas características constituye la evidencia generada por la revisión. Este segundo tipo de evidencia *no* es susceptible de apoyar supuesto alguno de causalidad; tan sólo permite establecer relaciones de asociación. Ello se debe a que los estudios no se asignan aleatoriamente a las condiciones y, en consecuencia, no pueden descartarse efectos de variables moderadoras. En palabras de Cooper (1984): «En investigación primaria es esta asignación aleatoria la que nos permite asumir que terceras variables están representadas por igual

en las condiciones experimentales» (página 28). Si el investigador no tiene en cuenta esta limitación de la evidencia generada por la revisión, sus interpretaciones pueden apoyar incorrectamente supuestos de causalidad.

4.5.5. Protección de la validez

El problema de la independencia de los datos constituye una de las principales amenazas contra la validez de la conclusión estadística (Cooper, 1984; Glass *et al.*, 1981; Hunter *et al.*, 1982; Rosenthal, 1984). Cuando se acumulan niveles de probabilidad no independientes, Strube (1985a) ha propuesto una técnica de análisis adecuado. Si la integración de los resultados de los estudios se hace en base a estimaciones del TE, ya sean diferencias medias tipificadas o coeficientes de correlación, Glass *et al.* (1981) recomiendan afrontar la falta de independencia mediante la técnica *jackknife* y Tukey Mosteller y Tukey, (1968), que consiste en obtener un pseudovalor para cada estudio representativo del conjunto de resultados dentro del estudio. Rosenthal y Rubin (1986) han desarrollado un procedimiento que proporciona una estimación combinada del TE a partir de un grupo de TEs pertenecientes a un mismo estudio. Esta técnica supone que, dado un estudio individual, tenemos dos condiciones (experimental y control) evaluadas sobre r variables dependientes mediante r pruebas t .

Suponiendo que se conoce la intercorrelación típica de las r variables dependientes, ρ , es posible obtener una estimación combinada e_c de los r TEs mediante

$$e_c = \frac{\sum_i \lambda_i t_i / I}{\left[\rho \left(\sum_i \lambda_i \right)^2 + (1 - \rho) \sum_i \lambda_i^2 \right]^{1/2}} \quad [4.85]$$

donde t_i es el resultado de la prueba t para la i -ésima variable dependiente, λ_i es el peso específico o ponderación que el revisor decide asignar a cada variable dependiente en función de su importancia, calidad de la medida o cualquier otro criterio. Obviamente, los valores λ_i deben especificarse de antemano. I representa un índice del tamaño del estudio y varía según la diferencia media tipificada definida. Si se utiliza la d de Cohen (1977), $I = [(n-1)/2]^{1/2}$; si es la g de Hedges (1981), $I = (n/2)^{1/2}$. El índice e_c es así una estimación combinada y ponderada de un grupo de diferencias medias tipificadas y dependientes entre sí y pueden expresarse en términos del coeficiente de correlación mediante

$$r_c = \frac{e_c}{\left[e_c^2 + (g.l.) / I^2 \right]^{1/2}}, \quad [4.86]$$

donde g.l. son los grados de libertad de la media cuadrática de error.

Una vez obtenida una combinación de los TEs, es posible aplicar las técnicas de integración de TEs descritas anteriormente sin preocuparnos por el problema de la independencia.

En todo lo que se refiere a la interpretación de los resultados, es aconsejable que los revisores de investigación hagan públicas sus reglas de inferencia para que el lector potencial del estudio pueda juzgar la validez de sus conclusiones. Además, si se ha hecho algún estudio sobre la validez de las reglas de interpretación, también debe ser presentado.

Por último, es preciso tener siempre presente la distinción entre evidencia generada por el estudio y la generada por la revisión al interpretar la varianza de las estimaciones del TE. Por regla general, los resultados de un meta-análisis basados en evidencia generada por la revisión no deben interpretarse como evidencia de alguna relación causal, sino como una guía para dirigir investigaciones futuras.

4.6. Publicación del estudio

Una característica esencial del método científico es la disponibilidad pública de los datos y del proceso de investigación. Este registro es particularmente importante para el meta-análisis, debido a que la investigación meta-analítica está basada, a su vez, sobre información ya publicada. De hecho, una de las principales contribuciones del meta-análisis a la revisión de la literatura es que puede ser objetivamente verificada (Bullock y Svyantek, 1985).

El informe de un estudio meta-analítico debe recoger aspectos tales como la lista de estudios utilizada en el análisis, las reglas de codificación aplicadas para extraer las características de los estudios, una copia de los datos y de los análisis efectuados, etc. No existe, sin embargo, un conjunto de normas específicas que guíen la publicación de un meta-análisis. La carencia de una normativa es perjudicial para la validez de los informes, ya que los diferentes criterios de las editoriales pueden provocar variabilidad en las revisiones (Cooper, 1984; Guskin, 1984). La mejor solución es utilizar la ya tradicional división de las investigaciones primarias: introducción, método, resultados y discusión.

En la introducción debe presentarse una panorámica de los problemas teóricos y conceptuales en torno al tema de la revisión y una descripción de las controversias que las revisiones anteriores han dejado sin resolver.

La sección destinada a método de un estudio meta-analítico será bastante diferente de la de una investigación primaria. Su objetivo es describir cómo se llevó a cabo la investigación. La organización de esta sección es muy desigual. A título orientativo podrían incluirse las siguientes subsecciones:

a) Presentación de los detalles de búsqueda de la literatura, incluyendo una descripción de las fuentes utilizadas, las palabras-clave y el intervalo temporal abarcado. El informe detallado del proceso de búsqueda de los estudios cumple dos finali-

dades (Hunter *et al.*, 1982), a saber: en primer lugar, permite al lector juzgar la representatividad de las fuentes objeto de la revisión y en segundo lugar, posibilita que las futuras revisiones del mismo tema puedan extender la revisión sin duplicarla.

b) Descripción del proceso de codificación de las características de los estudios. Ello implicará recoger las reglas de codificación aplicadas y la definición de las características incluidas. Si se ha hecho un estudio de la fiabilidad de la codificación con múltiples codificadores, deberá especificarse también.

c) Definición de la unidad de análisis del estudio y el tipo de medida elegido para la integración cuantitativa de los resultados, justificando además las razones de su elección. Ya sean niveles de probabilidad o índices del TE, se deberán especificar los criterios para la identificación de los contrastes de hipótesis independientes. Si los estudios incluidos son muy numerosos, es incluso conveniente presentar y analizar en detalle las metodologías prototípicas de los estudios.

d) Descripción y justificación de las técnicas de análisis estadístico seleccionadas por el revisor para confirmar o desconfirmar la hipótesis (o las hipótesis) propuesta(s) en la introducción.

En la sección destinada a resultados, el revisor debe presentar un resumen de la base estadística obtenida en la integración y comparación de los datos. En primer lugar, se especificará el número de datos extraídos junto con las características codificadas. En segundo lugar, se presentarán índices estadísticos globales que representen a todo el conjunto de datos, acompañados de índices de homogeneidad para determinar la magnitud de la varianza manifestada en los datos. A continuación, se pueden exponer los resultados obtenidos al clasificar los datos en función de las características para detectar posibles variables moderadoras.

El informe de un estudio meta-analítico finalizará con la discusión de los resultados obtenidos. En esta sección se presentará un resumen de los principales resultados de la

20

revisión, formulando una interpretación sustantiva de la importancia de los TEs obtenidos. También resultará interesante explicar las razones por las que los resultados difieren de los obtenidos en revisiones pasadas. Por último, se concluirá proponiendo las directrices más fructíferas que, en opinión del revisor, deberá guiar la investigación futura.

4.6.1. Amenazas contra la validez

La principal amenaza contra la validez del informe de un estudio meta-analítico proviene de la no inclusión de información relevante, sin la cual no será posible evaluar la validez del estudio ni su posible replicación. Esta amenaza se aplica tanto a la omisión de detalles importantes acerca de cómo se realizó la revisión como a la omisión de evidencia acerca de los moderadores potenciales de las relaciones (Cooper, 1984).

4.6.2. Protección de la validez

No es posible ofrecer recomendaciones concretas para proteger la validez del informe de una revisión cuantitativa. Difícilmente un revisor puede anticipar o pronosticar qué variables moderadoras podrían afectar a la validez de sus resultados si no se concluyen en el informe. Sin embargo, a lo largo de la exposición que hemos seguido de las etapas de un estudio meta-analítico, el lector podrá recoger sugerencias generales al respecto.

5. EVALUACIÓN

Entendido como una nueva perspectiva para la búsqueda de conocimiento, el meta-análisis es, hoy por hoy, una realidad. Basta hojear cualquier revista de pedagogía o de psicología para comprobar su éxito «comercial» (Pillemer, 1984). Pero el meta-análisis es también un enfoque novedoso y, como tal,

está sujeto a imperfecciones e inconvenientes que, con el tiempo, es de esperar que se superen. Buena parte de las críticas que ha desatado la aplicación del meta-análisis provienen del campo de la psicología clínica. En efecto, el impulso que los trabajos de Glass y colaboradores (Smith y Glass, 1977; Smith, Glass y Miller, 1980) sobre la efectividad de la psicoterapia ejerció sobre el desarrollo del meta-análisis ha sido reconocido por unos (Cooper y Arkin, 1981; Fiske, 1983; Kendall y Maruyama, 1985; Strube, Gardner y Hartmann, 1985; Strube y Hartmann, 1983, 1983) y criticado por otros (Cook y Leviton, 1980; Eysenck, 1978, 1983; Gallo, 1978; Kazdin, 1985; Kazdin y Wilson, 1978; Kazrin, Durac y Agteros, 1979; Leviton y Cook, 1981; Mansfield y Bussey, 1977; Presby, 1978; Rachman y Wilson, 1980; Wilson, 1985; Wilson y Rachman, 1983).

Se hace, por tanto, obligado abordar una valoración de las ventajas y limitaciones que presenta esta metodología de la revisión.

5.1. Ventajas

Un modo de evaluar la utilidad del enfoque meta-analítico a la integración del conocimiento científico consiste en describir sus ventajas frente a las revisiones tradicionales (Green y Hall, 1984; Pillemer, 1984).

1. *Eficiencia.* El meta-análisis es un método más eficiente que las revisiones tradicionales para integrar un área de investigación abundante. En efecto, la capacidad de un revisor puede verse superada por la gran cantidad de investigaciones que existe en prácticamente todas las áreas de investigación al uso (Curllette y Cannella, 1985; Glass *et al.*, 1981; Green y Hall, 1984).

2. *Rigor científico.* El meta-análisis exige mayor objetividad en el desarrollo de una revisión que las prácticas tradicionales. Un meta-análisis especificará bajo qué criterios fueron seleccionados los estudios, presenta-

rá los resultados globales sobre los que se basan sus conclusiones y utilizará técnicas de análisis estadístico. En el peor de los casos, es posible que el lector potencial de un estudio meta-analítico discrepe en los criterios de selección, en el análisis estadístico aplicado o en cualquier otro aspecto de la revisión; pero, al menos, conocerá todos los detalles del proceso de revisión, algo que las revisiones tradicionales no suelen especificar (Light y Pillemer, 1984; Sindelar y Wilson, 1984). Ello permite que el meta-análisis cumpla con el supuesto de *replicabilidad* (Fiske, 1983; Walberg y Haertel, 1980).

