

**Metodología en la investigación  
sobre discapacidad. Introducción al  
uso de las ecuaciones estructurales**  
VI Simposio Científico SAID, 2008

Salamanca 5 y 6 de junio de 2008



Miguel Ángel Verdugo  
Manuela Crespo  
Marta Badía  
Benito Arias  
(Coordinadores)

**Colección Actas 5/2008**

Los contenidos de esta publicación son propiedad intelectual de sus respectivos autores

El INICO no se hace responsable de las ideas manifestadas por los mismos en esta publicación



**Publicaciones del INICO**



**Colección Actas 5/2008**





Metodología en la investigación  
sobre discapacidad.  
Introducción al uso de las ecuaciones  
estructurales



Metodología en la investigación  
sobre discapacidad.  
Introducción al uso de las ecuaciones  
estructurales

VI Seminario Científico  
SAID, 2008

MIGUEL ÁNGEL VERDUGO  
MANUELA CRESPO  
MARTA BADÍA  
BENITO ARIAS  
(Coordinadores)

Publicaciones del INICO  
Colección Actas  
Salamanca, 2008

ISBN: 978-84-691-5852-4

Depósito Legal: S. 1.436-2008

Imprime KADMOS

Salamanca, 2008

## ÍNDICE

PRESENTACIÓN .....	9
PRÓLOGO .....	11
1. <i>Introducción a la metodología SEM: Concepto y propósitos fundamentales.</i> BEGO- ÑA ORGAZ BAZ .....	13
2. <i>Modelos de medida y análisis factorial confirmatorio.</i> RAMÓN FERNÁNDEZ PULIDO ....	29
3. <i>Introducción a los modelos de ecuaciones estructurales.</i> MIGUEL RUIZ DÍAZ .....	43
4. <i>Desarrollo de un ejemplo de análisis factorial confirmatorio con LISREL, AMOS y SAS.</i> BENITO ARIAS MARTÍNEZ .....	75
5. <i>Meta-análisis de la investigación.</i> JULIO SÁNCHEZ MECA .....	121
6. <i>Metodologías cualitativas: Características, procesos y aplicaciones.</i> M <sup>a</sup> TERESA AN- GUERA ARGILAGA .....	141
7. <i>Informática y análisis cualitativo.</i> M <sup>a</sup> CRUZ SÁNCHEZ GÓMEZ .....	127
8. <i>La ética en la investigación sobre discapacidad.</i> M <sup>a</sup> Teresa ANGUERA ARGILAGA .....	177
ANEXO. Programas SAID .....	193





## PRESENTACIÓN

En el estudio y tratamiento de la discapacidad se entrelazan, prácticamente, todos los dominios del saber, desde las ciencias más experimentales (la física o la biología) hasta las ciencias humanas y sociales (como el derecho o las ciencias de la información). Sin embargo, hay áreas de conocimiento que tienen mayor tradición de investigación porque las condiciones de discapacidad plantean problemas o cuestiones que atañen directamente a los objetivos propios de estas áreas. Las disciplinas de la psicología, la educación y la medicina son ejemplos paradigmáticos en este sentido.

Las aportaciones y avances de la psicología, la medicina y la educación en el ámbito de la discapacidad, se entrecruzan y relacionan entre sí, hasta el punto de llegar a formar un cuerpo de conocimiento que es indispensable conocer si se pretende orientar y diseñar líneas futuras de investigación.

Las relaciones que establecen entre sí estas áreas pueden ser analizadas partiendo de la naturaleza y el tamaño de los sistemas que estudian (neurofisiológico, mental, familiar, social, educativo, etc.) y la metodología de investigación en que basan su desarrollo y conocimientos actuales (encuestas, experimentos sobre la naturaleza de los trastornos, etc.).

Además de los dominios o tradiciones de investigación y la tipología de las investigaciones, el tercer elemento que se cruza con los anteriores se refiere a la metodología de la investigación que, por supuesto, está estrechamente relacionada con las tres tradiciones citadas.

Una de las líneas de actuación del INICO es la investigación en el ámbito de la discapacidad y, por ello y desde sus inicios, se ha interesado en la formación metodológica de los investigadores. Con este objetivo en 1998 surgió el I Seminario Científico de Actualización Metodológica, también conocido como SAID, y desde entonces y, aproximadamente, cada dos años este seminario se viene desarrollando periódicamente. A lo largo de esta década se han abordado diversos aspectos metodológicos relacionados con la discapacidad, siendo los objetivos generales los siguientes:

- Analizar los procedimientos tradicionales llevados a cabo por la medicina, la educación y la psicología para investigar en discapacidad.
- Ofrecer nuevas vías metodológicas (desde perspectivas conceptuales, cuantitativas y cualitativas) para abordar el estudio de problemas relativos a la discapacidad.
- Revisar algunas aportaciones metodológicas recientes en este campo de estudio.
- Revisar y discutir los problemas metodológicos de índole aplicada más frecuentes con que se encuentra la investigación sobre problemas relacionados con la discapacidad.

- Fomentar las habilidades y el debate metodológico entre los participantes.
- Proporcionar una visión actualizada de las herramientas conceptuales y analíticas fundamentales de las Metodologías de Investigación empleadas en los estudios de discapacidad.
- Capacitar para la intervención en procesos de aprendizaje, orientación, evaluación, organización, investigación y planificación en contextos de discapacidad.
- Transmitir la idea, a los futuros profesionales en discapacidad, de que la esencia de la metodología es expandir conocimiento, comprender, interpretar y resolver problemas psico-educativos.
- Formar y motivar hacia la utilización de herramientas que permitan comprender, interpretar, tener visión ética y crítica de investigaciones realizadas en el ámbito de la discapacidad.

La VI edición del programa ha tenido un carácter monográfico y abordó los fundamentos de la metodología de ecuaciones estructurales. Sus objetivos son los siguientes:

- Proporcionar a los asistentes las bases teóricas de la metodología de ecuaciones estructurales –SEM– (path analysis, modelos de medida / análisis factorial confirmatorio y modelos estructurales).
- Orientar sobre los problemas genéricos y específicos que pueden investigarse, dentro del ámbito de la discapacidad, mediante la metodología SEM.
- Facilitar la información y práctica necesarias para el manejo de programas de análisis de modelos SEM (AMOS y/o LISREL).

Con motivo de la celebración del VI SAID, y coincidiendo con los 10 años de su puesta en marcha, editamos esta publicación en la que se recogen las ponencias de este último seminario así como una selección de las presentadas en las anteriores ediciones, todo ello con el objetivo de que la publicación sirva de referencia en este campo.

MIGUEL ÁNGEL VERDUGO

## PRÓLOGO

Con sumo gusto prologo esta obra coordinada por Miguel Ángel Verdugo, a quien agradezco enormemente este ofrecimiento, y patrocinada por el INICO de la Universidad de Salamanca. Con su publicación se pone de manifiesto el interés y empeño del Instituto Universitario de Integración en la Comunidad por la investigación en el ámbito de la discapacidad y, como lógica consecuencia, por la formación metodológica de los investigadores, pieza clave para augurar una producción científica de prestigio.

Prueba clara de este interés son los clásicos Seminarios Científicos de Actualización Metodológica que el INICO viene organizando bianualmente desde 1998, y que han tratado monográficamente temáticas sumamente relevantes desde una perspectiva metodológica. Es una iniciativa magnífica, digna de todo encomio, y nos congratulamos por ella, con la seguridad de que, en un razonable plazo de tiempo, empezará a producir sus frutos.

Buena parte de los contenidos de esta obra corresponden a las contribuciones presentadas al **VI Seminario Científico: Actualización Metodológica en Investigación sobre Discapacidad (SAID 2008)**, que en la actual edición versa monográficamente sobre *Introducción a la metodología de ecuaciones estructurales en la investigación sobre discapacidad*, dirigido en esta edición por mis estimados colegas Marta Badía y Benito Arias.

El interés del tema propuesto para la edición de 2008 es de elevada relevancia en todos aquellos estudios en los cuales se dispone de un marco teórico consistente y se pretende(n) contrastar una o varias hipótesis que habitualmente se fundamentarán en relaciones de causalidad. Si de forma ciega el lector tuviese delante de sus ojos el título de este Seminario Científico, seguro que se interesaría sobremanera, por lo sugerente de la expresión, e incluso de desafiante en su sentido más positivo. Se trata de un feliz enclave estadístico-psicométrico que tiene un uso y aplicabilidad crecientes en muchos ámbitos de la investigación, y por supuesto, en concreto, en el de la investigación sobre discapacidad.

La obra se nutre en buena medida de las conferencias/ponencias del VI Seminario Científico antes mencionado, y configuran una masa crítica integrada de elevado nivel metodológico, de lo cual me congratulo con gran satisfacción. Por una parte, se presentan los trabajos *Introducción a la metodología SEM: Concepto y propósitos fundamentales* e *Introducción a los modelos de ecuaciones estructurales*, que introducen al lector sobre las cuestiones esenciales que deben conocerse en el ámbito de las ecuaciones estructurales; está también el capítulo *Modelos de medida y análisis factorial confirmatorio*, que pretende poner en contacto al investigador interesado con los fundamentos teóricos de esta técnica analítica, de forma que sea capaz de realizar una lectura crítica de trabajos en los cuales se utilice, y además aprenda a aplicarla. Y está el trabajo *Desarrollo de un ejemplo de análisis factorial confirmatorio con LISREL, AMOS y SAS*, que ofrece un primer desarrollo conceptual, y una exposición

detallada de un ejemplo-réplica respecto a los programas informáticos indicados en el título. Juntamente con otras ponencias que se presentarán, constituye un excelente augurio para el VI Seminario Científico.

Pero en la obra se incluyen igualmente otros capítulos seleccionados a partir de las aportaciones de anteriores Seminarios Científicos, todos ellos centrados en actualizaciones metodológicas determinadas, que atienden y dan soporte procedimental en diversas direcciones a los investigadores en el amplio ámbito de la discapacidad. Destacamos el capítulo '*Meta-análisis de la investigación*', tan interesante para poder integrar cuantitativamente los resultados obtenidos en diversos tópicos sobre discapacidad, los capítulos *La investigación cualitativa: su aplicación a la investigación y análisis de la realidad* y el de '*Metodologías cualitativas: Características, procesos y aplicaciones*', sobre una metodología en rápido proceso de expansión como es la cualitativa, tanto considerada de forma aislada como en su creciente complementariedad con la cuantitativa, y, finalmente, se incorpora la temática ética en el capítulo '*La ética en la investigación sobre discapacidad*', cuestión que debe considerarse como el referente común que rija en todas las investigaciones, sea cual sea el área de conocimiento, pero de forma superlativa en temas sensibles como el de la discapacidad.

La estructura de la obra, muy bien equilibrada y compensada en cuando a la extensión de cada uno de sus ocho capítulos, resultará útil, sin duda, a los lectores, y me congratulo sinceramente por ello, especialmente por la nitidez y perspectivas con que ha sido concebida y desarrollada.

Agradezco de nuevo a Miguel Ángel Verdugo que me haya brindado la oportunidad de escribir este prólogo, a la vez que reitero mi satisfacción por la preocupación del INICO respecto a la formación metodológica de los investigadores dedicados al ámbito de la discapacidad, y hago votos por este fortalecimiento, como semilla plantada en una tierra fértil y en un campo arado, y de la cual se esperan ubérrimos frutos. El paso del tiempo, este tiempo inexorable que marca nuestras vidas, espero que me dé la razón.

M<sup>a</sup> TERESA ANGUERA ARGILAGA  
*Universidad de Barcelona*  
Mayo de 2008

# INTRODUCCIÓN A LA METODOLOGÍA SEM: CONCEPTO Y PROPÓSITOS FUNDAMENTALES

M<sup>a</sup> BEGOÑA ORGAZ BAZ

*Área de Metodología de las Ciencias del Comportamiento  
Facultad de Psicología (Universidad de Salamanca)*

Los Modelos de Ecuaciones Estructurales o Modelos de Estructuras de Covarianza constituyen los modelos de análisis en los que mejor queda reflejado ese objetivo último de la Ciencia y que, por tanto, es también el fin último de las Ciencias Comportamentales y Sociales, que es la construcción de esquemas explicativos, es decir, teorías que posibiliten una mejor comprensión de la realidad y de los fenómenos observados. Su objetivo último es determinar mediante pruebas cuantitativas, en qué medida los datos de la muestra apoyan un modelo teórico de múltiples relaciones de dependencia entre variables propuesto a contraste por el investigador.

Así, surgieron ante la necesidad de abordar cuestiones de causalidad e inferencia causal con datos no experimentales. En esta línea, estos modelos constituyen un nexo de unión entre las dos metodologías utilizadas en la Psicología: la metodología experimental y la no experimental, en otras palabras de las dos “psicologías”: la “Psicología Experimental” y la “Psicología Correlacional” (López Feal, 1982).

Además, constituyen una síntesis de diferentes aproximaciones metodológicas, por una parte, de la Estadística y la Econometría que trata de analizar las relaciones entre una estructura o sistema de variables, comprobando los posibles efectos mediadores y espurios de determinadas variables; y, por otra, de la Psicometría, que aporta el tema del error de medida, y la posibilidad de estudiar los denominados constructos psicológicos, que no pueden ser observados y medidos directamente, sino a partir de diferentes indicadores observables. De esta forma, estos modelos representan los desarrollos más avanzados del Modelo de Regresión, y permiten el análisis de una realidad más compleja.

En el desarrollo del tema, comenzaremos exponiendo cómo surgen estos modelos ante la necesidad de abordar cuestiones de causalidad e inferencia causal con datos no experimentales. Haremos especial hincapié en cómo se hace frente en estos modelos a los tres criterios que se considera se han de cumplir para hablar de relaciones causa-efecto.

A continuación, plantearemos cómo han surgido estos modelos, partiendo del Modelo de Regresión, hasta llegar a los modelos más complejos de Estructuras de Covarianza, integrando los Modelos de Medida. En cada caso, iremos indicando qué aporta cada uno de los modelos en la explicación de los fenómenos comportamentales y a qué objetivos o hipótesis de investigación se puede dar respuesta con cada uno de ellos.

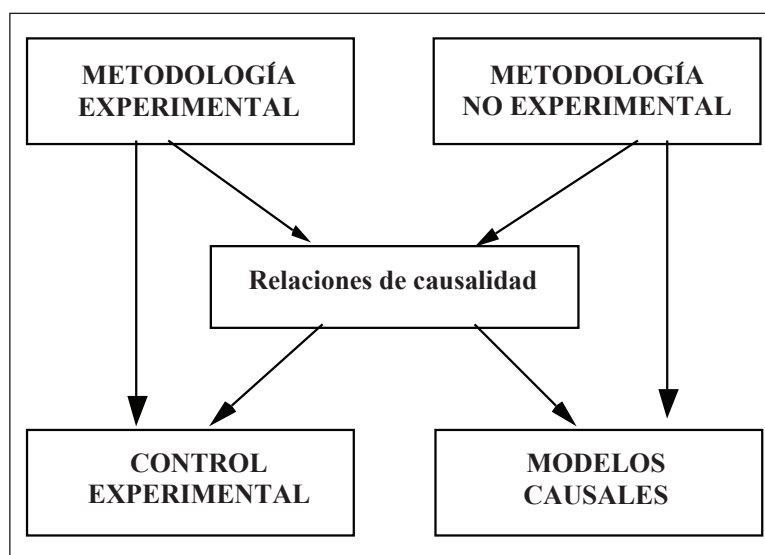
Y, por último, presentaremos los elementos y el proceso de evaluación de todo modelo causal, que son los mismos que en cualquier modelo de análisis. En el caso de los elementos, la teoría y el modelo matemático aunque en este caso, como veremos, la teoría adquiere una especial relevancia. Y, finalmente, desarrollaremos los pasos generales a seguir en el análisis de los datos generados con estos modelos, que también son los mismos que en cualquier análisis, los que corresponden al proceso de modelado estadístico, aunque en este caso adquiere una especial relevancia la fase de identificación del modelo.

## LAS RELACIONES DE CAUSALIDAD A PARTIR DE DATOS CORRELACIONALES

El desarrollo de estos modelos ha permitido a los estudios correlacionales ir más allá del descubrimiento de meras asociaciones entre variables y acercarse cada vez más al descubrimiento de relaciones causales. Durante muchos años se consideró que únicamente, cuando se utilizaba la Metodología Experimental se podían establecer relaciones causales entre variables, porque en este caso se cumplían los criterios o requisitos planteados por los diferentes autores (Blalock, 1985; Cook y Campbell, 1986; Gómez Benito, 1986; Kenny, 1979; Mulaik, 1986) para hablar con rigor de relaciones causa-efecto: la asociación, relación funcional o covariación entre las variables implicadas; el orden temporal de los eventos, es decir, la precedencia en el tiempo de la causa al efecto; y la eliminación de otras posibles variables causales, es decir, la ausencia de espuriedad.

En este sentido, y como hemos señalado previamente, estos modelos han constituido un nexo de unión entre las dos metodologías: Metodología Experimental vs. Metodología No Experimental o Correlacional, porque han permitido establecer modelos causales dentro de la Metodología Correlacional, lo que antes estaba circunscrito al ámbito de la Metodología Experimental.

Figura 1. Las relaciones de causalidad en la Metodología Experimental y No Experimental



A continuación abordaremos cómo se hace frente en estos modelos a los tres criterios citados anteriormente. En cuanto al primero, la asociación entre las variables, puede ser satisfecho tanto en la Metodología Experimental como en la No Experimental. En la Metodología

Experimental, se pueden observar los cambios en la variable dependiente, cuando se introducen diferentes niveles de la variable independiente. En la Metodología No Experimental, las técnicas estadísticas como la Correlación y el Análisis de Regresión ponen en evidencia la existencia o no de asociación.

En relación con el segundo criterio, la estrategia experimental es la única que garantiza plenamente este requisito de precedencia temporal de la causa al efecto, puesto que introduce variaciones sistemáticas en uno o varios aspectos de la “causa” –variable independiente o factor–, y observa posteriormente las variaciones concomitantes en uno o varios aspectos del “efecto” –variable dependiente–. En el contexto no experimental, donde los datos son generalmente recogidos en un único momento temporal, es difícil establecer el orden temporal de los eventos sobre la base de la evidencia empírica solamente, se requiere la existencia o disponibilidad de una perspectiva conceptual o una teoría plausible que permita la especificación de la direccionalidad de las relaciones.

Por último, una de las condiciones más importantes es el “supuesto de no espuriedad”, que establece que, una relación entre dos variables es “no espuria” sólo si no se da una tercera variable que sea causa de ambas. La Metodología Experimental, debido a su propia estructura de control, es la única que garantiza una relación causal, ya que elimina la presencia de otras variables concomitantes, que podrían constituir posibles causas explicativas. Los procedimientos de control aseguran que quede eliminada la posibilidad de un “factor causal común”. En resumen, el control asegura la validez interna y ésta permite establecer relaciones causales. En la Metodología No Experimental, estas variables pueden ser eliminadas como posibles causas sobre la base de un modelo teórico que establezca que estas variables son irrelevantes al proceso causal, o bien mediante el control estadístico, si la teoría en principio no la descarta.

En conclusión y en relación con el tema de la causalidad, en la Metodología No Experimental se debe hacer frente fundamentalmente a dos aspectos, la determinación de la dirección y la posible existencia de relaciones espurias. Para afrontar estos problemas, en esta metodología se realiza una cuidadosa revisión de hipótesis alternativas y el ajuste de modelos con varias ecuaciones, modelos causales, que implican y se basan, de alguna manera, en la idea de causalidad. La aplicación de estos modelos, como técnica para realizar inferencias causales a partir de simples correlaciones, ha de basarse prioritariamente en la teoría. Por ello, como ya mencionamos, en estos modelos causales, la teoría adquiere una especial relevancia. En definitiva, el análisis causal a partir de datos correlacionales no permite descubrir causas, sino que se limita a evaluar modelos causales elaborados por el investigador basándose en una teoría o modelo teórico, y, aún cuando se hayan controlado todas las posibles variables que teóricamente podrían ser consideradas de confundido, no hay completa seguridad de haber descartado totalmente la espuriedad (Asher, 1983).

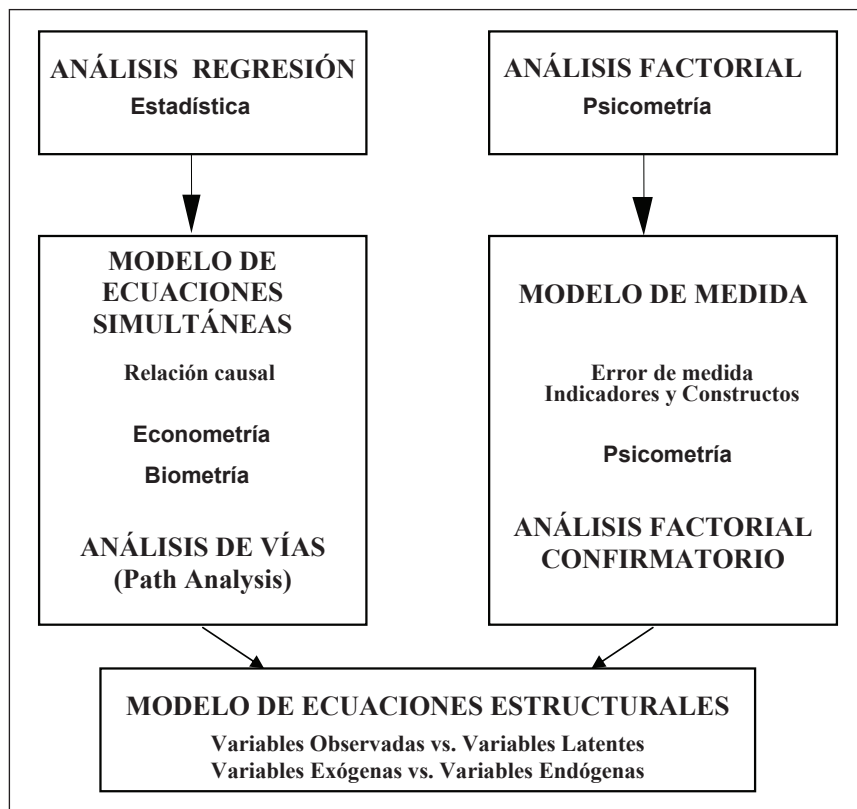
## DEL MODELO DE REGRESIÓN A LOS MODELOS DE ECUACIONES ESTRUCTURALES

Una vez planteado el objetivo de estos modelos desarrollaremos como surgen y evolucionan desde los modelos más simples, lo que permite una más fácil y rápida comprensión de los modelos más complejos. Partiendo del modelo más sencillo el Modelo de Regresión Lineal Simple, basado en el Modelo Lineal General, incrementando la complejidad del mismo, e incorporando las aportaciones del Modelo de Medida podemos llegar a los complejos Modelos de Ecuaciones Estructurales o de Estructuras de Covarianza.



El Modelo final constituye la integración de modelos y técnicas surgidos de diferentes campos como la Estadística, la Biometría, la Econometría y la Psicometría. La Estadística contribuyó con el Análisis de Regresión. El concepto de variables latentes y errores de medida, así como la relación entre los indicadores o variables observables y las variables latentes o constructos proviene del corpus teórico de la Psicometría a través del Análisis Factorial y la Teoría de la Fiabilidad. Por su parte, los econométricos y biométricos aportan el análisis de las influencias direccionales y simultáneas de unas variables sobre otras (Díez Medrano, 1992; Long, 1983b; Maruyama, 1998; Visauta, 1986).

Figura 2. El Modelo de Ecuaciones Estructurales como resultado de la integración de modelos y técnicas de diferentes campos.



Comenzaremos por el **Modelo de Regresión Lineal Simple** en el que se trata de explicar una variable dependiente o criterio en función de una única variable independiente o predictora, siendo ambas variables susceptibles de observación, y se asume que son medidas sin error. En este modelo no podemos hablar de relaciones causa-efecto.

Este modelo presenta bastantes limitaciones porque en las investigaciones actuales y, sobre todo en la investigación aplicada no experimental, raramente se analiza la relación entre un único predictor y un criterio. En la mayoría de las ocasiones, para explicar una determinada variable criterio se utiliza un conjunto de variables predictoras, todas ellas observables, como en el caso anterior, siendo necesario el planteamiento de un **Modelo de Regresión Lineal Múltiple**.

Este tipo de explicaciones tan sencillas, que se pueden obtener a partir del Modelo de Regresión, Simple o Múltiple, únicamente permiten el conocimiento de una parcela muy re-

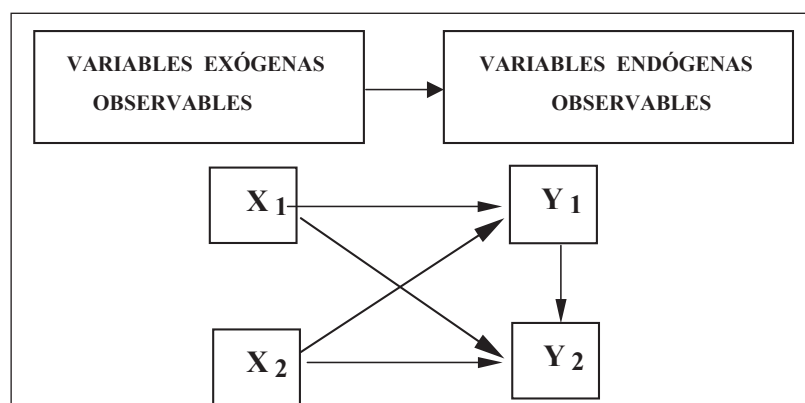
ducida de la realidad, y resultan insuficientes para las Ciencias Comportamentales y Sociales, que tratan de conocer y entender una realidad compleja y multideterminada, y por lo tanto precisa explicaciones de nivel más amplio y complejo. En la mayoría de las ocasiones, nos encontramos con conjuntos de variables predictoras y criterios relacionadas entre sí.

Así por ejemplo, nos encontramos con los denominados efectos mediadores o indirectos de determinadas variables. Se dice que una variable ejerce un efecto mediador cuando transmite el efecto de una variable independiente sobre la dependiente, es decir, se encuentra dentro de la secuencia causal de dos variables. Y se denomina efecto indirecto al que ejerce la variable independiente sobre la dependiente a través de la variable mediadora. En los estudios en los que se analizan estos efectos, una vez que se ha encontrado relación entre dos variables, se incluye una tercera en el análisis de la relación entre las dos primeras con el fin de mejorar la comprensión de la relación o determinar si dicha relación es espuria.

En estos casos, el estudio de los determinantes aislados no permite obtener conclusiones sobre los efectos de todos ellos actuando conjuntamente sobre la conducta, ya que el comportamiento de la configuración global de los distintos componentes y sus interacciones no es igual a la suma de los efectos de los componentes aislados. Y nos puede llevar a conclusiones erróneas, pues no reflejaría las verdaderas relaciones existentes entre todas las variables.

Para poder considerar simultáneamente las relaciones entre las diferentes variables implicadas en un fenómeno, el Modelo de Regresión resulta insuficiente y es necesario incorporar a las Ciencias Sociales y Comportamentales modelos nacidos y desarrollados en otros campos como la biometría y la econometría, los denominados **Modelos de Ecuaciones Simultáneas**, Modelos Multiecuacionales o Modelos Causales, que nos permiten poner en juego al mismo tiempo las interrelaciones entre varias variables, que pueden ser en las distintas ecuaciones del modelo en unos casos explicadas y en otros explicativas, como en el caso de los efectos mediadores.

Figura 3. Ejemplo de Modelo de Ecuaciones Simultáneas o Análisis de Vías



Los primeros desarrollos de estos modelos es el denominado Análisis de Sendas o **Análisis de Vías**, más conocido por el término **"Path Analysis"** (Asher, 1983). Esta técnica de análisis se utiliza cuando se dispone únicamente de una medida observada o indicador para cada variable o constructo teórico y de una hipótesis a priori basada en la teoría sobre la relación causal o dirección de la relación entre las variables. En estos modelos, como provienen de otros campos, no se habla de variables independientes o predictoras y variables dependientes o criterios, sino de variables exógenas y endógenas, respectivamente. Las variables endógenas son aquellas que el modelo pretende explicar a partir de otras variables exógenas o en-

dógenas, en tanto que las exógenas no son explicadas por otras variables dentro del modelo, es decir, sus causas no se representan en el modelo porque se consideran desconocidas, y las relaciones entre ellas no se analizan. En este modelo tanto las variables exógenas como las endógenas son variables observables.

Kenny (1979) diferencia entre variables puramente exógenas, cuya variabilidad no es explicada por el modelo, y variables que sí son explicadas en alguna medida por el modelo, a partir de aquella ecuación en la que figuran como endógenas, pero que en otra u otras ecuaciones pueden actuar como exógenas; como ya hemos señalado estos modelos admiten que una misma variable sea a la vez explicativa y explicada en dos ecuaciones distintas. Por ello, mientras el Análisis de Regresión únicamente permite estudiar efectos directos, es decir, la conexión entre el determinante y determinado se concibe como única y sin intermediarios, estos modelos permiten también analizar efectos indirectos de una variable sobre otra a través de una o más variables mediadoras, es decir, determinaciones mediadas.

Además, estos modelos permiten especificar y contrastar modelos causales que reflejan supuestas o hipotetizadas asociaciones espurias. Estas relaciones espurias se representan especificando causas comunes. En el ejemplo, Y1 e Y2 tienen dos causas comunes X1 y X2. El coeficiente de vías que representa el efecto directo de Y1 sobre Y2 controla las causas comunes. Si el valor de este coeficiente es 0 o próximo a 0, podríamos considerar que la relación entre Y1 e Y2 es espuria; si por el contrario, este coeficiente es elevado, se puede rechazar la hipótesis de espuriedad.

En definitiva, el objetivo de este tipo de Modelos Multiecuacionales es comprobar en qué medida los datos de la muestra apoyan los efectos directos e indirectos entre variables observadas o indicadores hipotetizados por un modelo teórico. Para ello, se analizan simultáneamente múltiples relaciones de dependencia a partir de los coeficientes de vías ("path coefficients") entre variables observadas: múltiples variables independientes (variables exógenas) y múltiples variables dependientes (variables endógenas), lo que permite modelizar relaciones complejas.

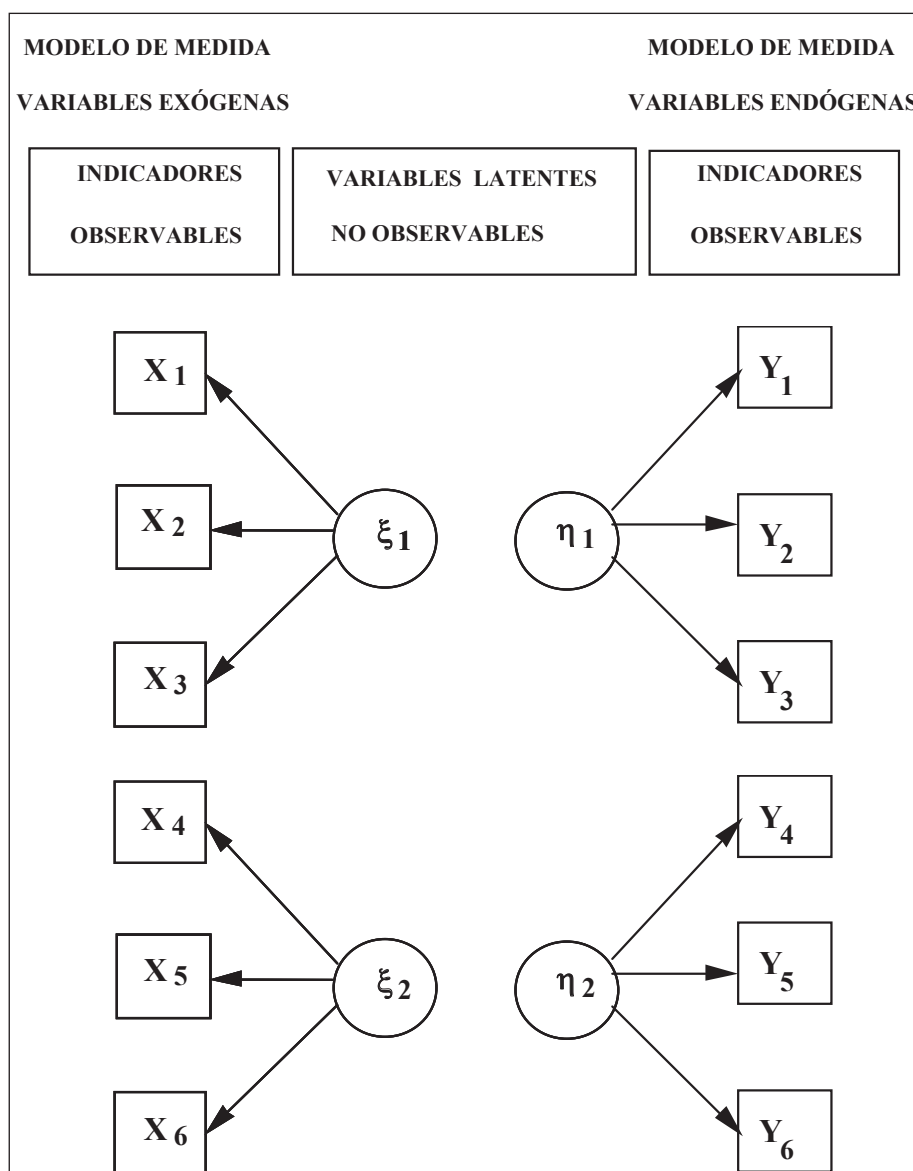
No obstante, la mayoría de las variables de interés teórico y con las que se trabaja en la mayoría de los casos en las Ciencias Comportamentales y Sociales no pueden ser observadas directamente, son los llamados constructos teóricos, factores, variables latentes, etc. que son inferidas a partir de la medida de unos indicadores. Y es en este punto donde se encuentra una de las principales limitaciones del Análisis de Vías, que utiliza una única medida o indicador por cada constructo que se incluye en el modelo. Esto obliga al investigador a optar entre las diferentes medidas del mismo constructo por una sola para incluirla en el modelo, y es poco realista esperar que un constructo teórico pueda ser medido adecuadamente por un único indicador, éste no reflejará las diferentes facetas del constructo, lo cual implicará que la medida no será válida.

Además, ese indicador del constructo se verá afectado por los errores de medida, lo que significa que no será fiable (Asher, 1983). Y, como ya hemos dicho, una de las condiciones de aplicación tanto de los Modelos de Regresión como de los Modelos de Ecuaciones Simultáneas o Análisis de Vías es que las variables, sobre todo las exógenas, estén medidas sin error. Una posible solución es utilizar múltiples indicadores para cada constructo. Un conjunto de medidas será más fiable que una única medida y evaluarán diferentes facetas del constructo lo que incrementará la validez.

En este punto, se hace necesario plantear un Modelo de Medida (Long, 1983a) en el que se especifican las relaciones que se hipotetizan entre los constructos teóricos que en estos modelos se denominan variables latentes y sus correspondientes indicadores o variables

observables. El objetivo de estos modelos es comprobar, mediante los datos obtenidos en la muestra, modelos teóricos que hipotetizan que determinados conjuntos de variables observadas o indicadores definen o configuran constructos, factores o variables latentes.

Figura 4. Ejemplo de Modelo de Medida o Análisis Factorial Confirmatorio



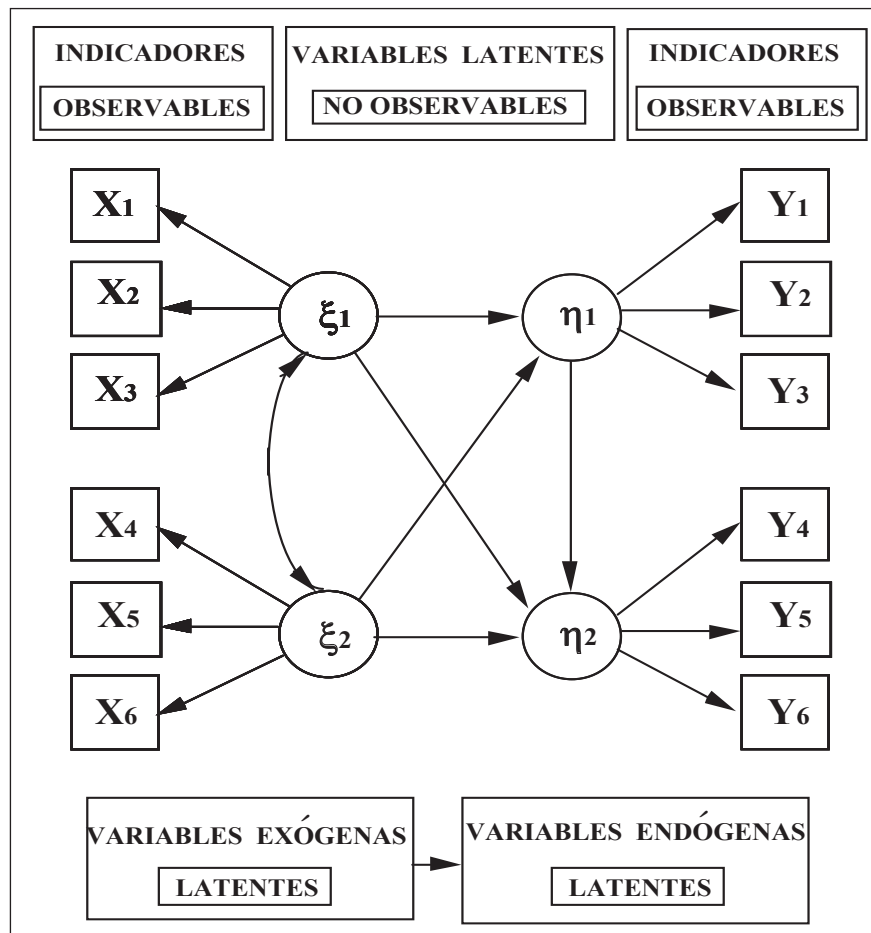
En el **Análisis Factorial Confirmatorio**, que es la técnica de análisis que se utiliza para contrastar los Modelos de Medida, se hipotetiza que una serie de variables observadas o indicadores miden una o más variables latentes (independientes o dependientes). Para ello, se especifican las relaciones entre las variables observables o indicadores y las variables latentes a partir de unos coeficientes que en este caso se denominan cargas factoriales (“factor loading”), y representan los efectos de las variables latentes o factores sobre los indicadores, es decir, en qué medida el factor se refleja en la puntuación del indicador.