3. *Detección de pequeños efectos.* La investigación en ciencias sociales y del comportamiento se caracteriza por la obtención de efectos experimentales y relaciones de baja magnitud (Rosenthal, 1983c). El uso de procedimientos estadísticos es capaz de identificar efectos sutiles que pasarían inadvertidos ante una inspección casual de un revisor cualitativo (Cooper y Rosenthal, 1980). No obstante, un énfasis excesivo en los pequeños efectos puede inducir a «glorificar lo trivial» (Pillemer, 1984). No hay un criterio absoluto para determinar cuándo un efecto es importante y cuándo debe ser desechado, pero la opinión de expertos en el área y la comparación con los resultados de estudios previos puede ayudar a su valoración (Wortman, 1983). Por otra parte, Rosenthal y Rubin (1982a) han propuesto un interesante método para evaluar la importancia social de los efectos. El que denominan *BESD* (*Binomial effect size display*) consiste en convertir la proporción de varianza explicada por un efecto en una tabla de contingencia 2x2. La pregunta a la que trata de responder es: ¿cuál es el efecto sobre la tasa de éxito (p. ej., tasa de supervivencia, de mejoría, de selección, etc.) manifestado por la implementación de un nuevo tratamiento, un nuevo mecanismo de selección o una variable predictora? Supóngase, por ejemplo, que el coeficiente de correlación medio global de un conjunto de estudios sobre la efectividad de un tratamiento fuera de 0,32, lo que implica una proporción de varianza explicada de tan sólo 0,10.

En la tabla 4 reproducimos el *BESD* correspondiente.

TABLA 4

Presentación del *TE* binomial (*BESD*) para un $r=0,32$ que explica el 10 por 100 de la varianza

Condición	Resultado del tratamiento		
	Vivo	Muerto	
Tratamiento	66	34	100
Control	34	66	100
	100	100	200

Resultaría absurdo calificar de modesto a un *TE* equivalente a un aumento en la tasa de éxito del 34 al 66 por 100. El uso, pues, del *BESD* ayuda a valorar la importancia real de los efectos encontrados en ciencias sociales.

4. *Aumento de la potencia.* La combinación de resultados de estudios diferentes implica un aumento del tamaño muestral y, como consecuencia, de la potencia estadística de la prueba aplicada. Esta propiedad es particularmente ventajosa cuando los tamaños muestrales de los estudios individuales son pequeños (Pillemer y Light, 1980b). No obstante, es preciso tomar precauciones contra la acumulación ciega de datos. No hay que olvidar que los resultados combinados deben referirse a una misma hipótesis conceptual y que han de ser independientes.

5. *Obtención de estimaciones precisas del TE.* A pesar de los esfuerzos de los metodólogos por aconsejar el uso de índices del *TE* junto con los resultados de las pruebas de significación, es raro encontrar estudios que informan del *TE*. El meta-análisis, sin embargo, obliga a trabajar con índices del *TE* y a interpretar el efecto global mediante alguna estimación media. Las revisiones tradicionales, por el contrario, cuando realizan algún tipo de análisis estadístico, se basan únicamente en el recuento de los resultados significativos *versus* no significativos (Pillemer y Light, 1980b; Sindelar y Wilson, 1984).

6. *Aprovechamiento de resultados contradictorios.* Cuando la evidencia empírica en un área de investigación no es coherente, el meta-análisis ofrece unas ventajas inmejorables para explorar las razones de tales discrepancias (Glass *et al.*, 1981). Los resultados contradictorios pueden deberse a diversas causas, todas ellas amenazas contra la validez de constructo (Cook y Campbell, 1979; Judd y Kenny, 1981): *a)* los tratamientos catalogados como similares pueden no serlo; *b)* el tipo de diseño de investigación puede afectar a los resultados; *c)* las interacciones ambiente-tratamiento pueden afectar a su eficacia, o *d)* las estrategias de análisis utilizadas en estudios diferentes pueden arrojar resultados diferentes (Pillemer y Light, 1980b). El efecto moderador de estos factores puede ser evaluado estadísticamente mediante diversas técnicas meta-analíticas.

7. *Seguridad.* Al basarse en un análisis detallado y objetivo de la evidencia empírica existente, el meta-análisis puede adoptar interpretaciones y conclusiones más seguras que en las revisiones tradicionales (Curllette y Cannella, 1985; Green y Hall, 1984). Esta ventaja posibilita que pueda llenar vacíos en la literatura y proponer convenientemente direcciones futuras para la investigación (Cook y Leviton, 1980; Fiske, 1983; Strube y Hartmann, 1983).

Resumen, pues, la principal aportación del meta-análisis a la revisión cuantitativa de investigaciones proviene del carácter científico con que impregna todo el proceso de revisión (Cooper, 1984). Parafraseando a Fiske (1983): «Los estudios meta-analíticos son claramente superiores a las revisiones cualitativas convencionales de dominios de investigación, simplemente porque son más científicas y porque se acercan más a los ideales del quehacer científico» (p. 69).

5.2. Limitaciones

Aunque el enfoque meta-analítico ha ejercido con enorme impacto sobre las prácti-

cas de las revisiones de investigación, también ha sufrido duras críticas (Eysenck, 1978; Gallo, 1978; Mansfield y Bussey, 1977; Presby, 1978; Searles, 1985). La polémica que este tema suscita ha acaparado la atención de editores de revistas (vid. p. ej., *American Psychologist*, 1984, 36, 71-75; *Clinical Psychology Review*, 1985, 5(1), completa; *Journal of Consulting & Clinical Psychology*, 1983, 51, completa; *Journal of Special Education*, 1984, 18(1), completa; *Personnel Psychology*, 1985, 38, 697-801; *Behavioral & Brain Sciences*, 1978, 3, p. 377-415).

A lo largo de la descripción de las etapas que hemos desarrollado en páginas anteriores, pueden deducirse algunas de las críticas y limitaciones que ha recibido este enfoque. Resumimos a continuación las más importantes.

1. *Propensión a cometer errores.* Debido a que los estudios primarios no suelen incluir medidas del TE, el trabajo meta-analítico es propenso a cometer errores (Green y Hall, 1984). Para la realización de un estudio meta-analítico se supone que: *a)* todas las condiciones que tienen un efecto importante sobre la variable dependiente se han descrito con exactitud en los informes de investigaciones primarias, y *b)* los resultados y las pruebas de significación se han obtenido e informado correctamente (Curllette y Cannella, 1985). Estos supuestos pueden ser violados debido a errores en los registros de los datos, a descripciones inadecuadas de los estudios y a análisis incorrectos de datos (Jackson, 1980; Searles, 1985). Por último, los informes incompletos inducen asimismo a error. Este problema es más serio en la medida en que la supresión de información sea sistemática (Strube, Gardner y Hartmann, 1985). No obstante, hay procedimientos estadísticos que permiten abordar el problema de los datos incompletos (Hedges, 1984a; Hedges y Olkin, 1985).

2. *Heterogeneidad.* Una de las críticas más frecuentes consiste en afirmar que no tiene sentido mezclar resultados de estudios que poseen características diferentes (Eysenck,

1978; Gallo, 1978; Presby, 1978; Searles, 1985). No es correcto agrupar estudios con técnicas de medida diferentes, con tipos de personas distintos, etc. La crítica de la heterogeneidad de los estudios (también denominada «el problema de las manzanas y las naranjas» ha sido hábilmente desarticulada por Glass *et al.* (1981; vid. también Glass, 1978; Glass y Kliegl, 1983) con dos argumentos. En primer lugar, «... la afirmación de que sólo estudios idénticos en todos los aspectos pueden ser comparados es una contradicción; *no hay necesidad de compararlos*, ya que tendrán el mismo resultado salvo errores estadísticos» (Glass *et al.*, 1981, pp. 218-219; el subrayado es nuestro). En segundo lugar, por analogía Glass y colaboradores argumentan que si en investigación primaria es legítimo generalizar sobre las personas, también es lícito hacerlo con investigaciones primarias. Esta estrategia no sólo permite alcanzar conclusiones generales (acerca de la fruta) sino que también es capaz de comparar resultados más específicos (sobre manzanas y naranjas) en base a determinadas características.

3. *Calidad de los datos.* El meta-análisis descarta la calidad de los estudios como un criterio para seleccionarlos. Como consecuencia, Eysenck (1978), Gallo (1978) y Wilson (1982, 1985; Wilson y Rachman, 1983) entre otros, afirman que las conclusiones descansan sobre cantidades indiscriminadas de datos más que sobre la calidad de los estudios. La defensa (Glass *et al.*, 1981; Hunter *et al.*, 1982; Rosenthal, 1984) argumenta que la eliminación de estudios a priori por deficiencias metodológicas es un error porque está sometida al sesgo del revisor y porque, en última instancia, la calidad de los datos es una característica cuyos efectos sobre los resultados pueden ser valorados formalmente en el meta-análisis. El problema de la calidad de los datos implica dos decisiones: a) cómo definir la «calidad», y b) qué hacer con los estudios de «baja calidad». Según Strube y Hartman (1982, 1983), el examen de la calidad de los datos requiere considerar tres tipos de validez: conceptual (validez de cons-

tructo), metodológica (validez interna) y estadística (validez de la conclusión estadística). El examen de la validez conceptual afecta a las decisiones sobre el ajuste entre las definiciones operativas y el constructo subyacente. La valoración de la validez metodológica implica determinar la calidad del contraste de hipótesis, lo que incluye la evaluación de las amenazas clásicas a la validez interna (Campbell y Stanley, 1966; Cook y Campbell, 1979; Ato, 1984), así como las variables metodológicas de un área particular. La validez de la conclusión estadística se refiere a la adecuación de los procedimientos estadísticos aplicados.

Estas ideas pueden ayudar a definir la calidad de los datos de un estudio meta-analítico. Pero ¿qué hacer con estudios de «baja calidad»? Los críticos recomiendan aplicar una estrategia todo-o-nada y eliminar tales estudios. Una postura más recomendable, sin embargo, consiste en considerar la calidad de los datos no como una dicotomía, sino como un continuo susceptible de ser evaluado por expertos. Así, cada estudio puede ser ponderado en función de su calidad (Kendall y Maruyama, 1985; Strube *et al.*, 1985).

4. *Representatividad.* Es prácticamente imposible que la población de estudios localizados se identifique exactamente con la población objetivo. Esto no sería problema si todos los estudios tuvieran la misma probabilidad de ser localizados. Pero la evidencia experimental ha demostrado que éste no es el caso (vid., p. ej., Lane y Dunlap, 1978). Es más fácil localizar estudios publicados que no publicados y, a su vez, aquéllos constituyen una muestra sesgada del efecto experimental, que puede modificar tanto la magnitud global de los efectos como su dirección (Sackett *et al.*, 1985; Searles, 1985; Wilson, 1982). El problema del sesgo de publicación ya se ha abordado anteriormente al tratar la etapa de la búsqueda de la literatura. La crítica es correcta, pero del mismo modo lo es para las revisiones tradicionales (Rosenthal, 1984). La única solución exige efectuar una búsqueda lo más exhaustiva po-

sible que alcance incluso al autor de los trabajos solicitando estudios no publicados sobre el tema, en base a la mayor probabilidad de que autores de estudios publicados tengan información de estudios no publicados que autores desconocidos en el área de investigación de que se trate (Bullock y Svyantek, 1985; Schmidt *et al.*, 1985; Strube *et al.*, 1985).

5. *Independencia.* Un solo estudio primario suele aportar más de una estimación del TE (o más de un nivel de significación) al meta-análisis. Ello resta independencia a los datos y provoca una impresión equivocada de la fiabilidad de los resultados. Glass y colaboradores (1981, p. 229) reconocen que ésta es la crítica más sólida de todas las formuladas contra el meta-análisis. La principal consecuencia de la falta de independencia es una disminución en la fiabilidad. Landman y Dawes (1982) discutieron cinco fuentes potenciales de dependencia, a saber: *a)* respuestas múltiples de los mismos sujetos; *b)* medidas tomadas sobre los mismos sujetos en momentos diferentes; *c)* dependencia de las puntuaciones dentro de una misma medida (p. ej., tomar como medidas distintas la puntuación en un test y la puntuación en uno de los subtests que lo componen); *d)* dependencia de los estudios dentro de un mismo artículo, y *e)* dependencia entre estudios (p. ej., dos estudios publicados en artículos diferentes utilizan los mismos sujetos).

El problema de la falta de independencia puede ser abordado desde diferentes perspectivas, aunque no hay ninguna definitiva. Una alternativa consiste en promediar las variables dependientes para obtener una estimación media (vid., p. ej., Kulik, Kulik y Cohen, 1979). Otra alternativa consiste en seleccionar a priori la «mejor» variable dependiente y obtener la estimación a partir de ella (Cooper, 1979). Una tercera solución consiste en utilizar modelos de ANOVA (Arnau, 1981, p. 137 y ss.; Walberg y Haertel, 1980). Glass y colaboradores (1981) recomiendan el uso del método *jackknife*. Pero, sin lugar a dudas, el problema de la no independencia irá

solucionándose en la medida en que la metodología meta-analítica vaya desarrollando técnicas específicas de análisis, como las propuestas por Strube (1985a) y las del Rosenthal y Rubin (1986) que describimos más arriba.

6. *Especificación del modelo.* Las inferencias en los estudios meta-analíticos se basan en el supuesto de que los datos integrados se adecúan a un determinado modelo estadístico; pero si los supuestos no se cumplen o se utilizan métodos de análisis inapropiados, las inferencias pueden ser erróneas. Las críticas esgrimidas por Searles (1985) y por Kazdin y Wilson (1978) argumentan el uso inapropiado del análisis de regresión en el meta-análisis sobre la efectividad de la psicoterapia. Pero estas críticas no han provocado un debilitamiento de la confianza del enfoque meta-analítico, sino que han desencadenado la búsqueda de modelos estadísticos específicos para el tratamiento óptimo de los datos. En la aplicación de tales modelos es obligado comprobar la plausibilidad de los mismos (Hedges y Olkin, 1985; Hunter *et al.*, 1982; Raudenbush y Bryk, 1985).

6. CONCLUSIÓN

No existe duda de que los métodos cuantitativos para integrar resultados de diferentes estudios constituyen, como hemos visto a lo largo de estas páginas, una valiosa innovación metodológica. El meta-análisis ha surgido precisamente como una perspectiva cuyo objetivo es elevar a categoría científica el proceso de revisión introduciendo el carácter sistemático y replicable propio del método científico.

Sin embargo, del mismo modo que cualquier otra metodología de la investigación, el enfoque meta-analítico no es un procedimiento acabado ausente de limitaciones. Buena parte de las críticas comentadas más arriba son erróneas o han sido mal enfocadas, pero hay otros problemas que no pueden obviarse. Es preciso distinguir tres tipos de

problemas en la realización de una revisión de investigación, sea o no cuantitativa, a saber: *a*) problemas propios de la literatura revisada, independientemente del método de revisión; *b*) problemas que se presentan en la realización de cualquier revisión, sea o no meta-analítica, y *c*) problemas propios de un estudio meta-analítico (Strube *et al.*, 1985). Un breve comentario de estas tres fuentes de discusión pueden ayudar a esclarecer las limitaciones intrínsecas de la metodología que tratamos.

En primer lugar, problemas tales como la falta de calidad de los estudios, la dependencia de los resultados o los informes incompletos son, sin duda, propios de toda literatura revisada y no de la metodología de la revisión. De hecho, el meta-análisis permite estudiar estos problemas objetivamente y aconsejar nuevas directrices para evitar tales deficiencias en la investigación que se realice en el futuro (Glass *et al.*, 1981; Rosenthal, 1984; Strube *et al.*, 1985).

Otro grupo de inconvenientes son aplicables a cualquier método de revisión. Tal es el caso del problema de la representatividad de la muestra de estudios localizados. Cualquier revisión, cuantitativa o no, ha de basarse en un conjunto de estudios representativos de la población objetivo, pero el meta-análisis tiene la ventaja de exigir del revisor una búsqueda exhaustiva del área y proporcionar métodos estadísticos para determinar su grado de fiabilidad (Orwin, 1983; Rosenthal, 1979; Strube, 1985b). El problema de la credibilidad en la interpretación de los resultados también es común a cualquier método de revisión, pero el enfoque meta-analítico aventaja a las revisiones tradicionales al introducir el análisis estadístico de los resultados (Hedges, 1984b; Strube *et al.*, 1985).

La única limitación es la que se refiere a la adecuación de los procedimientos estadísticos. La característica más peculiar del enfoque es precisamente el uso de métodos estadísticos para integrar resultados. Si los datos no cumplen con los supuestos del modelo postulado, la interpretación y las con-

clusiones derivadas del estudio meta-analítico pueden no ser adecuadas. Es preciso, por tanto, comprobar los supuestos estructurales del modelo y adoptar las máximas precauciones en la aplicación e interpretación de las técnicas estadísticas.

20

7. PERSPECTIVAS DE FUTURO

Desde nuestro punto de vista, el enfoque meta-analítico constituye hoy *una importante alternativa a la metodología de la revisión tradicional de la literatura*. El principal problema de que adolece es, y continuará siendo, el desarrollo y refinamiento de procedimientos estadísticos apropiados para el tratamiento de datos procedentes de investigaciones sobre un área común. Pero su utilidad se verá acrecentada en el futuro si las políticas editoriales y los investigadores primarios se concientizan de la importancia de la *disponibilidad pública* de los datos. Un acceso más abierto a los datos de la investigación primaria mejorará, sin duda, el empleo de las técnicas meta-analíticas: *a*) aumentando la disponibilidad de estudios primarios; *b*) evitando la ausencia de estadísticos descriptivos en los estudios originales; *c*) permitiendo al revisor mayor flexibilidad en la elección de los estadísticos, y *d*) posibilitando que el revisor utilice la misma unidad de análisis de los estudios primarios.

En consecuencia, la utilidad del enfoque se verá fortalecida en la medida en que se consigan avances en tres aspectos esenciales: 1. Rigor en la publicación de estudios primarios. 2. Sistematización a lo largo del proceso de revisión (p. ej., asegurando la representatividad de los estudios localizados y formulando «reglas de decisión» más objetivas). 3. Desarrollo de procedimientos estadísticos específicos que permitan un ajuste óptimo a los datos empíricos.

Hoy por hoy, las ventajas superan ampliamente a los inconvenientes. Y aunque resulta prematuro valorar su importancia en la acumulación del conocimiento científico, no

es en modo alguno arriesgado asegurar su necesidad en aquellos casos en que se desea obtener un conocimiento cabal acerca de un

área de investigación específica, frente al empleo de procedimientos subjetivos tradicionales.

BIBLIOGRAFÍA

- ANDERSON, D. C., y BORKOWSKI, J. G.: *Experimental Psychology: Research tactics and their applications*, Dallas, TX, Scott, Foresman & Co., 1978.
- ANGUERA, M. T., y SÁNCHEZ, P.: *Meta-análisis: conceptualización y posibilidades en la investigación experimental*, Comunicación presentada al VII Congreso Nacional de Psicología, Santiago de Compostela, abril, 1982.
- ARNAU, J.: *Psicología Experimental*, México, Trillas, 1978.
- : *Diseños experimentales en Psicología y Educación*, vol. I, México, Trillas, 1978.
- ATKINSON, D. R.; FURLONG, M. J., y WAMPOLD, B. E.: «Statistical significance, reviewer evaluations and the scientific process: Is there a (statistically significant relationship?)», *J. Couns. Psych.*, 29, 189-194, 1982.
- ATO, M.: *Validez de la investigación psicopedagógica*, Manuscrito no publicado, Universidad de Murcia, 1984.
- ATO, M.; LÓPEZ, J. A., y SERRANO, J. M.: *Fundamentos de Estadística Inferencial*, Murcia, Yerba, 1981.
- BANGERT-DROWNS, R. L.: «Review of developments in meta-analytic method», *Psych. Bull.*, 99(3), 388-399, 1986.
- BANGERT-DROWNS R. L.; KULIK, J. A., y KULIK, C. C.: «Effects of coaching programs on achievement test performance», *Rev. Educ. Res.*, 53(4), 571-585, 1983a.
- : «Synthesis of research on the effects of coaching for aptitude and admission tests», *Educ. Leadersh.*, 25, 80-82, 1983b.
- BEACHT, G., y GLASS, G. V.: «The external validity of experiments», *Am. Educ. Res. Journal*, 5, 437-474, 1968.
- BEEDERMAN, T.: «Process curricula in elementary school science», *Evaluation in Education: An International Review Series*, 4, 43-44, 1980.
- : «Effects of activity-based elementary science on student outcomes: A quantitative synthesis», *Rev. Educ. Res.*, 53(4), 499-518, 1983.
- : «Laboratory programs for elementary school science: A meta-analysis of effects on learning», *Sc. Educ.*, 69(4), 577-591, 1985.
- BULLOCK, R. J., y SVYANTEK, D. J.: «Analyzing meta-analysis: potential problems, an unsuccessful replication and evaluation criteria», *J. Appl. Psych.*, 70(1), 108-115, 1985.
- CAMPBELL, D. T., y STANLEY, J. C.: *Experimental and quasi-experimental designs for research*, Chicago, Il., Rand McNally, 1966. (Versión en castellano: Amorrortu, 1973.)
- CARROLL, R. M., y NORDHOLM, L. A.: «Sampling characteristics of Kelley's ϵ^2 and Hays' ω^2 », *Educ. Psych. Meas.*, 35, 541-544, 1975.
- CARVER, R. P.: «The case against statistical significance testing», *Harv. Educ. Rev.* 48(3), 378-399, 1978.
- COHEN, J.: «The statistical power of abnormal-social psychological research: a review», *J. Abn. Soc. Psych.*, 65, 145-153, 1962.
- : *Statistical power analysis for Behavioral Sciences*, New York, NY, Academic Press, 2 ed.: 1977.
- : «Eta-squared and partial eta-squared in fixed factor ANOVA designs», *Educ. Psych. Meas.*, 33, 107-112, 1973.
- CONGER, A. J.: «Integration and generalization of kappa for multiple raters», *Psych. Bull.*, 88, 322-328, 1980.
- COOK, T. D.: «The potential and limitations of secondary evaluations», en M. W. APPLE et al. (eds): *Educational evaluation: analysis and responsibility* (pp. 155-234), Berkeley, CA, McCutchan, 1974.
- COOK, T. D., y CAMPBELL, D. T.: *Quasiexperimentation: design and analysis issues for field settings*, Chicago, Il., Rand McNally, 1979.
- COOK, T. D., y LEVITON, L. C.: «Reviewing the literature: a comparison of traditional methods with meta-analysis», *J. Person.*, 48, 449-471, 1989.
- COOPER, H. M.: «Statistically combining independent studies a meta-analysis of sex differences in conformity research», *J. Pers. Soc. Psych.*, 37, 131-146, 1979.
- : «On the significance of effects and the effects of significances», *J. Pers. Soc. Psych.*, 41(5), 1013-1018, 1981.
- : «Scientific guidelines for conducting integrative research reviews», *Rev. Educ. Res.*, 552(2), 291-302, 1982.
- : *The integrative research review: a systematic approach*, Beverly Hills, CA, Sage, 1984.
- COOPER, H. M., y ARKIN, R. M.: «On quantitative reviewing», *J. Person.*, 49, 225-230, 1981.
- COOPER, H. M., y ROSENTHAL, R.: «Statistical versus traditional procedures for summarizing research findings», *Psych. Bull.*, 87(3), 442-449, 1980.
- COURSOL, A., y WAGNER, E. E.: «Effect of positive findings on submission and acceptance rates: a note on meta-analysis bias», *Prof. Psych. Res. & Pract.*, 17(2), 136-137, 1986.