En un Modelo de Medida únicamente se plantean relaciones entre los factores pero no efectos entre ellos. Estos efectos, como hemos visto, se especifican y contrastan en el Modelo

de Ecuaciones Simultáneas pero esta técnica, como ya hemos señalado, no permite incorporar diferentes indicadores del mismo constructo e incluye únicamente variables observables. Por ello, se necesita un modelo que combine o integre las características del Modelo de Medida y Modelo de Ecuaciones Simultáneas. Este modelo es el **Modelo de Ecuaciones Estructurales** o **Modelo de Estructuras de Covarianza** en el que es posible especificar un modelo híbrido que incorpora un componente estructural, como un Modelo de Ecuaciones Simultáneas, que trata de estudiar o probar hipótesis sobre los efectos directos e indirectos, pero ahora entre las variables latentes o factores; y un componente de medida, como un Modelo de Medida, que analiza las relaciones entre las variables observables o indicadores y las latentes o factores.

Este modelo híbrido es un Análisis de Vías con variables latentes que no son medidas directamente sino inferidas a través de variables observadas o indicadores, es decir, distingue entre variables observadas o indicadores y variables latentes o constructos que se presupone miden los indicadores (Bollen, 1989). En este modelo, se formulan explicaciones causales sobre las variables latentes a partir de las relaciones observadas entre los indicadores. Y, por otra parte, tiene en cuenta la fiabilidad de las medidas, ya que incorpora o incluye los errores de medida en el modelo, lo que permite controlarlos directamente. Lo cual constituye importante ventaja sobre el Modelo de Ecuaciones Simultáneas que no tiene en cuenta los posibles errores de medida cuando evalúan las relaciones causales entre las variables.

Figura 5. Ejemplo de Modelo de Ecuaciones Estructurales



Un Modelo de Estructuras de Covarianza se plantea cuando se pretende estudiar cómo un conjunto de indicadores o variables observadas definen constructos o variables latentes –Modelo de Medida–, y los efectos directos e indirectos o relaciones mediadas entre estos constructos –Modelo de Ecuaciones Simultáneas–. Además, permiten mejorar la representación de los conceptos teóricos a partir de múltiples indicadores o medidas y tener en cuenta el error de medida, es decir, la fiabilidad. En este caso, a los coeficientes que representan las relaciones de dependencia entre las variables exógenas y endógenas latentes se les denomina coeficientes estructurales.

## COMPONENTES Y DESARROLLO DE UN MODELO DE ECUACIONES ESTRUCTURALES

Una vez comprobado como los Modelos Causales constituyen una generalización del Modelo de Regresión en los que se pueden analizar los efectos directos e indirectos entre variables tanto observables como latentes, vamos a ver como en lo que respecta a sus componentes y desarrollo tampoco son tan diferentes de los modelos más sencillos.

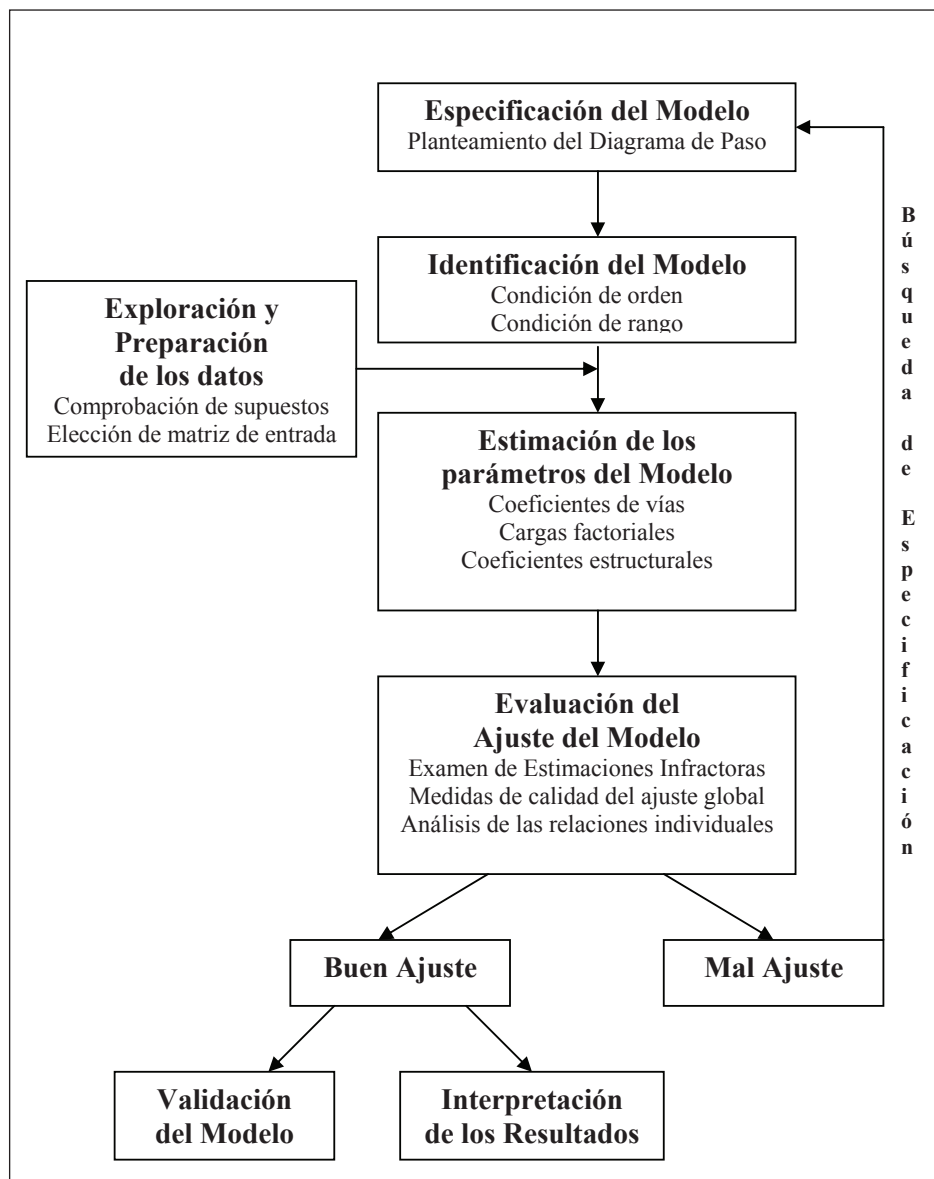
En primer lugar, hay que destacar la importancia que, como ya hemos señalado al inicio, tiene la teoría en el planteamiento de estos modelos. Por ello, se puede decir que en los modelos causales son fundamentales tanto el modelo matemático que se plantea, y que establece las relaciones entre las variables, como la teoría en la que se basa el sistema de ecuaciones.

Dentro del **modelo matemático**, como en todos los casos, hay que tener en cuenta los elementos que constituyen el modelo, y el proceso que se sigue en el desarrollo del modelo matemático, lo que se conoce con el nombre de Modelado Estadístico.

Por lo que respecta a los **componentes o elementos**, en este caso el modelo está constituido por un sistema de ecuaciones matemáticas, es decir, se trata de un modelo multiecuacional. Pero los elementos son los característicos que incluyen todos los modelos lineales. Por una parte, las variables, en este caso, exógenas y endógenas, sean observadas o latentes. Por otra, un conjunto de constantes o parámetros no observables pero potencialmente estimables, los diferentes coeficientes, y cuya estimación es una de las fases del proceso de modelización estadística. Por último, el modelo como nunca va a ser un reflejo totalmente exacto de la realidad incorpora también una o más variables aleatorias, no observables ni estimables, que representan el componente de error aleatorio –perturbación, error de medida–, que refleja la variación debida a diferencias individuales, a errores de medición y a la influencia de otras variables desconocidas u omitidas por parsimonia y que no han sido incorporadas al modelo. Sobre este componente descansa el proceso de crítica y evaluación del modelo, que constituye otro de los pasos dentro del proceso de modelado estadístico.

En relación con el desarrollo del modelo, vamos a hablar del denominado **Proceso de Modelado Estadístico**, proceso a desarrollar en el análisis de datos de cualquier modelo, y que consta de tres fases fundamentales: la Especificación del Modelo, la Estimación de los Parámetros del Modelo, y la Evaluación del Ajuste del Modelo propuesto. Pero además, en este caso adquiere, como veremos, una especial relevancia una fase denominada Identificación del Modelo. Pasamos a exponer cada una de estas fases en cada uno de los Modelos (Kline, 1998; Mueller, 1997; Schumacker & Lomax, 1996; Visauta, 1986).

Figura 6. Proceso de Modelado Estadístico en la evaluación los Modelos Causales



La fase de **Especificación del Modelo** tiene como objeto plantear un modelo estadístico, que sirva de puente entre la teoría y los datos. Para ello, a partir de las teorías existentes en el campo de estudio y de investigaciones previas, se propone un modelo teórico o modelos alternativos que se van a contrastar, es decir, se determina qué variables se van a incluir en el modelo, y cómo se relacionan estas variables, es decir, los parámetros de interés. En general, supone tomar importantes decisiones, acerca de qué variables intervienen en el fenómeno o aspecto de la realidad que se está estudiando, qué relaciones existen entre estas variables, y la dirección y magnitud de estas relaciones. Una vez tomadas estas decisiones, se especifican, en el diagrama de secuencias de relaciones de dependencia o diagrama de paso, y matemáticamente mediante las correspondientes ecuaciones, la naturaleza y la forma de las relaciones entre las variables planteadas por la teoría o el modelo teórico.

En el *Modelo de Ecuaciones Simultáneas* en el diagrama de paso se especifica qué variables se van a incluir en el modelo y la dirección de los efectos causales entre ellas. En estos



modelos, se plantean relaciones “no causales” y relaciones “causales” que expliquen las covarianzas de las variables exógenas y endógenas.

Entre las relaciones “no causales”, se incluyen relaciones que se estiman pero no se analizan, como relaciones entre variables exógenas, se asume que estas dos variables covarían pero se desconoce por qué covarían o si tienen causas comunes; y relaciones entre las perturbaciones, que reflejan el supuesto de que las variables endógenas comparten al menos una causa común que se ha omitido en el modelo. Por lo tanto, la ausencia de estas asociaciones en el diagrama de paso refleja el supuesto de independencia de las posibles causas no incluidas. Llegados a este punto nos podemos preguntar, ¿es posible plantear en el diagrama de paso la relación entre una perturbación y una variable exógena? Desde el punto de vista teórico es posible, y esta relación implicaría la presencia de una variable omitida que causa tanto la variable exógena como la endógena. En otras palabras, parte de la covarianza de las variables exógena y endógena es espuria debido a una causa común. No obstante, normalmente en estos modelos se asume que las variables exógena y las perturbaciones no están relacionadas, son independientes. Y las relaciones espurias que se plantean, especificando causas comunes, representan que al menos parte de la correlación observada entre las dos variables endógenas se presupone espuria.

Entre las relaciones “causales” se incluyen los efectos unidireccionales - directos e indirectos-. Para representar estos efectos se especifican en el modelo los coeficientes de vías (“path coefficient”). Y, por último, los errores que en este caso se denominan perturbaciones (“disturbances”) y representan los efectos combinados de todas las posibles variables exógenas que podrían afectar a una variable endógena, pero que han sido omitidas, es decir, que no han sido medidas e incluidas en el modelo.

En la especificación del *Modelo de Medida* hay que tener en cuenta tres aspectos. En primer lugar, que cada indicador recibe dos influencias, por una parte la del factor o constructo que se supone que mide, y, por otra, del término de error de medida. Las relaciones entre las variables observables o indicadores y las variables latentes se especifican a partir de unos coeficientes que en este caso se denominan cargas factoriales (“factor loading”), y representan, como ya hemos dicho, los efectos de las variables latentes o factores sobre los indicadores. Estos coeficientes se representan en un sistema de ecuaciones que reciben el nombre de “ecuaciones factoriales”. En segundo lugar, los términos de error que se incluyen, representan los errores aleatorios del tipo estimado por los coeficientes de fiabilidad, es decir, los “errores de medida” de las variables observables y todas las fuentes de varianza residual que no están explícitamente representadas en el modelo, es decir, otros aspectos que mide el indicador además del factor. Estos términos de error de medida se asume que son independientes entre sí y de los factores. Y, en tercer lugar, también se pueden incluir asociaciones entre las variables latentes que se estiman pero no se analizan en el modelo.

En el *Modelo de Ecuaciones Estructurales* hay que tener en cuenta que habrá que especificar por una parte el componente de medida y por otra el componente estructural. Este modelo está compuesto por tres sistemas de ecuaciones. Primero, dos sistemas de “ecuaciones factoriales” en los que se especifican las cargas factoriales correspondientes al componente de medida de las variables exógenas y endógenas. Y el tercero, un sistema de “ecuaciones estructurales” en el que se especifican los coeficientes estructurales que se incluyen en el componente estructural donde se plantean los efectos de las variables latentes exógenas sobre las variables latentes endógenas.

En este caso, se incorporan al modelo las perturbaciones correspondientes a cada una de las variables endógenas latentes, y los errores de medida correspondientes a todas y cada una de las variables observables o indicadores. Por lo tanto, y a diferencia del Modelo de



Ecuaciones Simultáneas, las perturbaciones de las variables endógenas representan o incluyen únicamente las posibles fuentes de variación o causas omitidas de las variables endógenas, y no los errores de medida, que ahora sí se representan en el modelo.

Una vez que se ha especificado el modelo, se aborda el problema de la **Identificación del Modelo** que es sin duda uno de los aspectos más arduos y difíciles en este tipo de modelos, pero absolutamente imprescindible para poder llevar a cabo las fases siguientes. De un modo muy general, la identificación del modelo se traduce en la exigencia de que el número de parámetros a estimar ha de ser igual o inferior al número de unidades de información de que disponemos en la matriz de varianzas-covarianzas obtenida a partir de los datos muestrales, lo que se denomina condición de orden. Este aspecto estaría en relación con la necesidad de plantear modelos parsimoniosos. Además, en los Modelos con variables latentes se añade la condición de que cada factor o variable latente ha de ser medida por tres o más indicadores, y como no son medidas directamente se les ha de asignar una escala de medida o métrica, lo que se denomina condición de rango.

Por lo que respecta al Modelo de Ecuaciones Estructurales, la evaluación de si el modelo es identificado y la subsecuente estimación se ha de realizar separadamente para cada submodelo, siguiendo los pasos que hemos señalado en cada caso, y teniendo en cuenta que se precisa un Modelo de Medida válido antes de evaluar el componente estructural. Por ello, se sigue lo que se denomina modelado en dos pasos. El primer paso consiste en buscar un Modelo de Medida aceptable, y una vez encontrado, el segundo paso es abordar la evaluación del Modelo Estructural.

En caso de seguir adelante con un modelo no adecuadamente identificado no se podrán obtener estimaciones únicas de los parámetros a partir de los datos obtenidos en la matriz de covarianzas muestral, sino que se detendrá el proceso de estimación o se producirán resultados ilógicos.

Una vez que se comprueba que el modelo está identificado, se ha de comprobar si el modelo es apropiado para los datos particulares a los que se va a aplicar, es decir, es necesario comprobar si los datos cumplen los supuestos y las condiciones de aplicación del modelo, lo que se denomina fase de diagnóstico (Apodaka y Páez, 1992). En estos Modelos, los supuestos son básicamente los mismos que en los Modelos de Regresión salvo dos excepciones: la primera se refiere a las perturbaciones o errores, que aunque en ambos casos generalmente se asume que las variables exógenas son independientes de las perturbaciones, en estos modelos se pueden plantear correlaciones entre variables exógenas y las perturbaciones de otras variables endógenas a las que afectan las primeras; y la segunda excepción es que en estos modelos se asume normalidad multivariada de las variables observadas endógenas y exógenas.

En la siguiente fase, se planteará la **Estimación de los Parámetros del Modelo** o coeficientes planteados en los diagramas de paso a partir de los datos obtenidos en las variables observadas en la muestra del estudio, concretamente a partir o bien de la matriz de varianzas-covarianzas o bien de la matriz de correlaciones. Habrá que optar por una matriz u otra en función de si el objetivo prioritario de la investigación es la contrastación de un modelo teórico o el análisis de las relaciones individuales entre las variables.

En estos modelos, los métodos de estimación de parámetros más utilizados son aquellos que asumen la normalidad multivariada. En concreto, el más utilizado es el de Máxima Verosimilitud (ML) que realiza una estimación simultánea de todos los coeficientes de forma iterativa hasta que las diferencias entre las covarianzas basadas en las estimaciones y las covarianzas observadas son mínimas. Otros procedimientos de estimación que también se

utilizan son los de Mínimos Cuadrados Generalizados (GLS) y Mínimos Cuadrados no Ponderados (ULS).

La última y definitiva fase del proceso será la ***Evaluación del Ajuste del Modelo***. En este último paso, se trata de comprobar la compatibilidad entre el modelo propuesto y la información empírica recogida, es decir, en qué medida el modelo teórico propuesto es apoyado por los datos muestrales obtenidos. Será el momento de examinar las estimaciones infractoras (errores estándar elevados, coeficientes estandarizados que sobrepasan la unidad, correlaciones elevadas entre dos estimaciones, etc.). También habrá que tener en cuenta unas medidas o índices de la calidad del ajuste modelo en su conjunto, como son las medidas de ajuste absoluto o global, las de ajuste incremental y las de parsimonia.

Las medidas de ajuste absoluto evalúan el ajuste global del modelo, sin tener en cuenta el posible "sobreajuste". De ellas, quizá el estadístico y la prueba de significación más utilizada en estos modelos sea chi-cuadrado. Sin embargo, este estadístico presenta dos problemas, primero como índice no tiene un límite superior, y, por lo tanto, su valor no se puede interpretar de una forma estandarizada; y, segundo, como prueba de significación es muy sensible al tamaño de la muestra. Por ello, se han propuesto (Jöreskog, 1974, 1977; Jöreskog&Sörbom, 2001) otros índices de ajuste estandarizados que oscilan entre 0 y 1, y que son menos sensibles al tamaño de la muestra como el Índice de Bondad de Ajuste (GFI: "Goodness of Fit Index"), análogo en su significado e interpretación al coeficiente de determinación del Análisis de Regresión; y otros índices basados en los residuos como el Residuo Cuadrático Medio (RMR o RMSR: "Root Mean Square Residual") o Error de Aproximación Cuadrático Medio (RMSEA: "Root Mean Square Error of Approximation").

Por su parte, las Medidas de Ajuste Incremental o Comparativo, como su nombre indica, comparan el modelo propuesto con el modelo de independencia o de ausencia de relación entre las variables. Se han propuesto además, del Índice Ajustado de Bondad de ajuste (AGFI), que es una corrección del índice GFI, otros índices (Bentler, 1988, 1995): El Índice de Ajuste Comparativo (CFI: "Comparative Fit Index"), el Índice de Ajuste Normalizado (NFI: "Bentler-Bonnet Normed Fit Index") y el Índice de Ajuste No Normalizado (NNFI: "Bentler-Bonnet Non-Normed Fit Index"). Todos ellos indican la proporción de mejora que supone el ajuste global del modelo propuesto en relación con el modelo nulo en el que se asume que las variables observadas no están relacionadas. Por esta razón, estos índices se denominan índices de incremento del ajuste.

Por último, las Medidas de Ajuste de Parsimonia, que evalúan la calidad del ajuste del modelo en función del número de coeficientes estimados para conseguir dicho nivel de ajuste (Akaike, 1987). Cuanto mejor ajuste pueda conseguirse con un menor número de coeficientes mayor confianza se puede tener en que los resultados no son producto del sobreajuste de los datos. En este caso, se compara el modelo propuesto con un modelo sobreidentificado o saturado, en el que se plantean todas las posibles relaciones entre las variables. Este último tipo de medida se utiliza sobre todo para la comparación de modelos alternativos. Se considera como mejor índice de parsimonia el Criterio de Información de Akaike (AIC: "Akaike Information Criterion"), además de las modificaciones de los estadísticos e índices anteriores como: el índice basado en la chi-cuadrado normalizada (NC: "Normed Chi-Squared"), el Índice de Ajuste Parsimonioso (PFI: "Parsimonius Fit Index"), basado en el índice NFI, o el Índice de Calidad de Ajuste de Parsimonia (PGFI: "Parsimony Goodness of Fit Index"), a su vez basado en el Índice GFI.

Como cada uno de estos índices refleja un aspecto particular del ajuste del modelo, un buen valor de un índice no indica por sí solo un buen ajuste. Es necesario revisar los resultados de diferentes índices que aportan informaciones complementarias. En resumen, para

poder hablar de un buen ajuste, la prueba de significación chi-cuadrado sobre el ajuste del modelo ha de ser no significativa; los índices de bondad de ajuste que indican la proporción absoluta o relativa de covarianza observada explicada (GFI, AGFI, NFI, NNFI, CFI) han de ser superiores a 0,90; y los índices basados en los residuos estandarizados (RMR, RMSA) inferiores a 0,10.

No obstante, hay que tener en cuenta tres limitaciones de estos índices. Primero, sus valores indican únicamente el ajuste global del modelo. Puede darse el caso de que algunas partes del modelo se ajusten mal a los datos, incluso aunque el valor del índice general sea bueno. Segundo, los índices de ajuste no indican si los resultados son teóricamente significativos. Por ejemplo, los signos de algunos coeficientes pueden ser de signo contrario al esperado, e incluso aunque los valores de ajuste sean buenos, estos resultados anómalos han de ser explicados. Finalmente, unos buenos índices de ajuste no indican que el poder predictivo del modelo sea también elevado. Por ejemplo, las perturbaciones de modelos con índices de ajuste buenos pueden ser elevadas, lo que significa que el modelo refleja la relativa ausencia de validez predictiva entre las variables.

Será el momento de examinar las relaciones individuales planteadas (los coeficientes de determinación de cada ecuación estructural, los coeficientes estructurales estandarizados y la significación estadística de los parámetros estimados), porque aunque las medidas de ajuste del modelo en su conjunto indiquen que es aceptable, algunas relaciones planteadas pueden carecer de sentido. Para ello, es necesario revisar las diferentes partes específicas del modelo a partir de los residuos de las correlaciones, que son las diferencias entre las correlaciones observadas y las implicadas en el modelo. Una regla que se utiliza en estos modelos es que valores absolutos de los residuos de las correlaciones superiores a 0,10 sugieren que el modelo no explica bien la correlación observada asociada.

Si después de esta fase, el modelo especificado inicialmente es apoyado por los datos muestrales, procederemos a su **Validación e Interpretación** de acuerdo con las hipótesis propuestas y la teoría de partida, y lo utilizaremos para tomar decisiones o realizar previsiones en la variable (Apodaka y Páez, 1992).

En caso contrario, se reformula el modelo inicial para encontrar un modelo con un mejor ajuste, pero siempre teniendo en cuenta la teoría de la que se parte inicialmente. Se sigue el procedimiento de **Búsqueda de Especificación** con el fin de detectar los errores de especificación que hemos cometido. Estos errores pueden ir en dos líneas, por una parte, la inclusión en el modelo de una relación irrelevante, que puede ser detectada a partir de las pruebas de significación de los coeficientes estructurales; y, por otra, la omisión de variables relevantes, que puede determinarse a partir los residuos estandarizados y de los índices de modificación. No obstante, destacar que además del criterio estadístico de la significación, para la eliminación o inclusión de las relaciones entre variables se ha de tener en cuenta el interés teórico o sustantivo. Así, aunque una relación no sea significativa estadísticamente si su interés teórico es suficiente debería mantenerse en el modelo; y del mismo modo, este criterio teórico ha de primar para incluir nuevas relaciones, sólo se incluirán aquellas que además de ser significativas tengan un significado sustantivo y práctico.

En conclusión, y como hemos podido comprobar, la fase de mayor importancia es la de especificación del modelo, ya que los modelos lineales son bastante robustos a las violaciones de los supuestos con dos excepciones, el error de especificación y el de medida. La plausibilidad del desarrollo de las fases posteriores depende de una correcta especificación, aspecto éste que está en manos del investigador, que ha de establecer el diagrama de secuencias de relaciones de dependencia o diagrama de paso y, posteriormente y en función de este diagrama, especificar los parámetros que se han de estimar. La hipótesis del investigador se expresa

en la forma de una serie de ecuaciones. Estas ecuaciones definen los parámetros del modelo, que corresponden a las relaciones que se plantean entre las variables y que el programa estadístico (AMOS: "Analysis of Moment Structure" (Arbuckle, 1997), EQS: "Structural EQUations Program" (Bentler, 1995) o LISREL: "Linear Structural Relationship" (Jöreskog&Sörbom, 2001)) estimará con los datos de la muestra.

En resumen, una característica de este enfoque es que se centra en la teoría y su contraste, integrando distintos métodos multivariantes. El análisis causal no permite descubrir causas, ni trata de establecer relaciones de causalidad entre variables, sino que se limita a evaluar modelos causales elaborados por el investigador basándose en una teoría o modelo teórico, y establecer su plausibilidad comparándolos con otros modelos alternativos. Cuando dos o más modelos son congruentes con un conjunto de datos, la decisión acerca de cual de los modelos o patrones es más plausible se apoya en la teoría que está en la base del modelo, no en los datos. Únicamente se puede concluir que un conjunto de relaciones observadas son consistentes con un determinado modelo causal propuesto, pero no se puede descartar que exista otro modelo que reproduzca las relaciones observadas.

En definitiva, el requerimiento fundamental se sitúa a dos niveles: a nivel teórico, debe haber una base suficientemente sólida que justifique el paso de relaciones de covarianza a relaciones causa-efecto; y a nivel metodológico, el investigador debe ser capaz de especificar matemáticamente modelos que representen de forma adecuada los mecanismos teóricos involucrados en los datos, decidiendo qué parámetros se han de fijar a partir de la teoría y cuáles han de estimarse a partir de los datos.

## REFERENCIAS BIBLIOGRÁFICAS

- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, 52, 317-332.
- Apodaka, P. y Paez, D. (1992). Los modelos estructurales: posibilidades y limitaciones. En D. Paez, J. Valencia, J.F. Morales, B. Sarabia y N. Ursua (Eds.), *Teoría y Métodos en Psicología Social* (pp. 249-287). Barcelona: Anthropos.
- Arbuckle, J. (1997). *AMOS User's Guide. Version 3.6*. Chicago, IL: Smallwaters Corporation.
- Asher, H.B. (1983). *Causal modeling (2nd. ed.)*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-003. Beverly Hills, CA: Sage.
- Bentler, P. M. (1988). Comparative fit indices in structural models. *Psychological Bulletin*, 107, 238-246.
- Bentler, P. M. (1995). *EQS Structural Equations Program Manual*. Encino, CA: Multivariate Software.
- Blalock, H.M., jr. (Ed.) (1985). *Causal models in the Social Sciences (2nd. ed.)*. New York, NY: Aldine Publishing Company.
- Bollen, K.A. (1989). *Structural Equations with Latent Variables*. New York, NY: John Wiley & Sons.
- Cook, T.D. y Campbell, D.T. (1986). The casual assumptions of quasi-experimental practice. *Synthese*, 68, 141-160.
- Díez Medrano, J. (1992). *Métodos de análisis causal. Cuadernos metodológicos, nº 3*. Madrid: Centro de Investigaciones Sociológicas.
- Gómez Benito, J. (1986). *Los modelos causales como metodología de validez de constructo*. Barcelona: Alamex.

- Jöreskog, K.G. (1974). Analyzing psychological data by structural analysis of covariance matrices. En R.C. Atkinson, D.H. Krantz, R.D. Luce y P. Suppes (Eds.), *Contemporary developments in Mathematical Psychology. Volume II* (pp. 1-56). San Francisco: Freeman.
- Jöreskog, K.G. (1977). Structural equations models in the Social Sciences: Specification, estimation and testing. En P.R. Krishnaiah (Ed.), *Applications of Statistics* (pp.265-287). Amsterdam: North-Holland.
- Jöreskog, K.G.&Sörbom (2001). *LISREL VIII. Analysis of Linear Structural Relationships by Maximum Likelihood and Least Squares Methods. User's Guide*. Lincolnwood, IL : Scientific Software International.
- Kenny, D.A. (1979). *Correlation and causality*. New York, NY: John Wiley & Sons.
- Kline, R.B. (1998). *Principles and Practice of Structural Equation Modeling*. New York, NY: Guilford Press.
- Long, J.S. (1983a). *Confirmatory Factor Analysis*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-033. Beverly Hills, CA: Sage.
- Long, J.S. (1983b). *Covariance Structure Models: An Introduction to LISREL*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-034. Beverly Hills, CA: Sage.
- López Feal, R. (1982). Los modelos causales asociados a estructuras de covarianza: Una vía para la unificación de los "dos métodos de la psicología científica". *Anuario de Psicología*, 2, 2-23.
- Maruyama, G.M. (1998). *Basics of Structural Equation Modeling*. London, UK: SAGE.
- Mueller, R.O. (1997). Structural equation modeling: Back to basics. *Structural Equation Modeling*, 4, 353-369.
- Mulaik, S.A. (1986). Toward a synthesis of deterministic and probabilistic formulations of causal relations by functional relation concept. *Philosophy of Science*, 53, 313-332.
- Schumacker, R.E. & Lomax, R.G. (1996). *Beginner's guide to Structural Equation Modeling*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Visauta, B. (1986). *Técnicas de Investigación Social. Modelos Causales*. Barcelona: Hispano Europea.



# MODELOS DE MEDIDA Y ANÁLISIS FACTORIAL CONFIRMATORIO

RAMÓN FERNÁNDEZ PULIDO

Área de Metodología de las Ciencias del Comportamiento  
Facultad de Psicología (Universidad de Salamanca)

## INTRODUCCIÓN

El Análisis Factorial se refiere a un grupo de técnicas de análisis de datos diseñadas para identificar factores o dimensiones inobservables empíricamente y que pueden explicar las relaciones que existen entre un grupo de variables observadas. Cada uno de esos factores o dimensiones constituyen un constructo y las variables observadas que están relacionadas con él son sus indicadores.

Como planteamiento teórico general podríamos decir que nos encontramos con un conjunto de variables manifiestas u observadas empíricamente (puntuaciones en ítems, tests o cualquier otro tipo de fuente de variación) y otro conjunto de variables latentes (inobservables empíricamente) que son lo que llamamos factores o constructos. Se supone que esos constructos son los que determinan o “causan” el comportamiento de los sujetos en las variables observadas (por eso cuando, en un diagrama, establecemos un posible modelo de relaciones entre variables manifiestas y latentes, las flechas parten de las latentes y llegan hasta las manifiestas). Los factores serían en este sentido como las variables independientes de la regresión múltiple, con la diferencia de que son latentes, por lo que hay que estimar su número y características; y para hacerlo utilizamos las técnicas del Análisis Factorial (Borsboom, Mellenbergh y van Haerden, 2004).

Los orígenes de estas técnicas, en su formalización conceptual, se remontan a principios del siglo pasado cuando Spearman (1904) los desarrolla para probar su teoría del *factor general de la inteligencia*. Sin embargo, debido a la complejidad de los cálculos exigidos por estas técnicas, no es hasta bien entrada la segunda mitad del siglo XX cuando su uso se generaliza continuando su auge en la actualidad gracias, principalmente, al fácil acceso de los investigadores al uso de los ordenadores y al desarrollo de programas informáticos asequibles.

Dentro de la investigación psicológica podríamos clasificar en tres grandes grupos los tipos de estudios para los que se emplea el Análisis Factorial (Thompson, 2004):

- Estudios de evidencias acerca de la validez de constructo de los instrumentos de medida.
- Estudios diseñados para el desarrollo de teorías sobre constructos psicológicos (gran cantidad de teorías de la inteligencia y de la personalidad se han desarrollado con la ayuda del Análisis Factorial).

- Estudios orientados a reducir una gran cantidad de variables a un número menor de ellas con el fin de hacer más manejable e inteligible su uso en otros estudios subsecuentes.

Aunque la familia de técnicas que constituye el Análisis Factorial es mucho más amplia, en líneas generales conviene distinguir entre dos grandes tipos: **El Análisis Factorial Exploratorio** (en adelante **AFE**) y el **Análisis Factorial Confirmatorio** (en adelante **AFC**).

Aún existiendo muchas semejanzas entre las dos aproximaciones, sus diferencias, tal y como se desprende de la comparativa entre las figuras 1 y 2, también son notorias siendo las siguientes las más significativas (Pedhazur y Pedhazur, 1991; Martínez, Hernández y Hernández, 2006):

- La primera y fundamental se desprende del propio nombre de ambas orientaciones: El AFE trata de *explorar* la posible existencia de una estructura de factores o dimensiones latentes partiendo de las correlaciones entre las variables observadas o empíricas; por su parte el AFC se focaliza en la *comprobación* de la existencia de una teoría o hipótesis establecida a priori mediante la formalización de un modelo de medida que se pone a prueba. Un ejemplo nítido de ámbito de aplicación del AFC se encuadra dentro del proceso de adaptación de un test transcultural. Generalmente la evidencia de validez de un test adaptado suele ser la replicación de la estructura factorial en la población de destino para comprobar si se ajusta a la estructura planteada en origen; el AFC es una técnica idónea para este objetivo.
- En segundo lugar conviene subrayar que, dado su carácter exploratorio, el AFE no presupone un número determinado de factores. Como punto de partida se asume, en líneas generales, que puede haber tantos factores como variables observadas, aunque se espera que ese número sea menor porque si no no tendría sentido el análisis, y que todas las variables observadas tienen pesos factoriales en todos los factores (independientemente de que esos pesos sean o no significativos). En la figura 1 se denota contemplando los factores 1, 2... y k, entendiendo por k un número indeterminado y, como podemos observar, lanzando vectores desde todos los factores a todas las variables. Por el contrario, el AFC, tal y como observamos en la figura 2, plantea un número concreto de factores (en este caso 3) y además el investigador especifica qué indicadores o variables observadas pesan en cada uno de esos factores. De hecho, uno de los usos específicos del AFC es como prueba complementaria de validación de los resultados de un AFE. Muchos autores, aunque también existen detractores (Pérez-Gil, Chacón y Moreno, 2000), recomiendan que, para validar los resultados de un AFE, se proceda a someter a prueba la solución obtenida mediante la aplicación posterior de un AFC.
- En tercer lugar, en el AFE la posibilidad de que los factores estén relacionados entre sí o sean independientes se contempla como un todo. Si se realiza una rotación ortogonal se supone que los factores son independientes entre sí y si se lleva a cabo una rotación oblicua lo que se está asumiendo es que todos los factores están interrelacionados. En el AFC se puede plantear como modelo hipotético uno en el que algún factor guarde relación con otro u otros y algún otro sea independiente (en la figura 2 se plantea un modelo que establece que los factores 1 y 2 están relacionados mientras que el factor 3 es independiente).
- En cuarto lugar, en el AFE se asume que los factores que corresponden al término de errores (o unicidades) no están correlacionados. Por el contrario, en el AFC (en la figura 2 podemos observar los conectores entre los términos de error de las variables

Figura 1. Ejemplo de Análisis Factorial Exploratorio

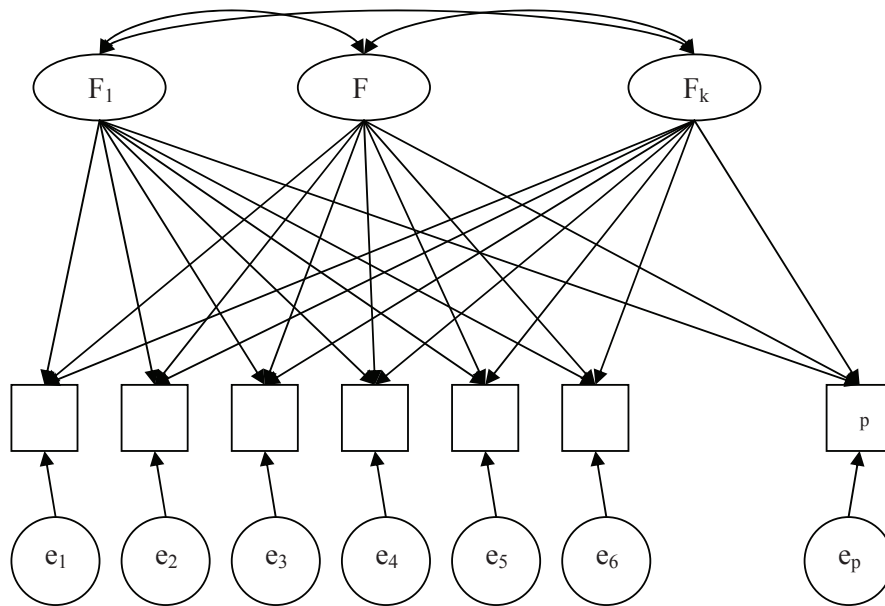
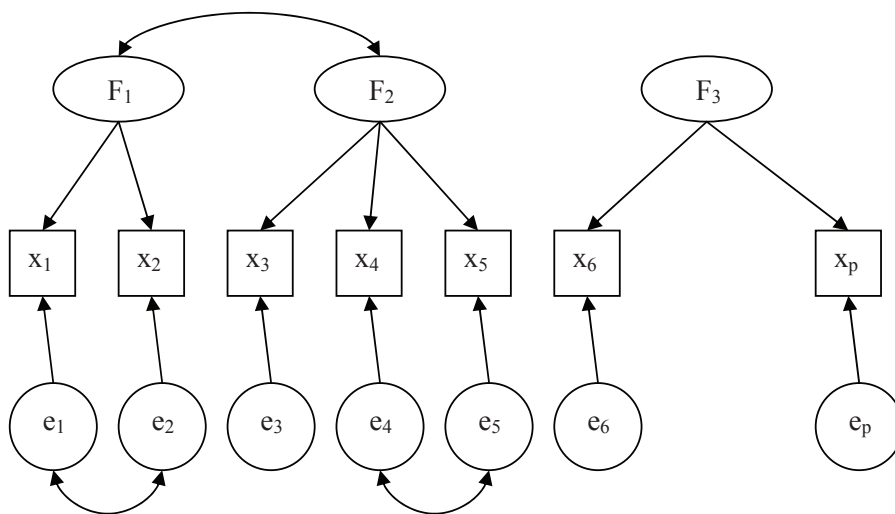


Figura 2. Ejemplo de Análisis Factorial Confirmatorio



observadas 1 por parte, los errores 4 por otra) las correlaciones entre los términos de error se pueden contemplar como parte del modelo que se somete a prueba.