- COWGER, C. D.: «Statistical significance tests: scientific ritualism or scientific method?», *Soc. Serv. Rev.*, 58, 358-372, 1984.
- COZBY, P. C.: *Methods in Behavioral Research*, Palo Alto, CA, Mayfield, 2d.ed., 1981.
- CRAIG, J. R.; EISON, C. L., y METZE, L. P.: «Significance tests and their interpretation: an example utilizing published research and ω^2 », *Bull. Psych. Soc.*, 7(3), 280-282, 1986.
- CROLL, P.: «Statistical inference in educational research», *Educ. Res.*, 26(3), 217-219, 1984.
- CURLETTE, W. L., y CANNELLA, K. S.: «Going beyond the narrative summarization of research findings: the meta-analysis approach», *Res. Nurs. & Health*, 8, 293-301, 1985.
- CHARTER, R. A.: «Practical formulae of strength of association measures», *Educ. Psych. Meas.*, 42, 969-974, 1982.
- CHRISTENSEN, L. B.: *Experimental methodology*, Boston, MA, Allyn & Bacon, 2 ed., 1980.
- DAWES, R. M.; LANDMAN, J., y WILLIAMS, M.: «Reply to Kurosawa», *Amer. Psych.*, 39(1), 74-75, 1984.
- DODD, D. H., y SCHULTZ, F. jr.: «Computational procedures for estimating magnitude of effects for some ANOVA designs», *Psych. Bull.*, 79(6), 391-395, 1973.
- DOMINOWSKI, R. L.: *Research Methods*. Englewood Cliffs, NJ, Prentice-Hall, 1980.
- DWYER, J. H.: «ANOVA and the magnitude of effects», *Psych. Bull.*, 81 (10), 731-737, 1974.
- EAGLY, A. H., y CARLI, L. L.: «Sex of researchers and sex-typed communications as determinants of sex differences in influenciability: a met-analysis of social influence studies», *Psych. Bull.*, 90(1), 1-20, 1981.
- EATON, W. O.: «On obtained unpublished data for research integrations», *Amer. Psych.* 39(11), 1325-1326, 1984.
- EDGINTON, E. S.: «An additive method for combining probability values from independent experiments», *J. Psych.*, 80, 351-363, 1972a.
- : «A normal curve method for combining probability values from independent experiments», *J. Psych.*, 82, 85-89, 1972b.
- EVANS, J. D.: *Invitation to Psychological Research*, New York, NY, CBS College Publications, 1985.
- EYSENCK, H. J.: «An exercise in mega-silliness», *Amer. Psych.*, 33, 517, 1978.
- : «Special review», *Beh. Res. & Ther.*, 21, 315-320, 1983.
- FAGLEY, N. S.: «Reviewer bias unresolved», *J. Couns. Dev.*, 63, 123, 1984.
- : «Applied statistical poser analysis and the interpretation of nonsignificant results by research consumers», *J. Couns. Psych.*, 32(3), 391-396, 1985.
- FAGLEY, N. S., y MCKINNEY, I.J.: «Reviewer bias for statistically significant results: a reexamination», *J. Couns. Psych.*, 30(2), 298-300, 1983.
- FELDMAN, K. A.: «Using the work of others», *Soc. Educ.*, 44, 86-102, 1971.
- FELDT, L. S.: «The use of extreme groups for the presence of a relationship», *Psychometrika*, 26, 307-316, 1981.
- FINDLEY, M., y COOPER, M. H.: «Locus of control and academic achievement: a literature review», *J. Pers. Soc. Psych.*, 44, 419-427, 1983.
- FISHER, R. A.: *Statistical methods for research workers* (4ª. ed.), London, Oliver & Boyd, 1983.
- FISKE, D. W.: «The meta-analytic revolution in outcome research», *J. Cons. Clin Psych.*, 51(1), 65-70, 1983.
- FITZ-GIBBON, C. T.: «Meta-analysis: an explanation», *Brit. Educ. Res. J.*, 37(2), 137-140, 1984.
- FLEISS, J. L.: «Estimating the magnitud of experimental effects», *Psych. Bull.*, 72(4), 273-276, 1969.
- FOWLER, R. L.: «Robust estimates and confidence intervals in measures of association», *Psych. Bull.*, 98(1), 160-165, 1985.
- FRIEDMAN, H.: «Magnitude of experimental effect and a table for its rapid estimation», *Psych. Bull.*, 70(4), 245-251, 1978.
- : «Simplified determinations of statistical power, magnitude of effect and research sample sizes», *Educ. Psych. Meas.*, 42, 521-526, 1982.
- GAEBELIN, J. W.; SODERQUIST, D. R., y POWERS, L.: «A note on variance explained in the mixed ANOVA model», *Psych. Bull.*, 83(6), 1110-1112, 1976.
- GAGE, N. L.: *The scientific basic of the art of teaching*, New York, NY, Teachers College Press, 1978.
- GALLO, P. S., jr.: «Meta-analysis: a mixed meta-phor?», *Amer. Psych.*, 33, 515-517, 1978.
- GEORGE, E. O.: *Combining independent one-sided and tow-sided statistical tests: Some theory and applications*, Doctoral dissertation, University of Rochester (cit. en Hedges y Olkin, 1985, p. 40), 1977.
- GLASS, G. V.: «Primary, secondary and meta-analysis of research», *Educ. Res.*, 5, 3-8, 1976.
- : «Integrating findings: the meta-analysis of research», *Rev. Res. Educ.*, 5, 351-371, 1977.
- : «In defense of generalization», *Beh. Br. Sciences*, 3, 394-395, 1978.
- GLASS, G. W.; MCGAW, B., y SMITH, M. L.: *Meta-analysis in social research*, Beverly Hills, CA, Sage, 1981.
- GLASS, G. V., y KLI EGL, R. M.: «An apology for research integration in the study of psychotherapy», *J. Cons. Clin. Psych.*, 51(1), 28-41, 1983.
- GLASS, G. V., y SMITH, M. L.: «Meta-analysis of the relationships between class size and achievements», *Educ. Eval. Pol. Analysis*, 1, 2-16, 1979.

- GLASS, G. V., y STANLEY, J. C.: *Statistical methods in Education and Psychology*, Englewood Cliffs, NJ: Prentice Hall, 1970. (Versión castellana: Prentice-Hall, 1974.)
- GORMAN, B. S.; PRIMAVERA, L. N., y KARRAS, A.: «A microcomputer program package for meta-analysis», *Beh. Res. Meth. Instrum.*, 15(6), 617-618, 1983.
- GREEN, B. F., y HALL, J. A.: «Quantitative methods for literature reviews», *Ann. Rev. Psych.*, 35, 37-53, 1984.
- GUSKIN, S. L.: «Problems and promises of meta-analysis in special education», *J. Spec. Educ.*, 18(1), 73-80, 1984.
- HAASE, R. F.: «Classical and partial eta-squared in multifactor ANOVA designs», *Educ. Psych. Meas.*, 43, 35-39, 1983.
- HALL, J. A.: «Gender effects in decoding nonverbal cues», *Psych. Bull.*, 85, 845-857, 1978.
- HANSFORD, B. C. y HATTIE, J. A.: «The relationship between self and achievement/performance measures», *Rev. Educ. Res.*, 52, 133-142, 1982.
- HARING, M. J.; OKUN, M. A., y STOCK, W. A.: «A quantitative synthesis of literature on work status and subjective well being», *J. Vocat. Beh.*, 25, 316-324, 1984.
- HARRIS, M. J. y ROSENTHAL, R.: «Mediation of interpersonal expectancy effects: 31 meta-analyses», *Psych. Bull.*, 97(3), 363-386, 1985.
- HATTIE, J. A., y HANSFORD, B. C.: «Meta-analysis: a reflection on problems», *Aust. J. Psych.*, 36(2), 239-254, 1984.
- HAYS, W. L.: *Statistics*, New York, NY, Holt, Rinehart & Winston, 1973.
- HEDGES, L. V.: «Distribution theory for Glass' estimator of effect size and related estimators», *J. Educ. Stat.*, 6(2), 197-228, 1981.
- : «Estimation of effect size from a series of independent experiments», *Psych. Bull.*, 92(2), 490-499, 1982a.
- : «Fitting categorical models to effect sizes from a series of experiments», *J. Educ. Stat.*, 7(2), 119-137, 1982b.
- : «Fitting continuous models to effect size data», *J. Educ. Stat.*, 7(4), 245-270, 1982c.
- : «A random effects model for effect sizes», *Psych. Bull.*, 93(2), 388-395, 1983.
- : «Estimation of effect size under nonrandom sampling; the effects of censoring studies yielding statistically insignificant means differences», *J. Educ. Stat.*, 9(1), 61-85, 1984a.
- : «Research synthesis: The state of the art», *Int. J. Ag. Hum. Dev.*, 19(2), 85-93, 1984b.
- HEDGES, L. V., y OLKIN, I.: «Vote-counting methods in research synthesis», *Psych. Bull.*, 88(2), 359-369, 1980.
- : «Clustering estimates of effects magnitude from independent studies», *Psych. Bull.*, 93(3), 563-573, 1983a.
- : «Regression models in research synthesis», *Amer. Stat.* 37(2), 137-140, 1983b.
- : «Nonparametric estimators of effect size», *Psych. Bull.* 96(3), 573-580, 1984.
- : *Statistical methods for meta-analysis*, Orlando, FL, Academic Press, 1985.
- HILLOCKS, G. Jr.: «What works in teaching composition: a meta-analysis of experimental treatment studies», *Am. J. Educ.*, 93(1), 133-170, 1984.
- HUNTER, J. E.; SCHMIDT, F. L., y JACKSON, G. B.: *Meta-analysis: comulating research findings across studies*, Beverly Hills, CA, Sage, 1982.
- HUMPHREYS, L. G.: «The statistics of failure to replicate: a comment on Buriels (1978) conclusions», *J. Educ. Psych.*, 72, 71-75, 1980.
- HYDE, J. S.: «How large are cognitive gender differences: a meta-analysis using ω^2 and d », *Amer. Psych.*, 36(8), 892-901, 1981.
- : «How large are gender differences in aggression? A developmental meta-analysis», *Dev. Psych.*, 20(4), 722-736, 1984.
- JACKSON, G. B.: «Methods for integrative review», *Rev. Educ. Res.*, 59(3), 438-460, 1980.
- JUDD, C. M., y KENNY, D. A.: *Estimating the effects of social interventions*, Cambridge, MA, Cambridge University Press, 1981.
- KATZ, B. M.; MARASCUULO, L. A., y MCSWEENEY, M.: «Nonparametric alternatives for testing main effects hypotheses: a model for combining data across independent studies», *Psych. Bull.*, 98(1), 200-208, 1985.
- KADIN, A. E.: «The role of meta-analysis in the evaluation of psychotherapy», *Clin. Psych. Rev.*, 5(1), 49-62, 1985.
- KADIN, A. E., y WILSON, G. T.: *Evaluation of behavior therapy: issues, evidence research strategies*, Cambridge, MA, Ballinger, 1978.
- KAZRIN, A.; DURAC, J., y AGTEROS, T.: «Meta-meta analysis: a new method for evaluating therapy outcome», *Beh. Res. Ther.*, 17, 397-399, 1979.
- KENDALL, M. G., y STUART, A.: *The advanced theory of statistics*, vol. II, London, Griffin, 1967.
- KENDALL, P. C., y MARUYAMA, G.: «Meta-analysis: on the road to a synthesis of knowledge?», *Clin. Psych. Rev.*, 5(1), 79-89, 1985.
- KENNEDY, J. J.: «The eta coefficient in complex ANOVA designs», *Educ. Psych. Meas.* 30, 885-889, 1970.
- KENNY, D. A.: *Correlation and causality*, New York, NY, Wiley, 1979.
- KEREN, G., y LEWIS, C.: Partial ω^2 for ANOVA designs *Educ. Psych. Meas.*, 39, 119-128, 1979.
- KERLINGER, F. N.: *Foundations of behavioral research*, New York, NY, Holt, Rinehart & Winston, 1973. (Versión castellana: Interamericana, 1975).