Todas estas diferencias pueden resumirse en la clasificación que la mayoría de los autores establecen de las técnicas multivariantes (Hair, Anderson, Tatham y Black, 1999) en las que incluyen tanto al AFE como al AFC:

- El AFE se contempla como una técnica multivariante especialmente indicada para el estudio de las relaciones de *interdependencia* (modelo correlacional) entre un conjunto de variables.



- Por el contrario, el AFC se enmarca dentro de los *Modelos de Ecuaciones Estructurales* (generalmente conocidos por las siglas en inglés, SEM) como una técnica multivariante diseñada para el estudio de las múltiples relaciones de *dependencia* (modelos causales) entre diferentes conjuntos de variables dependientes e independientes. Dentro de estos modelos, también conocidos como modelos de estructuras de covarianza, el AFC se refiere específicamente al llamado *Modelo de Medida* que trata de explicar de qué manera un conjunto de variables medidas empíricamente son un reflejo de otras variables latentes, es decir, no observables empíricamente (Martínez et al., 2006).

En esta pequeña charla introductoria del AFC pretendemos poner al interesado en contacto con los fundamentos teóricos de esta técnica sin entrar en los entresijos más formales ni especificaciones técnicas que exigen un cierto dominio de álgebra de matrices y teorías estadísticas de cierta complejidad. Perseguimos dos grandes objetivos: (a) facilitarles la lectura crítica y la comprensión de aquellos problemas de investigación que se recogen en la literatura científica y en los que se usa este tipo de técnicas y (b) ayudarles a empezar a aprender a aplicarlas. Para una mayor profundidad del tema recomendamos lecturas tales como los trabajos de Bohrnstedt (1983), Bollen (1989), Long (1983), o Batista y Coenders (2000).

Iniciaremos la especificación del modelo con una mención especial a la matriz de varianzas-covarianzas y la elección de la matriz de análisis. En segundo lugar, desarrollaremos aspectos relacionados con la identificación del modelo. Posteriormente, pasaremos al punto central de estimación de los parámetros del modelo deteniéndonos específicamente en el procedimiento de Máxima Verosimilitud por ser el más utilizado entre los investigadores. Continuaremos con la evaluación del ajuste de los modelos para terminar con la reespecificación de los modelos y la comparación de modelos anidados.

## ANÁLISIS FACTORIAL CONFIRMATORIO

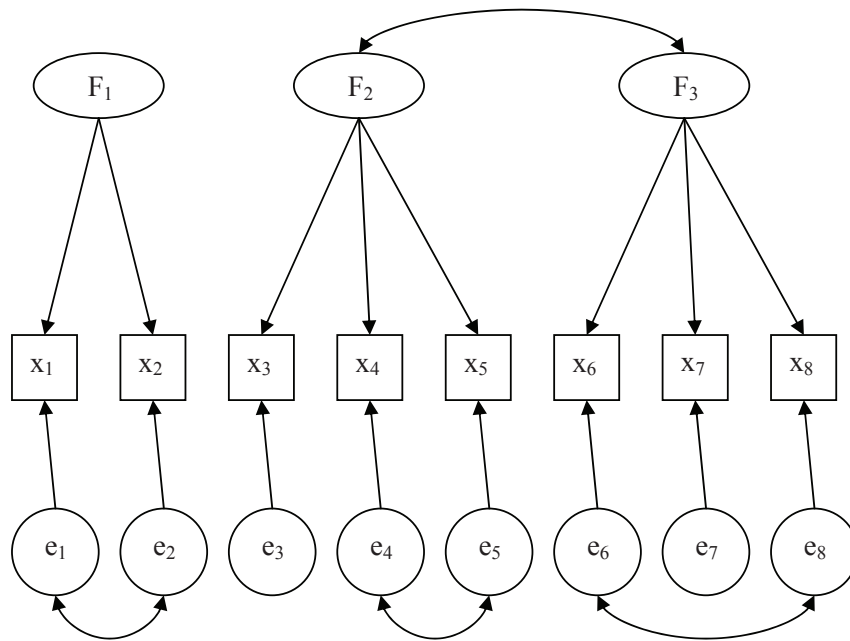
### 1. Especificación del modelo

Como hemos dicho anteriormente, en el AFC tratamos de probar una hipótesis. Esta hipótesis consiste en el planteamiento de un modelo teórico que recoja las relaciones específicas que se establecen entre factores comunes (variables latentes), factores únicos (errores) y medidas empíricas (variables empíricas medidas u observadas). Estas expectativas deben plasmarse en un modelo que establezca aquellos parámetros que deben ser estimados (de todos los posibles que se puedan estimar).

A modo de ejemplo, supongamos la figura siguiente:

En este ejemplo disponemos de 8 variables empíricas ( $x_1 \dots x_8$ ). Suponemos que esas variables pueden explicarse, en parte, a partir de 3 factores latentes ( $F_1 \dots F_3$ ) y lo que no explican esos factores constituye lo que se conoce como error (factores únicos o unicidad de cada variable:  $e_1 \dots e_8$ ). En teoría, podríamos estimar (liberar) un total de 63 parámetros:  $24 + 3 + 8 + 28$ ; en este desglose: 24 son las posibles saturaciones de las 8 variables en los 3 Factores comunes; 3 son las posibles correlaciones entre esos 3 Factores; 8 son los diferentes valores de las unicidades ( $e$ ) y 28 las posibles covarianzas entre esos factores únicos o unicidades ( $8 \times 7 / 2$ ). De estos 63 posibles parámetros, en la especificación del ejemplo, sin embargo, sólo hemos liberado (para que sean estimados) 20 parámetros:  $8 + 1 + 8 + 3$ ; es decir, 8 saturaciones de las variables en los 3 factores (2 variables saturan el factor 1, 3 el factor 2 y otras 3 el factor 3); sólo especificamos una relación significativa entre los factores 2 y 3; contemplamos

Figura 3. Ejemplo de Path Diagram o Diagrama de Relaciones



los 8 posibles valores de las unicidades ( $e_1$  con  $x_1$ ,  $e_2$  con  $x_2$  etc.) y finalmente sólo recogemos 3 posibles covarianzas entre los errores ( $e_1$  con  $e_2$ ,  $e_4$  con  $e_5$  y  $e_6$  con  $e_8$ ).

En definitiva, la especificación de un modelo requiere establecer qué parámetros, de todos los posibles, queremos liberar para que sean estimados y sean los definitorios de ese modelo. Por lo tanto, la especificación de un modelo requiere:

- Establecer el número de factores latentes (en el ejemplo, 3).
- Fijar las relaciones que suponemos significativas entre las variables empíricas y los factores latentes (en el ejemplo, 8)
- Fijar la posible relación significativa que puede existir entre los distintos factores latentes (en el ejemplo, 1)
- Establecer las relaciones entre los factores únicos (error) y las variables empíricas (observadas) (en el ejemplo, 8).
- Fijar la posible existencia de covarianzas entre los factores únicos (en el ejemplo, 3).

Cualquier modelo puede especificarse de dos formas:

- Una forma gráfica llamada *Path Diagram* o *Diagrama de Relaciones* (tal y como hemos planteado en la figura 3 que nos sirve de ejemplo) que convencionalmente recoge, dentro de círculos u óvalos, las variables latentes o factores y en cuadrados o rectángulos, las variables observadas; las relaciones de dependencia se establecen mediante flechas unidireccionales que van de los factores a las variables (ya que como modelos causales se espera que dichas variables dependan de ellos) y las correlaciones o las covarianzas se representan mediante flechas bidireccionales.
- Un conjunto de ecuaciones lineales que, siguiendo la notación clásica, puede definirse en la siguiente *ecuación matricial*:

$$x = \Lambda_x \xi + \delta$$

donde  $x$  es un vector p.1 de puntuaciones diferenciales (distancia a su respectiva media) en las variables observadas;  $\Lambda_x$  (Lambda) es una matriz p.k de saturaciones de las p variables en los k factores comunes;  $\xi$  (Ksi) es un vector k.1 (también con puntuaciones diferenciales) de factores comunes y  $\delta$  (delta) es un vector p.1 de factores únicos (componentes específicos y del error de cada una de las variables observadas).

Desarrollando esta ecuación matricial, junto con las matrices de varianzas-covarianzas de los factores comunes  $\xi$  (matriz  $\Phi$  (Fi) de coeficientes  $\phi$ ) y de los factores únicos  $\delta$  (matriz  $\Theta$  (Zeta) de coeficientes  $\theta$ ), aplicándola a nuestro ejemplo, tendríamos lo siguiente (ver Tabla 1):

Tabla 1. Desarrollo de la ecuación matricial correspondiente al ejemplo planteado

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 & 0 \\ \lambda_{21} & 0 & 0 \\ 0 & \lambda_{32} & 0 \\ 0 & \lambda_{42} & 0 \\ 0 & \lambda_{52} & 0 \\ 0 & 0 & \lambda_{63} \\ 0 & 0 & \lambda_{73} \\ 0 & 0 & \lambda_{83} \end{bmatrix} \cdot \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \end{bmatrix}$$

$x \qquad \Lambda_x \qquad \cdot \qquad \xi \qquad + \qquad \delta$

$$\begin{bmatrix} \phi_{11} & & \\ 0 & \phi_{22} & \\ 0 & \phi_{32} & \phi_{33} \end{bmatrix}$$

$\Phi$

$$\begin{bmatrix} \theta_{11} & & & & & & & & \\ \theta_{21} & \theta_{22} & & & & & & & \\ 0 & 0 & \theta_{33} & & & & & & \\ 0 & 0 & 0 & \theta_{44} & & & & & \\ \theta & 0 & 0 & \theta_{54} & \theta_{55} & & & & \\ 0 & 0 & 0 & 0 & 0 & \theta_{66} & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \theta_{77} & & \\ 0 & 0 & 0 & 0 & 0 & \theta_{86} & 0 & \theta_{88} \end{bmatrix}$$

$\Theta$

Con la aplicación de esta ecuación, junto con las matrices Fi y Zeta, se establece una matriz teórica "Sigma" ( $\Sigma$ ) de varianzas-covarianzas deducidas a partir del modelo propuesto. En la medida en que esta matriz se ajuste a la matriz empírica de varianzas-covarianzas (S) de las variables observadas podremos afirmar que el modelo factorial propuesto es bueno.

A la hora de elegir la matriz para el análisis de los datos podemos optar entre la matriz de varianzas-covarianzas o la matriz de correlaciones. Aunque en el fondo son lo mismo (no

olvidemos que una matriz de correlaciones no es más que una matriz de varianzas-covarianzas tipificadas, o lo que es lo mismo, una correlación es una covarianza tipificada), la estimación de los parámetros del modelo puede variar según se elija una u otra. La mayoría de los autores no se ponen de acuerdo en cuál es la mejor como punto de partida. Desde un punto de vista más pragmático, nosotros recomendamos partir de la matriz de varianzas-covarianzas si tenemos en cuenta que la mayoría del software comercial que ejecuta estos análisis permite la flexibilidad de operar tanto con la matriz de covarianzas como con la de correlaciones, con lo que partiendo de la matriz de covarianzas, siempre podremos transformarla en matriz de correlaciones, mientras que si partimos de la matriz de correlaciones directamente nunca podremos transformar a la de covarianzas.

Así, tal y como hemos planteado con anterioridad, si queremos hacer una valoración del ajuste del modelo, dado que este ajuste se realiza comparando la matriz de covarianzas reproducida o teórica con la empírica, parece lo más lógico que se use para el análisis la matriz de covarianzas. Posteriormente, si interesa un análisis de las relaciones (pesos factoriales) que existen entre los factores y las variables observadas, lo indicado será operar con la matriz de correlaciones porque al estar tipificada facilita la interpretación.

## 2. Identificación del modelo

Aunque la mayor parte del software disponible en la actualidad aporta siempre información sobre el grado de identificación del modelo antes de proceder a la estimación de los parámetros, es interesante que comprendamos en qué consiste este paso de la identificación si tenemos en cuenta que, en el caso de encontrarnos con un modelo no identificado, deberemos imponer las restricciones pertinentes al número de parámetros liberados para conseguir un modelo *sobreidentificado* (que es el que en la práctica podemos someter a prueba).

La identificación tiene que ver con el concepto de grados de libertad. Para entendernos de una forma sencilla diremos que a cada parámetro que propongamos para su estimación (liberemos para ser calculado) le corresponde un grado de libertad. En nuestro ejemplo, tal y como dijimos anteriormente, hemos liberado 20 parámetros, por lo que consumimos 20 grados de libertad. Esos grados de libertad hay que deducirlos del total posible de grados de libertad disponibles según nuestros datos empíricos y que vienen dados por el número de combinaciones que permitan las variables empíricas u observadas en función de la matriz

empírica de varianzas y covarianzas  $GL_{disponibles} = \frac{v(v+1)}{2}$ . En nuestro ejemplo, si plantea-

mos un análisis con 8 variables observadas, el número de grados de libertad disponibles es de 36 ( $8 \times 9 / 2$ ). Pues bien, los grados de libertad con los que operamos en este modelo concreto vienen dados por la diferencia entre los grados de libertad disponibles, y los consumidos en función del número de parámetros especificados. En nuestro caso, el modelo que sometemos a prueba tendría 16 grados de libertad (36 disponibles menos 20 consumidos).

Cuando los grados de libertad del modelo son *negativos*, es porque definimos un modelo con un número de parámetros que exceden los grados de libertad disponibles, dando lugar a los llamados modelos *No Identificados* (los parámetros pueden tomar infinitos valores con lo que se hace imposible su estimación). Si los grados de libertad del modelo fuesen 0, es decir, el número de parámetros propuesto coincide con los grados de libertad disponibles, nos encontraríamos con un modelo *Saturado*. La particularidad de este modelo es que reproduce exactamente los datos empíricos que se están analizando. No cumple su cometido de simplificación de los datos observados y es un modelo único, sin alternativa de ningún tipo. Sólo cuando los grados de libertad del modelo es un número mayor que 0, nos encontramos con

un modelo *Sobreidentificado*. Esto implica que el modelo propuesto es uno más entre varios modelos posibles alternativos. Cabrían varias soluciones para los mismos datos y en esta situación es cuando tiene objeto plantear el análisis de los datos para comprobar si la solución propuesta (el modelo que sometemos a prueba) minimiza las diferencias entre la matriz de varianzas y covarianzas empírica y la teórica que se deduce del modelo propuesto.

Como vemos, en el caso de que el programa informático utilizado nos informe de un modelo no identificado, es tarea del investigador (siempre que desde el punto de vista teórico se justifique) redefinir el modelo propuesto imponiendo restricciones a algunos parámetros liberados en el modelo previo hasta conseguir un modelo sobreidentificado.

Debido a que los factores propuestos en el modelo son variables latentes, como ya hemos apuntado con anterioridad, su métrica (escala de medida) está indeterminada. En esta fase de la identificación del modelo hay que especificar, también, la métrica de esas variables. Generalmente se escogen escalas de medida fáciles de interpretar. En el AFC suelen seguirse dos procedimientos: el primero de ellos consiste en fijar, para cada factor propuesto, un valor para uno de sus coeficientes  $\lambda$  (es decir el correspondiente a una de las variables asociadas a ese factor). Generalmente se atribuye el valor 1 a ese coeficiente, con el fin de escalar las variables latentes en valores similares a los tipificados para los coeficientes de correlación; la variable elegida para fijar ese valor es indiferente aunque la mayoría de los autores recomiendan que sea la que, a priori, se suponga que sea un mejor reflejo del factor definido. El segundo procedimiento consiste en fijar la varianza de los factores en un número adecuado dentro de la matriz de varianzas-covarianzas de los factores. El valor preferido es la unidad, ya que así nos dejará una matriz  $\Phi$  de varianzas-covarianzas de los factores con unos en la diagonal (varianzas) y con coeficientes de correlación fuera de la diagonal (covarianzas). Con cada procedimiento se obtienen las mismas estimaciones, por lo que es indiferente usar una u otra solución; si bien algunos autores consideran que, con vistas a contrastar la teoría, el segundo procedimiento (estimar directamente la varianza del constructo) es más recomendable.

### 3. Estimación de los parámetros del modelo

En este paso subsiguiente se procede a la estimación de los parámetros propuestos en el modelo de medida. De lo que se trata es de encontrar aquellos posibles estimadores de los parámetros que reproduzcan de una forma, lo más ajustada posible, la matriz de varianzas-covarianzas calculada a partir de los datos empíricos (matriz  $S$ ). Se asume que a nivel poblacional existe una matriz de varianzas-covarianzas (matriz  $\Sigma$ ) que viene determinada por los parámetros estructurales y que reproduce perfectamente la matriz empírica  $S$ . El problema es que de todos los posibles parámetros que pueden dar lugar a la matriz  $\Sigma$ , ponemos a prueba la estimación de unos determinados por el modelo definido previamente. Esos parámetros serán estimados a partir de las matrices generadas con el modelo de medida propuesto ( $\Lambda_x$ ,  $\Phi$  y  $\Theta_\delta$ ). Todo el problema radica en encontrar unos estimadores  $\hat{\Lambda}_x$ ,  $\hat{\Phi}$  y  $\hat{\Theta}_\delta$  que generen una matriz  $\hat{\Sigma}$  lo más parecida posible a la matriz empírica  $S$ .

Para evaluar el grado de ajuste entre la matrices  $\hat{\Sigma}$  y  $S$  es necesario definir una función de ajuste (algunos autores también la denominan función de discrepancia) que lo que hace es evaluar las distancias entre la matriz de varianzas-covarianzas muestral y la matriz de varianzas-covarianzas ajustada (estimada) en función del modelo especificado. En los modelos SEM existen varias funciones de ajuste dependiendo del procedimiento utilizado para la estimación de los parámetros. Los tres métodos de estimación de parámetros más utilizados son el de **Mínimos Cuadrados no Ponderados** (OLS: Ordinary Least Squares), **Mínimos Cuadrados Generalizados** (GLS: Generalized Least Squares) y **Máxima Verosimilitud** (ML: Maximum Likelihood).

El método de *Mínimos Cuadrados no Ponderados* tiene la limitación de que los valores dependen de la escala de medida de las variables. Al utilizarlo se recomienda trabajar con la matriz de correlaciones para que se unifiquen las variables al tipificar su escala de medida.

El método de *Mínimos Cuadrados Generalizados*, soluciona el problema del anterior dando distintos pesos a las varianzas y covarianzas. Es un método más adecuado para el tipo de análisis propuesto.

En cualquier caso, la mayoría de los autores recomiendan el método de *Máxima Verosimilitud* por sus propiedades asintóticas de carencia de sesgo, consistencia y eficiencia. A medida que el tamaño muestral aumenta, su distribución se aproxima a la distribución normal lo que permite realizar contrastes que se ajusten a ese tipo de distribución. La idea que subyace a un estimador de máxima verosimilitud, es que asumimos que conocemos la forma general de la distribución de una población de la que se extrae una muestra más o menos representativa. Lo que no conocemos son los parámetros de la población que dan a esa distribución esa forma particular entre todas las posibles. Ante ese desconocimiento lo que se hace es tomar valores arbitrarios tratándolos como si fueran los parámetros poblacionales y analizando hasta qué punto son verosímiles esos parámetros. Cuando en este proceso llegamos a establecer aquellos valores de parámetros que son los estimadores más verosímiles para explicar las observaciones empíricas, finaliza el proceso (Mulaik, 1972).

Por lo antedicho, las funciones de ajuste requieren procedimientos iterativos que sólo pueden realizarse con Programas implementados en ordenador. Las iteraciones son procesos parsimoniosos en los que en pasos sucesivos, se van obteniendo los distintos valores paramétricos, hasta que el salto entre dos valores es lo suficientemente pequeño, por lo que se detiene el proceso en un punto de máxima convergencia (máxima verosimilitud). Cuando no se consigue esa convergencia en un número de iteraciones preestablecido, de antemano, conviene corregir técnicamente el proceso, ya que puede depender de la estimación inicial de parámetros, de la diferencia establecida para cada salto o iteración o del número de iteraciones prefijado. Si aún así persiste la no convergencia, puede ser necesario revisar el modelo especificado.

En este punto estamos en disposición de llevar a cabo la estimación de los parámetros del modelo de medida especificado. Una vez obtenidos los parámetros, lo primero que conviene realizar es una inspección visual de la solución obtenida, por si se ha producido una infracción en la estimación. El caso más corriente es el llamado fenómeno *Heywood* (Hair et al., 1999). Este caso consiste en la identificación de valores de parámetros inapropiados; se constata un valor de parámetro  $\lambda$  para una variable superior a la unidad o, como complementario, un valor negativo de parámetro  $\delta$  para esa variable. Este fenómeno no debería producirse porque nos está indicando que el término error (unicidad de esa variable es negativo) o lo que es lo mismo, que el peso del factor en esa variable (en términos de correlación) supera la unidad; lo que empíricamente es imposible desde el punto de vista conceptual del modelo.

La presencia de este fenómeno obliga a reespecificar el modelo original. Esta reespecificación pasa por una solución drástica que es eliminar del análisis la variable infractora (si bien esta solución es bastante inapropiada porque debilita el modelo y lo altera con respecto a la hipótesis de partida). Otra solución pasa por fijar un valor mínimo para la unicidad de la variable (próximo a cero) y mantenerla en el análisis; la lógica de esta segunda solución es que si el peso del factor en esa variable es tan grande que llega a una solución de sobrevaloración del parámetro y sabemos que lo máximo a lo que se puede llegar es a uno, forcemos la situación para que ese valor se fije en uno haciendo que la unicidad, el término de error sea prácticamente cero. En la mayoría de los AFC se opta por esta segunda solución.



Una vez el modelo está reespecificado, lo sometemos a un nuevo proceso de estimación de los parámetros y ya, con la nueva solución, estamos en disposición de analizar y evaluar el grado de ajuste del mismo.

#### 4. Evaluación del ajuste del modelo

Como hemos apuntado, el grado de ajuste se evalúa comparando la semejanza existente entre las matrices de varianzas-covarianzas del modelo y muestral. Existen multitud de pruebas estadísticas de Bondad de Ajuste y aunque han sido objeto de múltiples estudios en la literatura especializada (Jöreskog y Sörbom, 1979; Bentler, 1990; Bollen, 1990; Bollen y Long, 1993), no se ha llegado a un consenso sobre cuál o cuáles de los indicadores estadísticos son los mejores. Incluso los autores no se ponen de acuerdo en cuál es el valor de cada indicador que supone un buen ajuste. Nuestra recomendación es que se evalúe el ajuste del modelo basándose en varios indicadores y prestando especial atención a las limitaciones de cada uno de ellos

La evaluación del ajuste se puede clasificar en dos grandes aproximaciones: Globales e Individuales. Los índices de ajuste globales evalúan el grado de ajuste del modelo en su conjunto. Los índices individuales, lo que aportan es una evaluación particular de los parámetros del modelo de medida. Pasamos a continuación a describir muy someramente los más importantes.

##### A) Evaluación del ajuste del modelo conjunto:

1. Medidas de ajuste absoluto: Tratan de determinar el grado en que el modelo propuesto predice la matriz de correlaciones o covarianzas observadas:

1.1. **Prueba de significación estadística de  $\chi^2$** . Históricamente fue el primer indicador propuesto para evaluar el grado de ajuste entre los modelos. Para concluir un ajuste adecuado hay que aceptar  $H_0$  ( $p > .05$ ; óptimo:  $.10 < p < .20$ ) lo que indica que las diferencias entre las matrices observadas y estimadas no son estadísticamente significativas. Esto no garantiza la identificación del modelo "correcto", podría haber otro mejor. Muy sensible al tamaño de la muestra (ideal entre 100 y 200). Si  $n > 200$  se tiende a rechazar  $H_0$  y si  $n < 100$  se tiende a aceptar  $H_0$  incluso sin ser significativa ninguna relación del modelo. Por ello se requiere complementar con otros índices de ajuste.

1.2. Otros índices de bondad de ajuste absolutos: **Índice de Bondad de Ajuste** (Goodness of Fit Index, **GFI**). Representa el grado de ajuste conjunto pero no está ajustado por los grados de libertad. No hay un umbral absoluto de aceptabilidad, su valor oscila entre 0 y 1, siendo 1 el valor óptimo; **Residuo Cuadrático Medio** (Root Mean Square Residual **RMSR**) y **Error de Aproximación Cuadrático Medio** (Root Mean Square Error of Approximation, **RMSEA**). RMSR y RMSEA analizan la discrepancia entre las matrices de correlaciones estimada y observada. RMSR la mide en términos de la muestra y RMSEA en términos poblacionales. La evaluación del ajuste se basa en el análisis de los residuos o errores (cuanto más bajos mejor). No hay umbrales de aceptabilidad. Intentan corregir la tendencia del estadístico Chi cuadrado de rechazar cualquier modelo especificado cuando la muestra es grande. RMSEA es preferible a RMSR cuando las muestras son grandes (aceptables entre .05 y .08)

2. Medidas de Ajuste Incremental: Evalúan el ajuste incremental del modelo propuesto comparado con un modelo nulo (modelo de un único factor sin error de medida). Todos ellos aceptables en torno a .90. Los más importantes son el **Índice de Tucker-Lewis** (**TLI**) o los Índices de Bentler-Bonett, tanto el de **Ajuste Normalizado** (Normed Fit Index, **NFI**) como el de **Ajuste No Normalizado** (Non-Normed Fit Index, **NNFI**).

Otros índices en la misma línea de los anteriores son los de Bollen: **Índice de Ajuste Relativo (RFI)** e **Índice de Ajuste Incremental (IFI)** y el de Bentler: **Índice de Ajuste Comparado (CFI)**.

3. Medidas de Ajuste de Parsimonia: Relacionan la calidad del ajuste del modelo con el nº de coeficientes estimados necesarios para conseguir ese nivel de ajuste. Son índices que penalizan a los modelos en función del número de parámetros liberados. Cuanto más parámetros, más decrecen estos indicadores. Por eso reciben el nombre de parsimoniosos. No existe ninguna prueba de significación estadística, su uso se limita a la comparación entre modelos. Son un tipo de indicadores de ajuste que tienen su razón de ser cuando se plantean modelos alternativos al que se trata de probar. Los más importantes son los siguientes:

- **Índice de Bondad de Ajuste Incremental** (Adjusted Goodness of Fit Index, **AGFI**). Para que sea adecuado su valor debe ser superior a 0,90 (es un índice que oscila entre 0 y 1)
- $\chi^2$  **Normalizado**. Son valores aceptables los comprendidos entre 1 y 2 ó 3 (incluso 5 siendo menos estrictos). Si obtenemos valores inferiores a 1, es que el modelo puede estar sobreajustado. Si se encuentran valores mayores que el umbral superior, es que el modelo no es verdaderamente representativo de los datos observados y es necesario su mejora. Sin embargo, es una medida poco fiable y es necesario complementarla con otros índices de bondad de ajuste.
- **Índice de Bondad de Ajuste de Parsimonia** (Parsimony Goodness of Fit Index, **PGFI**): Puede oscilar entre 0 y 1. Cuanto más alto mayor parsimonia, es decir mejor ajuste.
- **Índice de Ajuste Normalizado de Parsimonia** de James-Mulaik-Brett (Parsimony Normed of Fit Index, **PNFI**). Cuanto más bajo mejor. Cuando se comparan dos modelos alternativos se proponen valores de .06 a .09 para que sean indicativos de diferencias sustanciales entre modelos.

B) Ajuste del modelo de medida. Evaluación de los constructos estimados.

1. Examen de las ponderaciones estimadas de los indicadores: Se trata de valorar la significación estadística de los parámetros de los modelos. En los Programas informáticos los valores de los parámetros estimados suelen aparecer con su significación estadística representada por la probabilidad asociada. Si esa probabilidad es  $p < .01$  ó  $.05$  (para los niveles de confianza del 99% ó del 95% respectivamente), las variables están significativamente relacionadas con sus constructos específicos, verificando las relaciones propuestas entre los indicadores y los constructos. Si  $p > .05$  se puede eliminar el indicador o intentar trasformarlo para un mejor ajuste del constructo.

2. Estimación de la Fiabilidad del constructo y del porcentaje de la varianza extraída: Evaluación de si los indicadores especificados son suficientes en su representación de los constructos.

– **Fiabilidad Compuesta**: Concepto similar al de **Fiabilidad Psicométrica**. Viene a reflejar el grado en que los indicadores o variables observadas son consistentes en su medida del constructo latente tal y como el modelo se ha especificado (mínimo recomendado = .70).

$$FC = \frac{(\Sigma R)^2}{(\Sigma R)^2 + \Sigma e}$$

donde *FC* es el coeficiente de fiabilidad.



$R$  es el peso factorial tipificado (parámetro) de cada variable observada en un constructo o factor determinado.

$e$  es el peso factorial tipificado (parámetro) del error de cada una de esas variables observadas.

– Varianza extraída: Porcentaje de la varianza de los indicadores que interviene en la explicación del constructo. (mínimo recomendado = 50% ; si es inferior, se deben explorar ponderaciones adicionales de estos indicadores sobre los otros constructos, siempre que se justifique desde el punto de vista teórico)

$$S_{extraída}^2 = \frac{\sum R^2}{p}$$

donde  $S_{extraída}^2$  es la proporción de varianza común extraída para un constructo determinado, o porcentaje de varianza explicado por ese factor o constructo.

$R$  es el peso factorial tipificado (parámetro) de cada variable observada en un constructo o factor determinado.

$p$  es el número de variables que definen ese constructo o factor.

3. Evaluación de la relación propuesta entre los constructos y entre los errores. Si en el modelo sometido a prueba hemos especificado relación entre los constructos (en nuestro ejemplo, plasmado en la figura 3, recordemos que establecimos una posible relación entre el factor 2 y el factor 3) o entre algunos términos de error (en nuestro ejemplo relaciones entre  $e_1$  y  $e_2$ ,  $e_4$  y  $e_5$ , y  $e_6$  y  $e_8$ ), deberemos analizar también su significación. Si alguna de estas relaciones no fuese significativa convendría redefinir el modelo eliminando esas relaciones no significativas y volviendo a analizarlo de nuevo.

##### 5. Reespecificación del modelo y comparación de modelos anidados

Para terminar esta breve exposición, vamos a dedicar unas líneas a la reespecificación de modelos. Reespecificar un modelo significa someter a prueba otro u otros alternativos. Generalmente se hace cuando la valoración del ajuste del modelo inicial no es buena. De hecho, la mayoría de los Programas informáticos ofrecen alternativas al modelo sometido a prueba aportando sugerencias de modificación (liberar algún nuevo parámetro que inicialmente no se había contemplado) y cómo afectarían esas modificaciones al ajuste del modelo alternativo. Normalmente, esas sugerencias no suelen alterar el número de factores de la estructura inicial y se limitan a alterar las relaciones (parámetros) entre las variables observadas y los factores establecidos a priori. Estos modelos alternativos establecen una relación de anidamiento con el modelo original. Diríamos que el primer modelo (que se supone que libera menos parámetros que el alternativo) está anidado en el segundo y así sucesivamente en un proceso parsimonioso. En este tipo de situación es frecuente redefinir ese modelo con los cambios sugeridos para mejorar su ajuste.

Sin embargo, cuando reespecificamos un modelo inicial alterando el número de factores o constructos, en realidad lo que hacemos es someter a prueba un nuevo modelo de medida que poco o nada tiene que ver con el original.

En cualquier caso, cuando nos encontramos ante la tesitura de reespecificar un modelo, sea anidado o no, lo primero que hay que tener en cuenta es que esa reespecificación se justifique desde el punto de vista teórico. No olvidemos que cuando planificamos y realizamos un AFC estamos sometiendo a prueba una hipótesis determinada que debe estar avalada por

un *corpus teórico* subyacente que la sustente. Todo lo que sea alterar el modelo sometido a prueba significa estrictamente alterar la hipótesis de partida y puede ocurrir que la búsqueda de modelos más ajustados a los datos empíricos que manejamos altere sustancialmente el objetivo inicial por el que se recabaron esos datos y por el que se realizó ese AFC.

Con esto no queremos decir que sistemáticamente deban rechazarse las prácticas de reespecificación de modelos. En los modelos de Ecuaciones Estructurales es una práctica bastante habitual. Lo único que queremos es subrayar la necesidad de que cuando se vaya a reespecificar un modelo, se justifique plenamente este proceder desde el punto de vista teórico. Si no tenemos especial cuidado en esto estaremos, a la postre, convirtiendo el Análisis Factorial Confirmatorio en un mero Análisis Factorial Exploratorio, y como hemos insistido, los modelos que se someten a prueba pueden ser infinitos con lo que la reespecificación puede llegar a convertirse en un círculo demasiado vicioso.

## REFERENCIAS BIBLIOGRÁFICAS

- Batista, J.M., y Coenders, G. (2000). *Modelos de ecuaciones estructurales*. Madrid: Hespérides.
- Bentler, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238-246.
- Bohrstedt, G.W. (1983). Measurement. En P.H. Rossi, J.D. Wright y A.B. Anderson (Eds.), *Handbook of survey research* (pp. 69-121). New York: Academic Press.
- Bollen, K.A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bollen, K.A. (1990). Overall fit in covariance structure models. Two types of sample size effects. *Psychological Bulletin*, 107, 256-259.
- Bollen, K.A. y Long, J.S. (1993). *Testing structural equation models*. Newbury Park, CA: Sage.
- Borsboom, D., Mellenbergh, G.J., y van Haerden (2004). The concept of validity. *Psychological Review*, 111, 1061-1071.
- Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C. (1999). *Análisis Multivariante* (5ª Ed.). Madrid: Prentice Hall Iberia.
- Jöreskog, K.G. y Sörbom, D. (1979). *Advances in factor analysis and structural equation models*. Cambridge, MA: Abt Associates.
- Long, J.S. (1983). *Confirmatory factor analysis*. Newbury Park, CA: Sage
- Martínez, M.R., Hernández, M.J., Hernández, M.V. (2006). *Psicometría*. Madrid: Alianza Editorial.
- Mulaik, S.A. (1972). *The foundations of factor analysis*. New York: McGraw-Hill.
- Pedhazur, E.J. y Pedhazur, L. (1991). *Measurement, Design and Analysis*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Pérez-Gil, J.A., Chacón, S., y Moreno, R. (2000). Validez de constructo: el uso del análisis factorial exploratorio-confirmatorio para obtener evidencias de validez. *Psicothema*, 12-2, 442-446.
- Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15, 201-293.
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis*. Washington DC: American Psychological Association.



# INTRODUCCIÓN A LOS MODELOS DE ECUACIONES ESTRUCTURALES

MIGUEL RUIZ DÍAZ  
*Departamento de Psicología Social y Metodología*  
*Facultad de Psicología (Universidad Autónoma de Madrid)*

## INTRODUCCIÓN

Los modelos de ecuaciones estructurales nacieron de la necesidad de dotar de mayor flexibilidad a los modelos de regresión. En palabras de Bollen (1989):

“Podemos mirar estos modelos de diversos modos. Son ecuaciones de regresión con supuestos menos restrictivos, que permiten errores de medida tanto en las variables criterio (independientes) como en las variables dependientes. Consisten en análisis factoriales que permiten efectos directos e indirectos entre los factores. Habitualmente incluyen múltiples indicadores y variables latentes. Resumiendo, engloban y extienden los procedimientos de regresión, el análisis econométrico y el análisis factorial.” (Página V).

Matemáticamente, estos modelos son más complejos de estimar que los modelos de Regresión o Análisis Factorial. Su uso no se extendió hasta 1973, momento en el que apareció el programa de análisis LISREL (*Linear Structural Relations*; Jöreskog, 1973) –originalmente denominado ACOUS–. El LISREL fue perfeccionado, dando lugar al LISREL VI (Jöreskog y Sörbom, 1983), que ofrecía una mayor variedad de métodos de estimación. El EQS (Bentler, 1985) es el otro paquete utilizado tradicionalmente para este tipo de análisis. En la actualidad, existen otros programas de estimación en entorno gráfico, como el AMOS. Tal ha sido la influencia de los programas de estimación en la posibilidad de desarrollo de los modelos de ecuaciones estructurales, que no es infrecuente que se denomine a los modelos estructurales (por metonimia) como modelos LISREL.

La gran ventaja de este tipo de modelos es que permiten proponer el tipo de relaciones que se espera encontrar entre las diversas variables, para pasar posteriormente a estimar los parámetros que vienen especificados por las relaciones propuestas. Por este motivo se denominan también *modelos confirmatorios*, ya que el interés fundamental es “confirmar” mediante el análisis de la muestra (o muestras) las relaciones propuestas a nivel teórico. Como veremos más adelante, la especificación teórica del modelo permite proponer estructuras causales entre las variables, de manera que unas variables causen un efecto sobre otras variables que, a su vez, pueden trasladar estos efectos a otras variables.

## CONCEPTOS BÁSICOS

En la base de los modelos estructurales se encuentra el concepto de *causalidad* que, en este contexto, tiene un significado especial. Por otra parte, existe un acuerdo generalizado en *representar* estos modelos de manera gráfica, siguiendo una serie de convenciones. Antes de entrar a discutir con detalle las características de los modelos de ecuaciones estructurales vamos a revisar brevemente estas dos cuestiones.

### *El concepto de causalidad*

El concepto de causalidad ha sido discutido ampliamente por filósofos y científicos, sin que exista un acuerdo unánime sobre su alcance. Aquí vamos a acotar su significado al tema que nos ocupa.

Intuitivamente, entendemos que existe *causalidad* cuando la aparición de un fenómeno (la causa) determina, en mayor o menor medida, la aparición de un segundo fenómeno (el efecto). Así por ejemplo, decimos que la fatiga es efecto del exceso de ejercicio, por lo que pensamos que el exceso de ejercicio causa fatiga. Además, si realmente existe causalidad, la desaparición de la causa, implicará la desaparición del efecto.

En el plano científico, sólo se admite la formulación de la causalidad a partir de la manipulación experimental de las variables que se pretende relacionar (Campbell y Stanley, 1963). Para determinar el efecto de la variable *X* sobre la variable *Y* deberemos manipular diversos niveles (intensidades) del estímulo *X*, para cuantificar su efecto (o ausencia de éste) sobre la variable de respuesta *Y*. Sólo después de esta manipulación, preferiblemente a través de varios experimentos repetidos, podremos asegurar la causalidad de *X* sobre *Y*, así como su magnitud.

Por desgracia, en la mayoría de las ciencias sociales es imposible llevar a cabo un estudio de tipo experimental cuando se pretende estudiar la relación jerárquica entre variables. De hecho, incluso en el terreno experimental se recurre al control estadístico de las variables, cuando el control experimental excede el alcance del experimentador, como sucede en los diseños de bloques, o en el análisis de covarianza (ANCOVA).

Los trabajos de Boudon (1965) y Duncan (1966) propusieron una nueva posibilidad de acercamiento a la causalidad, el análisis de dependencias o *análisis de rutas* (path analysis). En este tipo de análisis se estudia una teoría causal mediante la especificación de todas las variables importantes para dicha teoría. Posteriormente, se pueden derivar las relaciones entre los efectos causales a partir de la teoría causal para, en último término, estimar el tamaño de estos efectos.

La generalización lógica del análisis de rutas fueron los modelos de ecuaciones estructurales y el contraste de teorías o lo que es lo mismo, de modelos causales. La lógica de estos modelos se basa en que, a partir de la teoría que fundamenta el modelo, será posible derivar las medidas de covariación entre las variables a partir de los efectos causales del modelo. Si la teoría es correcta, las medidas de covariación derivadas del modelo y las medidas de covariación obtenidas a partir de los datos deberán ser iguales.

Por tanto, la causalidad contenida en los modelos estructurales debe entenderse en el sentido de *control estadístico* y no en el sentido *determinista* de la manipulación experimental.

La estructuración causal de las relaciones teóricas es una manera de resumir nuestro conocimiento de un fenómeno y permite un abordaje racional lógico en su estudio. Además,

como señalan Dillon y Goldstein (1984), las teorías y su formulación provienen de fuera de la estadística. Si no existe una teoría consistente y explícita, poco sentido tiene analizar o intentar interpretar, ni tan siquiera una correlación o una media.

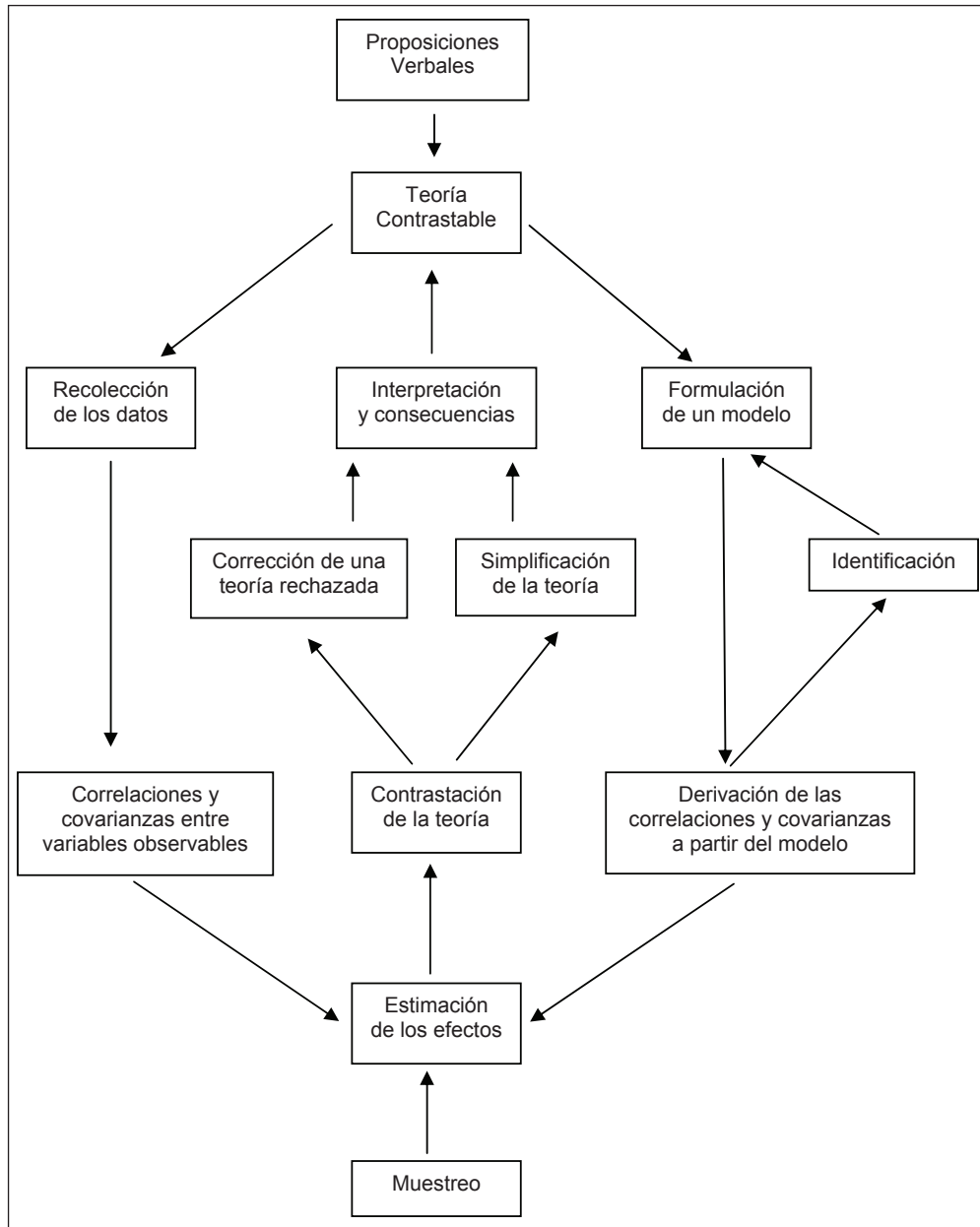
Precisamente esta necesidad de teorías explícitas es lo que confiere valor a los modelos causales. Para poner a prueba los modelos estructurales, el investigador deberá especificar con sumo detalle las variables de interés y el tipo de relaciones existentes entre las variables. Si no se incluyen en el modelo todas las variables relevantes para el tema investigado o las relaciones se definen de manera errónea o, lo que es más común, las relaciones entre las variables se encuentran medidas con escasa precisión, será difícil conseguir el ajuste buscado entre la teoría y los datos.

Bien es cierto que muchos estudios en los que se utilizan estos modelos abusan de un excesivo ajuste y reajuste de las posibles relaciones teóricas, a partir de los indicios de falta de ajuste matemático que aparecen en los datos. Estos modelos sobre-manipulados suelen ser muy inestables y pierden sus buenas propiedades de ajuste cuando se utilizan otras muestras para su réplica. Por desgracia los estudios de replicación son escasos, por lo que es recomendable mantener un cierto escepticismo cuando en un estudio no se nos informe detalladamente de las manipulaciones que hayan podido sufrir los datos.

En cualquier caso, no está de más recordar que el hecho de que se cumplan todas las propiedades estadísticas deseables en los ejemplos que se proponen en esta monografía, no quiere decir que sea sencillo obtener modelos como estos. Por el contrario, el desarrollo de un modelo estructural suele ser bastante laborioso.

El proceso de construcción y desarrollo de teorías causales mediante la utilización de modelos de ecuaciones estructurales puede resumirse en el siguiente cuadro.

Figura 1. Diagrama de flujo para contrastar hipótesis causales de datos no experimentales



### Los diagramas estructurales: convenciones y definiciones

Para representar un modelo causal y las relaciones que se desea incluir en él, se acostumbra a utilizar diagramas similares a los diagramas de flujo. Estos diagramas se denominan *diagramas causales* o *diagramas estructurales*. Los diagramas estructurales siguen unas convenciones particulares.

- Las variables observables se representan encerradas en rectángulos.
- Las variables no observables (latentes) se representan encerradas en óvalos o círculos.

- Las variables que representan los errores de cualquier tipo de variable dependiente se representan sin rectángulos ni círculos. (Aunque algunos programas los dibujan como variables latentes).
- Las relaciones entre variables de tipo covariante (las correlaciones y las covarianzas) se representan como vectores curvos con una flecha en cada extremo.
- Cualquier efecto estructural se representa como una flecha recta, cuyo origen es la variable predictora y cuyo final, donde se encuentra el extremo de la flecha, es la variable dependiente.
- Se utilizan letras griegas para denominar a las variables latentes ( $\xi$ ,  $\eta$ ).
- Se utilizan letras latinas para denominar a las variables observables ( $x$ ,  $y$ ).
- Se utilizan letras griegas para denominar los efectos de unas variables sobre otras ( $\gamma$ ,  $\phi$ ,  $\lambda$ , etc.), si bien cada tipo de efecto utiliza un símbolo determinado (ver más adelante).
- En cada símbolo de una relación se indican, mediante subíndices, las variables implicadas en él. El primer subíndice corresponde a la variable dependiente y el segundo subíndice indica la variable predictora para ese efecto.
- Cualquier variable que reciba efecto de otras variables del modelo deberá incluir también un término de error.

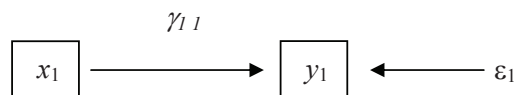
Tabla 1. Letras griegas utilizadas en los diagramas estructurales

$\xi$	Ksi	$\beta$	Beta	$\delta$	Delta
$\eta$	Eta	$\phi$	Phi	$\varepsilon$	Epsilon
$\lambda$	Lambda	$\theta$	Theta	$\psi$	Psi
$\gamma$	Gamma	$\Sigma$	Sigma	$\zeta$	Zeta

### Ejemplo 1

El siguiente diagrama estructural representa un modelo de regresión simple.

Figura 2. Modelo de regresión lineal bivalente



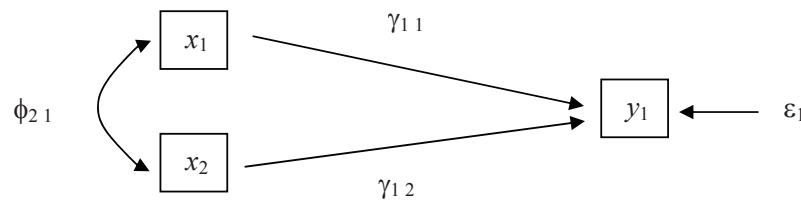
Las variables dependiente ( $y_1$ ) e independiente ( $x_1$ ) son observables. La variable dependiente está afectada por un error de medida ( $\varepsilon_1$ ), mientras que la variable independiente está medida sin error. El efecto de la variable independiente sobre la dependiente viene reflejada por un vector cuyo parámetro es  $\gamma_{11}$ . Este modelo carece de variables latentes, si bien el error puede considerarse técnicamente como una variable latente.

### Ejemplo 2

El siguiente diagrama representa un modelo de regresión múltiple con dos variables predictoras.



Figura 3. Modelo de regresión lineal múltiple



La variable dependiente recibe el efecto de las dos variables independientes. Cada efecto está representado por su parámetro correspondiente ( $\gamma_{ij}$ ). La covarianza entre las dos variables predictoras se representa como un vector curvo de doble punta. Su parámetro ( $\phi_{21}$ ) corresponde a la covarianza observada entre<sup>1</sup>  $x_1$  y  $x_2$ .

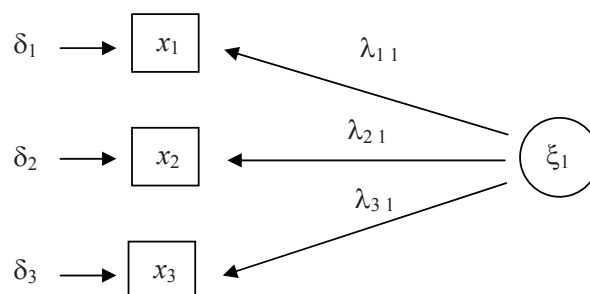
Siguiendo el modelo de regresión múltiple lineal clásico, las variables predictoras carecen de error de medida.

El error de medida ( $\epsilon_1$ ) de la variable dependiente, afecta a dicha variable dependiente, pero la variación es aleatoria e independiente de la de la propia variable dependiente. Por ello no se acostumbra a representarla como una variable más del modelo, encerrada en un rectángulo o un círculo. El error de la variable dependiente es también independiente de las variables predictoras, por lo que no manda flecha hacia ninguna de ellas. El error de medida recoge las fluctuaciones en la variable dependiente que no son atribuibles a las variables del modelo. No es una variable observable. Algunos programas como el AMOS la representan como una variable latente (no observable) encerrada en un círculo. También es relativamente frecuente que las variables de error no se representen en el diagrama para simplificar la representación del modelo cuando existen múltiples variables, o incluso que sólo se represente la flecha del efecto del error, pero no se representa el símbolo de la variable.

### Ejemplo 3

El siguiente diagrama representa un modelo factorial con tres variables y un factor latente.

Figura 4. Modelo de análisis factorial confirmatorio



En este modelo existe un único factor, representado por una variable latente ( $\xi_1$ ) que no es observable. Este factor latente es estimado por tres indicadores observables ( $x_1$ ,  $x_2$  y  $x_3$ ), por lo que reciben su efecto del factor latente.

1 En el texto se utilizan letras minúsculas para las variables observadas por dos motivos: por encontrarse la variable en escala diferencial y por ser variables aleatorias que pertenecen a un vector.

Las variables observables no son indicadores perfectos del factor latente, por lo que tendrán un cierto grado de error, representado por sus términos correspondientes  $\delta_i$ . Como veremos, este término de error no es exactamente lo mismo que la unicidad de la variable correspondiente, aunque está estrechamente relacionado con ella.

El factor latente no tiene error, por lo que no recibe ninguna flecha de efecto.

En el mismo gráfico podrían representarse simultáneamente las correlaciones entre las variables observables, mediante flechas curvas.

Como hemos visto, la representación de diversos modelos causales es relativamente simple, utilizando las convenciones mencionadas más arriba. Los nombres de los parámetros del modelo (las letras griegas de las flechas) se concretarán a medida que sea necesario y el desarrollo de la monografía lo permita.

### *El concepto de "ajuste"*

Para entender la fundamentación de los modelos de ecuaciones estructurales, es necesario reorientar nuestro conocimiento de lo que significa el concepto de *ajuste* de un modelo. En regresión lineal, cuando hablamos de las estimaciones de los parámetros, escogemos aquellas estimaciones que mejor ajustan el modelo a los datos, en el sentido de que minimizan los errores de predicción cometidos con el modelo para el conjunto de sujetos de la muestra (en el método de mínimos cuadrados). Por el contrario, en los modelos de ecuaciones estructurales, lo que se pretende ajustar son las covarianzas entre las variables, en vez de buscar el ajuste de los casos. En lugar de minimizar la diferencia entre los valores pronosticados y los observados a nivel individual, se minimiza la diferencia entre las covarianzas muestrales y las covarianzas pronosticadas por el modelo estructural. Este es el motivo por el que también se han denominado modelos de estructura de covarianza (*covariance structure models*; Long, 1983). En el caso que nos ocupa, los residuos son la diferencia entre las covarianzas observadas y las covarianzas reproducidas (pronosticadas) por el modelo estructural teórico.

El ajuste de un modelo se puede recoger en una hipótesis fundamental, que propone que, *si el modelo es correcto* y conociéramos los parámetros del modelo estructural, la matriz de covarianzas poblacional podría ser reproducida exactamente a partir de la combinación de los parámetros del modelo. La idea de ajuste se resume en la siguiente ecuación

$$H_0: \Sigma = \Sigma(\theta) \quad (1)$$

donde  $\Sigma$  es la matriz de covarianzas poblacionales entre las variables observables,  $\theta$  es un vector que contiene los parámetros del modelo y  $\Sigma(\theta)$  es la matriz de covarianzas derivadas como una función de los parámetros contenidos en el vector  $\theta$ .

Veamos el significado y extensión de esta hipótesis con un ejemplo (Bollen, 1989). Tomemos el modelo mostrado en el ejemplo 1. La ecuación de regresión que lo define es

$$y = \gamma x + \varepsilon \quad (2)$$

siendo  $\gamma$  el coeficiente de regresión<sup>2</sup> y  $\varepsilon$  la variable que representa el término de error, que es independiente de  $x$  y cuyo valor esperado,  $E(\varepsilon)$ , es cero. La matriz de varianza-covarianza entre las variables aleatorias observadas  $x$  e  $y$  será

$$\Sigma = \begin{pmatrix} \text{VAR}(y) & \text{COV}(x, y) \\ \text{COV}(x, y) & \text{VAR}(x) \end{pmatrix} \quad (3)$$

2 Dado que no existe confusión posible, se han eliminado los subíndices por simplicidad.

Esta es la matriz que obtenemos directamente al analizar descriptivamente los datos y representa las relaciones existentes entre las variables en la muestra. Ahora bien, la variable  $y$  es función de las variables  $x$  y  $\varepsilon$  y del parámetro  $\gamma$ . Podemos volver a escribir los elementos de la matriz  $\Sigma$  en función de la ecuación (2). Trabajando, por simplicidad en diferenciales, sabemos que  $\text{VAR}(x) = E(x^2) - E(x)^2 = E(x^2)$ , ya que el valor esperado de  $x$ , en diferenciales, es cero. La varianza de la variable  $y$  puede expresarse como

$$\begin{aligned}\text{VAR}(y) &= E(y^2) - E(y)^2 = E(y^2) = \\ &= E[(\gamma x + \varepsilon)^2] = E(\gamma^2 x^2 + \varepsilon^2 + 2\gamma x\varepsilon) = E(\gamma^2 x^2) + E(\varepsilon^2) = \\ &= \gamma^2 E(x^2) + E(\varepsilon^2) = \gamma^2 \text{VAR}(x) + \text{VAR}(\varepsilon)\end{aligned}\quad (4)$$

donde hemos utilizado las propiedades del término error para simplificar la igualdad.

De manera similar, podemos derivar la covarianza entre  $x$  e  $y$ .

$$\begin{aligned}\text{COV}(x, y) &= E(xy) = \\ &= E[x(\gamma x + \varepsilon)] = E(\gamma x^2 + x\varepsilon) = E(\gamma x^2) + E(x\varepsilon) = \\ &= \gamma E(x^2) = \gamma \text{VAR}(x)\end{aligned}\quad (5)$$

Sustituyendo en la ecuación 3 las expresiones derivadas que hemos obtenido en las ecuaciones 4 y 5, llegamos a la matriz

$$\Sigma(\theta) = \begin{pmatrix} \gamma^2 \text{VAR}(x) + \text{VAR}(\varepsilon) & \gamma \text{VAR}(x) \\ \gamma \text{VAR}(x) & \text{VAR}(x) \end{pmatrix}\quad (6)$$

que no es más que la matriz de varianza-covarianza poblacional escrita en función de los parámetros del modelo. A esta matriz también se la denomina la matriz de covarianza *implícita*. Podemos sustituir ahora en la ecuación (1) y volver a expresar la hipótesis básica como

$$\begin{aligned}H_0: \Sigma &= \begin{pmatrix} \text{VAR}(y) & \text{COV}(x, y) \\ \text{COV}(x, y) & \text{VAR}(x) \end{pmatrix} = \\ &= \begin{pmatrix} \gamma^2 \text{VAR}(x) + \text{VAR}(\varepsilon) & \gamma \text{VAR}(x) \\ \gamma \text{VAR}(x) & \text{VAR}(x) \end{pmatrix} = \Sigma(\theta)\end{aligned}\quad (7)$$

En esta igualdad, los elementos de la parte derecha y los de la parte izquierda se corresponden, dadas las especificaciones del modelo que hemos propuesto. Si el modelo es el correcto y conociéramos los valores de los parámetros de la parte derecha de la igualdad, no sería difícil comprobar la igualdad de los términos. Cómo estimar los parámetros del modelo es un problema que ha necesitado que se desarrollen programas informáticos para poder ser resuelto y que veremos más adelante.

Los diferentes modelos de ecuaciones estructurales se basan en una hipótesis similar a la igualdad expresada en (7), si bien la derivación de las ecuaciones correspondientes puede ser más compleja. En todos los casos, las ecuaciones básicas del modelo deben ser lineales, tanto en los parámetros como en las incógnitas. Por ejemplo, la ecuación del modelo del ejemplo 1 es  $y = \gamma x + \varepsilon$ , la ecuación del modelo del ejemplo 2 es  $y_1 = \gamma_{11}x_1 + \gamma_{12}x_2 + \varepsilon_1$ , y las ecuaciones del ejemplo 3 son,  $x_1 = \lambda_{11}\xi_1 + \delta_1$ ,  $x_2 = \lambda_{21}\xi_1 + \delta_2$ ,  $x_3 = \lambda_{31}\xi_1 + \delta_3$ . Esto no quiere decir

que las igualdades derivadas de estas ecuaciones, que se encuentran en la matriz de covarianza implícita, también deberán ser lineales.

Este planteamiento de entender los modelos de ecuaciones estructurales como modelos teóricos que representan una estructura propuesta en función de unos parámetros, no es tan novedosa como cabría pensar. En los modelos factoriales, como ya hemos visto, uno de los criterios de la bondad del modelo factorial es su capacidad para reproducir la matriz de correlaciones que se esté analizando. De hecho, el término *matriz de correlaciones reproducida* del análisis factorial exploratorio hace referencia a un planteamiento similar al propuesto en la hipótesis de la ecuación (1).

## NOTACIÓN

La notación que utilizaremos en este capítulo es la desarrollada por Jöreskog (1973, 1977), Wiley (1973) y Keesling (1972), popularizada posteriormente por la utilización del LISREL.

Se denomina *modelo completo* a un sistema de ecuaciones estructurales, es decir, un conjunto de ecuaciones lineales que expresan las relaciones entre las variables. Las ecuaciones contienen variables aleatorias y parámetros estructurales (que expresan las relaciones entre las variables).

- Las variables aleatorias pueden ser de tres tipos: observables, latentes y de error. Las variables *observables*, o indicadores, son aquellas cuya variabilidad y relaciones con otras variables observables pueden ser cuantificadas a partir de las medidas obtenidas del mundo real. Son variables observables las preguntas de un cuestionario, las variables socio-demográficas de una encuesta y los índices econométricos.
- Las variables *latentes*, no observables o factores, son variables cuya medida directa es imposible de obtener y que debe ser inferida a partir de su relación con variables observables que sean sus indicadores. La inteligencia, el nivel socioeconómico, el autoritarismo o los factores comunes de un modelo factorial son variables latentes.
- Los *errores de medida* son variables aleatorias no observables que aglutinan todos los efectos no considerados en el sistema y que puedan estar afectando a la medida de la variable a la que influyen. Estas variables son equivalentes al término error de un modelo de regresión.

Las variables que son causa de otras variables y cuya variabilidad se asume que es causada por otras variables no consideradas en el modelo se denominan variables *exógenas* (las independientes). Puesto que las relaciones entre las variables exógenas (sus correlaciones) no son tomadas en cuenta dentro del modelo, no se intenta explicar sus interrelaciones. Las variables cuya variación es explicada por las variables exógenas, o por otras variables del sistema, se denominan variables *endógenas* (las dependientes).

- Las variables exógenas se representan como  $X$  si son observables y como  $\xi$  si son latentes.
- Las variables endógenas observables se representan como  $Y$  mientras que las endógenas latentes se representan como  $\eta$ .
- Los errores de medida de las variables observables exógenas se representan como  $\delta$ , los de las variables observables endógenas como  $\epsilon$  y los errores que afectan a las variables latentes endógenas se representan como  $\zeta$ . Las variables latentes exógenas nunca están afectadas por un error.

- Los coeficientes de regresión que representan la relación de una variable latente (endógena o exógena) con sus indicadores se representa mediante  $\lambda$ , acompañada de un subíndice que indique el tipo de variable (X o Y) si fuera necesario distinguirlos.
- Los efectos de una variable endógena sobre otra endógena se representan por un coeficiente  $\beta$ , los efectos de una variable exógena sobre otra endógena se representan por un coeficiente  $\gamma$ .
- Las covarianzas entre los errores de medida se representan como  $\theta$ , acompañado del subíndice  $\delta$  o  $\varepsilon$ , según al error que se refieran. Las covarianzas entre variables latentes exógenas se representan como  $\phi$ .
- Las covarianzas entre los errores de variables endógenas latentes se representan como  $\psi$ . Por último, las relaciones entre variables observables exógenas se representan mediante líneas curvas con una flecha en cada extremo.

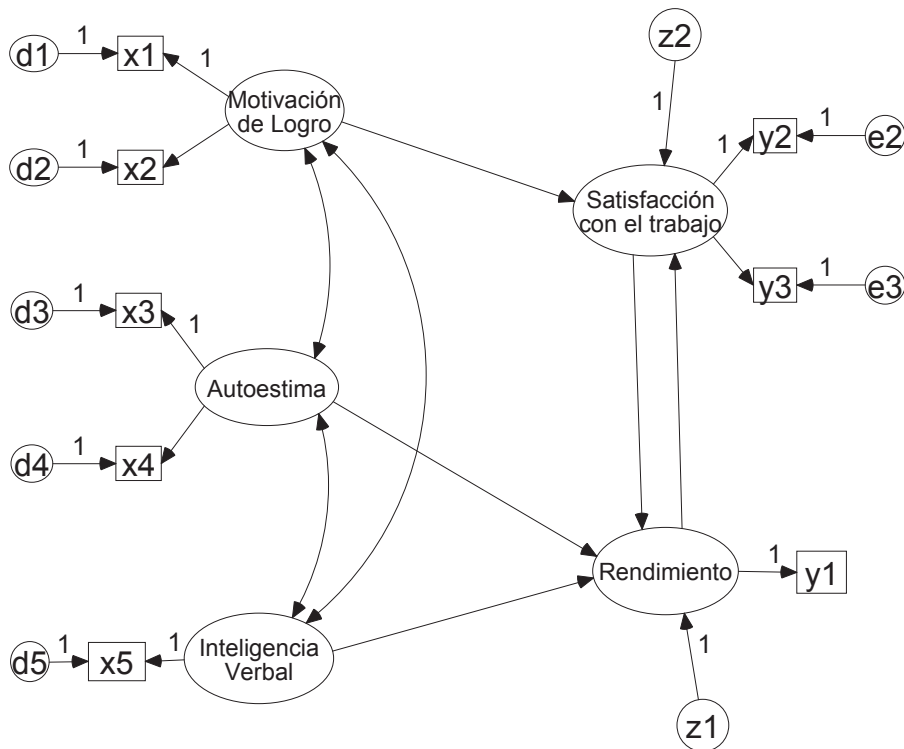
Estas convenciones se acompañan de las mencionadas en la página 47 para confeccionar los gráficos de efectos del modelo.

Familiarizarse con esta nueva terminología requiere alguna práctica. Para ejemplificar representaremos el modelo propuesto por Bagozzi (1980).

#### Ejemplo 4

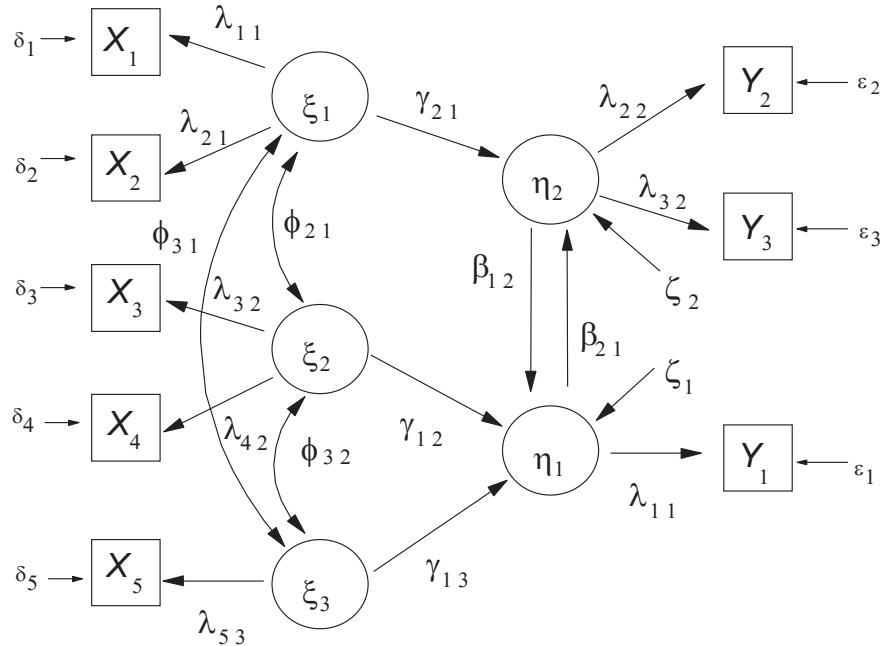
El modelo que recogemos a continuación fue propuesto por Bagozzi (1980) y se discute también en Jöreskog y Sörbom (1989) y Dillon y Goldstein (1984). El modelo pretende explicar la relación entre la satisfacción y el rendimiento del equipo de ventas de una empresa.

Figura 5. Representación estructural del Modelo de Bagozzi



En la figura 5 se muestra un diagrama estructural con los nombres de las variables latentes y los efectos pertinentes para definir el modelo. En la figura 6 se muestra el mismo diagrama utilizando la nomenclatura desarrollada anteriormente para la denominación de los parámetros y de las variables latentes.

Figura 6. Modelo estructural de la satisfacción



El modelo final propuesto por el autor contiene tres variables latentes exógenas ( $\xi_1$ ,  $\xi_2$  y  $\xi_3$ ), así como dos variables latentes endógenas ( $\eta_1$  y  $\eta_2$ ). Estas variables representan medidas perfectas de sus respectivos conceptos o constructos. Cada una de estas variables latentes es estimada por sus correspondientes indicadores, o medidas, las X y las Y.

Las variables latentes corresponden a los conceptos

- $\xi_1$  = Motivación de Logro
- $\xi_2$  = Autoestima ligada a la Tarea
- $\xi_3$  = Inteligencia Verbal
- $\eta_1$  = Rendimiento del vendedor
- $\eta_2$  = Satisfacción Laboral

La Motivación de logro está medida por dos indicadores,  $X_1$  = una escala tipo Likert de 4 preguntas y  $X_2$  = una escala tipo Likert de 4 preguntas. La Autoestima ligada a la tarea se encuentra medida por dos indicadores,  $X_3$  = una escala de 3 ítems y  $X_4$  = una escala de 3 elementos. La Inteligencia verbal se encuentra medida por un único indicador,  $X_5$  = la puntuación obtenida en una escala de 30 ítems emparejados. El Rendimiento de los vendedores se cuantifica mediante un indicador  $Y_1$  = el volumen de ventas anual de cada individuo en dólares. La Satisfacción laboral está medida por dos indicadores,  $Y_2$  = la puntuación en una escala de 4 ítems y  $Y_3$  = la puntuación en una escala de 4 preguntas tipo Likert.

Como puede observarse en la figura 5, cada variable latente manda una flecha a su indicador o indicadores correspondientes, como si se tratara de un modelo factorial, en el que

la variable latente (el concepto) fuera el factor y los indicadores los ítems observables del factor.

Además, el modelo propone una serie de rutas causales muy específicas, que han sido determinadas a partir de la teoría. La Motivación de logro influye sólo sobre la Satisfacción laboral y tanto las Habilidades verbales como la Autoestima afectan solamente al Rendimiento. A su vez, tanto el Rendimiento como la Satisfacción influyen una sobre la otra mediante los efectos correspondientes. El modelo es no recursivo, ya que ambas variables latentes endógenas influyen una sobre la otra. Si una de las  $\beta_{ij}$  fuera igual a 0, esa flecha no existiría y el modelo sería recursivo.

Cuando se estudia un modelo como el de la figura 5, se acostumbra a descomponer los conjuntos de ecuaciones estructurales que lo especifican en dos niveles de análisis, el modelo de las variables latentes y el modelo de medida.

## FORMULACIÓN DE UN MODELO

Para poder estimar los parámetros de un modelo es necesario definir primero sus ecuaciones. Tradicionalmente era necesario escribir primero las ecuaciones estructurales, para luego proceder con la comprobación de la identificabilidad, la estimación y la verificación de las propiedades del modelo. En la actualidad los programas de estimación permiten comenzar con la formulación gráfica del modelo, mediante la elaboración del diagrama causal, para luego proceder con las distintas fases del análisis. El modelo causal es la representación gráfica de las ecuaciones estructurales.

El conjunto de ecuaciones estructurales que definen un modelo puede llegar a ser muy extenso. Para organizar estas ecuaciones se recurre a la agrupación de las mismas en dos subconjuntos denominados el *modelo de medida* y el *modelo de las variables latentes*. Ambos subconjuntos se complementan y recogen todas las relaciones contempladas por el modelo.

### *Modelo de Medida*

El modelo de medida es el que contiene las ecuaciones correspondientes a las relaciones entre las variables latentes y las variables observables que constituyen sus correspondientes indicadores.

En el modelo de medida tendremos tantas ecuaciones como variables observables. Se acostumbra a agrupar las ecuaciones en dos conjuntos, uno para las variables exógenas y otro para las variables endógenas. Para simplificar la notación se asume que las variables están medidas en su forma diferencial, respecto a la media.

En el modelo de la figura 5, las ecuaciones que definen el modelo de medida serán, para las variables exógenas:

$$\left. \begin{aligned} X_1 &= \lambda_{11}\xi_1 + \delta_1 \\ X_2 &= \lambda_{21}\xi_1 + \delta_2 \\ X_3 &= \lambda_{32}\xi_2 + \delta_3 \\ X_4 &= \lambda_{42}\xi_2 + \delta_4 \\ X_5 &= \lambda_{53}\xi_3 + \delta_5 \end{aligned} \right\} \quad (8)$$

y para las variables endógenas:



$$\left. \begin{aligned} Y_1 &= \lambda_{11}\eta_1 + \varepsilon_1 \\ Y_2 &= \lambda_{22}\eta_2 + \varepsilon_2 \\ Y_3 &= \lambda_{32}\eta_2 + \varepsilon_3 \end{aligned} \right\} \quad (9)$$

Estas ecuaciones pueden expresarse de manera más compacta utilizando la notación matricial. El conjunto de ecuaciones (8) puede expresarse como

$$\mathbf{X} = \Lambda_x \boldsymbol{\xi} + \boldsymbol{\delta} \quad (10)$$

donde  $\mathbf{x}$  es un vector  $q \times 1$  que contiene las variables aleatorias correspondientes a las variables observables;  $\Lambda_x$  es una matriz de orden  $q \times n$  que contiene las saturaciones de las variables observables en las variables latentes independientes;  $\boldsymbol{\xi}$  es un vector, de dimensiones  $n \times 1$ , que contiene las variables latentes independientes (exógenas); y el vector  $\boldsymbol{\delta}$ , de dimensiones  $q \times 1$ , contiene las variables aleatorias correspondientes a los errores de medida.