- : *Behavioral Research: a conceptual approach*, New York, NY: Holt, Rinehart & Winston, 1979. (Versión castellana: Interamericana, 1979).
- KRAEMER, H. C., y ANDREWS, G.: «A nonparametric technique for meta-analysis effect size calculation», *Psych. Bull.*, 91(2), 190-192, 1982.
- KRAUTH, J.: «Nonparametric effect size estimation: a comment on Kraemer & Andrews», *Psych. Bull.*, 94(1), 190-192, 1983.
- KRUSKAL, W. H.: «Ordinal measures of association», *J. Amer. Stat. Ass.*, 53, 814-864, 1958.
- KULIK, J. A.; BANGERT-DROWNS, R. L., y KULIK, C. C.: «Effectiveness of coaching for aptitude tests», *Psych. Bull.*, 95(2), 179-188, 1984.
- : «Effects of practice on aptitude and achievement test scores», *Am. Educ. Res. Journal*, 21(2), 435-447, 1984.
- KULIK, J. A.; KULIK, C. C., y COHEN, P. A.: «A meta-analysis of outcome studies of Keller's personalized system of instruction», *Am. Psych.*, 34, 307-318, 1979.
- KUROSAWA, K.: «Meta-analysis and selective publication bias», *Amb. Psych.*, 39(1), 73-74, 1984.
- LANDMAN, J. C., y DAWES, R. M.: «Psychotherapy outcome: Smith and Glass' conclusions stand up under scrutiny», *Am. Psych.*, 37(5), 504-516, 1982.
- : «Reply to Orwin and Cordray», *Am. Psych.*, 39(1), 72-73, 1984.
- LANE, D., y DUNLAP, W.: «Estimating effect sizes: bias resulting from the significance criterion in editorial decisions», *Br. J. Math. Stat. Psych.*, 31, 107-112, 1978.
- LEVITON, L. C., y COOK, T. D.: «What differentiates meta-analysis from other forms of review», *J. Pers.*, 49(2), 231-236, 1981.
- LIGHT, R. J.: «Capitalizing on variation: how conflicting research findings can be helpful for policy», *Educ. Researcher*, 8(9), 7-11, 1979.
- LIGHT, R. J., y PILLEMER, D. B.: «Numbers and narrative: combining their strengths in research reviews», *Harv. Educ. Rev.*, 52, 1-26, 1982.
- : *Summing up: the science of reviewing research*, Cambridge, MA, Harvard University Press, 1984.
- LIGHT, R. J., y SMITH, P. V.: «Accumulating evidence: procedures for resolving contradictions among different research studies», *Harv. Educ. Rev.*, 41(4), 429-471, 1971.
- LÓPEZ, J. A.; SÁNCHEZ, J., y LÓPEZ, J. J.: «METAAN-1: un programa BASIC de meta-análisis para el ajuste de modelos categóricos a los tamaños del efecto», *Psicológica*, en prensa.
- : «METAAN-2: un programa BASIC para la aplicación del análisis de regresión múltiple ponderado en meta-análisis», *Psicológica*, en prensa.
- LYKKEN, D. T.: «Statistical significance in psychological research», *Psych. Bull.*, 70(3), 151-159, 1968.
- MAGNUSSON, D.: *Test theory*, Reading, MA, Addison-Wesley, 1966. (Versión castellana: Trillas, 1977).
- MAHONEY, M. J.: «Publication prejudices: an experimental study of confirmatory bias in the peer review system», *Cogn. Ther. Res.*, 1(2), 161-175, 1977.
- MANSFIELD, R. S., y BUSSEY, T. V.: «Meta-analysis in research: a rejoinder to Glass», *Educ. Researcher*, 6, 3, 1977.
- MAXWELL, S. E.; CAMP, C. J., y ARVEY, R. D.: «Measures of strength of association: a comparative examination», *J. Appl. Psych.* 66(3), 1981.
- MCDANIEL, M. A.: «Computer programs for calculating meta-analysis statistics», *Ed. Psych. Meas.*, 46, 175-177, 1986.
- MCGAW, B., y GLASS, G. V.: «Choice of the metric for effect size in meta-analysis», *Am. Educ. Res. Journal*, 17(3), 325-337, 1980.
- MCGUIGAN, F. J.: *Experimental psychology: a methodological approach*, Englewood Cliffs, NJ, Prentice Hall, 3d ed., 1978.
- MCNEMAR, Q.: «At random: sense and nonsense», *Am. Psych.*, 15, 295-300, 1960.
- MILLER, R. C., y BERMAN, J. S.: «The efficacy of cognitive behavior therapies: A quantitative review of the research evidence», *Psych. Bull.*, 94(1), 39-53, 1983.
- MOSTELLER, F., y BUSH, R. R.: «Selected quantitative techniques», en G. Lindzey (Ed.): *Handbook of Social Psychology: Theory and method*, vol I, Cambridge, MA, Addison-Wesley, 289-334, 1954.
- MOSTELLER, F. M., y TUKEY, J. W.: «Data analysis, including statistics», en G. LINDZEY & E. ARONSON (eds.), *Handbook of Social Psychology*, vol. II, Reading, MA, Addison-Wesley, 89-203, 1968.
- MULLEN, B., y ROSENTHAL, R.: *Basic Meta-analysis*, Hillsdale, NJ, Erlbaum, 1985.
- NICHOLSON, R. A., y BERMAN, J. S.: «Is follow-up necessary in evaluating psychotherapy?», *Psych. Bull.*, 93(2), 261-278, 1983.
- O'GRADY, K. E.: «Measures of explained variance: cautions and limitations», *Psych. Bull.*, 92(3), 766-777, 1982.
- ORTEGA, J. E., y FERNÁNDEZ DOLS, J. M.: *Fuentes documentales en Psicología*. Madrid, Debate, 1980.
- ORWIN, R. G.: «A fail-safe N for effect size in meta-analysis», *J. Educ. Stat.*, 8(2), 157-159, 1983.
- ORWIN, R. G., y CORDRAY, D. S.: «Smith & Glass' psychotherapy conclusions need further probing: on Landman & Dawes' reanalysis», *Am. Psych.*, 39(1), 71-72, 1984.

- : «Effects of deficient reporting on meta-analysis: a conceptual framework and reanalysis», *Psych. Bull.*, 97(1), 134-147, 1985.
- PEARSON, K.: «On a method of determining whether a samples of given size n supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random», *Biometrika*, 25, 379-410, 1933.
- PEECE, P. F. W.: «A measure of experimental effect size based on success rate», *Educ. Psych. Meas.*, 43, 763-766, 1983.
- PILLEMER, D. B.: «Conceptual issues in research synthesis», *J. Spec. Educ.*, 18(1), 27-40, 1984.
- PILLEMER, D. B., y LIGHT, R. J.: «Benefiting from variation in study outcomes», en R. Rosenthal (Ed.): *New directions for methodology of social and behavioral sciences*, vol. V, pp. 1-12, San Francisco, CA, Jossey-Bass, 1980a.
- : «Synthesizing outcomes: how to use research evidence from many studies», *Harv. Educ. Review*, 50(2), 176-195, 1980b.
- PLUTCHIK, R.: *Foundations of Experimental Research*, Cambridge, MA, Harper & Row, 3ª ed., 1983.
- POSAVAC, E. J., y SINACORE, J. M.: «Improving the understanding of statistical significance», *Knowledge: creation, diffusion & utilization*, 5(4), 503-508, 1984.
- PRESBY, S.: «Overly broad categories obscure important differences between therapies», *Am. Psych.*, 33, 514-515, 1978.
- RACHMAN, S. J., y WILSON, G. T.: *The effects of psychological therapy*, New York, NY, Pergamon Press, 2ª ed., 1980.
- RAUDENBUSH, S. W.: «Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy induction: a synthesis of findings from 18 experiments», *J. Educ. Psych.*, 76(1), 85-97, 1985.
- RAUDENBUSH, S. W., y BRYK, A. S.: «Empirical Bayes meta-analysis», *J. Educ. Stat.*, 10(2), 75-98, 1985.
- RONIS, D. L.: «Comparing the magnitude of effects in Anova designs», *Educ. Psych. Meas.*, 41, 993-1000, 1981.
- ROSENTHAL, R.: *Experimenter effects in behavioral research* (enlarged edition), New York, NY, Irvington, 1976.
- : «Combining results of independent studies», *Psych. Bull.*, 85(1), 185-193, 1978.
- : «The "file drawer problem" and tolerance for null results», *Psych. Bull.*, 86(3), 638-641, 1979.
- (ed.): *New directions for methodology of social and behavioral science*, vol V, San Francisco. CA, Jossey-Bass, 1980.
- : «Valid interpretation of quantitative research results», en D. BRINBERG y L. KIDDER (Eds.): *New directions for methodology of social and behavioral science: forms of validity in research*, 12, pp. 59-75, San Francisco, CA, Jossey-Bass, 1982.
- : «Assessing the statistical and social importance of the effects of psychotherapy», *J. Cons. Clin. Psych.*, 51(1), 4-13, 1983a.
- : «Improving meta-analytic procedures for assessing the effects of psychotherapy versus placebo», *Beh. Br. Sciences*, 6, 298-299, 1983b.
- : «Meta-analysis: toward a more cumulative social science», en L. Bickman (Ed.): *Applied social psychology annual*, vol. IV, pp. 65-93, Beverly Hills, CA, Sage, 1983c.
- : *Meta-analytic procedures for social research*, Beberly Hills, CA, Sage, 1984.
- ROSENTHAL, R., y ROSNOW, R. L.: *Essentials of behavioral research: methods and data analysis*, New York, NY, McGraw Hill, 1984.
- ROSENTHAL, R., y RUBIN, D. B.: «Interpersonal expectancy effects: the first 345 studies», *Beh. Br. Sciences*, 3, 377-415, 1978.
- : «A note on percent variance explained as a measure of the importance of effects», *J. Appl. Soc. Psych.*, 9(5), 395-396, 1979a.
- : «Comparing significance levels of independent studies», *Psych. Bull.*, 86(5), 1165-1168, 1979b.
- : «A simple, general purpose display of magnitude of experimental effect», *J. Educ. Psych.*, 74(2), 166-169, 1982a.
- : «Comparing effect sizes of independent studies», *Psych. Bull.*, 92(2), 500-504, 1982b.
- : «Further meta-analytic procedures for assessing cognitive gender differences», *J. Educ. Psych.*, 74(5), 708-712, 1982c.
- : «Meta-analytic procedures for combining studies with multiple effect sizes», *Psych. Bull.*, 99(3), 400-406, 1986.
- ROSSI, J. S.: «Tables of effect size for z score tests of differences between proportions and between correlation coefficients», *Educ. Psych. Meas.*, 45, 737-742, 1985.
- SACKETT, P. R.; TENOPYR, M. L.; SCHMITT, N., y KEHOE, J.: «Comentary on forty questions about validity generalization and meta-analysis», *Pers. Psych.*, 38(4), 697-798, 1985.
- SÁNCHEZ, J.: *Etapas iniciales del procesamiento de información visual en BL y ML: un estudio meta-analítico*. Comunicación presentada al «Symposium sobre Actividad Humana y Procesos Cognitivos», Madrid, 1984.
- : *Memoria icónica e informe parcial: un estudio meta-analítico*. Tesis doctoral no publicada. Universidad de Murcia, 1985.
- : «La revisión cuantitativa: una alternativa a las revisiones tradicionales», *Anales de psicología* (en prensa), 1986.