Por su parte, el conjunto de ecuaciones (9) que vincula las variables dependientes, puede expresarse como

$$\mathbf{X} = \Lambda_y \boldsymbol{\eta} + \boldsymbol{\varepsilon} \quad (11)$$

donde  $\mathbf{y}$  es un vector  $p \times 1$  que contiene las variables aleatorias correspondientes a las variables observables;  $\Lambda_y$  es una matriz de orden  $p \times m$  que contiene las saturaciones de las variables observables en las variables latentes dependientes;  $\boldsymbol{\eta}$  es un vector, de dimensiones  $m \times 1$ , que contiene las variables latentes dependientes (endógenas); y el vector  $\boldsymbol{\varepsilon}$ , de dimensiones  $p \times 1$ , contiene las variables aleatorias correspondientes a los errores de medida.

### *Modelo de las Variables latentes*

El modelo de variables latentes, el modelo de ecuaciones estructurales propiamente dicho, contiene las ecuaciones correspondientes a las relaciones entre las variables no observables. Es el modelo que vincula las variables hipotetizadas, haciendo corresponder las variables independientes con las dependientes.

Las ecuaciones estructurales del modelo de la figura 2 serán

$$\begin{aligned} \eta_1 &= \beta_{12}\eta_2 + \gamma_{12}\xi_2 + \gamma_{13}\xi_3 + \zeta_1 \\ \eta_2 &= \beta_{21}\eta_1 + \gamma_{21}\xi_1 + \zeta_2 \end{aligned} \quad (12)$$

o alternativamente como

$$\begin{aligned} \eta_1 - \beta_{12}\eta_2 &= \gamma_{12}\xi_2 + \gamma_{13}\xi_3 + \zeta_1 \\ -\beta_{21}\eta_1 + \eta_2 &= \gamma_{21}\xi_1 + \zeta_2 \end{aligned} \quad (12b)$$

En notación matricial, puede expresarse como

$$\boldsymbol{\beta}\boldsymbol{\eta} = \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \quad (13)$$

donde  $\boldsymbol{\eta}$  es un vector  $m \times 1$  que contiene las variables latentes endógenas;  $\boldsymbol{\xi}$  es un vector  $n \times 1$  que contiene las variables latentes exógenas;  $\boldsymbol{\beta}$  es una matriz  $m \times m$  que contiene los coeficientes de regresión de las variables endógenas sobre las variables endógenas;  $\boldsymbol{\Gamma}$  es una matriz  $m \times n$  que contiene los coeficientes de regresión de las variables exógenas sobre las variables endógenas; y  $\boldsymbol{\zeta}$  es un vector  $m \times 1$  que contiene los errores de regresión que resultan de



predecir las variables endógenas a partir de las exógenas. Se asume que las variables latentes contenidas en  $\eta$  y  $\xi$  no están correlacionadas y cualquier relación entre ellas debe de estar contenida en los pesos de regresión correspondientes. Además, se asume que la matriz  $\beta$  no es singular.

En algunos textos, la ecuación estructural (13) se expresa de manera ligeramente diferente, reorganizando los términos para que las variables endógenas figuren sin coeficientes en el lado izquierdo de la ecuación. De manera que

$$\eta = \beta\eta + \Gamma\xi + \zeta \quad (14)$$

Existen otras cuatro matrices implícitas en el modelo que representan covarianzas, o correlaciones, entre los términos del mismo. Estas son: la matriz de correlaciones entre las variables latentes exógenas (las  $\xi$ ) denominada  $\Phi$ , de dimensiones  $n \times n$ ; la matriz de correlaciones entre los errores de las variables endógenas latentes  $\Psi$ , de dimensiones  $m \times m$ ; la matriz de covarianzas entre las variables exógenas observables  $\Theta_\delta$ , de dimensiones  $q \times q$ ; y la matriz de covarianzas entre las variables endógenas observables  $\Theta_\epsilon$ , de dimensiones  $p \times p$ .

### Ejemplo 5

Veamos de manera explícita las matrices que genera el modelo del ejemplo 4. El modelo estructural matricial es

$$\begin{pmatrix} 1 & -\beta_{12} \\ -\beta_{21} & 1 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} 0 & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & 0 & 0 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} + \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix} \quad (15)$$

$$\mathbf{\beta} \quad \mathbf{\eta} = \quad \mathbf{\Gamma} \quad \mathbf{\xi} + \mathbf{\zeta}$$

Mientras que el modelo de medida está definido por las ecuaciones

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} \lambda_{11} & 0 & 0 \\ \lambda_{21} & 0 & 0 \\ 0 & \lambda_{32} & 0 \\ 0 & \lambda_{42} & 0 \\ 0 & 0 & \lambda_{53} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \end{pmatrix} \quad (16)$$

$$\mathbf{x} = \mathbf{\Lambda}_x \mathbf{\xi} + \mathbf{\delta}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} \lambda_{11} & 0 \\ 0 & \lambda_{22} \\ 0 & \lambda_{32} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{pmatrix} \quad (17)$$

$$\mathbf{y} = \mathbf{\Lambda}_y \mathbf{\eta} + \mathbf{\epsilon}$$

Las matrices que faltan para completar el modelo son: la matriz de varianza-covarianza entre las variables latentes exógenas,

$$\mathbf{\Phi} = \begin{pmatrix} \phi_{11} & & \\ \phi_{21} & \phi_{22} & \\ \phi_{31} & \phi_{32} & \phi_{33} \end{pmatrix} \quad (18)$$

la matriz de varianza-covarianza entre los términos de error de las variables latentes endógenas

$$\Psi = \begin{pmatrix} \Psi_{11} & \\ \Psi_{21} & \Psi_{22} \end{pmatrix} \quad (19)$$

la matriz de varianza-covarianza de los errores de medida de las variables independientes

$$\Theta_{\delta} = \begin{pmatrix} \theta_{\delta_{11}} & & & & \\ 0 & \theta_{\delta_{22}} & & & \\ 0 & 0 & \theta_{\delta_{33}} & & \\ 0 & 0 & 0 & \theta_{\delta_{44}} & \\ 0 & 0 & 0 & 0 & \theta_{\delta_{55}} \end{pmatrix} \quad (20)$$

y la matriz de varianza-covarianza de los errores de medida de las variables dependientes

$$\Theta_{\varepsilon} = \begin{pmatrix} \theta_{\varepsilon_{11}} & & \\ 0 & \theta_{\varepsilon_{22}} & \\ 0 & 0 & \theta_{\varepsilon_{33}} \end{pmatrix} \quad (21)$$

En este ejemplo las matrices de varianzas-covarianzas de los errores de medida son diagonales, ya que no se hipotetiza ninguna relación entre las variables observables que no sea debida al modelo de las variables latentes.

## TIPOS DE RELACIONES

En las técnicas multivariantes estamos acostumbrados a estudiar la relación simultánea de diversas variables entre sí. En estas técnicas las relaciones entre variables dependientes e independientes son todas del mismo nivel o de las mismas características. En un modelo de ecuaciones estructurales podemos distinguir distintos tipos de relaciones. Entender estos distintos tipos de relaciones puede ser de gran ayuda a la hora de formular los modelos a partir de las verbalizaciones en lenguaje común. A continuación vamos a discutir estos tipos de relaciones, siguiendo el esquema propuesto por Saris y Stronkhorst (1984).

### *Covariación vs Causalidad*

Decimos que dos fenómenos *covarian*, o que están correlacionados, cuando al observar una mayor cantidad de uno de los fenómenos también se observa una mayor cantidad del otro fenómeno que covaría con el primero. De igual forma, a niveles bajos del primer fenómeno se asocian niveles bajos del segundo. Así, por ejemplo, cuando decimos que inteligencia y rendimiento correlacionan entre sí, esperamos que los sujetos con un mayor nivel de inteligencia manifiesten un mejor rendimiento y viceversa. Sin embargo, ya hemos enfatizado que covariación y causalidad no son la misma cosa. Cuando se observa una alta relación (covariación) entre dos variables, no debemos interpretarla como una relación causal entre ambas. Pueden existir otras variables que no hemos observado y que potencien o atenúen esta relación. Por ejemplo, en la relación existente entre inteligencia y rendimiento, es esperable que exista también la influencia positiva del tiempo dedicado al estudio sobre el rendimiento y el número de horas de estudio puede afectar a la relación de la inteligencia con el rendimiento (potenciando o atenuando la relación original). Un ejemplo tal vez más claro es el propuesto por Saris. Si recolectamos datos sobre el número de vehículos y el número de

aparatos telefónicos en distintas poblaciones, es seguro que encontraremos una covariación entre ambas variables. No por ello aseguraremos que un mayor número de vehículos es el causante de haya un mayor número de aparatos telefónicos.

Otro nivel de análisis es la causalidad. Si recogemos información sobre el número de fumadores en una habitación y la cantidad de humo existente en la habitación, observaremos que existe una alta covariación entre ambas variables. Parece razonable dar un paso más en la interpretación del resultado de la covariación y argumentar, conceptualmente, que la cantidad de fumadores *causa* la cantidad de humo y que los cambios en la cantidad de fumadores causarán un cambio en la cantidad de humo observada en la habitación.

El cambio de perspectiva desde la covariación observada a la causalidad atribuida a dos variables lo lleva a cabo el investigador, que es quien hipotetiza la causalidad. Es una buena costumbre que las verbalizaciones, o enunciados, sean explícitos respecto al tipo de relación que deseamos probar entre dos variables.

### Ejemplo 6

Los ejemplos que hemos expuesto en este apartado pueden representarse mediante la representación gráfica que hemos desarrollado hasta aquí.

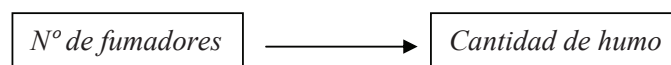
Si estamos estudiando la correlación entre Inteligencia y Rendimiento deberemos representarla como una flecha curva entre ambas variables.

Figura 7. Relación de covariación



Por el contrario, la relación causal entre el número de fumadores y la cantidad de humo la representaremos como un vector que apunte de la causa hacia el efecto.

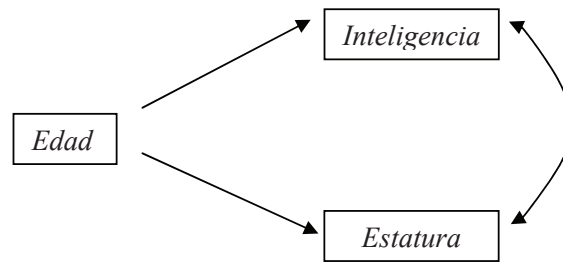
Figura 8. Relación de tipo causal



### Relación espúrea

Una relación causal básica o una relación de covariación involucran a dos variables. En una relación espúrea la relación debe comprender al menos tres variables. Una relación espúrea se refiere a la existencia de covariación entre dos variables que es debida, total o parcialmente, a la relación común de ambas variables con una tercera. Por tanto, mientras que la covariación entre dos variables puede ser muy elevada, su relación causal puede ser nula. El ejemplo típico es la covariación entre estatura e inteligencia en preescolares. Si medimos ambas variables a niños de preescolar es muy posible que encontremos una alta covariación entre ellas, sin embargo, a nadie se le ocurrirá plantear que la estatura causa la inteligencia. Existe una tercera variable, el desarrollo del niño (la edad) que es causa de ambas variables y que hace que se observe la covariación. Gráficamente se puede representar de la siguiente forma.

Figura 9. Relación espúrea



Para estudiar la presencia de este fenómeno se utiliza el coeficiente de correlación parcial, que mide la relación entre dos variables tras eliminar el efecto de una tercera (dependiendo del orden de la correlación parcial también puede eliminarse simultáneamente el efecto de más de una variable). En nuestro ejemplo, la correlación entre las tres variables será alta y positiva, mientras que la correlación parcial entre la inteligencia y la estatura, tras eliminar el efecto estadístico de la edad, será prácticamente nula.

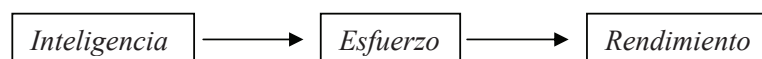
En general, podemos decir que la relación causal entre dos variables implica que ambas variables coviarán, permaneciendo constantes el resto de las variables, pero lo contrario no es cierto. La covariación entre dos variables no implica necesariamente que exista una relación causal entre ambas. La relación puede ser espúrea (aleatoria).

#### *Relación causal directa e indirecta*

Hasta ahora sólo hemos mencionado relaciones causales directas. Una relación causal indirecta implica la presencia de tres variables. Existe una relación indirecta entre dos variables cuando una tercera variable mediatiza el efecto entre ambas. Es decir, cuando el efecto entre la primera y la segunda pasa a través de la tercera. A las variables que median en una relación indirecta se las denomina también variables moduladoras.

Volviendo al ejemplo de la relación entre la inteligencia, el rendimiento y el esfuerzo, podemos pensar en el esfuerzo (medido como las horas de estudio) como una variable moduladora. Existiendo un efecto indirecto entre la inteligencia y el rendimiento. Esta relación puede representarse gráficamente como:

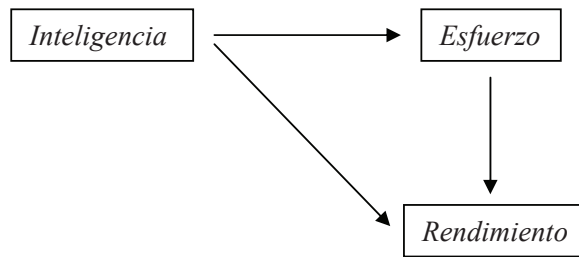
Figura 10. Relación causal indirecta



El efecto indirecto entre la variable Inteligencia y el Rendimiento puede ser potenciado (o atenuado) por la variable moduladora Esfuerzo.

La existencia de un efecto indirecto entre dos variables no anula la posibilidad de que también exista un efecto directo entre ellas. Así, el ejemplo que acabamos de utilizar puede hacerse más complejo de la siguiente forma.

Figura 11. Relaciones directa e indirecta



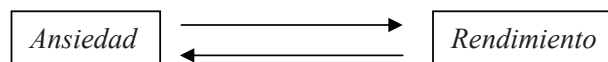
Una vez más, será el investigador quién debe explicitar el tipo de relaciones que su teoría es capaz de justificar.

### *Relación causal recíproca*

La relación causal entre dos variables puede ser recíproca o unidireccional. Cuando la relación es recíproca (bidireccional) la variable causa es a su vez efecto de la otra. Este tipo de relaciones se representa como dos flechas separadas orientadas en sentidos contrarios. Una relación recíproca es en definitiva un bucle de retroalimentación entre dos variables. La relación causal recíproca puede ser directa o indirecta, implicando a otras variables antes de cerrarse el bucle.

La relación entre niveles altos de Ansiedad y el Rendimiento puede representarse como un bucle recíproco. A mayor ansiedad peor será el rendimiento y cuanto más empeore el rendimiento más aumentará la ansiedad.

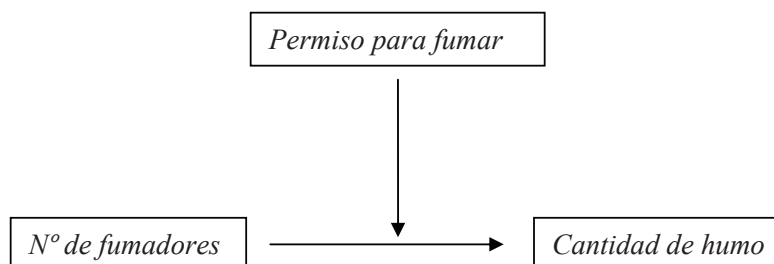
Figura 12. Relación causal recíproca



### *Relación causal condicionada*

Existen situaciones en las que la presencia (o valor) de una variable puede determinar la existencia de relación entre otras dos variables. Este tipo de relaciones se denomina relación causal condicionada y, aunque conceptualmente asumibles, son difíciles de formular en un modelo causal. Una variable condicional determina el tamaño del efecto existente entre otras dos variables. Por ejemplo, en el ejemplo de la relación entre el número de fumadores y la cantidad de humo existente en la habitación, esta relación podrá observarse si está permitido fumar en espacios cerrados, caso contrario es difícil que a los fumadores se les llegue a permitir crear cantidad alguna de humo. Gráficamente es fácil distinguir este tipo de relaciones de las relaciones indirectas. En nuestro ejemplo,

Figura 13. Relación causal condicionada



La manera de estimar estas relaciones condicionadas es mediante el análisis por subgrupos.

### *Efectos totales*

Como hemos visto, cada tipo de relación causal se representa mediante un tipo de efecto. Existe un último tipo de efecto (o relación) que no hemos mencionado, son los efectos *no analizados*. Este tipo de efectos pueden ser debidos a dos motivos. Por un lado puede ser debido a un error de especificación del modelo, en el que no se ha tenido en cuenta otras variables que pueden explicar también la covariación observada entre las variables. Por otro, puede ser debido a que se asume que el resto de las variables no consideradas en el modelo se compensan entre sí, incorporándose su efecto en los términos de error del modelo.

A la suma de efectos directos e indirectos entre dos variables se les denomina *efecto total*. A la suma de los efectos espúreos más los efectos no analizados se les denomina *efectos no causales*. Una vez que el modelo está definido, los efectos espúreos aparecen cuando las variables endógenas están implicadas en la correlación. Las matrices del modelo que contienen estos efectos son  $\Psi$  y  $\Phi$ . Los efectos no analizados aparecen cuando las variables observables están correlacionadas y se especifican en el modelo en las matrices  $\Theta_{\delta}$  y  $\Theta_{\epsilon}$ .

## IDENTIFICACIÓN

Una vez especificado el modelo que se desea ajustar a la matriz de varianza-covarianza muestral, deberemos derivar (en teoría) todas las ecuaciones necesarias para estimar los parámetros del modelo, al estilo como se hizo en las páginas 49-50. Una vez derivadas las ecuaciones que permiten expresar los elementos de la matriz de varianza-covarianza como una función de los parámetros del modelo obtendremos tantas ecuaciones como elementos diferentes posea la matriz de varianza-covarianza (recordemos que es simétrica y la submatriz triangular superior es redundante con la triangular inferior). Para que el modelo esté identificado el número de ecuaciones deberá ser al menos igual al número de parámetros contenidos en el modelo, de lo contrario será imposible obtener una solución única para el sistema y los parámetros no podrán ser estimados debidamente. Los modelos causales pueden estar:

- *identificados* (just identified o exactamente identificados): el número de ecuaciones es igual al número de parámetros por estimar y la solución es única.
- *Sobreidentificados*: el número de ecuaciones posibles es superior al número de parámetros.
- *Infraidentificados*: el número de ecuaciones es menor que el número de parámetros y existen infinitas soluciones.

Los modelos exactamente identificados no admiten contrastes estadísticos, ya que reproducen exactamente la matriz de correlaciones. Estos modelos, también denominados completamente recursivos, se dan cuando todas las variables del modelo se encuentran relacionadas con otras variables mediante líneas curvas (para las variables exógenas) o líneas rectas (para las endógenas). Para poder contrastar estadísticamente un modelo, deberemos imponer restricciones sobre algunos de los parámetros, anulándolos o igualándolos a otros parámetros del modelo, dando lugar a un modelo sobreidentificado. En ese caso la reproducción de la matriz de correlaciones no será perfecta, y será posible evaluar estadísticamente el ajuste del modelo. La cuestión, en definitiva, es similar a lo que sucede cuando se ajusta un modelo loglineal saturado. Los parámetros del modelo reproducen exactamente las frecuencias de

las casillas y no es posible interpretar la significación de los parámetros, ya que el ajuste es perfecto. El problema de la identificación ha sido tratado extensamente en la literatura económica. Véanse Fischer (1966) y Johnston (1972).

Para los modelos complejos, evaluar la identificación es prácticamente imposible, ya que derivar todas las ecuaciones necesarias para expresar la matriz de varianza-covarianza en función de los parámetros resulta una tarea titánica. Sin embargo, existen reglas heurísticas para inferir si el modelo está o no identificado.

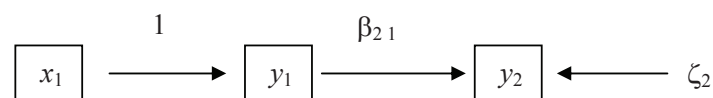
El número de elementos distintos de una matriz de dimensiones  $n \times n$  es  $\frac{1}{2} n(n + 1)$ . Se define como los grados de libertad del modelo, la diferencia entre el número de ecuaciones posibles, que es igual al número de elementos distintos de la matriz de varianza-covarianza, y el número de parámetros del modelo. Para que el modelo esté identificado los grados de libertad deben ser mayores o iguales que 0. Esta es una condición *necesaria* para la identificación del modelo, pero no es *suficiente*.

A la hora de determinar el número de parámetros del modelo deberemos observar el diagrama del modelo con cierta cautela, ya que se tiende a no representar algunos de los parámetros para no dificultar la claridad del gráfico. Así por ejemplo, el número de parámetros  $\phi$  será igual a  $\frac{1}{2} q(q + 1)$ , donde  $q$  es el número de variables  $x$ . El número de parámetros  $\psi$  es al menos  $p$ . Y este número deberá aumentarse a medida que existan más elementos no nulos que representen covarianzas entre los términos de error endógenos. Algo similar sucede con las correlaciones entre los errores de las variables observables, contenidas en las matrices  $\Theta$ , que pueden aumentar rápidamente si se proponen relaciones entre los errores (relaciones espúreas).

Una manera de evitar el número excesivo de parámetros es desarrollar el modelo en puntuaciones típicas, es decir, tratando sólo con correlaciones. De esta forma, las varianzas de algunas variables pueden ser fijadas de antemano, ganando grados de libertad, por ejemplo la diagonal de la matriz  $\Phi$ .

### Ejemplo 7

Sea el modelo definido por el siguiente diagrama causal.



Las ecuaciones estructurales correspondientes a este modelo son:

$$\begin{aligned} y_1 &= x_1 \\ y_2 &= \beta_{21}y_1 + \zeta_2 \end{aligned} \quad (22)$$

Las matrices de parámetros de este modelo son

$$\begin{aligned} \mathbf{B} &= \begin{pmatrix} 0 & 1 \\ \beta_{21} & 0 \end{pmatrix}, & \mathbf{\Gamma} &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ \mathbf{\Psi} &= \begin{pmatrix} 0 & 0 \\ 0 & \psi_{22} \end{pmatrix}, & \mathbf{\Phi} &= (\phi_{11}) \end{aligned} \quad (23)$$



Donde se asume que  $\text{COV}(\zeta_2, x_1) = 0$ , y algunos de los parámetros han sido fijados a valores específicos (constantes) de 0 y 1.

El número de ecuaciones posibles es igual a  $\frac{1}{2} n(n + 1) = \frac{1}{2} 3(3 + 1) = 6$  y el número de parámetros por estimar es 3. El modelo cumple la condición necesaria de identificación.

Si no hubiéramos impuesto restricciones sobre los parámetros, el número de parámetros por estimar hubiera sido 8 (ya que, por definición,  $\psi_{12} = \psi_{21}$  y  $\beta_{11} = \beta_{22} = 0$ ) y el modelo no estaría identificado.

Sin embargo, el hecho de que se cumpla esta regla necesaria para la identificación no quiere decir que las ecuaciones permitan obtener soluciones únicas para los parámetros. Deben cumplirse otras condiciones *suficientes* para que los parámetros estén completamente identificados. Por ello, conviene no plantear el problema de la identificación desde la perspectiva del modelo global, sino más bien desde la perspectiva de cada parámetro. Dado un modelo particular, algunos de los parámetros pueden estar sobre-identificados, mientras que otros pueden estar infra-identificados. A partir de la igualdad  $\Sigma = \Sigma(\theta)$ , Bollen (1989) plantea la identificación de una manera más completa:

“Si un parámetro desconocido de la  $\theta$  puede ser expresado como una función de uno o más elementos de  $\Sigma$ , ese parámetro estará identificado. Si todos los parámetros desconocidos de  $\theta$  están identificados, entonces el modelo está identificado.” (Página 89).

Una manera distinta de ver la identificación es considerando dos vectores de parámetros, de dimensiones  $t \times 1$ ,  $\theta_1$  y  $\theta_2$ , cada uno de ellos conteniendo valores particulares de  $\theta$ . A partir de ellos se pueden desarrollar las matrices de varianza-covarianza implícitas  $\Sigma(\theta_1)$  y  $\Sigma(\theta_2)$ , para cada una de las soluciones del vector. Si el modelo está identificado, todas las soluciones de  $\theta_1$  y  $\theta_2$  en las que  $\Sigma(\theta_1) = \Sigma(\theta_2)$ , deberán tener  $\theta_1 = \theta_2$ . Si existe un par de vectores  $\theta_1$  y  $\theta_2$  que cumplan que  $\Sigma(\theta_1) = \Sigma(\theta_2)$  mientras que  $\theta_1 \neq \theta_2$ , entonces  $\theta$  no estará identificado.

Como ya hemos comentado, comprobar algebraicamente la identificación, mediante el desarrollo de las ecuaciones correspondientes, es prácticamente inviable, especialmente cuando el modelo es suficientemente complejo o cuando el número de variables contenidas en la matriz de varianza-covarianza es elevado. Para evitar este paso se aplican algunas reglas para tipos particulares de modelos que aseguran la identificabilidad. A continuación se mencionan dos de ellas.

Una regla *suficiente* de identificación es la de **B** nula. Cuando en un modelo las variables endógenas no influyen unas sobre otras la matriz  $\mathbf{B} = \mathbf{0}$ . Se puede demostrar que estos modelos siempre están identificados, asumiendo además que no existen errores de medida y que no existe correlación entre las variables endógenas ( $x_i$ ) y los errores del modelo ( $\zeta_j$ ).

Otra regla suficiente es la de los modelos recursivos. Para un modelo recursivo la matriz **B** será triangular y la matriz  $\Psi$  será diagonal, si esto es así, el modelo estará identificado. En ocasiones, para los modelos recursivos, es necesario reordenar las variables endógenas para comprobar que estas reglas se cumplen. Esta condición es suficiente pero no necesaria.

Existen otras reglas similares a estas, como la condición de orden y la condición de rango, cuya revisión excede el propósito de esta monografía. Para una exposición más detallada consúltese Bollen (1989), Fox (1984), Blalock (1971) o Bekker y Pollock (1986).

Memorizar estas reglas carece de sentido, y comprobar la aplicabilidad de algunas de ellas puede resultar bastante complejo. La manera más frecuente de proceder, por parte de los analistas, es someter el modelo a su estimación mediante los programas al uso y dejar que estas reglas sean comprobadas directamente por el programa. Aunque no sea un proce-



dimiento muy ortodoxo, sí resulta tremendamente eficaz. Si a pesar de haber realizado una especificación cuidadosa y detallada del modelo, antes de su implementación en el programa de estimación, el resultado es que el modelo no está debidamente identificado, la manera de proceder más aconsejable es establecer restricciones sobre el modelo para de esta forma ganar grados de libertad. Este proceso es lo que se ha denominado *corrección teórica* o *calibración de la teoría* (Theory trimming). En este proceso se eliminan aquellos efectos que no son significativos desde un punto de vista conceptual o estadístico, por ejemplo, a partir de los coeficientes de regresión. El problema que puede surgir es que los datos puedan dictar el modelo y no la propia teoría. Otra posibilidad es comenzar con submodelos sobreidentificados y contrastar la pérdida de ajuste a medida que se van liberando parámetros.

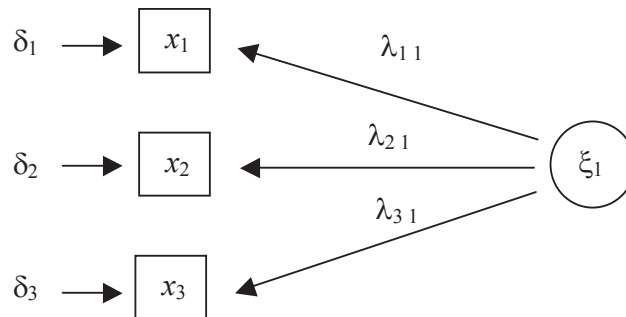
En cualquier caso, dado que los programas de estimación siguen procedimientos numéricos iterativos para encontrar la mejor estimación de los parámetros, es una buena costumbre comprobar que el proceso de estimación de cada parámetro ha seguido una trayectoria monótona, sin grandes cambios bruscos de unos valores a otros, antes de converger en la solución final (por ejemplo a través de las derivadas parciales de los parámetros aislados).

### Ejemplo 8

En este ejemplo se estudia cómo identificar un modelo de medida con un solo factor latente. En el ejemplo se utiliza un recurso muy frecuente en la identificación de los modelos que consiste en establecer una de las saturaciones con el valor unitario.

Sea un modelo de un factor con tres indicadores como el de la figura 14.

Figura 14. Modelo factorial de un factor con tres indicadores



La matriz de varianzas-covarianzas poblacional para las tres variables observables es la siguiente:

$$\Sigma = \begin{pmatrix} \text{VAR}(x_1) & & \\ \text{COV}(x_2, x_1) & \text{VAR}(x_2) & \\ \text{COV}(x_3, x_1) & \text{COV}(x_3, x_2) & \text{VAR}(x_3) \end{pmatrix} \quad (24)$$

La matriz de varianza-covarianza implícita (en diferenciales) tendrá la forma general

$$\begin{aligned} \Sigma(\theta) &= E(\mathbf{xx}') = \\ &= E[(\Lambda_x \xi + \delta)(\xi' \Lambda_x' + \delta')] = \\ &= \Lambda_x E(\xi \xi') \Lambda_x' + \Theta_\delta = \\ &= \Lambda_x \Phi \Lambda_x' + \Theta_\delta \end{aligned} \quad (25)$$

Las ecuaciones estructurales del modelo de un factor con tres indicadores, así como los supuestos del modelo son las siguientes:

$$\begin{aligned}x_1 &= \lambda_{11}\xi_1 + \delta_1 \\x_2 &= \lambda_{21}\xi_1 + \delta_2 \\x_3 &= \lambda_{31}\xi_1 + \delta_3\end{aligned}\tag{26}$$

$$\begin{aligned}\text{COV}(\xi_1, \delta_i) &= 0, \quad \text{para } i = 1, 2, 3 \\ \text{COV}(\delta_i, \delta_j) &= 0, \quad \text{para } i \neq j\end{aligned}$$

Las matrices pertinentes para los parámetros del modelo son:

$$\begin{aligned}\mathbf{x} &= \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} & \Lambda_x &= \begin{pmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \end{pmatrix} & \xi &= (\xi_1) & \Phi &= (\phi_{11}) \\ \delta &= \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{pmatrix} & \theta_\delta &= \begin{pmatrix} \text{VAR}(\delta_1) & 0 & 0 \\ 0 & \text{VAR}(\delta_2) & 0 \\ 0 & 0 & \text{VAR}(\delta_3) \end{pmatrix}\end{aligned}\tag{27}$$

Sustituyendo en la fórmula 25, que determina la construcción de la matriz de varianzas-covarianzas implícita, con las matrices contenidas en 27, obtenemos la siguiente matriz función de los parámetros estructurales:

$$\Sigma(\theta) = \begin{pmatrix} \lambda_{11}^2\phi_{11} + \text{VAR}(\delta_1) & & \\ \lambda_{21}\lambda_{11}\phi_{11} & \lambda_{21}^2\phi_{11} + \text{VAR}(\delta_2) & \\ \lambda_{31}\lambda_{11}\phi_{11} & \lambda_{31}\lambda_{21}\phi_{11} & \lambda_{31}^2\phi_{11} + \text{VAR}(\delta_3) \end{pmatrix}\tag{28}$$

Igualando los elementos de la matriz de varianzas-covarianzas de las variables observables (24) con los términos de la matriz implícita (28) obtenemos el siguiente conjunto de 6 ecuaciones

$$\begin{aligned}\text{VAR}(x_1) &= \lambda_{11}^2\phi_{11} + \text{VAR}(\delta_1) \\ \text{VAR}(x_2) &= \lambda_{21}^2\phi_{11} + \text{VAR}(\delta_2) \\ \text{VAR}(x_3) &= \lambda_{31}^2\phi_{11} + \text{VAR}(\delta_3) \\ \text{COV}(x_2, x_1) &= \lambda_{21}\lambda_{11}\phi_{11} \\ \text{COV}(x_3, x_1) &= \lambda_{31}\lambda_{11}\phi_{11} \\ \text{COV}(x_3, x_2) &= \lambda_{31}\lambda_{21}\phi_{11}\end{aligned}\tag{29}$$

Con lo que disponemos de 6 ecuaciones y 7 incógnitas. El sistema no está determinado. Para poder obtener una solución para los parámetros, debemos imponer al menos una restricción. Habitualmente se acostumbra a fijar una de las saturaciones a 1 o bien fijar a 1 la

varianza del factor latente cuando se trabaja con la solución estandarizada (y por ello con la matriz de correlaciones).

Imponiendo la restricción  $\lambda_{11} = 1$ , la matriz de varianzas-covarianzas implícitas se simplifica a la siguiente

$$\mathbf{\Sigma}(\boldsymbol{\theta}) = \begin{pmatrix} \phi_{11} + \text{VAR}(\delta_1) & & & \\ \lambda_{21}\phi_{11} & \lambda_{21}^2\phi_{11} + \text{VAR}(\delta_2) & & \\ \lambda_{31}\phi_{11} & \lambda_{31}\lambda_{21}\phi_{11} & \lambda_{31}^2\phi_{11} + \text{VAR}(\delta_3) & \\ & & & \end{pmatrix} \quad (30)$$

Dando lugar a las siguientes ecuaciones,

$$\begin{aligned} \text{VAR}(x_1) &= \phi_{11} + \text{VAR}(\delta_1) \\ \text{VAR}(x_2) &= \lambda_{21}^2\phi_{11} + \text{VAR}(\delta_2) \\ \text{VAR}(x_3) &= \lambda_{31}^2\phi_{11} + \text{VAR}(\delta_3) \\ \text{COV}(x_2, x_1) &= \lambda_{21}\phi_{11} \\ \text{COV}(x_3, x_1) &= \lambda_{31}\phi_{11} \\ \text{COV}(x_3, x_2) &= \lambda_{31}\lambda_{21}\phi_{11} \end{aligned} \quad (31)$$

Operando se obtiene el siguiente conjunto de soluciones,

$$\begin{aligned} \lambda_{21} &= \frac{\text{COV}(x_2, x_3)}{\text{COV}(x_3, x_1)} \\ \lambda_{31} &= \frac{\text{COV}(x_2, x_3)}{\text{COV}(x_2, x_1)} \\ \phi_{11} &= \frac{\text{COV}(x_2, x_1)\text{COV}(x_3, x_1)}{\text{COV}(x_3, x_2)} \\ \text{VAR}(\delta_1) &= \text{VAR}(x_1) - \phi_{11} \\ \text{VAR}(\delta_2) &= \text{VAR}(x_2) - \lambda_{21}^2\phi_{11} \\ \text{VAR}(\delta_3) &= \text{VAR}(x_3) - \lambda_{31}^2\phi_{11} \end{aligned} \quad (32)$$

El modelo está ahora identificado y existe una sola solución única.

ESTIMACIÓN Y CONTRASTE DE HIPÓTESIS<sup>3</sup>

Derivando algebraicamente la hipótesis fundamental de ajuste,  $\Sigma = \Sigma(\theta)$ , se puede demostrar que la matriz de varianza-covarianza implícita (la expresada en función de los parámetros) de cualquier modelo de ecuaciones estructurales es una matriz dividida de la forma

$$\Sigma(\theta) = \begin{pmatrix} (\mathbf{I} - \mathbf{B})^{-1}(\Gamma\Phi\Gamma' + \Psi)^{-1}(\mathbf{I} - \mathbf{B})^{-1'} & (\mathbf{I} - \mathbf{B})^{-1}\Gamma\Phi \\ \Gamma\Phi'(\mathbf{I} - \mathbf{B})^{-1'} & \Phi \end{pmatrix} \quad (33)$$

Si el modelo es correcto y los parámetros poblacionales son conocidos, entonces  $\Sigma$  será igual a  $\Sigma(\theta)$ .

*Ejemplo 9*

Sea por ejemplo la ecuación estructural

$$y_1 = x_1 + \xi_1 \quad (34)$$

Donde  $\gamma_{11}$  ha sido fijado a 1. La matriz de varianza-covarianza poblacional para las variables  $y_1$  y  $x_1$  es

$$\Sigma = \begin{pmatrix} \text{VAR}(y_1) & \text{COV}(y_1, x_1) \\ \text{COV}(x_1, y_1) & \text{VAR}(x_1) \end{pmatrix} \quad (35)$$

y la matriz implícita en función de los parámetros es

$$\Sigma(\theta) = \begin{pmatrix} \phi_{11} + \psi_{11} & \phi_{11} \\ \phi_{11} & \phi_{11} \end{pmatrix} \quad (36)$$

Asumiendo que el modelo es correcto y que los parámetros poblacionales son conocidos, cada elemento de  $\Sigma$  deberá ser igual a cada elemento de la misma posición de  $\Sigma(\theta)$ . Obsérvese que el parámetro  $\phi_{11}$  está sobreidentificado, ya que es igual, simultáneamente a  $\text{COV}(x_1, y_1)$  y a  $\text{VAR}(x_1)$ . Eso no quiere decir que el parámetro tome dos valores diferentes, sino que, en la población la varianza de  $x_1$  y la covarianza deben ser iguales.

En la práctica, no conoceremos ni los valores de los parámetros ni los valores poblacionales de las varianzas y las covarianzas. Debemos estimar estos valores a partir de la muestra. La matriz de varianza-covarianza muestral será

$$\mathbf{S} = \begin{pmatrix} \text{var}(y_1) & \text{cov}(y_1, x_1) \\ \text{cov}(x_1, y_1) & \text{var}(x_1) \end{pmatrix} \quad (37)$$

La cual tomará valores concretos para una muestra dada.

A continuación deberemos seleccionar valores para los parámetros del modelo,  $\phi_{11}$  y  $\psi_{11}$ , los cuales representaremos, con su notación de estimadores, como  $\hat{\phi}_{11}$  y  $\hat{\psi}_{11}$ , respectivamente. Una vez elegidos estos valores, podemos sustituirlos en la matriz de covarianza implícita estimada  $\Sigma(\hat{\theta})$ , que suele representarse también como  $\hat{\Sigma}$ .

3 En este apartado seguiremos la argumentación de Bollen (1989).

$$\hat{\Sigma} = \begin{pmatrix} \hat{\phi}_{11} + \hat{\psi}_{11} & \hat{\phi}_{11} \\ \hat{\phi}_{11} & \hat{\phi}_{11} \end{pmatrix} \quad (38)$$

Igualando la matriz implícita a la matriz muestral obtenemos el siguiente conjunto de ecuaciones:

$$\left. \begin{aligned} \text{var}(y_1) &= \hat{\phi}_{11} + \hat{\psi}_{11} \\ \text{var}(x_1) &= \hat{\phi}_{11} \\ \text{cov}(y_1, x_1) &= \hat{\phi}_{11} \end{aligned} \right\} \quad (39)$$

El problema de la estimación es ahora obtener unos valores adecuados para los parámetros del modelo, de forma que sea lo más parecida posible a la matriz S.

Más adelante comentaremos los métodos posibles para estimar los valores de los parámetros. Por ahora, a modo ilustrativo, supongamos que la matriz de covarianza muestral es

$$\mathbf{S} = \begin{pmatrix} 10 & 6 \\ 6 & 4 \end{pmatrix} \quad (40)$$

Además, hemos decidido que las mejores estimaciones posibles son 7 para  $\hat{\phi}_{11}$  y 3 para  $\hat{\psi}_{11}$ . Por tanto, la matriz de covarianza implícita será

$$\hat{\Sigma} = \begin{pmatrix} 7+3 & 7 \\ 7 & 7 \end{pmatrix} = \begin{pmatrix} 10 & 7 \\ 7 & 7 \end{pmatrix} \quad (41)$$

Si formamos la matriz residual ( $\mathbf{S} - \hat{\Sigma}$ ), ésta contendrá los residuos obtenidos de intentar sustituir la matriz de covarianzas muestrales por la matriz implícita. Esta matriz nos indicará lo próximas que se encuentran las matrices S y  $\hat{\Sigma}$ . En nuestro caso:

$$(\mathbf{S} - \hat{\Sigma}) = \begin{pmatrix} 10-10 & 6-7 \\ 6-7 & 4-7 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ -1 & -3 \end{pmatrix} \quad (42)$$

Las estimaciones de los parámetros que hemos utilizado permiten un ajuste perfecto de la  $\text{var}(y_1)$ , mientras que se ajustan peor en el pronóstico de  $\text{cov}(x_1, y_1)$  y de  $\text{var}(x_1)$ , sobrestimando su valor.

Supongamos ahora que las estimaciones que seleccionamos para los parámetros son  $\hat{\phi}_{11} = 5$  y  $\hat{\psi}_{11} = 5$ . La matriz implícita reproducida por estos parámetros es

$$\hat{\Sigma} = \begin{pmatrix} 10 & 5 \\ 5 & 5 \end{pmatrix} \quad (43)$$

Mientras que la matriz residual obtenida es

$$(\mathbf{S} - \hat{\Sigma}) = \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} \quad (44)$$

Aunque en este caso el ajuste tampoco es perfecto, el nuevo conjunto de valores para los parámetros parece que dan lugar a un mejor ajuste a la matriz S.

Cuando las matrices que se desea comparar son de reducidas dimensiones, como en el ejemplo, la matriz residual ofrece una sensación bastante clara de la bondad de las estimaciones de los parámetros que se seleccionan en cada ocasión. Sin embargo, es más frecuente utilizar una función que dé como resultado un único escalar y que resuma el ajuste de todos

los elementos de la matriz residual. A esta función se la representa, de manera genérica, como  $F(\mathbf{S}, \Sigma(\theta))$ , ya que compara los valores de la matriz de varianza-covarianza muestral con la matriz implícita desarrollada a partir de los parámetros del modelo. Si sustituimos un conjunto de valores para el vector de parámetros  $\theta$ , la *función de ajuste* se puede escribir como  $F(\mathbf{S}, \hat{\Sigma})$ .

Para modelos más complejos, la matriz implícita estimada  $\hat{\Sigma}$ , será desarrollada a partir de un mayor número de parámetros, en particular los contenidos en  $\mathbf{B}$ ,  $\Gamma$ ,  $\Phi$  y  $\Psi$ , pero la función de ajuste seguirá siendo de la misma forma. La elección de una función de ajuste particular es la que da lugar a los distintos métodos de estimación de los parámetros. Algunos de los métodos de estimación más extendidos son el de mínimos cuadrados ordinarios, el de máxima verosimilitud y el de mínimos cuadrados generalizados. Cualquiera que sea el método de estimación utilizado, y por tanto, cualquiera que sea la función de ajuste seleccionada, las funciones de ajuste deben cumplir las propiedades:

1. El resultado de  $F(\mathbf{S}, \Sigma(\theta))$  debe ser un escalar
2.  $F(\mathbf{S}, \Sigma(\theta)) \geq 0$
3.  $F(\mathbf{S}, \Sigma(\theta)) = 0$  sólo si  $\Sigma(\theta) = \mathbf{S}$ .
4.  $F(\mathbf{S}, \Sigma(\theta))$  es una función continua en  $\mathbf{S}$  y  $\Sigma(\theta)$ .

Browne (1984) demuestra que al minimizar las funciones que cumplan estas propiedades se obtienen estimadores consistentes de los parámetros contenidos en  $\theta$ .

#### *Máxima verosimilitud*<sup>4</sup>

La estimación de máxima verosimilitud es posiblemente la más extendida en la estimación de modelos estructurales. Su función de ajuste es

$$F_{ML} = \log|\Sigma(\theta)| + \text{tr}(\mathbf{S}\Sigma(\theta)^{-1}) - \log|\mathbf{S}| - (p+q) \quad (45)$$

donde  $p$  es el número de variables endógenas observables y  $q$  es el número de variables exógenas observables. Se asume que las matrices  $\mathbf{S}$  y  $\Sigma(\theta)$  son positivas definidas para evitar que sus determinantes sean nulos y que no existan sus logaritmos. Además, se asume que los vectores de variables observadas  $\mathbf{x}$  e  $\mathbf{y}$  siguen una distribución normal multivariante.

Es fácil demostrar que esta función de ajuste cumple la tercera propiedad del apartado anterior. Si  $\mathbf{S} = \Sigma(\theta)$ , la función  $F_{ML}$  se reduce a

$$F_{ML} = \log|\mathbf{S}| + \text{tr}(\mathbf{I}) - \log|\mathbf{S}| - (p+q) \quad (46)$$

como  $\text{tr}(\mathbf{I}) = p+q$ , entonces  $F_{ML}$  toma el valor cero. Luego la función de hecho alcanzará un valor de cero sólo cuando el ajuste es perfecto.

Una vez definida la función de ajuste el problema se centra en obtener un mínimo absoluto para la función. Para ello bastará con obtener las derivadas primera y segunda respecto al vector de parámetros  $\theta$ . Sin embargo, obtener analíticamente estas derivadas puede ser prácticamente imposible, por lo que se recurre a métodos numéricos de estimación. Es decir, métodos que de manera iterativa van sustituyendo los valores de los parámetros por valores tales que la función se aproxime rápidamente a su mínimo. Algunos métodos numéricos de

<sup>4</sup> El término Máxima verosimilitud se expresa en Inglés como Maximum Likelihood que, abreviado, se acostumbra a escribir como ML.

estimación son el de Newton-Raphson, el de Fletcher y Powell (una modificación de éste es el utilizado por el *lisrel*) y el de Gauss-Newton (el utilizado por el *eqs*).

La estimación de máxima verosimilitud tiene una serie de propiedades estadísticas deseables. Estas propiedades son asintóticas, lo que quiere decir que se cumplen para muestras grandes. Aunque la estimación puede estar sesgada en muestras pequeñas, asintóticamente es insesgada. La estimación es consistente, es asintóticamente eficiente y, a medida que el tamaño de la muestra aumenta, la distribución de las estimaciones tiende a una distribución normal. Esta última propiedad implica que es posible obtener los errores típicos de estimación, con lo que es fácil obtener sus intervalos de confianza y con ellos comprobar si puede asumirse que su valor es nulo en la población. Además, las estimaciones, y la función, son invariantes respecto a la escala de medida.

Existen también algunos problemas asociados a la utilización de este método de estimación. El primero es que deben cumplirse los supuestos en los que se basa la estimación, es decir, la normalidad, el tamaño de la muestra y la especificación correcta del modelo. La segunda es que los procesos de estimación numérica pueden encontrar un mínimo local de la función de ajuste, lo que no asegura la obtención de las estimaciones adecuadas para los parámetros. En tercer lugar, las restricciones impuestas al modelo, fijando los parámetros a valores inadecuados o imponiendo igualdades para sus valores, pueden dar lugar a un modelo inadecuadamente definido. Por último, los procedimientos numéricos (y este es el caso del *lisrel*) pueden dar lugar a estimaciones impropias, como puedan ser varianzas negativas o valores para los parámetros que se encuentran fuera del espacio admisible para los parámetros.

Una propiedad importante de la función FML es que permite desarrollar una prueba de contraste sobre el ajuste global del modelo. Bajo la hipótesis nula  $H_0: \Sigma = \Sigma(\theta)$ , la distribución asintótica del estadístico  $(N - 1)$  FML sigue una distribución  $\chi^2$  con grados de libertad  $\frac{1}{2}(p + q)(p + q + 1) - t$ , donde  $t$  es el número de parámetros libres para su estimación. Por tanto, los valores del estadístico de contraste, para una muestra dada, pueden ser comparados con su distribución teórica y calcular su nivel crítico (su significación). Si el estadístico de contraste toma valores extremadamente altos, deberemos concluir que las restricciones impuestas en el modelo para su sobreidentificación no son adecuadas y que, por tanto,  $\Sigma \neq \Sigma(\theta)$ . A la hora de estudiar los resultados de este contraste debemos tener en cuenta que su utilización está justificada sólo cuando el tamaño de la muestra es grande, las variables se distribuyen normalmente y el modelo especificado es correcto.

### *Otros métodos de estimación*

Además del método de máxima verosimilitud existen otros métodos de estimación que son utilizados con cierta frecuencia.

El método de *mínimos cuadrados no ponderados* (Unweighted Least Squares, ULS) utiliza como función de ajuste la función:

$$F_{ULS} = \frac{1}{2} \text{tr}[(\mathbf{S} - \Sigma(\theta))^2] \quad (47)$$

que equivale a minimizar un medio de la suma de cuadrados de los residuos. A diferencia del método de mínimos cuadrados utilizado en regresión, aquí los residuos se refieren a las matrices de covarianzas y no a los individuos.

Este método de ajuste es fácil de entender intuitivamente y también da lugar a estimaciones consistentes de  $\theta$ , sin necesidad de asumir la multinormalidad de las variables, siempre que los parámetros de  $\theta$  estén identificados. Sin embargo,  $F_{ULS}$  es menos eficiente que la

estimación  $F_{ML}$  y no es invariante respecto a la escala de medida, con lo que las soluciones obtenidas de analizar la matriz de correlaciones y la de varianza-covarianza diferirán.

El método de *mínimos cuadrados generalizados* (Generalized Least Squares, GLS) pondera la función de ajuste por los inversos de las varianzas y covarianzas de las variables. Su función de ajuste es

$$F_{GLS} = \frac{1}{2} \text{tr} \left[ \left\{ (\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta})) \mathbf{W}^{-1} \right\}^2 \right] \quad (48)$$

donde la matriz  $\mathbf{W}$  es una matriz de constantes o cualquier matriz positiva definida. La elección más habitual es  $\mathbf{W}^{-1} = \mathbf{S}^{-1}$ . Cuando  $\mathbf{W}^{-1} = \mathbf{I}$ ,  $F_{GLS}$  es un caso particular de  $F_{ML}$ . Este método de ajuste es preferible cuando la distribución de las variables observadas se desvía de la normalidad.

$F_{GLS}$  es invariante respecto a la escala y da lugar a estimaciones consistentes de distribución multinormal con covarianzas conocidas, lo que permite realizar contrastes estadísticos sobre los valores de los parámetros. Cuando la distribución de las variables observables es demasiado apuntada o demasiado aplanada (leptocúrtica o platicúrtica, respectivamente) la estimación de los errores típicos de estimación puede ser inadecuada y el estadístico de bondad de ajuste global puede ser inapropiado. Esto es debido a que la matriz de varianza-covarianza asintótica de los estimadores es de difícil obtención y, caso de encontrarnos distribuciones anómalas, la derivación de los errores típicos puede ser incorrecta. Si este es el caso, puede estimarse la matriz asintótica  $\mathbf{W}$  de manera externa al propio programa de estimación (por ejemplo con el programa PRELIS) y utilizar esta estimación previa en el proceso de estimación de la función de ajuste.

## ESTADÍSTICOS DE BONDAD DE AJUSTE

Una vez que se ha obtenido una estimación de los parámetros del modelo se deberán efectuar diagnósticos sobre la adecuación de la solución y la bondad del modelo.

En cuanto a la adecuación de la solución, se debe comprobar la significación de los parámetros estimados y las propiedades de las estimaciones. Así por ejemplo, se debe comprobar que no existen casos Heywood, es decir, parámetros que toman valores impropios, como puedan ser varianzas negativas o correlaciones mayores que 1 o menores que -1. También será sintomático un parámetro estimado con un signo contrario al esperado por la teoría que justifica el modelo.

En cuanto a la bondad del modelo, se utilizan estadísticos globales que contrastan la capacidad del modelo estimado para reproducir la matriz de parámetros muestrales (las varianzas y covarianzas calculadas en la muestra). Existen distintos estadísticos de bondad de ajuste. A continuación se mencionan algunos de ellos

### *Chi-cuadrado*

El estadístico de bondad de ajuste chi-cuadrado es una derivación directa de la propia función de ajuste de máxima verosimilitud.

$$F_{ML}(\mathbf{S}, \hat{\boldsymbol{\Sigma}}) = \log |\hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta})| + \text{tr}(\mathbf{S} \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}) - \log |\mathbf{S}| - (p+q) \quad (49)$$



Ponderando la función de ajuste por el tamaño muestral se obtiene el siguiente estadístico.

$$X_2 = (N - 1) F_{ML} \quad (50)$$

Cuando se cumplen los supuestos de estimación, se puede demostrar que el estadístico  $X_2$  sigue una distribución  $\chi^2$  con  $\frac{1}{2}(p + q)(p + q + 1) - t$  grados de libertad, donde  $t$  es el número de parámetros libres para su estimación. Con el estadísticos  $X_2$  se puede contrastar la hipótesis nula  $H_0: \Sigma = \Sigma(\theta)$ . Es decir, que todos los residuos son nulos.

### GFI

Jöreskog y Sörbom (1986) propusieron los estadísticos GFI (Goodness of Fit Index, Estadístico de bondad de ajuste) y AGFI (Adjusted Goodness of Fit Index, Estadístico de bondad de ajuste corregido).

Los estadísticos de bondad de ajuste GFI para las funciones de ajuste de máxima verosimilitud y mínimos cuadrados ordinarios son las siguientes:

$$GFI_{ML} = 1 - \frac{\text{tr}[(\Sigma^{-1}(\theta)\mathbf{S} - \mathbf{I})^2]}{\text{tr}[(\Sigma^{-1}(\theta)\mathbf{S})^2]} \quad (51)$$

$$GFI_{OLS} = 1 - \frac{\text{tr}[(\Sigma^{-1}(\theta) - \mathbf{S})^2]}{\text{tr}(\mathbf{S})^2} \quad (52)$$

Estos estadísticos cuantifican la cantidad de varianza y covarianza contenida en la matriz  $\mathbf{S}$  que es pronosticada por  $\Sigma(\theta)$ . Los valores del estadístico GFI son habitualmente muy elevados y se suele exigir que al menos alcancen el valor 0,9 para ser indicativos de un buen ajuste.

### AGFI

El estadístico de bondad de ajuste corregido AGFI corrige el estadístico GFI por los grados de libertad del modelo y por el número de variables contenidas en el mismo. Este estadístico busca corregir la excesiva benignidad del estadístico GFI.

El estadístico AGFI para las estimaciones de máxima verosimilitud y mínimos cuadrados ordinarios son los siguientes.

$$AGFI_{ML} = \left[1 - \frac{q(q+1)}{2gl}\right] [1 - GFI_{ML}] \quad (53)$$

$$AGFI_{OLS} = \left[1 - \frac{q(q+1)}{2gl}\right] [1 - GFI_{OLS}] \quad (54)$$

El estadístico AGFI favorece los modelos que, con igual ajuste global, contienen un menor número de variables  $y$ , por ello, son más simples. Tanto el estadístico GFI como el AGFI alcanzan su valor máximo de 1 cuando todos los residuos son nulos.

## REFERENCIAS BIBLIOGRÁFICAS

- Arbuckle, J. L. (1997). *Amos Users' Guide. Version 3.6*. Chicago: SmallWaters Corporation.
- Arbuckle, J. L. & W. Wothke (1999). *Amos 4.0 User's Guide*. Chicago: SmallWaters Corporation.
- Bagozzi, R. P. (1980). Performance and satisfaction in an industrial sales force: An examination of their antecedents in simultaneity. *Journal of Marketing*, 44: 65-77.
- Bekker, P. A. & D. S. G. Pollock (1986). Identification of lineal stochastic models with covariance restrictions. *Journal of Econometrics*, 31: 179-208.
- Bentler, P. M. (1985). *Theory and implementation of EQS: A structural equations program*. Los Angeles: BMDP Statistical Software.
- Blalock, H. M., Jr. (Ed.) (1971). *Causal Models in the Social Sciences*. Chicago: Aldine Publishing Co.
- Bollen, K. (1989). *Structural Equations with Latent Variables*. New York: John Wiley & sons.
- Boudon, R. (1965). A method of linear causal analysis: Dependence analysis. *American Sociological Review*, 30: 365-373.
- Browne, M. W. (1984). Asymptotic distribution free methods in analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37: 62-83.
- Campbell, D. y J. Stanley (1973). *Diseños Experimentales y Cuasiexperimentales en la Investigación Social*. Buenos Aires: Amorrortu.
- Dillon, W. R. & M. Goldstein (1984). *Multivariate Analysis: Methods and Applications*. New York: John Wiley & sons.
- Duncan, O. D. (1966) Path analysis: Sociological examples. *American Journal of Sociology*, 72: 1-12.
- Fischer, F. M. (1966). *The identification problem in Econometrics*. New York: McGraw-Hill.
- Fox, J. (1984). *Linear Statistical Models and Related Methods*, New York: Wiley.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., Tatham, R. L. (2006). *Multivariate Data Analysis* (6th Edition). Upper Saddle River, NJ: Pearson Prentice-Hall
- Johnson, R. A. & D. W. Wichern (1982). *Applied Multivariate Statistical Analysis*. Englewood Cliffs: Prentice-Hall.
- Johnston, J. (1972). *Econometric Methods*. New York: McGraw-Hill.
- Johnston, J. (1984). *Econometric Methods*. New York: McGraw-Hill.
- Jöreskog, K. G. (1973) A general method for estimating a linear structural equation system, pp. 85-112 in A. S. Goldberger and O. D. Duncan (eds.) *Structural Equation Models in the Social Sciences*. New York: Seminar.
- Jöreskog, K. G. (1977). Structural equation models in the social sciences: especification, estimation and testing. In P. R. Krishnaiah (Ed.), *Applications of Statistics*. Amsterdam: North-Holland, pp. 265-287.
- Jöreskog, K. G. & D. Sörbom (1986). *LISREL VI: Analysis of Linear Structural Relationships by Maximum Likelihood and Least Squares Methods*. Mooresville, IN: Scientific Software, Inc.
- Jöreskog, K. G. & D. Sörbom / SPSS (1988). *LISREL® 7: A guide to the Program and Applications*. 2<sup>nd</sup> ed. Chicago, IL: SPSS.
- Keesling, W. (1972). *Maximum likelihood approaches to causal flow analysis*. Ph D. dissertation University of Chicago.
- Long, J. S. (1983a). *Confirmatory Factor Analysis: A Preface to LISREL*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 007-033. Newbury Park, CA: Sage.

- Long, J. S. (1983b) *Covariance Structure Models: An introduction to LISREL*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-034. Beverly Hills and London: Sage Pubns.
- Long, J. S. (1990). *Covariance Structure Models: An Introduction to LISREL*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 007-034. Newbury Park, CA: Sage.
- MacLeod C. M., R. A. Jackson & J. Palmer (1986). On the Relation between Spatial Ability and Field Dependence. *Intelligence*, 10, 141-151.
- Saris, W. E. & L. H. Stronkhorst (1984). *Causal Modelling in Non-Experimental Research*. Amsterdam: Sociometric Research.
- SPSS (1990). *SPSS® LISREL® 7 and PRELIS®: User's Guide and Reference*. Chicago, IL: SPSS.
- SPSS (1993). *SPSS® LISREL® 7 and PRELIS®: User's Guide and Reference*. Chicago, IL: SPSS.
- Warren, R.D., J.K. White & W.A. Fuller (1974). An errors-in-variables analysis of managerial role performance. *Journal of the American Statistical Association*, 69, 886-893.
- Wiley, D. E. (1973) The identification problem for structural equation models with unmeasured variables, pp. 69-83 in A. S. Goldberger and O. D. Duncan (eds.) *Structural Equation Models in the Social Sciences*. New York: Seminar.

# DESARROLLO DE UN EJEMPLO DE ANÁLISIS FACTORIAL CONFIRMATORIO CON LISREL, AMOS Y SAS

BENITO ARIAS MARTÍNEZ  
Departamento de Psicología  
Facultad de Educación (Universidad de Valladolid)

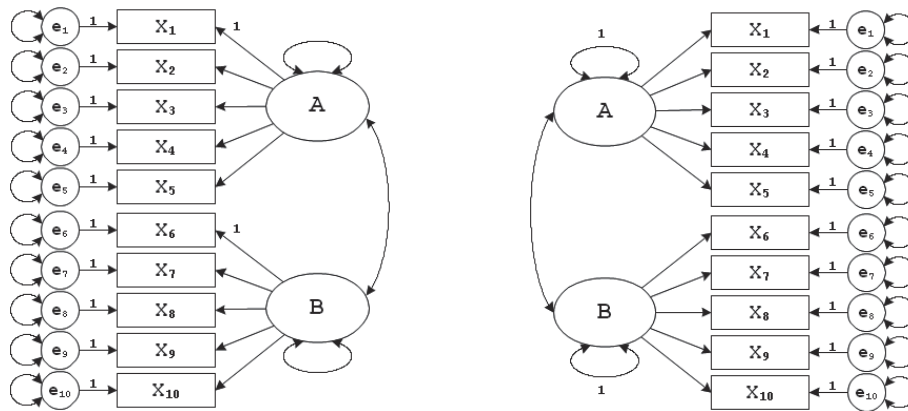
## INTRODUCCIÓN

El presente trabajo tiene por objeto revisar las principales características del análisis factorial confirmatorio y desarrollar de forma detallada un ejemplo de análisis con los programas informáticos LISREL v. 8.8 (Scientific Software International, 2006), AMOS v. 6.0 (SPSS, 2006) y SAS v. 9.1.3 (The SAS Institute, 2006).

El análisis factorial confirmatorio (AFC, en adelante) se ha convertido en los últimos años en uno de los procedimientos de análisis más utilizados en investigación en ciencias sociales. El AFC es un procedimiento de análisis encuadrado en los modelos de ecuaciones estructurales (*SEM, Structural Equation Models*), cuyo propósito se centra en el estudio de los *modelos de medida*, esto es, en analizar las relaciones entre un conjunto de *indicadores* o variables observadas y una o más *variables latentes* o factores. Los indicadores pueden ser, por ejemplo, los ítems de un test, las puntuaciones obtenidas por los sujetos en distintas escalas o los resultados provenientes de instrumentos de clasificación conductual. A diferencia de lo que sucede en el análisis factorial exploratorio (uno de cuyos objetivos tiene que ver con la determinación del número de factores que subyacen a los datos, y se permite que todos los indicadores saturan en todos los factores, y que todos los factores estén correlacionados), una característica esencial del AFC es que el investigador debe concretar *de antemano* todos los aspectos relevantes del modelo, aspectos que deben estar sólidamente fundamentados en la teoría previa y en la evidencia conocida. Así, deben especificarse con anterioridad al análisis, qué factores y qué indicadores forman el modelo, qué indicadores presentan saturaciones en cada factor, si existe o no relación entre los factores, y así sucesivamente. El AFC es, en consecuencia, una estrategia sumamente útil en el ámbito de la prueba de hipótesis y la confirmación de teorías<sup>1</sup>.

1 En algunas ocasiones, sin embargo, el AFC no tiene un carácter tan estrictamente 'confirmatorio'. Muchas autoridades en la materia (vid., por ejemplo, Arbuckle, 2000; Cribbie, 2007; Hancock, 1999; Loehlin, 2004) han abundado en la concepción de los modelos SEM como meramente exploratorios en aquellos casos en que no existen modelos previos claramente establecidos, ni la teoría es suficientemente sólida.

Figura 1. Modelo AFC con factores no estandarizados (izquierda) y estandarizados (derecha)



En la Figura 1 se presenta un ejemplo simple de AFC. Los indicadores están representados por la letra  $X$ , los errores de medida por  $e$ , y los factores por  $A$  y  $B$ . La hipótesis puesta a prueba por el modelo es la siguiente:  $X_1, X_2, X_3, X_4$  y  $X_5$  miden el factor  $A$ ;  $X_6, X_7, X_8, X_9$  y  $X_{10}$  miden el factor  $B$ ; ambos factores covarían; los errores de medida son independientes.

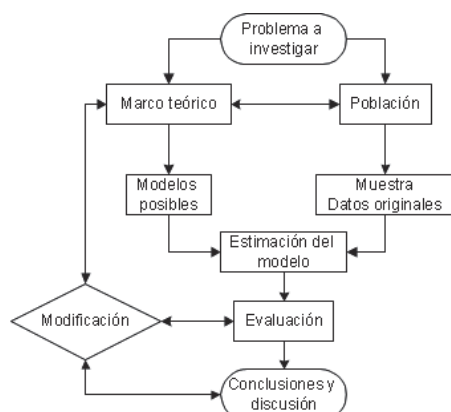
Cada indicador ( $X_1$ - $X_{10}$ ) es una variable continua (excepcionalmente ordinal) que tiene dos causas: por una parte, un factor único subyacente que el indicador se supone que mide y, por otra, cualesquiera otras fuentes únicas de causación que están representadas por el término de error ( $e$ ). Los errores de medida son independientes entre sí y con respecto a los factores y todas las asociaciones entre los factores (e.g., correlación entre  $A$  y  $B$ ) figuran en el modelo como no analizadas.

Las flechas desde los factores a los indicadores indican los efectos causales (i.e., saturaciones factoriales) de los factores sobre las variables observadas y, en general, se interpretan como coeficientes de regresión (no estandarizados en la primera figura, estandarizados en la segunda). Los indicadores que se suponen causados por los factores  $A$  y  $B$  son variables 'endógenas' toda vez que son influenciadas (directa o indirectamente) por las variables incluidas en el modelo, en tanto que los factores son variables 'exógenas' (i.e., son siempre variables independientes y por tanto no pueden estar influidas por ninguna otra variable del modelo). Las flechas que vinculan los errores de medida ( $e_1$  a  $e_{10}$ ) con los respectivos indicadores representan el efecto combinado de cualquier fuente de influencia (distinta a los factores) sobre las variables manifiestas. Puesto que sus causas nos son desconocidas, los errores de medida se consideran en el AFC como variables exógenas, y reflejan dos clases de varianza única: por una parte, el error aleatorio y, por otra, la varianza sistemática debida a otros elementos que el indicador mide además del factor subyacente, tales como los efectos de un particular método de medida.

## CÓMO LLEVAR A CABO UN AFC: PREPARACIÓN DE LOS DATOS Y CONDICIONES PREVIAS

En la Figura 2 (modificada a partir de Boomsna, 2000) se sintetizan los pasos básicos a seguir para llevar a cabo un AFC.

Figura 2. Diagrama de flujo con los pasos a seguir en un AFC



Partiendo del marco teórico en que se inscribe el estudio y de las características de la población a la que se desean generalizar los resultados, el investigador recoge los datos de la muestra y selecciona uno de los varios modelos posibles. Realiza la estimación del modelo y evalúa su ajuste a los datos originales. Decide, por último, si es preciso realizar alguna modificación del modelo, sin perder de vista la teoría de partida. Para llevar a término sin contratiempos un AFC resulta esencial preparar correctamente los datos a analizar, por lo que dedicaremos los párrafos que siguen a revisar algunas cuestiones básicas relativas a esta cuestión.

La preparación de los datos para su análisis con AFC resulta crucial por dos razones: en primer lugar, la mayor parte de métodos de estimación en SEM imponen ciertas asunciones sobre la distribución de los datos; en segundo lugar, ciertos problemas relacionados con los datos pueden hacer que el AFC fracase en encontrar una solución plausible. Para realizar un AFC, en general no son necesarios los datos originales: la mayor parte de los programas aceptan tanto datos en bruto como matrices de correlaciones o de covarianzas. Deberíamos seleccionar como datos de entrada las puntuaciones originales cuando su distribución no es normal y se pretenden analizar con un método que requiere normalidad, o cuando se va a utilizar un método de estimación que no asume distribución normal o se acomoda a casos con observaciones perdidas. En el resto de las ocasiones pueden usarse tanto datos originales como matrices considerando, en todo caso, que las matrices son mucho más 'económicas' en cuanto a memoria y recursos informáticos utilizados y, en consecuencia, su análisis mucho más rápido y eficiente. Por otra parte, muchos artículos que usan metodología SEM ofrecen suficiente información (e.g., matrices de correlación con medias y DT, matrices de covarianza), de suerte que el lector puede hacer análisis secundarios sobre esos datos o replicarlos sin tener que acudir a los datos originales.

Tabla 1. Condiciones necesarias para realizar un AFC

<i>Condición</i>	<i>Observaciones</i>
1. Nivel de medida	Indicadores en nivel de intervalo o de razón (excepcionalmente, ordinal).
2. Valores por indicador	Los indicadores deberían tener un mínimo de 4 valores.
3. Normalidad y <i>outliers</i>	Distribución normal de los datos, control de outliers.
4. Homocedasticidad	Corrección mediante normalización o transformaciones.
5. Datos perdidos	Tratamiento adecuado de los datos perdidos.
6. Tipo de relaciones	Relaciones lineales y aditivas.
7. Multicolinealidad	Ausencia de multicolinealidad.
8. Variables relevantes	Inclusión dentro del modelo de todas las variables relevantes.
9. Identificación del modelo	Modelo supraidentificado.
10. Número mínimo de observaciones	Al menos 150 observaciones, o 5 observaciones por cada parámetro a estimar.
11. Indicadores por variable latente	Preferible disponer de más de 2 (lo ideal es disponer de al menos 4 ó 5).
12. Número de indicadores	El número máximo de indicadores no debería exceder de 20-30.
13. Varianzas relativas	Ausencia de matrices <i>ill-scaled</i>

Conviene, una vez preparados los datos para su análisis, realizar una serie de comprobaciones para evitar problemas potenciales. En la Tabla 1 se sintetizan un conjunto de requisitos (Arbuckle, 2000; Diamantopoulos y Sigauw, 2000; Hatcher, 2006; Jöreskog, 1993; Jöreskog y Sörbom, 1996a, 1996b) que deben cumplirse para poder llevar a cabo un AFC relativos al nivel de medida, número mínimo de valores por indicador, distribución, tipo de relaciones, multicolinealidad, identificación del modelo, número mínimo de observaciones, número de indicadores por cada variable latente, etc.

Como se pone de manifiesto en las condiciones expresadas en la Tabla 1, el AFC es una técnica de análisis ciertamente exigente. Haremos en los párrafos que siguen algunos comentarios adicionales a lo expuesto en dicha tabla:

1. La exigencia de *nivel de medida* intervalar o continuo de los indicadores podría atenuarse, de modo que es posible usar indicadores medidos en escala ordinal (como sucede, por ejemplo, con las escalas probabilísticas sumatorias tipo Likert), si bien en este caso debería optarse por el uso de métodos de estimación apropiados (en ningún caso ML).
2. Se recomienda que el *número mínimo de valores* de cada indicador sea 4, con el fin de aumentar la probabilidad de que la distribución de los datos se acerque a la normalidad.
3. La distribución de los datos debería cumplir con el requisito de *normalidad multivariada*, toda vez que la mayoría de los métodos de estimación SEM exigen el cumplimiento de esta condición. Ello implica que a) todas las distribuciones univariadas son normales; b) la distribución conjunta de cualquier par de variables es normal bivariada y c) todos los dispersigramas (*scatterplots*) son lineales y presentan homocedasticidad.

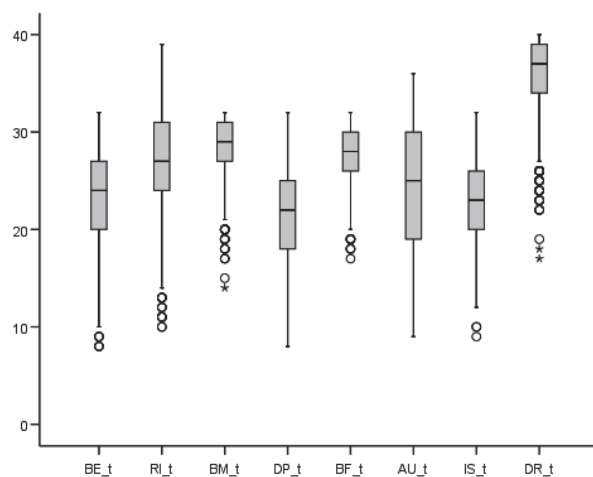


ticidad. Es preciso, en este punto, detectar los *outliers* o valores fuera de límites. En cuanto a la asimetría, valores por encima de  $|3.00|$  indican asimetría extrema; en el caso de la curtosis, valores entre  $|8.00|$  y  $|20.00|$  denotarían curtosis extrema; valores por encima de  $|20.00|$  indicarían un serio problema de normalidad.

Una forma de solucionar la falta de normalidad consiste en transformar los datos (i.e., las puntuaciones originales se convierten mediante una operación matemática en nuevos valores que pueden tener una distribución más próxima a la normal). Las transformaciones son útiles para el control de *outliers*. Entre las transformaciones posibles para tratar de remediar la asimetría figuran las siguientes (vid. Hair, Black, Babin, Anderson, & Tatham, 2006; Tabachnick & Fidell, 2001):  $X^{1/2}$ , función logarítmica ( $\log_{10} X$ ), o función inversa ( $1/X$ ). Para solventar problemas de curtosis, pueden utilizarse funciones como  $X^{1/3}$ , función seno, inversa (e.g.,  $1/X$ ,  $1/Y$ ), o transformaciones polinomiales ( $X^2$ ,  $X^3$ ). En términos generales, se recomienda probar todas las transformaciones mencionadas hasta llegar a aquella que mejor funcione con una distribución particular. Otra alternativa es utilizar puntuaciones normalizadas (*normal scores* calculadas directamente en el programa PRELIS). Desde una perspectiva práctica, podemos juzgar el impacto potencial de una transformación calculando la razón entre la media y la desviación típica de la variable (los efectos de la transformación serán previsiblemente mayores cuando dicha razón sea inferior a 4.00).

El caso de los *outliers* merece algún comentario adicional. Se trata de casos muy diferentes al resto (con puntuación extrema en una variable), alejados en torno a 3 DT o más de la media. La forma más sencilla de detectar *outliers* univariados es transformar las puntuaciones originales a puntuaciones típicas ( $z$ ) y examinar los valores superiores a  $|3.00|$  (pueden detectarse también con facilidad construyendo un gráfico de caja *–box-and-whisker–* para cada variable). En la Figura 3 puede verse un ejemplo de *outliers* univariados.