- SCHMIDT, F. L.; PEARLMAN, K.; HUNTER, J. E., y HIRSH, H. R.: «Forty questions about validity generalization and meta-analysis», *Pers. Psych.*, 38(4), 697-798, 1985.
- SEARLES, J. S.: «A methodological and empirical critique of psychotherapy outcome meta-analysis», *BehBr. Sciences*, 22(4), 453-463, 1985.
- SERLIN, R. C.: «A multivariate measure of association based on the Pillai-Barlett procedure», *Psych. Bull.*, 91(2), 413-417, 1982.
- SERLIN, R. C.; CARR, J., y MARASCUILO, L. A.: «A measure of association for selected nonparametrics procedures», *Psych. Bull.*, 92(3), 786-790, 1982.
- SHAPIRO, D. A., y SHAPIRO, D.: «Meta-analysis of comparative therapy outcome research: a critical appraisal», *Beh. Psychother.*, 10, 4-25, 1982a.
- : «Meta-analysis of comparative therapy outcome studies: a replication and refinement», *Psych. Bull.*, 92(3), 581-604, 1982b.
- : «Comparative therapy outcome research: Methodological implications of meta-analysis», *J. Cons. Clin. Psych.*, 51(1), 42-53, 1983.
- SINDELAR, P. T., y WILSON, R. J.: «The potential effects of meta-analysis on special education practice», *J. Spec. Educ.*, 18(1), 81-92, 1984.
- SMITH, M.L.: «Publication bias and meta-analysis», *Eval. Educ.*, 4, 22-24, 1980.
- : «Research integration», en H. MITZEL (Ed.): *Enciclopedia of Educational Research*, vol. III, pp. 1611-1618, London, The Free Press, 1982.
- SMITH, M.L., y GLASS, G. V.: «Meta-analysis of psychotherapy outcome studies», *Amer. Psych.*, 32, 752-760, 1977.
- SMITH, M. L.; GLASS, G. V., y MILLER, T. I.: *The benefits of psychotherapy*, Baltimore, MD, John Hopkins University Press, 1980.
- SMITH, P. L.: «Measures of variance accounted for: theory and practice», en G. KEREN (Ed.): *Statistical and methodological issues in behavioral sciences research*, pp. 101-130, Hillsdale, NJ: Erlbaum, 1982.
- STEINBRUECK, S. M.; MAXWELL, S. E., y HOWARD, G. S.: «A meta-analysis of psychotherapy and drug therapy in the treatment of unipolar depression with adults», *J. Cons. Clin. Psych.*, 51(6), 856-863, 1983.
- STOCK, W. A.; OKUN, M. A.; HARING, M. J.; MILLER, W.; KINNEY, C., y CEURVORST, R. W.: «Rigor in data synthesis: a case study of realivility in meta-analysis», *Educ. Researcher*, 11, 10-14, 1982.
- STOCK, W.A.; OKUN, M. A.; HARING, M. J., y WITTER, R. A.: «Race and subjective well-being in adulthood: a black-white research synthesis», *Hum. Dev.*, 28, 192-197, 1985.
- STOUFFER, S. A.; SUCHMAN, E. A.; DE VINNEY, L. C.; STAR, S. A., y WILLIAMS, R. M. jr.: *The American soldier*, vol. I, Princeton, NJ, Princeton University Press, 1949.
- STRAHAN, R. F.: «Assessing magnitude of effects from rank-order correlation coefficient», *Educ. Psych. Meas.*, 42, 763-765, 1982.
- STRUBE, M. J.: «Combining and comparing significance levels from nonindependent hypothesis tests», *Psych. Bull.*, 97(2), 334-341, 1985a.
- : «Power analysis from combining significance levels», *Psych. Bull.*, 98(3), 595-599, 1985b.
- STRUBE, M. J.; GARDNER, W., y HARTMANN, D. P.: «Limitations, liabilities and obstacles in reviews of the literature: the current status of meta-analysis», *Clin. Psych. Review*, 5(1), 63-78, 1985.
- STRUBE, M. J., y HARTMANN, D. P.: «A critical appraisal of meta-analysis», *Clin. Psych.*, 21, 129-139, 1982.
- : «Meta-analysis: techniques, applications and functions», *J. Cons. Clin. Psych.*, 51(1), 14-27, 1983.
- STRUBE, M. J., y MILLER, R. H.: «Comparison of power rates for combined probability procedures: a simulation study», *Psych. Bull.*, 99(3), 407-415, 1986.
- THOMAS, J. R., y FRENCH, K. E.: «Gender differences across age in motor performance: a meta-analysis», *Psych. Bull.*, 98(2), 260-282, 1985.
- TIPPETT, L. H.: *The method of statistics*, London, Williams & Norgate, 1931.
- TUKEY J. W.: *Exploratory data analysis*, Reading, MA, Addison-Wesley, 1977.
- UEBERSAX, J. S.: «A generalized kappa coefficient», *Educ. Psych. Meas.*, 42, 181-183, 1982.
- VAUGHAN, G. M., y CORBALLIS, M. C.: «Beyond tests of significance: estimating strength of effects in selected ANOVA designs», *Psych. Bull.*, 72(3), 204-213, 1969.
- WALBERG, H. J., y HAERTEL, E. H.: «Research integration: the state of the art», *Evaluation in Education*, 4(1), completo, 1980.
- WALKER, L. J.: «Sex differences in the development of moral reasoning: a critical review», *Ch. Developm.*, 55, 677-691, 1984.
- WANG, M. D.: «Estimation of ω^2 for a one-way fixed effects when sample size is disproportionate», *Educ. Psych. Meas.*, 42, 167-179, 1982.
- WEBB, E; CAMPBELL, D.; SCHWARTZ, R.; SECHREST, L., y GROVE, J.: *Nonreactive measures in the social sciences*, 2ª ed., Boston, MA, Houghton Mifflin, 1981.
- WHITE, K. R.: «Socio-economic status and academic achievement», *Evaluation in education*, 4, 79-81, 1980.

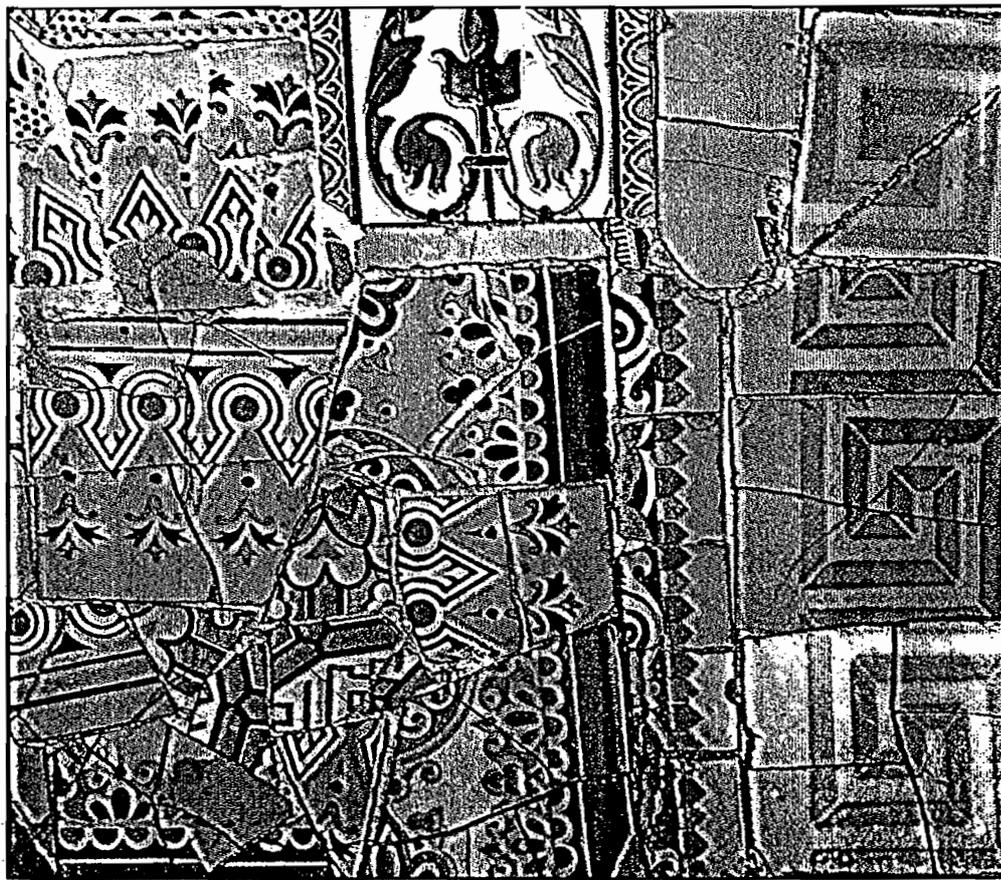
- WHITLEY, B. E. jr.: «Sex-role orientation and psychological well-being: two meta-analyses», *Sex roles*, 12(1/2), 207-225, 1985.
- WHITLEY, B. E. jr., y FRIEZE, I. H.: «Children's causal attributions for success and failure in achievement settings: a metaanalysis», *J. Educ. Psych.*, 77(5), 608-616, 1985.
- WILSON, G. T.: «How useful is meta-analysis in evaluating the effects of different psychological therapies», *Beh. Psychoth.*, 10, 221-231, 1982.
- : «Limitation of meta-analysis in the evaluation of the effects of psychological therapy», *Clin. Psych. Review*, 5(1), 35-48, 1985.
- WILSON, G. T., y RACHMAN, S. J.: «Meta-analysis and the evaluation of psychotherapy outcome: limitations and liabilities», *J. Cons. Clin. Psych.*, 51(1), 54-64, 1983.
- WILLSON, V. L.: «Critical values of the rank-biserial correlation coefficient», *Educ. Psych. Meas.*, 36, 297-300, 1976.
- : *Two statistical foundations for meta-analysis*. Paper presented in Annual Meeting of the American Educational Research Association, 1981.
- : «A meta-analysis of the relationship between science achievement and science attitude: Kindergarten through college», *J. Res. Sc. Teaching*, 20(9), 839-850, 1983.
- : «Adding results to a meta-analysis: theory and example», *J. Res. Sc. Teaching*, 21(6), 649-659, 1984.
- WINER, B. J.: *Statistical principles in Experimental Design*, New York, NY, McGraw-Hill, 2ª ed, 1971.
- WORTMAN, P. M.: «Evaluation research: a methodological perspective», *Ann. Rev. Psych.*, 34, 223-260, 1983.
- WORTMAN, P. M., y BRYANT, F. B.: «School desegregation and black achievement: an integrative review», *Soc. Meth. Research*, 13, 289-324, 1985.
- YEATON, W. H., y SECHREST, L.: «Meaningful measures of effect», *J. Cons. Clin. Psych.*, 49(5), 766-767, 1981.
- YEATON, W. H., y WORTMAN, P. M. (eds.): *Issues in data synthesis*, San Francisco, CA, Jossey-Bass, 1984.

TRATADO DE PSICOLOGÍA GENERAL

Juan Mayor - José Luis Pinillos

HISTORIA, TEORÍA Y MÉTODO

Jaume Arnau y Heliodoro Carpintero



Alhambra  Universidad

4.4

TRATADO DE PSICOLOGÍA GENERAL 1

Juan Mayor y José Luis Pinillos

A (091)

HIS

his

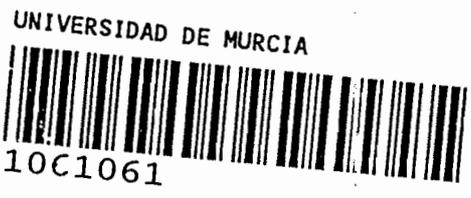
TRATADO DE PSICOLOGÍA GENERAL

1

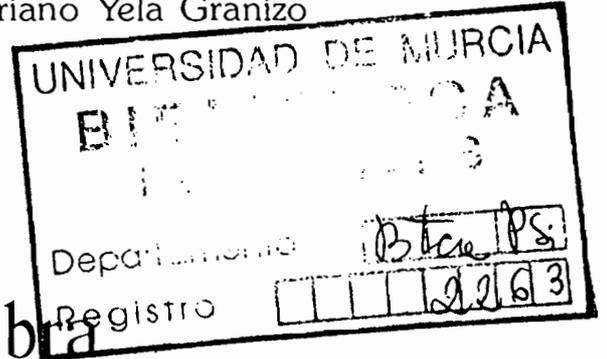
Juan Mayor y José Luis Pinillos

HISTORIA TEORÍA Y MÉTODO

Coordinadores: Jaume Arnau y Heliodoro Carpintero



- Blanca Anguera Domingo
- Jaume Arnau Gras • Manuel Ato García
- Antonio Caparrós Beneito • Enrique J. Carbonell Vayá
- Heliodoro Carpintero • Alejandra Ferrándiz Lloret
- Juan B. Fuentes Ortega • Fernando Gabuccio Cerezo
- Julia García Sevilla • Luis García Vega • Carmen Giménez Segura
- Jose María Gondra Rezola • Enrique Lafuente Niño
- Juan Mayor Sánchez • Vicenta Mestre Escrivá
- Luis Montoro González • Rafael Moreno Rodríguez
- José Moya Santoyo • M.ª José Pedraja Linares
- Esteban Pérez Delgado • Jesús Pérez Ríos
- Elena Quiñones Vidal • Julio Sánchez Meca
- Francisco M. Tortosa Gil • Mariano Yela Granizo



Primera edición, 1989

© EDITORIAL ALHAMBRA, S.A., para la presente edición
R.E. 182

España:
28001 Madrid. Claudio Coello, 76

México:
03100 México, D. F.
Editorial Alhambra Mexicana, S. A. de C. V.
Amores, 2027. Colonia del Valle

Delegaciones en España:
08008 Barcelona. Enrique Granados, 46
48014 Bilbao. Iruña, 12
15005 La Coruña. Pasadizo de Pernas, 13
18009 Granada. Pza. de las Descalzas, 2
28002 Madrid. Saturnino Calleja, 1
35003 Las Palmas. Tomás Morales, 48
38004 Santa Cruz de Tenerife. General Porlier, 14
41012 Sevilla. Reina Mercedes, 35
46003 Valencia. Cabillers, 5
47013 Valladolid. Julio Ruiz de Alda, 10
50005 Zaragoza. Concepción Arenal, 25

Representante:
07010 Palma de Mallorca. D. Francisco Molina
Francisco Suau, 14

Distribuidor exclusivo en Argentina:

Distribuidora Yenny
Avda. Rivadavia, 3860
1.204 Buenos Aires

nc 18010442

ISBN 84-205-1882-4 (obra completa)
ISBN 84-205-1789-5 (vol. I)

Depósito legal: M 7884-1989

© Es propiedad de los autores.