Figura 3. Ejemplo de outliers univariados (aparecen con los signos  $\circ$  y  $\star$  en el gráfico)

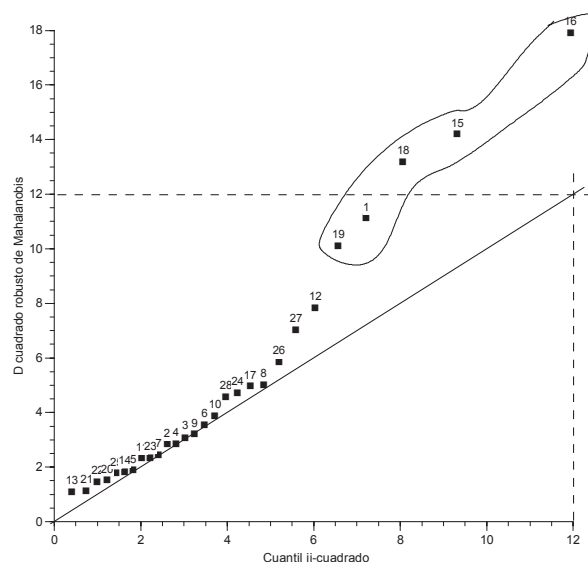


Los *outliers* multivariados son casos con puntuaciones extremas en dos o más variables (i.e., alejados de la media entre 2 y 3 DT). La detección de *outliers* multivariados es más complicada que la detección de univariados. En general, se utiliza la distancia de Mahalanobis al cuadrado ( $D^2$ ), que indica la distancia en unidades de DT entre un conjunto de puntuaciones (vector) para un caso individual y las medias de la muestra para todas las variables (centroides). Con muestras grandes,  $D^2$  se distribuye aproximadamente como  $\chi^2$  con tantos grados de libertad como número de variables. Un valor de  $D^2$  con un  $p$  bajo puede



conducir al rechazo de la  $H_0$  de que el caso proviene de la misma población que el resto. Una vez realizados los análisis (e.g., en SAS) se grafican los resultados, como se muestra en el ejemplo de la Figura 4.

Figura 4. Ejemplo de detección de outliers multivariados (los números corresponden al número de caso)



4. La *falta de homocedasticidad* es asimismo un aspecto relacionado con la normalidad multivariada, y una forma sencilla y útil de evaluarla es mediante el examen de dispersigramas bivariados. La heterocedasticidad puede ser causada por una falta de normalidad en X o Y, de modo que existe más error aleatorio en unos niveles de X o Y que en otros (por ejemplo, una prueba de habilidad cognitiva podría ser fiable en sujetos con nivel alto y no fiable en sujetos con nivel bajo). La forma más frecuente de corregir la heterocedasticidad es llevar a cabo una transformación de las puntuaciones (e.g. normalización en PRELIS). Debe advertirse, sin embargo, que esta estrategia puede ser útil para corregir la heterocedasticidad debida a la no normalidad, pero no resultar tan útil para corregir la heterocedasticidad debida a fiabilidades diferentes en X e Y.
5. Un problema omnipresente en la investigación tiene que ver con los *casos perdidos* (*missing data*). Es incuestionable que todas las investigaciones han de enfrentarse al problema de los datos perdidos debidos a fallos del equipo o la instrumentación, mortalidad (e.g., en investigaciones longitudinales), negativa de los sujetos a contestar o dificultad para hacerlo, tendencias extremas sistemáticas de las respuestas que convierten en inútiles los datos recopilados, o cualquier otra causa. Lo sorprendente es que casi ningún estudio informa adecuadamente sobre cómo ha tratado los datos perdidos (pese a ser, insistimos, un problema crucial en la investigación). Se han desarrollado diferentes métodos para el tratamiento de los datos perdidos, así como programas informáticos que implementan distintos procedimientos de imputación. La decisión a tomar (e.g., desechar los casos perdidos, utilizar algún método de imputación) depende de factores tales como el tamaño de la muestra, el número de casos perdidos y si son o no sistemáticos<sup>2</sup>.

<sup>2</sup> Para mayor información sobre el tratamiento de datos perdidos, puede consultarse McKnight, McKnight, Sidaniy, & Figueredo (2007).

6. El modelo supone *relaciones lineales y aditivas* entre las variables (debe advertirse, en este punto, que es posible someter a prueba modelos no lineales, pero este asunto sobrepasa el objetivo del presente trabajo)<sup>3</sup>.
7. Los datos deben estar libres de *multicolinealidad*. La multicolinealidad tiene lugar cuando variables diferentes miden de hecho el mismo constructo, lo que implica una correlación muy elevada entre ellas (del orden de .90 o superior). Esto supone que ciertas operaciones matemáticas se tornan imposibles o muy inestables (e.g., ocasionan matrices de covarianza singulares) debido a que algunos denominadores se aproximan a cero<sup>4</sup>. Existen dos modos principales de detectar la multicolinealidad: a) examinar la matriz de correlaciones y detectar variables con  $r_{xy} > .90$ ; b) calcular la correlación múltiple al cuadrado  $R^2_{smc}$  entre cada variable y todas las demás: valores superiores a .90 sugieren multicolinealidad. Es asimismo conveniente considerar, en este sentido, dos conceptos relacionados con la multicolinealidad: la *tolerancia* y el *VIF* o *factor de inflación de la varianza*. La *tolerancia* se calcula mediante la fórmula  $t = 1 - R^2_{smc}$ , e indica la proporción de la varianza estandarizada total que es única (i.e., no explicada por todas las demás variables). Valores de  $t < .10$  indican colinealidad. Por su parte, el *VIF* o *factor de inflación de la varianza* se obtiene mediante la fórmula  $VIF = 1/(1 - R^2_{smc})$ , es decir, la razón entre la varianza total estandarizada y la varianza única: si  $VIF > 10$ , la variable puede ser redundante. Debe recordarse, finalmente, que la multicolinealidad puede también suceder entre variables latentes (cuando la correlación entre ellas es tan alta que bien podría pensarse que no son variables diferentes).
8. El modelo debería incluir aquellas variables que resultaran *relevantes*, omitiendo el resto. En otras palabras, el modelo debería ser lo más parsimonioso posible.
9. El modelo debería ser *supraidentificado* (i.e., con grados de libertad positivos, lo que equivale a que el número de datos u observaciones sea mayor que el de parámetros a estimar).
10. El número mínimo de observaciones disponibles debería ser en torno a 150 o, al menos, 5 observaciones por cada parámetro a estimar. Se prefieren muestras grandes<sup>5</sup>, especialmente si se han de realizarse modificaciones en el modelo.
11. Se requiere disponer al menos de 3 indicadores por cada variable latente. Si bien es cierto que, desde el punto de vista técnico, bastarían 2 indicadores por factor, deberían usarse al menos 3 para evitar problemas de identificación y convergencia.

3 Puede consultarse información sobre efectos de interacción y no lineales en SEM en Schumaker y Marcoulides (1998).

4 Carece de sentido mantener en un análisis variables que presenten multicolinealidad. Si, por ejemplo, se detectara que X e Y son redundantes, se debería optar por a) eliminar una de ellas o b) combinarlas en una nueva variable compuesta (e.g., el total reemplazaría a X e Y). Un caso –nada infrecuente– ocurre cuando se incluye el total como variable. Imaginemos que tenemos un cuestionario de 10 ítems: si incluimos el total como variable en el análisis, aunque las correlaciones bivariadas entre cada ítem y el total no sean muy altas, la correlación múltiple entre el total y los 10 ítems es 1.00 (¡con lo que hemos alcanzado el caso máximo de multicolinealidad!)

5 No está demasiado claro, en la metodología SEM, qué significa ‘muestra grande’. *Grosso modo*, se considera (Kline, 2005, p. 15) que la muestra es ‘pequeña’ si tiene menos de 100 sujetos, ‘media’ si tiene de 100 a 200 sujetos, y ‘grande’ si la integran más de 200 sujetos. En el estudio de Breckler (1990) –citado en Kline, 2005, p. 15– sobre 72 artículos que utilizaron metodología SEM, la media fue de 198 sujetos, con un rango de 40 a 8650 casos. El 22% de tales estudios utilizaron menos de 100 sujetos. Con todo, debe tenerse en cuenta que el número de sujetos no es una cuestión que pueda decidirse en términos absolutos, dado que depende de la complejidad (i.e., parámetros a estimar) del modelo. Por otra parte, en SEMMET (<http://bama.ua.edu>) el asunto ha sido objeto de una amplísima discusión, y pueden consultarse cientos de entradas al respecto, con comentarios de autoridades tan señaladas en este campo como K. Bollen, L. Hayduk, S. Mulaik, E. Rigdon, P. Barrett, etc.

Una posición más conservadora recomienda utilizar al menos 4, para prevenir la posibilidad de tener que descartar alguno del análisis.

12. Un modelo eficaz no debería superar los 30 indicadores (Bentler & Chu, 1987; Lomax, 1982). Conjuntos muy grandes de datos con frecuencia resultan en valores enormes de  $\chi^2$ , de modo que el ajuste global del modelo a los datos se torna imposible. En estos casos, puede optarse por la formación de parcels, asunto que trataremos más adelante. Un modelo AFC ideal, en consecuencia con lo afirmado, debería incluir como máximo 30 variables observadas y 6 variables latentes.
13. Por último, conviene examinar las varianzas relativas. Las matrices de covarianza con una ratio entre el valor máximo y el valor mínimo superior a 10.00 son ill-scaled y pueden ocasionar problemas en los análisis. En el ejemplo que analizaremos más adelante (vid. Tabla 5), la razón equivale a  $2061.875/491.042=4.199$ , por lo que no tenemos el problema mencionado.

## IDENTIFICACIÓN DE MODELOS AFC

Dos son las condiciones básicas para identificar un modelo AFC. En primer lugar, el número de parámetros a estimar ha de ser igual o inferior al número de observaciones (dicho de otro modo, el número de grados de libertad del modelo ha de ser igual o superior a 0); en segundo lugar todas las variables latentes (i.e., factores y errores de medida) deberían tener una escala, lo que implica por lo general asignar el valor de 1.00 (ULI, *Unit Loading Identification*). Así, en la Figura 1, la especificación  $e_1 \rightarrow X_1=1.0$  asigna a  $e_1$  una métrica que corresponde a la de la varianza única de  $X_1$ . La misma estrategia se utiliza para escalar los factores, de modo que  $A_1 \rightarrow X_1=1.0$  y  $B_6 \rightarrow X_6=1.0^6$ . Para calcular el número de observaciones se sigue la fórmula

$$n = v(v + 1) / 2$$

donde  $v$  es el número de variables observadas. En el ejemplo de la Figura 1, el número de observaciones (datos de entrada) equivale, pues, a  $10(10+1)/2$ , esto es, 55 observaciones que corresponden a las 10 varianzas más las 45 covarianzas entre las variables observadas. Tal como se ilustra en la Tabla 2, con 5 variables observadas tendríamos 15 puntos de datos (5 varianzas y 10 covarianzas).

Tabla 2. Ejemplo de matriz de observaciones (varianzas en fondo gris)

	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>
X <sub>1</sub>	σ <sub>11</sub>				
X <sub>2</sub>	σ <sub>21</sub>	σ <sub>22</sub>			
X <sub>3</sub>	σ <sub>31</sub>	σ <sub>32</sub>	σ <sub>33</sub>		
X <sub>4</sub>	σ <sub>41</sub>	σ <sub>42</sub>	σ <sub>43</sub>	σ <sub>44</sub>	
X <sub>5</sub>	σ <sub>51</sub>	σ <sub>52</sub>	σ <sub>53</sub>	σ <sub>54</sub>	σ <sub>55</sub>

El número de parámetros a estimar, por su parte, corresponde al número de varianzas y covarianzas de las variables exógenas (factores y errores de medida), más los efectos directos de los factores sobre los indicadores. En el ejemplo de la Figura 1, tenemos, por consiguiente,

6 Una descripción detallada de las restricciones ULI (*Unit Loading Identification*) y UVI (*Unit Variance Identification*) puede consultarse en Kline, 2005 (pp. 170-171).

21 parámetros a estimar por el modelo (1 covarianza entre los factores, 2 varianzas de los factores, 10 varianzas de los errores de medida y 8 efectos directos de los factores sobre los indicadores). Así, disponemos de un modelo supraidentificado con 34 grados de libertad (55 puntos de datos menos 21 parámetros a estimar). De acuerdo con la notación LISREL, tendríamos los parámetros a estimar que se reflejan en la Tabla 3.

Tabla 3. Notación (en LISREL) de un modelo de los variables latentes y 10 indicadores

Nombre	Parámetro	Matriz	Tipo	Descripción	N
Lambda-X	$\lambda_x$	$\Lambda_x$	Regresión	Saturaciones factoriales	8
Theta-Delta	$\delta$	$\Theta_\delta$	Varianza-covarianza	Varianzas y covarianzas de error	10
Phi	$\phi$	$\Phi$	Varianza-covarianza	Varianzas y covarianzas de los factores	3
<b>Total parámetros a estimar</b>					<b>21</b>

## ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO AFC

El objetivo del AFC es obtener estimaciones de cada uno de los parámetros del modelo de medida (saturaciones factoriales, varianzas y covarianzas de los factores, varianzas y, en su caso, covarianzas de los errores de medida) que configuran una matriz ( $\hat{\Sigma}$ ) que reproduzca lo más fielmente posible la matriz ( $S$ ) observada. El método más común de estimación es el de máxima verosimilitud (*ML, Maximum Likelihood*), siempre que se cumpla la serie de supuestos a que más arriba hemos hecho referencia, tales como disponer de una muestra de tamaño suficiente, medición de los indicadores al menos en nivel intervalar, o distribución normal multivariada de los indicadores de las variables latentes<sup>7</sup>. En el caso de que no se cumpla el supuesto de normalidad multivariada, conviene usar el método de máxima verosimilitud robusta (*MLM, Maximum Likelihood Mean Adjusted*; Bentler, 1995). Si uno o más de los indicadores es categórico (o si la ausencia de normalidad es extrema), habrá de optarse por otros métodos de estimación tales como mínimos cuadrados ponderados (*WLS, Weighted Least Squares*), mínimos cuadrados ponderados diagonalizados (*DWLS, Diagonal Weighted Least Squares*), mínimos cuadrados ponderados robustos (*WLSMV, Weighted Least Squares Mean and Variance Adjusted*) o mínimos cuadrados no ponderados (*ULS, Unweighted Least Squares*)<sup>8</sup>. En estos casos, el uso del método ML puede producir estimaciones atenuadas de las relaciones entre los indicadores (en especial cuando existen efectos 'suelo' o 'techo'), conducir a la formación de pseudofactores (artefactos producidos por la dificultad o el carácter extremo de los ítems), producir errores estándar incorrectos o estimaciones erróneas de los parámetros, llegar a resultados imprecisos si parte de una matriz de correlaciones en lugar de covarianzas (Brown, 2006).

7 Lamentablemente, los paquetes estadísticos de propósito general (e.g., SPSS, SYSTAT, STATISTICA, etc.) no permiten la evaluación directa (vía menú o sintaxis) de la normalidad multivariada de los datos. Sí se puede llevar a cabo con facilidad mediante el programa PRELIS, o utilizando macros (e.g., en SPSS o SAS). En SAS puede utilizarse la macro 'multnorm.sas' (<http://support.sas.com/kb/24/983.html>). En SPSS, la macro 'normtest.sps' (<http://www.columbia.edu/~ld208/normtest.sps>).

8 Una referencia bastante completa de los métodos de estimación puede consultarse en Brown (2006), pp. 387 y ss.

## AJUSTE DEL MODELO AFC

El índice de ajuste por excelencia en los modelos AFC es  $\chi^2$ . Se acepta que  $S = \hat{\Sigma}$  en el caso de que  $\chi^2$  sea suficientemente pequeño (i.e., el nivel de significación asociado  $p$  sea superior a .05). Sin embargo, este índice raramente es utilizado como prueba única o concluyente de bondad del ajuste del modelo<sup>9</sup> (e.g., en muchas ocasiones la distribución de los datos no se ajusta a la distribución  $\chi^2$ ; su valor está influenciado por el tamaño de la muestra; se basa en la hipótesis –excesivamente restrictiva– de que  $S$  es igual a  $\Sigma$ ). Para superar estos inconvenientes, se han desarrollado multitud de índices parciales de ajuste, tanto de carácter absoluto (e.g., SRMR, GFI, AGFI, PGFI)<sup>10</sup>, como parsimonioso (e.g., RMSEA), predictivo (e.g., ECVI, CAIC, BIC) o incremental (e.g., CFI, TLI, NFI, PNFI, RNI, PCFI)<sup>11</sup>. En la Tabla 4 se muestran las principales características de un modelo de medida para considerar que se ha alcanzado un buen ajuste (volveremos sobre este asunto más adelante, al comentar la salida de los análisis LISREL).

Tabla 4. Características del ‘ajuste ideal’ en un modelo de medida

<i>Característica</i>	<i>Observaciones</i>
Nivel de significación de $\chi^2$	El valor de $p$ ( $\chi^2$ ) debería ser superior a .05
Razón $\chi^2$ /gl	Debería ser inferior a 2.00
CFI y NNFI	Deberían ser superiores a .95; mejor cuanto más próximos a 1.00
Valores de t	Los valores absolutos deberían ser superiores a 1.96
Saturaciones	Superiores a .30
Residuos	Distribución normal, simétrica en torno a 0, pocos residuos superiores a $ 2.00 $
Fiabilidad compuesta	Las fiabilidades compuestas de las VL deberían ser superiores a .60 (preferiblemente, superiores a .70)
Varianza media extractada	Las VME de las VL deberían ser superiores a .50
Validez discriminante	Se debería demostrar la VD entre pares de factores a través de la prueba de diferencias de $\chi^2$ , los intervalos de confianza y la varianza extractada
RMSEA	Inferior a .08 (preferiblemente, inferior a .06); el modelo debería rechazarse si $RMSEA > .10$
SRMR	Inferior a .08, mejor mientras más próximo a .00

## DESARROLLO DE UN EJEMPLO DE AFC

El ejemplo que se expone a continuación es una réplica del siguiente artículo: Williams, T. O., Eaves, R. C., & Cox, C. (2002). Confirmatory factor analysis of an instrument designed

9 Existe entre los expertos una notable controversia en torno a este requerimiento. El lector interesado puede consultar la lista SEMNET (<http://bama.au.edu>).

10 SRMR: *Standardized Root Mean Square Residual*; GFI: *Goodness of Fit Index*; AGFI: *Adjusted Goodness of Fit Index*; PGFI: *Parsimony GFI*; RMSEA: *Root Mean Square Error of Approximation*; ECVI: *Expected cross-validation Index*; CAIC: *Consistent Akaike Information Criterion*; BIC: *Bayes Information Criterion*; CFI: *Comparative Fit Index*; TLI: *Tucker-Lewis Index*; NFI: *Normed Fit Index*; PNFI: *Parsimony NFI*; RNI: *Relative Noncentrality Index*; PCFI: *Parsimony Comparative Fit Index*.

11 Una revisión relativamente exhaustiva de los índices de ajuste puede consultarse en Hu & Bentler (1995) o en Jackson (2007).

to measure affective and cognitive arousal. *Educational and Psychological Measurement*, 62(2), 264-283.

El VST II (Williams, Eaves, & Cox, 2002, p. 270) es un instrumento para evaluar los niveles de arousal en sujetos de 5 a 21 años. Tiene dos formas: Afectiva (trata de medir el arousal límbico con 30 ítems) y Cognitiva (pretende medir las funciones de arousal corticales y consta de 25 ítems). La prueba se administró a 216 niños de 10 a 12 años.

Los análisis se realizaron no sobre las puntuaciones directas en los ítems del test, sino sobre conjuntos homogéneos de ítems o *parcels*<sup>12</sup>. Un *parcel* es una puntuación total (i.e., puntuación compuesta lineal) de un conjunto de ítems homogéneos, y se trata como una variable continua. Una condición imprescindible de los *parcels* es que sean unidimensionales, lo que vale tanto como demostrar que miden un solo constructo. Existen diferentes procedimientos para la formación de *parcels* (e.g., asignación aleatoria de los ítems a cada *parcel*, agrupamiento de los ítems en función de su contenido, o agrupamiento atendiendo a la magnitud de las correlaciones entre los ítems, entre otros). Los *parcels* se usan, por lo general, con tres propósitos principales: en primer lugar, permiten reducir un elevado número de ítems a un conjunto más reducido y manejable; en segundo lugar, tienen más probabilidades que los ítems individuales de presentar una normalidad univariada (y, por tanto, multivariada), con lo que hacen posible la estimación de parámetros mediante máxima verosimilitud; en tercer lugar, los modelos basados en *parcels* son considerablemente menos complejos que los basados en ítems (la matriz de entrada es más reducida, y son menos los parámetros a estimar).

El uso de *parcels* entraña, también, algunos inconvenientes. El más serio ocurre cuando la estructura subyacente de los ítems que conforman un *parcel* no es unidimensional: en tales casos (i.e., uno o más *parcels* son multidimensionales) el uso de esta estrategia contribuirá más a la confusión que al esclarecimiento de la estructura de los datos. Lo mismo ocurre cuando las unicidades (varianzas de los errores de medida) de los ítems de un determinado *parcel* correlacionan con las de otros ítems pertenecientes a otros *parcels*. Por último, en algunas ocasiones puede ser inviable su uso: tal sería el caso de una escala formada por ítems dicotómicos; si consideramos que se necesitan no menos de 10 de estos ítems para acercarse a la distribución normal, si la escala tuviera, pongamos por caso, 40 ítems, el resultado sería una estructura de únicamente 4 *parcels*, lo que conduciría probablemente a una prueba incompleta de la estructura latente de los datos (West, Finch, & Curran, 1995). Con todo, y puesto que hoy disponemos de métodos de estimación para datos no normales con exigencias menos restrictivas respecto al tamaño de la muestra (como sucede, por ejemplo, con el método WLSMV en Mplus), el uso de *parcels* debería quedar restringido a los casos en que se parte de un número de ítems muy grande (digamos, 60 o más) y con falta de normalidad multivariada (lo que exigiría, de utilizar los ítems individuales, trabajar con matrices de covarianzas asintóticas de un tamaño difícilmente manejable por los medios informáticos comunes). Puede consultarse más información sobre este problema en Bandalos (2002), Bandalos & Finney (2001), Brown (2006), Kishton, & Widaman (1994) o Little, Cunningham, Shahar, & Widaman (2002).

En el citado artículo de Williams, Eaves, & Cox (2002) los autores utilizaron AFC para determinar si los 10 *parcels* citados medían dos constructos<sup>13</sup> (i.e., variables latentes o fac-

12 La conveniencia de usar 'parcels' en AFC ha suscitado una notable controversia, aún no resuelta. Pueden consultarse al respecto multitud de argumentos a favor y en contra del uso de esta estrategia en SEMNET (<http://bama.au.edu>).

13 El término 'constructo' refleja la hipótesis (con frecuencia incompleta) de que un conjunto de conductas correlacionarán mutuamente en estudios sobre diferencias individuales, a la par que se verán afectadas de forma similar por manipulaciones experimentales, de acuerdo con la concisa definición de Nunnally y Bernstein (1994).



tores) –arousal afectivo y arousal cognitivo–. Se sometieron a prueba cuatro modelos<sup>14</sup>: a) modelo unidimensional; b) modelo bidimensional de 2 factores correlacionados, c) modelo bidimensional de 2 factores no correlacionados y d) modelo jerárquico con un factor de segundo orden. Los análisis parten de la matriz de covarianzas entre los indicadores o variables observadas (incluida por los autores en el texto del artículo). En las Tablas 5 y 6 se reproducen, respectivamente, la matriz de correlaciones (también incluida en el texto) y de covarianzas entre los 10 *parcels* o miniescalas.

Tabla 5. Matriz de correlaciones y desviaciones típicas de los indicadores

	a1	a2	a3	a4	a5	c1	c2	c3	c4	c5
a1	1,000									
a2	0,822	1,000								
a3	0,852	0,839	1,000							
a4	0,814	0,831	0,779	1,000						
a5	0,840	0,846	0,910	0,819	1,000					
c1	0,582	0,637	0,643	0,580	0,680	1,000				
c2	0,634	0,660	0,662	0,623	0,704	0,731	1,000			
c3	0,605	0,607	0,628	0,603	0,673	0,784	0,757	1,000		
c4	0,587	0,575	0,622	0,570	0,674	0,750	0,758	0,703	1,000	
c5	0,674	0,662	0,685	0,659	0,716	0,724	0,782	0,771	0,789	1,000
DT	45,145	43,956	36,247	45,102	45,408	25,110	27,235	32,976	29,245	26,994

Como más arriba hemos señalado, para llevar a cabo el análisis, es indiferente partir de la matriz de correlaciones (con las desviaciones típicas incluidas) o de la de covarianzas (obviamente, en este caso no es preciso incluir las DT). De hecho, el análisis podría realizarse directamente sobre los datos originales. Asimismo, la matriz de datos puede incorporarse al conjunto de comandos del programa o almacenarse en un fichero independiente.

Tabla 6. Matriz de covarianzas de los indicadores

	a1	a2	a3	a4	a5	c1	c2	c3	c4	c5
a1	2038,035									
a2	1631,766	1932,163								
a3	1393,567	1336,871	1313,809							
a4	1657,299	1647,164	1273,261	2034,216						
a5	1721,830	1688,292	1498,401	1677,767	2061,875					
c1	659,795	702,969	585,019	656,527	775,118	630,518				
c2	779,519	790,448	653,734	764,755	871,211	500,128	741,767			
c3	899,912	879,179	750,037	897,441	1008,287	648,935	679,970	1087,409		
c4	775,235	739,157	659,867	751,860	895,616	550,476	603,95	677,865	855,272	
c5	821,170	785,419	669,951	802,825	877,681	491,042	574,775	686,391	622,830	728,674

### A. Modelo unidimensional

Construimos, en primer lugar, un modelo de un solo factor (i.e., todos los indicadores saturan en un factor único, vid. Figura 5). Cuando la teoría no ha especificado claramente el número de constructos, conviene comenzar sometiendo a prueba un modelo unidimensional, toda vez que, si los resultados conducen a la aceptación de dicho modelo, carece de sentido plantear modelos más complejos. Incluso si la teoría se decantara por un número determinado de factores, debería determinarse en qué medida un modelo unidimensional es plausible. Comenzaremos, pues, sometiendo a prueba el modelo unidimensional. La sintaxis

<sup>14</sup> Los autores únicamente sometieron a prueba los modelos aquí denominados como (a), (c) y (d). Incluimos, por su interés, uno más, el modelo (b), de dos factores correlacionados.

en SIMPLIS se muestra en la Tabla 7. Los comandos utilizados son los siguientes: OBSERVED VARIABLES (listado de variables observadas, A1-A5 para los *parcels* afectivos y C1-C5 para los cognitivos); LATENT VARIABLES (en este caso, únicamente tendremos la variable latente AROUSAL); escribimos a continuación los datos de la matriz de covarianzas; especificamos seguidamente el tamaño de la muestra (N=216) y las relaciones entre las variables manifiestas y la variable latente, fijando a 1.0 la saturación no estandarizada de la primera variable en el factor. Puesto que partimos de 10 indicadores, tenemos un total de 55 observaciones para estimar un total de 20 parámetros (1 varianza del factor, 10 varianzas de error y 9 saturaciones factoriales). El modelo es, pues, sobreidentificado con 35 grados de libertad.

Figura 5. Modelo AFC unifactorial

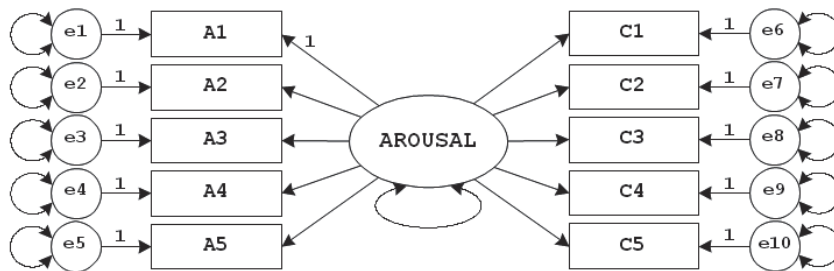


Tabla 7. Sintaxis en SIMPLIS para el modelo unidimensional

```
!Matriz de covarianzas Williams Eaves & Cox, modelo A (Unidimensional)
Observed Variables
A1 A2 A3 A4 A5 C1 C2 C3 C4 C5
Latent Variables
AROUSAL
Covariance Matrix
2038.035
1631.766 1932.163
1393.567 1336.871 1313.809
1657.299 1647.164 1273.261 2034.216
1721.830 1688.292 1498.401 1677.767 2061.875
659.795 702.969 585.019 656.527 775.118 630.518
779.519 790.448 653.734 764.755 871.211 500.128 741.767
899.912 879.179 750.037 897.441 1008.287 648.935 679.970 1087.409
775.235 739.157 659.867 751.860 895.616 550.476 603.950 677.865 855.272
821.170 785.419 669.951 802.825 877.681 491.042 574.775 686.391 622.830 728.674
Sample Size is 216
Relationships
A1 = 1*AROUSAL
A2 = AROUSAL
A3 = AROUSAL
A4 = AROUSAL
A5 = AROUSAL
C1 = AROUSAL
C2 = AROUSAL
C3 = AROUSAL
C4 = AROUSAL
C5 = AROUSAL
OPTIONS wp rs ss sc mi nd=2
Path Diagram
End of Problem
```



En la Tabla 8 se muestran los principales resultados obtenidos en el primer análisis. Estos resultados indican un pobre ajuste del modelo a los datos: el valor de  $\chi^2$  es muy elevado, RMSEA supera el umbral crítico de .10, la magnitud del índice RMR estandarizado es excesiva. En consecuencia, rechazamos la hipótesis de unidimensionalidad. Conviene, como antes se ha dicho, someter a prueba el modelo de un solo factor, toda vez que la imposibilidad de rechazar este modelo implicaría que las variables no demuestran validez discriminante, esto es, que no miden dominios distintos (Kenny, 1979).

Tabla 8. Resumen de resultados del modelo unidimensional

Goodness of Fit Statistics	
Degrees of Freedom = 35	
Minimum Fit Function Chi-Square = 409.68 (P = 0.0)	
Normal Theory Weighted Least Squares Chi-Square = 630.65 (P = 0.0)	
. . .	
Root Mean Square Error of Approximation (RMSEA <sup>15</sup> ) = 0.28	
. . .	
Normed Fit Index (NFI) = 0.92	
Non-Normed Fit Index (NNFI) = 0.90	
. . .	
Comparative Fit Index (CFI) = 0.92	
Incremental Fit Index (IFI) = 0.92	
. . .	
Standardized RMR = 0.083	
Goodness of Fit Index (GFI) = 0.63	

### B. Modelo de dos factores correlacionados

El segundo modelo que sometemos a prueba es un modelo de dos factores correlacionados. Para escalar los factores, fijamos a 1.0 las saturaciones no estandarizadas de A1 y C1 a los factores AFECT y COGNI respectivamente (vid. Figura 6). Al igual que en el primer modelo, tenemos un total de 55 observaciones para estimar un total de 21 parámetros (1 covarianza entre los factores, 2 varianzas de los dos factores, 10 varianzas de error y 8 saturaciones factoriales, como se puede observar en la Tabla 9). El modelo es, pues, sobreidentificado con 34 grados de libertad.

15 Advertirá el lector –si se toma el trabajo de contrastar la solución LISREL con las ofrecidas por otros programas SEM como EQS, AMOS o SAS– que tanto el valor RMSEA como el resto de los índices parciales de ajuste son ligeramente diferentes en LISREL. La razón estriba en que LISREL, con el método de estimación de máxima verosimilitud, no utiliza el ‘Minimum Fit Function Chi-Square’ sino el ‘Normal Theory Weighted Least Squares Chi-Square’ en el cálculo (vid. Tabla 8). Así, el valor RMSEA de la Tabla 8 es el resultado de  $[(630.65-35)/(35*(216-1))]^{1/2} = \mathbf{0.2813}$ , en tanto que el valor RMSEA obtenido en EQS, AMOS o SAS resulta de  $[(409.68-35)/(35*(216-1))]^{1/2} = \mathbf{0.2231}$ . Es este último valor el que ofrecen, precisamente, los autores del artículo que aquí se comenta (sus análisis se realizaron con AMOS).

Figura 6. Modelo AFC de dos factores correlacionados

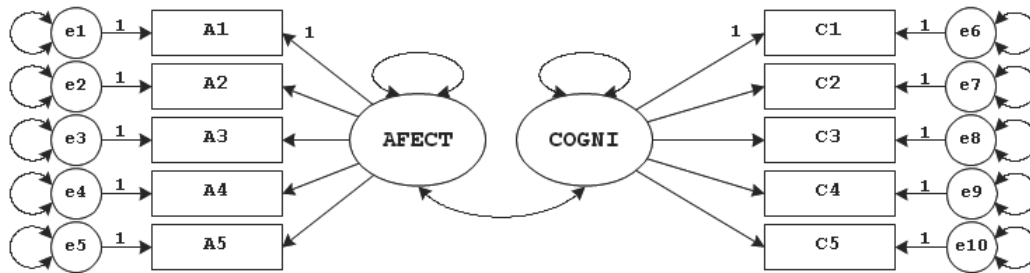


Tabla 9. Especificación de parámetros (modelo de dos factores correlacionados)

Parameter Specifications			PHI				
LAMBDA-X			AFECT		COGNI		
	AFECT	COGNI	-----	-----			
A1	0	0	AFECT	9			
A2	1	0	COGNI	10	11		
A3	2	0	THETA-DELTA				
A4	3	0	A1	A2	A3	A4	A5
A5	4	0	-----	-----	-----	-----	-----
C1	0	0	12	13	14	15	16
C2	0	5					
C3	0	6	C1	C2	C3	C4	C5
C4	0	7	-----	-----	-----	-----	-----
C5	0	8	17	18	19	20	21

La sintaxis en LISREL que puede consultarse en la Tabla 10, tiene las siguientes especificaciones: DA (datos); NI=10 (número de indicadores); NO=216 (número de sujetos); MA=CM (se analiza una matriz de covarianzas). LA indica las etiquetas de las variables observadas (A1-C5); MO (modelo a prueba); NX=10 (número de variables X del modelo); NK=2 (número de variables  $\xi$ )<sup>16</sup>; PH=SY,FR (la matriz  $\Phi$  de varianza-covarianza es simétrica y sus parámetros libres); LX=FU,FR (la matriz  $\Lambda_x$  de saturaciones factoriales es completa y sus parámetros libres); TD=SY,FR (la matriz  $\Theta_{\delta}$  de varianzas y covarianzas de error es simétrica y sus parámetros libres). Se especifican a continuación las variables independientes latentes del modelo: LK (AFECT, COGNI). Seguidamente, se especifican las matrices de los parámetros a estimar (utilizando 0 para los parámetros fijos y 1 para los libres). La primera matriz de parámetros es LX ( $\Lambda_x$ ) mediante el comando PA LX, a la que sigue la lista de valores de los elementos fijos de la matriz. Así, la línea VA 1.0 LX(1,1) LX(6,2) indica al programa que el parámetro se fija a 1.0 en la fila 1, columna 1 de la matriz, y en la fila 6, columna 2. Se especifica a continuación PA TD (matriz  $\Theta_{\delta}$ ) en la que todos los parámetros son fijos con excepción de las varianzas de error, y PA PF (matriz  $\Phi$ ) en la que todos los parámetros a estimar son libres. Finalmente, se establecen las opciones de salida OU: ME=ML indica que el método de estimación será el de máxima verosimilitud; RS (el programa ofrecerá los residuos, residuos estandarizados, gráfico QQ y matriz de covarianzas o correlaciones ajustadas); MI (índices de modificación); SC (solución completamente estandarizada) y ND=2 (número de decimales en los resultados).

16 En la notación LISREL, las variables  $\xi$  son variables latentes independientes (exógenas, no explicadas por otras variables del modelo), en contraposición a las variables  $\eta$  o variables latentes dependientes (endógenas o explicadas por otras variables del modelo).

Tabla 10. Sintaxis en LISREL para el modelo de dos factores correlacionados

```

TITLE MODELO B1 WILLIAMS EAVES & COX MODELO DE DOS FACTORES CORRELACIONADOS
DA NI=10 NO=216 MA=CM
LA
A1 A2 A3 A4 A5 C1 C2 C3 C4 C5
CM
[ESCRIBIR AQUÍ LA MATRIZ DE COVARIANZAS, VID. TABLA 6]
MO NX=10 NK=2 PH=SY,FR LX=FU,FR TD=SY,FR
LK
AFECT COGNI
PA LX
0 0
1 0
1 0
1 0
1 0
0 0
0 1
0 1
0 1
0 1
VA 1.0 LX(1,1) LX(6,2)
PA TD
1
0 1
0 0 1
0 0 0 1
0 0 0 0 1
0 0 0 0 0 1
0 0 0 0 0 0 1
0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 1
PA PH
1
1 1
OU ME=ML RS MI SC ND=2

```

En la Tabla 11 se muestra un resumen de los resultados de este modelo. Claramente se puede apreciar una mejora considerable del ajuste del modelo de dos factores correlacionados en comparación con el unifactorial. Estos resultados pueden considerarse en general favorables con excepción del valor de RMSEA, excesivamente elevado, si bien su límite inferior al nivel de confianza del 90% entra dentro del rango aceptable.

Podemos apreciar en la Tabla 12, que recoge los índices de modificación del modelo, que el único valor significativo corresponde a la saturación del *parcel* A5 en el factor COGNI (IM=4.97). Si volviésemos a especificar el modelo permitiendo que dicha variable saturara en el factor, se reduciría en consonancia el valor del  $\chi^2$  global. Resulta obvio que la reespecificación del modelo debería fundamentarse en sólidas razones teóricas. Aun sin llevar a cabo dicha reespecificación, hemos de concluir que el modelo de dos factores relacionados es claramente superior al modelo unifactorial.