Reservados todos los derechos. Ni la totalidad, ni parte de esta publicación pueden reproducirse, registrarse o transmitirse, por un sistema de recuperación de información, en ninguna forma ni por ningún medio, sea electrónico, mecánico, fotoquímico, magnético o electroóptico, por fotocopia, grabación o cualquier otro, sin permiso previo por escrito del editor. El préstamo, alquiler o cualquier otra forma de cesión de uso de este ejemplar requerirá también la autorización del editor o de sus representantes.

Fotocomposición: Estudio Enlace, S.A.
Papel: Kanguros
Encuadernación: Gómez Pinto, S.A.

Impreso en España - Printed in Spain

Lavel, S.A. - Los Llanos, nave 6- Humanes (Madrid)

ÍNDICE GENERAL

INTRODUCCIÓN

1	¿Psicología o psicologías? Un problema de identidad, por JUAN MAYOR SÁNCHEZ y JESÚS PÉREZ RÍOS	3
	1. Aproximaciones convencionales al concepto de psicología	3
	1.1. Perspectiva histórica	4
	1.2. Perspectiva sistemática	6
	1.3. Perspectiva pragmática	11
	2. De las tendencias centrífugas a la radical diversidad de la psicología ..	13
	2.1. Diversidad disciplinar (ámbito interdisciplinar)	13
	2.1.1. Las «psicologías filosóficas»	14
	2.1.2. Las «psicologías científicas»	15
	2.1.3. Diversidad de especialidades de la psicología científica ...	15
	2.2. Diversidad paradigmática (ámbito intradisciplinar)	21
	2.2.1. Diversidad de escuelas	21
	2.2.2. Diversidad de orientaciones o paradigmas	24
	3. De las tendencias centrípetas a la unificación	26
	3.1. La tentación del reduccionismo radical	26
	3.2. Tentativas de reducción dicotómica	28
	3.3. Propuestas de integración	33
	3.3.1. De la yuxtaposición a la síntesis	33
	3.3.2. La integración en una concepción, teoría o sistema general	36
	4. La articulación de la diversidad y la unidad	41
	4.1. El equilibrio inestable entre las tendencias centrífugas y las centrípetas	41
	4.2. Hacia una organización sistémica inter e intradisciplinar	47
	4.2.1. Organización interdisciplinar	47
	4.2.2. Organización intradisciplinar	48
	5. Conclusión	60
	Bibliografía	60
2	Unidad y diversidad de la psicología, por MARIANO YELA GRANIZO	71
	1. La diversidad de la psicología	71
	1.1. La situación actual	71
	1.2. Las razones de la diversidad	74
	1.3. El dualismo básico	74
	2. La unidad de la psicología. El desarrollo histórico	75

2.1. La posible unidad	76
2.2. El significado de la conducta	77
2.3. La verificación de la experiencia privada	78
2.4. El desarrollo convergente de la historia de la psicología	80
3. Hacia una ciencia psicológica unificada	85
Bibliografía	88

HISTORIA

3 La psicología filosófica. Desde Aristóteles hasta la época de Wundt, por ESTEBAN PÉREZ DELGADO	95
1. La psicología sustancialista: desde Aristóteles a Descartes	100
— 1.1. La psicología de Aristóteles	100
1.1.1. La psicología como ciencia del ser vivo	101
1.1.2. El método de la psicología aristotélica	102
1.1.3. Algunos temas más relevantes de la psicología de Aristóteles	103
1.1.3.1. Procesos de cognición	103
1.1.3.2. Procesos apetitivos o motivacionales	104
1.1.3.3. Aprendizaje y formación de la personalidad	105
1.2. La psicología en la Edad Media	106
1.2.1. La psicología de san Agustín	106
1.2.2. La psicología de santo Tomás de Aquino	107
1.2.2.1. El compuesto sustancial de alma y cuerpo	108
1.2.2.2. Los procesos cognitivos	108
1.2.2.3. Los procesos apetitivos	109
2. La psicología de Descartes (1596-1650)	110
2.1. La psicología cartesiana	111
2.1.1. El <i>cogito</i> y la fundamentación de una ciencia universal	111
2.1.2. La mente como sustancia pensante	112
2.1.3. La unión de alma y cuerpo	112
3. Empirismo y asociacionismo	114
3.1. Los inicios del empirismo inglés: J. Locke (1632-1704)	114
3.1.1. La psicología del ensayo acerca del entendimiento humano	115
3.2. El asociacionismo de Hume (1711-1776)	116
3.2.1. El origen de las ideas	117
3.2.2. Las pasiones	118
3.3. El asociacionismo psicológico	119
3.3.1. El asociacionismo de David Hartley	119
4. La psicología de las facultades en el siglo XVIII	121
4.1. La psicología de Thomas Reid y la escuela escocesa	121
4.2. La aportación de M. Kant a la psicología	122
5. La psicología en los filósofos de la primera mitad del siglo XIX	124
5.1. La psicología alemana de principios del siglo XIX	124
5.1.1. La psicología como ciencia fundamentada en la experiencia, en la metafísica y en la matemática de Herbart	124
6. Desarrollo de la filosofía y nacimiento de la psicofísica	127
6.1. Desarrollo de la fisiología	127
6.1.1. La aportación de la fisiología de Johannes Müller a la psicología	128
6.2. El nacimiento de la psicofísica	129
Bibliografía	131

4	Estructuralismo y funcionalismo, por FRANCISCO M. TORTOSA GIL	133
	1. Introducción	133
	2. Estructuralismo	134
	2.1. Wilhelm Max Wundt	134
	2.2. El desarrollo de la psicología experimental alemana	139
	2.3. El estructuralismo: E. B. Titchener (1867-1927)	143
	3. El funcionalismo	147
	3.1. El funcionalismo norteamericano: primeros desarrollos	148
	3.1.1. G. T. Ladd (1842-1921)	148
	3.1.2. W. James (1842-1910)	149
	3.1.3. G. S. Hall (1844-1924)	151
	3.1.4. J. McKeen Cattell (1860-1944)	152
	3.1.5. J. M. Baldwin (1861-1934)	153
	3.2. El funcionalismo americano	154
	3.2.1. J. Dewey (1859-1952)	154
	3.2.2. J. R. Angell (1869-1949)	155
	3.2.3. H. Carr (1873-1954)	157
	3.2.4. R. S. Woodworth (1869-1962)	159
	Bibliografía	160
5	Las escuelas de psicología profunda, por ALEJANDRA FERRÁNDIZ LLORET	167
	1. Introducción	167
	2. Sigmund Freud: la fundación del psicoanálisis	168
	2.1. Biografía profesional	168
	2.2. Comienzos del psicoanálisis: establecimiento de los principios	172
	2.3. Definición del psicoanálisis	174
	2.3.1. La interpretación de los sueños	175
	2.3.2. El aparato psíquico: la teoría topográfica	176
	2.3.3. Los instintos	179
	2.3.4. El desarrollo psicosexual	181
	2.3.5. De la perspectiva topográfica a la estructural	183
	2.4. La difusión del psicoanálisis	186
	2.5. Críticas al psicoanálisis	187
	3. Carl Gustav Jung: la psicología analítica	188
	3.1. Biografía profesional	188
	3.2. La teoría de la libido	190
	3.3. La estructura de la psique	191
	3.3.1. Las funciones básicas de la psique	193
	3.3.2. Los tipos de actitud	193
	3.4. La psicoterapia analítica	194
	3.5. La suerte del jungismo	195
	4. Alfred Adler: la psicología individual	196
	4.1. Biografía profesional	196
	4.2. El sentimiento de inferioridad	197
	4.3. El sentimiento social	199
	4.4. Etiología de la neurosis y psicoterapia	199
	4.5. Repercusiones de su obra	200
	Bibliografía	201
6	Las psicologías objetivas: reflexología, conductismo, por JOSÉ MARÍA GONDRA	205
	1. La reflexología rusa	206
	1.1. Orígenes: los reflejos cerebrales	206
	1.2. Los reflejos condicionados	209

1.3. El método de Pavlov	211
1.4. La teoría fisiológica de Pavlov	212
1.5. Psicopatología y personalidad	213
2. El conductismo	215
2.1. Antecedentes intelectuales	215
2.2. La gestación del conductismo	216
2.3. La evolución del conductismo	218
2.4. La psicología conductista	219
2.5. Los sistemas de hábitos	220
2.6. Conclusión	222
Bibliografía	223
7 La psicología soviética, por VICENTA MESTRE ESCRIVÁ	225
1. L. S. Vygotski (1896-1934)	226
1.1. El estudio de los procesos superiores	227
2. Alexander Romanovich Luria (1902-1984)	228
2.1. Datos biográficos	229
2.2. Su concepto de psicología	231
2.3. Visión evolucionista de la psicología. Importancia de los procesos psíquicos superiores	232
Bibliografía	234
8 La psicología de la Gestalt, por ANTONIO CAPARRÓS BENEDICTO, FERNAN- DO GABUCCIO CEREZO, BLANCA ANGUERA DOMENGO y M.^a CARMEN GIMÉNEZ SEGURA	235
1. En la tradición de Brentano	235
2. Surgimiento de la psicología de la Gestalt	237
3. Principales representantes y aportaciones	241
4. Otros desarrollos en la tradición gestaltista y fenomenológica	246
Bibliografía	250
9 Los neoconductismos, por JUAN B. FUENTES ORTEGA y ENRIQUE LAFUENTE NIÑO	251
1. Introducción. Los neoconductismos: tradición e innovación dentro del conductismo	251
2. El neoconductismo metodológico	252
2.1. El propósito común	252
2.2. El conductismo cognitivo y propositivo de Tolman	254
2.3. El conductismo formal de Hull	256
2.4. El conductismo asociativo de Guthrie	259
3. El neoconductismo radical skinneriano	261
3.1. El lugar de Skinner en el marco neoconductista: el carácter «radical» del conductismo radical	261
3.1.1. La concepción skinneriana del «método» como crítica de la concepción conductista metodológica de la «metodología» ..	262
3.1.2. La crítica skinneriana de las teorías como crítica de la con- cepción conductista metodológica de las teorías	263
3.1.3. La perspectiva pragmática radical de Skinner	263
3.2. El condicionamiento operante y el análisis experimental de la con- ducta	264
3.2.1. El condicionamiento operante	264
3.2.1.1. Principios básicos del condicionamiento operante	267
3.2.2. El análisis experimental de la conducta	272
3.2.2.1. El dato básico como construcción que permite una lectura directa	273

	3.2.2.2. El estudio del organismo individual y el rechazo de la estadística	273
	3.2.2.3. Carácter molar de la conducta y distal de los estímulos en el análisis de la conducta	274
	3.3. Las interpretaciones teóricas en la obra de Skinner	274
	3.4. Desarrollo institucional del análisis de la conducta	277
	Bibliografía	278
10	Psicología contemporánea, por HELIODORO CARPINTERO	281
	1. Introducción	281
	1.1. Gran desarrollo institucional	281
	1.2. Implantación de las técnicas aplicadas	282
	1.3. Pluralismo teórico	282
	1.4. Insuficiencias en el rol profesional del psicólogo	283
	2. Tendencias recientes	283
	3. El conductismo reciente	284
	3.1. La aportación de Skinner	284
	3.2. Psicología de la modificación de conducta	286
	4. Los planteamientos funcionalistas estructurales	287
	5. «Conductismo subjetivista»	290
	5.1. Procesamiento de información	292
	6. Piaget y su influencia	293
	7. Psicología del aprendizaje	296
	7.1. Hull y su tradición	296
	7.2. Desarrollos en torno al condicionamiento clásico	297
	7.3. Aprendizaje observacional o vicario. La teoría del aprendizaje social	298
	8. Personalidad: situación o rasgos	300
	9. Psicología cognitiva	302
	Bibliografía	308
11	Tabla cronológica y cuadro sinóptico de la historia de la psicología científica, por LUIS GARCÍA VEGA y JOSÉ MOYA SANTOYO	313
12	La psicología en España: una síntesis, por HELIODORO CARPINTERO	329
	1. Introducción	329
	2. Antecedentes de la psicología científica	329
	3. La introducción de la psicología científica en España (1880-1900)	332
	4. Los krausistas y la psicología	334
	5. La psicología en el marco de la Restauración	335
	6. Hacia una mentalidad positiva	338
	7. Pasos hacia el siglo XX	341
	8. Los comienzos del siglo XX	341
	9. La generación de 1886 y la psicología	342
	10. Los antecedentes de una profesión	345
	11. Mira y Germain, pioneros de la psicología científica en España	346
	Bibliografía	348
13	La psicología en Iberoamérica, por RUBÉN ARDILA ARDILA	353
	1. América y su psicología	353
	2. Comienzos	354
	3. Los pioneros	355
	4. La psicología en la universidad	357
	5. Características	360
	6. Enfoques de la psicología	361

	7. Asociaciones profesionales	362
	8. Publicaciones	364
	9. Congresos	366
	10. Investigación científica	367
	11. El futuro	370
	Bibliografía	370
14	El uso de instrumentos en la investigación psicológica , por ELENA QUIÑONES VIDAL, JULIA GARCÍA SEVILLA y M ^a JOSÉ PEDRAJA LINARES... Bibliografía	 373 390
15	La comunicación científica en psicología: producción y disseminación de la información , por LUIS MONTORO GONZÁLEZ y ENRIQUE J. CARBONELL VAYÁ	391
	1. Introducción	391
	2. Los congresos como cauce de comunicación científica	392
	3. Las revistas como canal de comunicación científica	398
	3.1. Dimensiones de una revista	399
	3.2. La interacción entre revistas	399
	3.3. Las revistas sobre psicología	401
	4. La disseminación de la información	404
	4.1. Revisiones (<i>reviews</i>)	404
	4.2. Resúmenes (<i>abstracts</i>)	407
	4.3. Índices y repertorios	409
	4.4. Centros de documentación	413
	Bibliografía	415

TEORÍA Y MÉTODO

16	El método científico en psicología , por JUAN MAYOR SÁNCHEZ	419
	1. Introducción	419
	2. El método científico	420
	2.1. Diversidad de métodos y psicología	420
	2.2. El fundamento y el desarrollo de la ciencia y del método científico	422
	2.2.1. La posición heredada	423
	2.2.2. La ciencia en sus contextos	426
	2.2.3. Hacia una nueva epistemología realista	429
	3. El método científico positivo en psicología	432
	3.1. La articulación de las fases del método en psicología	433
	3.1.1. Momento empírico	436
	3.1.2. Momento teórico	440
	3.2. Modalidades y técnicas metodológicas	448
	3.2.1. El «método» experimental	451
	3.2.2. Otros «métodos» no experimentales	466
	4. Alternativas metodológicas a la psicología como ciencia positiva	472
	4.1. El método fenomenológico	474
	4.2. El método hermenéutico	476
	4.2.1. La <i>Verstehen</i> de la psicología comprensiva	477
	4.2.2. La interpretación psicoanalítica	478
	4.2.3. La hermenéutica en sentido estricto	480
	4.3. Método estructuralista	482
	4.4. El método dialéctico	484

5.	El método de la psicología entre el reduccionismo y el pluralismo metodológico	486
5.1.	La psicología como ciencia positiva	486
5.2.	¿Método científico positivo o pluralismo metodológico?	489
5.3.	Pluralismo metodológico de la psicología	491
	Bibliografía	493
17	Explicación y causalidad: enfoques y alternativas, por RAFAEL MORENO RODRÍGUEZ	505
1.	Primera aproximación al concepto de explicación en ciencia: algunos elementos de referencia	505
2.	La explicación: noción integradora	509
2.1.	Noción clásica	509
2.2.	Hacia una reformulación: aspectos subsuntivo y validador	511
2.3.	La cuestión de la causalidad, análisis de los componentes de un problema de la ciencia	516
3.	Diversidad categorial de la explicación en psicología e integración conceptual	518
3.1.	Implicaciones de la subsunción bajo conocimiento general	518
3.2.	Explicaciones y pseudoexplicaciones	521
3.3.	La explicación según la validez del conocimiento general que la sustente	522
3.3.1.	Conceptos básicos	522
3.3.2.	Las relaciones y la noción subsuntiva	525
3.4.	Niveles de complejidad en las explicaciones	526
3.4.1.	Modo cuantitativo de complejización	526
3.4.2.	Modo cualitativo	530
3.4.3.	Recapitulación sobre los niveles de complejidad	531
3.5.	Naturaleza de las relaciones y la noción subsuntiva	532
4.	Epílogo	536
	Bibliografía	538
18	Hacia una representación conceptual: teorías y modelos, por M.^a TERESA ANGUERA ARGILAGA	543
1.	Introducción	543
2.	Niveles del lenguaje científico y sus relaciones	543
2.1.	Descripción	543
2.2.	Conceptualización	545
2.2.1.	Conceptos naturales	547
2.2.2.	Conceptos teóricos	548
2.3.	Operacionalismo	548
3.	Teoría científica	550
3.1.	Concepto y significado desde la filosofía de la ciencia	550
3.2.	Construcción de teorías	552
3.3.	Estructura y funciones	558
3.4.	Situación actual en la psicología teórica	561
4.	Modelos	563
4.1.	Concepto y su alcance	563
4.2.	Sistemas clasificatorios	564
4.3.	Funciones y criterios valorativos	567
4.4.	Construcción de modelos	569
4.5.	Modelo y teoría	571
5.	Sistemas alternativos de representación conceptual	572
6.	Conclusión	574
	Bibliografía	575

19	Metodología de la investigación y diseño, por JAIME ARNAU GRAS	581
1.	Metodología de la investigación y diseño	581
1.1.	Modelo de la investigación científica	581
1.2.	Niveles y estadios del modelo de investigación científica	582
1.2.1.	Nivel teórico-conceptual	583
1.2.2.	Nivel técnico-metodológico	585
1.3.	A modo de resumen	590
2.	Estructura interna del diseño e investigación no-experimental	591
2.1.	Estructura formal del diseño de investigación	591
2.2.	Modelo estadístico y diseño	592
2.3.	Inferencia de causalidad y diseño	593
2.4.	Aproximación al diseño no-experimental	594
2.4.1.	Diseño no-experimental transversal: modelos de explicación alternativos	595
2.4.2.	Diseño no-experimental longitudinal: hacia la búsqueda de in- dicios de causalidad	597
2.4.3.	Regresión y causalidad: dos términos de difícil conciliación en investigación no-experimental	600
3.	Modelo estructural paramétrico y diseño experimental	601
3.1.	Modelo estructural general de diseño experimental	601
3.2.	Modelo estructural paramétrico y diseño simple completamente al azar	604
3.2.1.	Análisis estructural del modelo	604
3.2.2.	Hipótesis sobre los parámetros del modelo estructural	605
3.2.3.	Estimación de los parámetros del modelo	606
3.2.4.	Análisis de la varianza	607
3.2.4.1.	Fórmulas generales para el cálculo de las SC's de los términos clave del modelo	608
3.2.4.2.	Ejemplo práctico del cálculo de las SC's	609
3.2.4.3.	Medias cuadráticas y pruebas de significación	610
3.3.	Hacia una categorización del diseño experimental	613
	Bibliografía	615
20	Meta-análisis: una alternativa metodológica a las revisiones tradicionales de la investigación, por JULIO SÁNCHEZ MECA y MANUEL ATO GARCÍA	617
1.	Introducción	617
2.	La revisión cualitativa	618
2.1.	Cuantificación en las revisiones tradicionales	619
2.2.	Rigor científico en las revisiones de la literatura	620
3.	Hacia una definición de meta-análisis	621
4.	Etapas de un estudio meta-analítico	622
4.1.	Formulación del problema	623
4.1.1.	Amenazas contra la validez	624
4.1.2.	Protección de la validez	624
4.2.	Búsqueda de la literatura	624
4.2.1.	Técnicas de localización de estudios	625
4.2.2.	Amenazas contra la validez	626
4.2.3.	Protección de la validez	627
4.3.	Codificación de los estudios	628
4.3.1.	Amenazas contra la validez	629
4.3.2.	Protección de la validez	629
4.4.	Medida de los resultados	631
4.4.1.	La unidad de análisis	631
4.4.2.	El problema de la medida	632
4.4.2.1.	El nivel de significación	632

4.4.2.2.	El tamaño del efecto	632
4.4.2.2.1.	La diferencia media tipificada	633
4.4.2.2.2.	El coeficiente de correlación	636
4.4.2.2.3.	Estimuladores no paramétricos	636
4.4.2.3.	Obtención del tamaño del efecto	637
4.4.3.	Amenazas contra la validez	639
4.4.4.	Protección de la validez	639
4.5.	Análisis e interpretación de datos	641
4.5.1.	Las pruebas inferenciales convencionales	642
4.5.2.	Técnicas de acumulación de niveles de probabilidad	642
4.5.2.1.	Suma de algoritmos	643
4.5.2.2.	Suma de probabilidades	643
4.5.2.3.	Suma de puntuaciones t	643
4.5.2.4.	Suma de puntuaciones z	644
4.5.2.5.	Suma de puntuaciones z ponderadas	644
4.5.2.6.	Contraste del nivel P medio	644
4.5.2.7.	Contraste de la puntuación z media	644
4.5.2.8.	El método de Tippett	644
4.5.2.9.	Método logit	645
4.5.2.10.	Suma de puntuaciones z dependientes	645
4.5.2.11.	Usos e interpretaciones	646
4.5.3.	Técnicas de integración de tamaños del efecto	647
4.5.3.1.	Ajuste al modelo de la media global	648
4.5.3.1.1.	Diferencia media tipificada	648
4.5.3.1.2.	Coeficiente de correlación	649
4.5.3.2.	Ajuste de modelos categóricos	650
4.5.3.3.	Ajuste de modelos continuos	652
4.5.3.3.1.	Diferencia media tipificada	652
4.5.3.3.2.	Coeficiente de correlación	653
4.5.4.	Amenazas contra la validez	654
4.5.5.	Protección de la validez	655
4.6.	Publicación del estudio	656
4.6.1.	Amenazas contra la validez	657
4.6.2.	Protección de la validez	657
5.	Evaluación	657
5.1.	Ventajas	657
5.2.	Limitaciones	659
6.	Conclusión	661
	Bibliografía	663
	Índice analítico	671
	Índice onomástico	681