17 Advertirá el lector –si se toma el trabajo de contrastar la solución LISREL con las ofrecidas por otros programas SEM como EQS, AMOS o SAS– que tanto el valor RMSEA como el resto de los índices parciales de ajuste son ligeramente diferentes en LISREL. La razón estriba en que LISREL, con el método de estimación de máxima verosimilitud, no utiliza el ‘Minimum Fit Function Chi-Square’ sino el ‘Normal Theory Weighted Least Squares Chi-Square’ en el cálculo (vid. Tabla 8). Así, el valor RMSEA de la Tabla 8 es el resultado de  $[(630.65-35)/(35*(216-1))]^{1/2}=\mathbf{0.2813}$ , en tanto que el valor RMSEA obtenido en EQS, AMOS o SAS resulta de  $[(409.68-35)/(35*(216-1))]^{1/2}=\mathbf{0.2231}$ . Es este último valor el que ofrecen, precisamente, los autores del artículo que aquí se comenta (sus análisis se realizaron con AMOS).

Tabla 11. Resumen de resultados del modelo de dos factores correlacionados

Goodness of Fit Statistics	
Degrees of Freedom = 34	
Minimum Fit Function Chi-Square = 88.47 (P = 0.00)	
Normal Theory Weighted Least Squares Chi-Square = 85.11 (P = 0.00)	
. . .	
Root Mean Square Error of Approximation (RMSEA) = 0.084	
90 Percent Confidence Interval for RMSEA = (0.062 ; 0.11)	
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.0076	
. . .	
Normed Fit Index (NFI) = 0.98	
Non-Normed Fit Index (NNFI) = 0.99	
Comparative Fit Index (CFI) = 0.99	
. . .	
Standardized RMR = 0.022	
Goodness of Fit Index (GFI) = 0.93	

Tabla 12. Índices de modificación del modelo de dos factores correlacionados

	Modification Indices for LAMBDA-X	
	AFECT	COGNI
	-----	-----
A1	- -	1.02
A2	- -	0.16
A3	- -	1.28
A4	- -	0.02
A5	- -	4.97
C1	0.00	- -
C2	0.51	- -
C3	0.69	- -
C4	1.78	- -
C5	1.74	- -

La superioridad del modelo bifactorial sobre el unifactorial puede determinarse asimismo mediante una prueba de las diferencias en  $\chi^2$ . Podría considerarse el modelo unifactorial como una versión ‘restringida’ del modelo de 2 factores fijando a 1.0 la correlación entre ellos (lo que sería lo mismo que reemplazar ambos factores por un único factor). Los resultados de la prueba de diferencias de  $\chi^2$  (vid. Tabla 13) ponen de manifiesto que la diferencia entre el ajuste del modelo bifactorial y el unifactorial es estadísticamente significativa).

Figura 7. Modelo AFC de dos factores correlacionados (solución estandarizada)

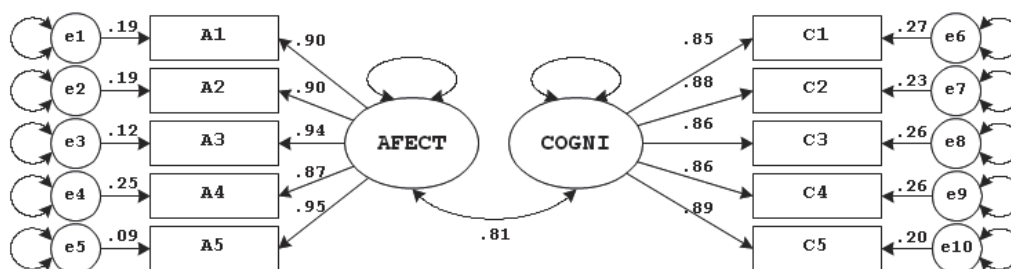


Tabla 13. Diferencias y significación de los valores  $\chi^2$  en los dos primeros modelos

Modelo	$\chi^2$	GL	p
<b>Unifactorial</b>	409.68	35	0.000
<b>Bifact. corr.</b>	88.47	34	0.000
<b>Diferencia</b>	321.21	1	0.000

La solución final (estimaciones estandarizadas) para el modelo analizado se representa gráficamente en la Figura 7. Comentamos a continuación con algún detalle la salida en SIMPLIS y LISREL para este modelo (Tabla 14). Incluiremos la salida completa en formato SIMPLIS, a la que añadiremos algunos fragmentos de la salida en LISREL para ilustrar las diferencias entre ambos formatos.

Se muestra, en primer lugar, la cabecera del programa y se reproduce a continuación toda la sintaxis utilizada (variables observadas y latentes, matriz de correlaciones o de covarianzas, relaciones entre las variables, etc.).

Tabla 14. Salida en LISREL del análisis del modelo de dos factores correlacionados

DATE: 3/27/2008
TIME: 19:36
L I S R E L 8.80
BY
Karl G. Jöreskog & Dag Sörbom
This program is published exclusively by
Scientific Software International, Inc.
7383 N. Lincoln Avenue, Suite 100
Lincolnwood, IL 60712, U.S.A.
Phone: (800)247-6113, (847)675-0720, Fax: (847)675-2140
Copyright by Scientific Software International, Inc., 1981-2006
Use of this program is subject to the terms specified in the
Universal Copyright Convention.
Website: www.ssicentral.com

Tabla 14 (cont.). Reproducción de la sintaxis

```

!Matriz de covarianzas Williams Eaves
Observed Variables
A1 A2 A3 A4 A5 C1 C2 C3 C4 C5
Latent Variables
AFECT COGNI
Covariance Matrix
2038.035
1631.766 1932.163
1393.567 1336.871 1313.809
1657.299 1647.164 1273.261 2034.216
1721.830 1688.292 1498.401 1677.767 2061.875
659.795 702.969 585.019 656.527 775.118 630.518
779.519 790.448 653.734 764.755 871.211 500.128 741.767
899.912 879.179 750.037 897.441 1008.287 648.935 679.970 1087.409
775.235 739.157 659.867 751.860 895.616 550.476 603.950 677.865 855.272
821.170 785.419 669.951 802.825 877.681 491.042 574.775 686.391 622.830
728.674
Sample Size is 216
Relationships
A1 = 1*AFECT
A2 = AFECT
A3 = AFECT
A4 = AFECT
A5 = AFECT
C1 = 1*COGNI
C2 = COGNI
C3 = COGNI
C4 = COGNI
C5 = COGNI
OPTIONS wp rs ef ss sc mi nd=2
Path Diagram
End of Program

Sample Size = 216

```

A continuación se calcula la matriz de covarianzas en el caso de que los datos de entrada hayan sido correlaciones; si han sido covarianzas, el programa se limita a reproducir la matriz. En la salida en formato LISREL se especifican los parámetros a estimar.

Tabla 14 (cont.). Matriz de covarianzas

Covariance Matrix										
	A1	A2	A3	A4	A5	C1	C2	C3	C4	C5
A1	2038.04									
A2	1631.77	1932.16								
A3	1393.57	1336.87	1313.81							
A4	1657.30	1647.16	1273.26	2034.22						
A5	1721.83	1688.29	1498.40	1677.77	2061.88					
C1	659.79	702.97	585.02	656.53	775.12	630.52				
C2	779.52	790.45	653.73	764.75	871.21	500.13	741.77			
C3	899.91	879.18	750.04	897.44	1008.29	648.93	679.97	1087.41		
C4	775.24	739.16	659.87	751.86	895.62	550.48	603.95	677.87	855.27	
C5	821.17	785.42	669.95	802.83	877.68	491.04	574.77	686.39	622.83	728.67

Tabla 14 (cont.). Especificaciones de los parámetros

[SALIDA EN FORMATO LISREL] Parameter Specifications										
LAMBDA-X										
	AFECT					COGNI				
	-----					-----				
A1						0				0
A2						1				0
A3						2				0
A4						3				0
A5						4				0
C1						0				0
C2						0				5
C3						0				6
C4						0				7
C5						0				8
PHI										
	AFECT					COGNI				
	-----					-----				
AFECT						9				
COGNI						10				11
THETA-DELTA										
A1	A2	A3	A4	A5						
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
12	13	14	15	16						
THETA-DELTA										
C1	C2	C3	C4	C5						
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
17	18	19	20	21						

Los parámetros estimados por el programa se muestran inmediatamente después de la matriz de covarianzas. Se presentan en forma de ecuaciones, en las que cada una de las variables observadas se expresa como una función lineal de su variable latente o factor. Si en el modelo hubiera variables dependientes latentes, se expresarían en las ecuaciones como funciones lineales de las respectivas variables latentes independientes. Para cada uno de los parámetros libres a estimar (8, en nuestro ejemplo) el programa proporciona a) la estimación del parámetro no estandarizado, b) su error estándar y c) el valor de t. Tomemos, por ejemplo, la ecuación que vincula la variable A3 con el factor AFECT: el parámetro estimado no estandarizado es 0.84; su error estándar, 0.036; el valor de t, 23.57. Los parámetros fijos (A1àAFECT y C1àCOGNI) no llevan aparejados, obviamente, ni error estándar ni valor t. La magnitud de los parámetros no estandarizados muestran el cambio (positivo o negativo) resultante en la va-

riable dependiente, producido cuando la variable independiente cambia una unidad en tanto que las demás variables independientes se mantienen constantes. Por ejemplo, la ecuación  $A3=0.96*AFECT$  indica que el cambio de una unidad en AFECT supone el incremento de 0.96 unidades en A3. El error estándar (0.036 para la variable A3) muestra con qué grado de precisión se ha estimado el parámetro: mientras más pequeño sea, más precisa habrá sido la medida. Si dividimos el valor del parámetro por el error estándar, tendremos el valor de t. Así,  $0.84/0.036=23.57$ . Los valores de t se utilizan para contrastar la hipótesis de que el parámetro es distinto de 0 en la población. Aceptamos la hipótesis si el valor absoluto es superior a 1.96, lo que supone una significación de  $p=.05$ . En la salida se ofrece, finalmente, otra información relevante: las varianzas de error (que, al ser parámetros libres, vienen acompañadas del correspondiente error estándar y valor de t), así como la correlación múltiple al cuadrado  $R^2$  (muestra la cantidad de varianza de la variable dependiente explicada por la variable independiente en cada ecuación). En el caso que vamos comentando, la varianza de error de la variable A3 es 155.67; el error estándar, 20.13; el valor de t, 7.73 y el valor de  $R^2$  0.88.

Tabla 14 (cont.). Estimaciones del modelo de medida

LISREL Estimates (Maximum Likelihood)		
Measurement Equations		
A1 = 1.00*AFECT, Errorvar.= 385.92, R <sup>2</sup> = 0.81	(43.48)	8.88
A2 = 0.98*AFECT, Errorvar.= 359.01, R <sup>2</sup> = 0.81	(0.046)	(40.61)
	21.10	8.84
A3 = 0.84*AFECT, Errorvar.= 155.67, R <sup>2</sup> = 0.88	(0.036)	(20.13)
	23.57	7.73
A4 = 0.96*AFECT, Errorvar.= 498.77, R <sup>2</sup> = 0.75	(0.050)	(53.61)
	19.21	9.30
A5 = 1.06*AFECT, Errorvar.= 192.58, R <sup>2</sup> = 0.91	(0.043)	(27.73)
	24.58	6.95
C1 = 1.00*COGNI, Errorvar.= 172.46, R <sup>2</sup> = 0.73	(19.70)	8.75
C2 = 1.11*COGNI, Errorvar.= 172.60, R <sup>2</sup> = 0.77	(0.065)	(20.66)
	17.05	8.36
C3 = 1.33*COGNI, Errorvar.= 277.30, R <sup>2</sup> = 0.74	(0.080)	(32.29)
	16.60	8.59
C4 = 1.17*COGNI, Errorvar.= 225.24, R <sup>2</sup> = 0.74	(0.071)	(25.99)
	16.44	8.67
C5 = 1.13*COGNI, Errorvar.= 147.55, R <sup>2</sup> = 0.80	(0.064)	(18.53)
	17.67	7.96

[SALIDA EN FORMATO LISREL]		
LISREL Estimates (Maximum Likelihood)		
LAMBDA-X		
	AFECT	COGNI
-----		
A1	1.00	- -
A2	0.98	- -
	(0.05)	
	21.10	
A3	0.84	- -
	(0.04)	
	23.57	
A4	0.96	- -
	(0.05)	
	19.21	
A5	1.06	- -
	(0.04)	
	24.58	
C1	- -	1.00
C2	- -	1.11
		(0.07)
		17.05
C3	- -	1.33
		(0.08)
		16.60
C4	- -	1.17
		(0.07)
		16.44
C5	- -	1.13
		(0.06)
		17.67



Tabla 14 (cont.). Estimación del modelo de medida

[SALIDA EN FORMATO LISREL]					
THETA-DELTA					
A1	A2	A3	A4	A5	C1
385.92	359.01	155.67	498.77	192.58	172.46
(43.48)	(40.61)	(20.13)	(53.61)	(27.73)	(19.70)
8.88	8.84	7.73	9.30	6.95	8.75
THETA-DELTA					
C2	C3	C4	C5		
172.60	277.30	225.24	147.55		
(20.66)	(32.29)	(25.99)	(18.53)		
8.36	8.59	8.67	7.96		
Squared Multiple Correlations for X - Variables					
A1	A2	A3	A4	A5	C1
0.81	0.81	0.88	0.75	0.91	0.73
Squared Multiple Correlations for X - Variables					
C2	C3	C4	C5		
0.77	0.74	0.74	0.80		

Como se ve, el formato de salida de SIMPLIS y LISREL es ligeramente diferente: en LISREL se adopta la denominación de las distintas matrices (e.g., LAMBDA-X, LAMBDA-Y, THETA-DELTA, PSI, etc.), lo que requiere que el lector esté familiarizado con dicha notación, en tanto que la salida SIMPLIS es más sencilla de interpretar.

En el siguiente apartado de la salida se muestra la matriz de varianzas y covarianzas de las variables independientes (exógenas) del modelo. Cada estimación viene de nuevo acompañada por su error estándar y el correspondiente valor de t.

Tabla 14 (Cont.). Matriz de covarianza de las VI

Covariance Matrix of Independent Variables		
	AFECT	COGNI
AFECT	1652.11 (194.31) 8.50	
COGNI	705.34 (86.35) 8.17	458.06 (59.44) 7.71
[SALIDA EN FORMATO LISREL]		
PHI		
	AFECT	COGNI
AFECT	1652.11 (194.31) 8.50	
COGNI	705.34 (86.35) 8.17	458.06 (59.44) 7.71

Los apartados de la salida que vienen a continuación se refieren al ajuste obtenido por el modelo. Como más arriba se ha dicho, el proceso de estimación da como resultado una matriz  $\hat{\Sigma}$  que reproduzca con la máxima fidelidad posible a la matriz observada  $S$ . Conviene, en este punto, aclarar al lector que los índices de ajuste no proporcionan garantía alguna respecto a la utilidad del modelo: de hecho, la información que *realmente* suministran se refiere más bien a la *falta de ajuste* del modelo. En segundo lugar, un buen ajuste no *prueba* nada en sentido estricto (podrían concebirse múltiples modelos que se ajustaran mejor a los datos observados que el que sometemos a análisis). En este sentido, resultaría más útil un modelo mal ajustado, en la medida en que dicho mal ajuste *sí* sería una evidencia concluyente de

que el modelo *no se ajusta a los datos*. En tercer lugar, la evaluación del ajuste debería realizarse desde perspectivas diversas y utilizando distintos índices tanto globales como parciales (dicho de otro modo, no disponemos de un índice incuestionable que determine si el modelo se ajusta o no a los datos empíricos). Por último, la evaluación del ajuste de un modelo en modo alguno deberá desvincularse de la teoría en la que se sustenta. Antes al contrario: un buen ajuste no vinculado a una sólida teoría de base se torna inútil.

Tabla 14 (Cont.). Estadísticos de bondad del ajuste

<i>Porción de la salida</i>	
Goodness of Fit Statistics	
Degrees of Freedom = 34 Minimum Fit Function Chi-Square = 88.47 (P = 0.00) Normal Theory Weighted Least Squares Chi-Square = 85.11 (P = 0.00) Estimated Non-centrality Parameter (NCP) = 51.11 90 Percent Confidence Interval for NCP = (27.72 ; 82.19)	
<i>Resultado</i>	<i>Comentarios</i>
Degrees of Freedom = 34	Grados de libertad del modelo. Evaluar el ajuste global significa determinar el grado en que el modelo considerado como un todo es consistente con los datos empíricos. Los índices de ajuste global evalúan en qué medida el modelo se ajusta perfectamente a la población: $\Sigma = \Sigma(\theta)$
Minimum Fit Function Chi-Square = 88.47 (P = 0.00)	El valor de $\chi^2$ es la medida tradicionalmente utilizada para evaluar el ajuste global del modelo (pone a prueba la hipótesis nula de que el modelo se ajusta perfectamente a los datos de la población). En el ejemplo ( $\chi^2=88.47$ , $p=.000$ ) debemos rechazar la hipótesis nula ( $p=.000$ ). Sin embargo, desde una perspectiva más pragmática, conviene no tanto examinar el nivel de significación estadística, cuanto la magnitud del valor de $\chi^2$ : valores grandes corresponderían a un ajuste deficiente, y valores pequeños a un mejor ajuste. Los grados de libertad podrían servir como estándar a través del cual juzgar si el valor de $\chi^2$ es grande o pequeño.
Normal Theory Weighted Least Squares Chi-Square = 85.11 (P = 0.00)	El valor de <i>Normal Theory Weighted Least Squares Chi-Square</i> se obtiene mediante una fórmula más complicada que el valor anterior. Con todo, su interpretación es la misma. Ambos valores son sensibles a las desviaciones de la normalidad multivariada (en especial a la curtosis), al tamaño de la muestra y asumen, además, que el modelo se ajusta perfectamente a la población.
Estimated Non-centrality Parameter (NCP) = 51.11	En lugar de preguntarnos si el modelo es correcto (i.e., se ajusta perfectamente a la matriz de covarianzas de la población), podríamos estimar su falta de ajuste. Esto es lo que pretende el índice <i>Estimated Non-centrality Parameter (NCP)</i> . El valor de la discrepancia entre $\Sigma$ y $\Sigma(\theta)$ obtenido es de 51.11, valor razonablemente bajo como para considerar la aceptación de la hipótesis de nulidad.
90 Percent Confidence Interval for NCP = (27.72 ; 82.19)	El intervalo de confianza al 90% del NCP es 27.72-82.19, lo que refrenda el comentario anterior (de hecho, LISREL no ofrece el intervalo de confianza NCP cuando su valor es demasiado alto). Así pues, la discrepancia entre $\Sigma$ y $\Sigma(\theta)$ no puede considerarse excesiva, lo que apoya el ajuste del modelo a los datos.

Tabla 14 (Cont.). Estadísticos de bondad del ajuste

<i>Porción de la salida</i>	
Minimum Fit Function Value = 0.41 Population Discrepancy Function Value (F0) = 0.24 90 Percent Confidence Interval for F0 = (0.13 ; 0.38) Root Mean Square Error of Approximation (RMSEA) = 0.084 90 Percent Confidence Interval for RMSEA = (0.062 ; 0.11) P-Value for Test of Close Fit (RMSEA < 0.05) = 0.0076	
<i>Resultado</i>	<i>Comentarios</i>
Minimum Fit Function Value = 0.41	El valor de función de ajuste mínimo se usa para el cálculo de $\chi^2$ con el método de máxima verosimilitud (ML) mediante la fórmula $\chi^2 = (n - 1) \cdot F_{ML}$ donde n es el tamaño de la muestra y $F_{ML}$ es el <i>Minimum Fit Function Value</i> ; en el ejemplo, $\chi^2 = (216 - 1) \cdot .4115 = 88.4725$
Population Discrepancy Function Value (F0) = 0.24	El <i>Population Discrepancy Function Value</i> es el valor estimado de la función de ajuste cuando el modelo se ajusta a la matriz de covarianza de la población ( $\Sigma$ ). Sirve para calcular el valor RMSEA (vid. más abajo) mediante la fórmula $RMSEA = (F_0 / GL)^{1/2}$ donde $F_0$ es el valor de función de discrepancia de la población y GL los grados de libertad. Así, en el ejemplo, $RMSEA = (.24 / 35)^{1/2} = .0840$
90 Percent Confidence Interval for F0 = (0.13 ; 0.38)	El valor verdadero de $F_0$ se sitúa, al nivel de confianza del 90%, entre 0.13 y 0.38.
Root Mean Square Error of Approximation (RMSEA) = 0.084	El valor RMSEA = 0.084 indica un ajuste pobre (inferior a .05 es bueno; entre .05 y .08 razonable; entre .08 y .10 mediocre; por encima de .10 el modelo debería ser rechazado). RMSEA, al igual que NCP, se centra en la discrepancia entre $\Sigma$ y $\Sigma(\theta)$ , pero en esta ocasión por grado de libertad, lo que supone tomar en consideración la complejidad del modelo.
90 Percent Confidence Interval for RMSEA = (0.062 ; 0.11)	El intervalo de confianza (0.062 ; 0.11) no incluye el valor .05, (tampoco .06) lo que refrenda el escaso ajuste.
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.0076	La probabilidad de que el valor RMSEA sea inferior a .05 es únicamente en nuestro ejemplo del 7.6%.

Tabla 14 (Cont.). Estadísticos de bondad del ajuste

Expected Cross-Validation Index (ECVI) = 0.59 90 Percent Confidence Interval for ECVI = (0.48 ; 0.74) ECVI for Saturated Model = 0.51 ECVI for Independence Model = 23.00	
Resultado	Comentarios
Expected Cross-Validation Index (ECVI) = 0.59	En tanto que NCP y RMSEA se centran en el error de aproximación, i.e., la discrepancia entre $\Sigma$ y $\Sigma(\theta)$ , el valor ECVI lo hace sobre el error global entre $\Sigma$ y $\hat{\Sigma}$ , esto es, la diferencia entre la matriz de covarianzas de la población y la ajustada por el modelo a la muestra. ECVI es, por tanto, un indicador ilustrativo del ajuste global del modelo. El valor obtenido en nuestro ejemplo, ECVI=0.59 sería útil si comparáramos el modelo con otros (no es informativo por sí mismo). Si comparáramos el modelo con otros, elegiríamos el que tuviera un valor ECVI más bajo. En la práctica, los 'otros' modelos son el modelo de <i>independencia</i> y el modelo <i>saturado</i> . El primero es un modelo de independencia completa entre todas las variables (i.e., todas las variables observadas son no correlacionadas y, por tanto, es el más restrictivo). Sus grados de libertad equivalen a $k(k-1)/2$ donde $k$ es el número de variables; en nuestro caso, $10(10-1)/2=45$ gl. El modelo saturado es aquél en el que el número de parámetros a estimar es igual al número de varianzas y covarianzas entre las variables observadas. Tiene, por tanto, $k(k+1)/2$ parámetros y 0 grados de libertad, es decir, es un modelo identificado ( <i>just-identified</i> ). En el ejemplo, los parámetros a estimar son $10(10+1)/2=55$ . Cualquier modelo hipotetizado, en consecuencia, se situará siempre en algún punto entre el modelo saturado y el modelo de independencia. Si el ajuste es bueno, el valor ECVI será inferior al ofrecido por ambos modelos. En nuestro ejemplo, ECVI=0.59 es menor que el ECVI del modelo de independencia (23.00) pero mayor que el del modelo saturado (0.51), lo que abunda en que el ajuste obtenido no es excesivamente bueno. Sin embargo, debe considerarse que el intervalo al 90% está entre 0.48 y 0.74, de modo que el valor más bajo del intervalo sí es inferior al ECVI del modelo saturado.
90 Percent Confidence Interval for ECVI = (0.48 ; 0.74)	Vid. explicación del ECVI
ECVI for Saturated Model = 0.51	Vid. explicación del ECVI
ECVI for Independence Model = 23.00	Vid. explicación del ECVI

Tabla 14 (Cont.). Estadísticos de bondad del ajuste

<i>Porción de la salida</i>	
Chi-Square for Independence Model with 45 Degrees of Freedom = 4924.37 Independence AIC = 4944.37 Model AIC = 127.11 Saturated AIC = 110.00 Independence CAIC = 4988.12 Model CAIC = 218.99 Saturated CAIC = 350.64	
<i>Resultado</i>	<i>Comentarios</i>
Chi-Square for Independence Model with 45 Degrees of Freedom = 4924.37	Valor de $\chi^2$ para el modelo de independencia (vid. más arriba). Este valor se usará para calcular otros índices de ajuste (e.g., NFI, NNFI).
Independence AIC = 4944.37	Los siguientes índices de ajuste ( <i>Akaike Information Criterion AIC</i> , <i>Consistent Akaike Information Criterion CAIC</i> ) se utilizan para comparar modelos (i.e., el hipotetizado frente al modelo de independencia y al saturado). Tratan de evaluar el ajuste incorporando la parsimonia del modelo, para lo que tienen en cuenta el número de parámetros estimados.
Model AIC = 127.11	En nuestro ejemplo, AIC=127.11 es inferior al AIC del modelo de independencia (4944.37), pero superior al AIC del modelo saturado (110.00): este resultado es conflictivo.
Saturated AIC = 110.00	Valor del AIC en el modelo saturado.
Independence CAIC = 4988.12	Valor del CAIC en el modelo de independencia.
Model CAIC = 218.99	En nuestro ejemplo, CAIC=218.99 es inferior tanto al AIC del modelo de independencia (4988.12), como al AIC del modelo saturado (350.64).
Saturated CAIC = 350.64	Valor del CAIC en el modelo saturado.

Tabla 14 (Cont.). Estadísticos de bondad del ajuste

<i>Porción de la salida</i>	
<p>Normed Fit Index (NFI) = 0.98  Non-Normed Fit Index (NNFI) = 0.99  Parsimony Normed Fit Index (PNFI) = 0.74  Comparative Fit Index (CFI) = 0.99  Incremental Fit Index (IFI) = 0.99  Relative Fit Index (RFI) = 0.98</p>	
<i>Resultado</i>	<i>Comentarios</i>
Normed Fit Index (NFI) = 0.98	Los siguientes índices evalúan en qué medida nuestro modelo es mejor que otros, especialmente el modelo de independencia. En el ejemplo, NFI = 0.9820; NNFI = 0.9852; PNFI = 0.7420; CFI = 0.9888; IFI = 0.9889; RFI = 0.9762. Con excepción de PNFI, estos índices indican un relativo buen ajuste del modelo comparado con el modelo de independencia.
Non-Normed Fit Index (NNFI) = 0.99	Vid. explicación NFI
Parsimony Normed Fit Index (PNFI) = 0.74	Vid. explicación NFI
Comparative Fit Index (CFI) = 0.99	Vid. explicación NFI
Incremental Fit Index (IFI) = 0.99	Vid. explicación NFI
Relative Fit Index (RFI) = 0.98	Vid. explicación NFI

Tabla 14 (Cont.). Estadísticos de bondad del ajuste

<i>Porción de la salida</i>	
Critical N (CN) = 137.24  Root Mean Square Residual (RMR) = 29.47 Standardized RMR = 0.022 Goodness of Fit Index (GFI) = 0.93 Adjusted Goodness of Fit Index (AGFI) = 0.88 Parsimony Goodness of Fit Index (PGFI) = 0.57	
<i>Resultado</i>	<i>Comentarios</i>
Critical N (CN) = 137.24	CN indica el tamaño que una muestra debe alcanzar para aceptar el ajuste de un determinado modelo sobre una base estadística. El límite inferior aceptable suele estar en 200. En este caso, CN = 137.2438, valor algo apartado de dicho límite.
Root Mean Square Residual (RMR) = 29.47	RMR = 29.47, indica el valor promedio de los elementos en $(S - \hat{\Sigma})$ (vid. explicación siguiente).
Standardized RMR = 0.022	Puesto que el tamaño de los residuos puede variar con la unidad de medida, es mejor utilizar el SRMR = 0.022 (los residuos estandarizados son los residuos ajustados divididos por su error estándar) y pueden interpretarse como desviaciones típicas normales (valores por encima de 2.58 se consideran altos). En el caso del SRMR, valores por debajo de 0.05 indican un buen ajuste (tal es el caso de nuestro modelo).
Goodness of Fit Index (GFI) = 0.93	Los índices GFI, AGFI y PGFI evalúan el grado en que las varianzas y covarianzas del modelo reproducen correctamente la matriz de varianzas y covarianzas original. El primero (GFI = 0.9266) lo hace globalmente, el segundo (AGFI = 0.8813) ajustado a los grados de libertad y el tercero (PGFI = 0.5728) ajustado a la complejidad del modelo. Los dos primeros deben superar el valor de 0.90; el tercero puede ser mucho más bajo y aun así el modelo tener un buen ajuste.
Adjusted Goodness of Fit Index (AGFI) = 0.88	Vid. explicación GFI
Parsimony Goodness of Fit Index (PGFI) = 0.57	Vid. explicación GFI

Como se ha visto, son muchos los posibles índices de ajuste, y ninguno de ellos por separado es suficiente para determinar que el modelo se ajusta a los datos. La combinación hoy día más utilizada es la siguiente:  $\chi^2$ , RMSEA, ECVI, SRMR, GFI y CFI: ese conjunto debería resultar suficiente para tomar una decisión respecto al ajuste del modelo (Boomsna, 2000; McDonald & Ho, 2002).

Los elementos de la matriz de covarianzas ajustadas (*Fitted Covariance Matrix*) son las varianzas y covarianzas de las variables manifiestas estimadas por el modelo. En otras palabras, conforman la matriz que se compara con la matriz observada,  $S$ .

Tabla 14 (Cont.). Matriz de covarianzas ajustadas

Fitted Covariance Matrix										
	A1	A2	A3	A4	A5	C1	C2	C3	C4	C5
A1	2038.04									
A2	1612.15	1932.16								
A3	1383.25	1349.79	1313.81							
A4	1592.71	1554.18	1333.51	2034.22						
A5	1757.35	1714.84	1471.36	1694.17	2061.88					
C1	705.34	688.28	590.55	679.98	750.27	630.52				
C2	786.24	767.22	658.29	757.97	836.33	510.60	741.77			
C3	938.02	915.33	785.36	904.29	997.77	609.16	679.03	1087.41		
C4	827.22	807.21	692.59	797.47	879.91	537.20	598.82	714.42	855.27	
C5	794.46	775.24	665.17	765.90	845.07	515.93	575.11	686.13	605.08	728.67

Los resultados de la comparación anterior se incluyen en el apartado *Fitted Residuals* que componen la matriz residual ( $S - \hat{\Sigma}$ ). Por ejemplo, el residuo entre A2 y C1 es 14.69, resultado de la diferencia entre la covarianza original (702.969) y la estimada por el modelo (688.28). Tenemos un caso de ‘infra-ajuste’ (*underfitting*) toda vez que el modelo lleva a cabo una infraestimación de la covarianza. Lo contrario sucede cuando los residuos ajustados son negativos.

Tabla 14 (Cont.). Matriz de residuos ajustados

Fitted Residuals										
	A1	A2	A3	A4	A5	C1	C2	C3	C4	C5
A1	0.00									
A2	19.62	0.00								
A3	10.32	-12.91	0.00							
A4	64.59	92.98	-60.25	0.00						
A5	-35.52	-26.55	27.04	-16.40	0.00					
C1	-45.54	14.69	-5.53	-23.45	24.85	0.00				
C2	-6.72	23.22	-4.56	6.78	34.88	-10.47	0.00			
C3	-38.11	-36.15	-35.33	-6.85	10.52	39.78	0.94	0.00		
C4	-51.98	-68.05	-32.73	-45.61	15.71	13.27	5.13	-36.55	0.00	
C5	26.71	10.17	4.78	36.93	32.61	-24.89	-0.34	0.26	17.75	0.00

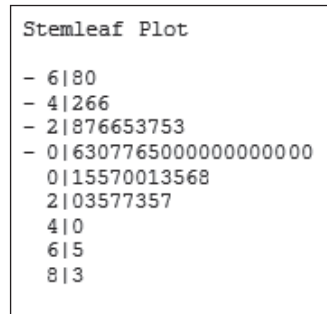
Summary Statistics for Fitted Residuals

Smallest Fitted Residual = -68.05  
 Median Fitted Residual = 0.00  
 Largest Fitted Residual = 92.98



Los residuos se presentan asimismo en un gráfico de tallo y hojas (*stem-and-leaf Plot*) para facilitar la inspección visual de su distribución.

Tabla 14 (Cont.). Gráfico de residuos ajustados



Los residuos vuelven a presentarse, en esta ocasión en formato estandarizado (*Standardized Residuals*); se adjunta la mediana de los residuos y los valores máximo y mínimo. Los residuos estandarizados se consideran ‘grandes’ si superan el valor de  $|2.58|$ .

Tabla 14 (Cont.). Matriz de residuos estandarizados

	A1	A2	A3	A4	A5	C1	C2	C3	C4	C5
A1	0.00									
A2	0.95	0.00								
A3	0.86	-1.12	0.00							
A4	2.55	<b>3.81</b>	<b>-4.15</b>	0.00						
A5	<b>-2.90</b>	-2.27	<b>4.23</b>	-1.10	0.00					
C1	-1.58	0.53	-0.27	-0.75	1.03	0.00				
C2	-0.23	0.82	-0.22	0.21	1.45	-1.16	0.00			
C3	-1.03	-1.01	-1.37	-0.17	0.34	<b>3.40</b>	0.08	0.00		
C4	-1.57	-2.12	-1.41	-1.27	0.57	1.25	0.50	<b>-2.76</b>	0.00	
C5	0.97	0.38	0.25	1.21	1.48	<b>-3.12</b>	-0.04	0.03	1.97	0.00
Summary Statistics for Standardized Residuals										
Smallest Standardized Residual = -4.15										
Median Standardized Residual = 0.00										
Largest Standardized Residual = 4.23										

Al igual que en el caso de los residuos no estandarizados, se ofrece un gráfico *stem-and-leaf* con la distribución de los residuos estandarizados<sup>18</sup>, y se incluye información sobre los valores más elevados, tanto positivos como negativos. En el ejemplo, el residuo negativo más elevado corresponde a A3-A4 (-4.15) en tanto que el positivo más elevado corresponde a C3-

18 En el gráfico ‘stem-and-leaf’ se muestra, a la izquierda de la línea punteada, el ‘tallo’ (i.e. primer dígito del residuo) y a la derecha de la línea las ‘hojas’ (i.e., segundo dígito). Así, tenemos un valor cuyo primer dígito es -4 (-4.2); otro valor cuyo primer dígito es -3 (-3.1); cuatro valores cuyo primer dígito es -2 (-2.9, -2.8, -2.3 y -2.1) y así sucesivamente, hasta completar los 55 residuos.

C1 (3.40). Deberían considerarse los residuos superiores a  $|2.58|$ . De nuestros 55 residuos, 7 (12.7%) presentan valores superiores a 2.58<sup>19</sup> (se señalan en negrita en la tabla anterior).

Si el modelo ajusta bien, se espera que el *stem-and-leaf* sea simétrico y los residuos se distribuyan en torno a cero, con la mayoría situados en la zona central. Un exceso de residuos en las colas sugiere que las covarianzas han sido sistemáticamente infraestimadas o sobreestimadas por el modelo. Residuos positivos grandes indican que el modelo infraestima la covarianza entre las variables manifiestas (*underfitting*) y por tanto debería modificarse añadiendo *paths* (e.g., a través de dejar parámetros libres) con el fin de explicar mejor la covarianza entre dos variables. En contraste, residuos negativos grandes indican que el modelo sobreestima la covarianza entre las variables manifiestas (*overfitting*) y por tanto debería modificarse eliminando *paths* (e.g., incluyendo parámetros fijos) que están asociados con las variables implicadas.

Tabla 14 (Cont.). Gráfico de residuos estandarizados

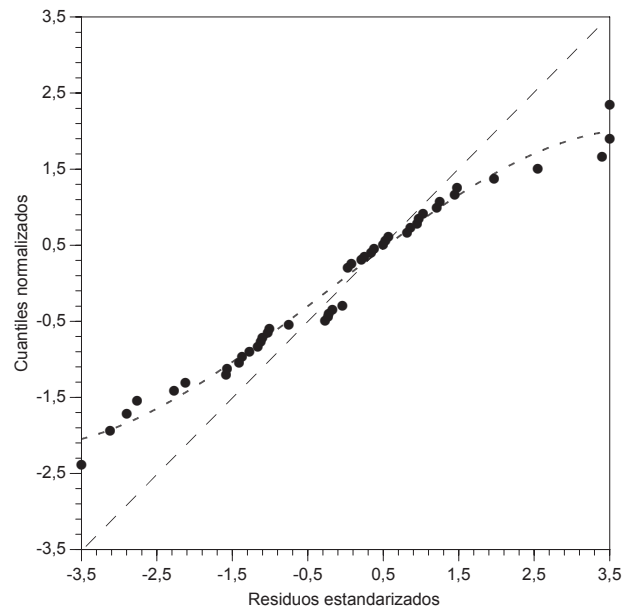
Stemleaf Plot			
-	4		2
-	3		1
-	2		9831
-	1		6644321100
-	0		7322200000000000
	0		1233455689
	1		0002355
	2		05
	3		48
	4		2
Largest Negative Standardized Residuals			
Residual for	A4 and	A3	-4.15
Residual for	A5 and	A1	-2.90
Residual for	C4 and	C3	-2.76
Residual for	C5 and	C1	-3.12
Largest Positive Standardized Residuals			
Residual for	A4 and	A2	3.81
Residual for	A5 and	A3	4.23
Residual for	C3 and	C1	3.40

El gráfico *stem-and-leaf* se completa, en la salida de LISREL, con el de residuos estandarizados contra los cuantiles de la distribución normal<sup>20</sup> (vid. Figura 8).

19 Algunos autores (e.g., Diamantopoulos & Siguaaw, 2000) estiman el valor de  $|4|$  como punto de corte para considerar 'grandes' los residuos. Adoptando esa opción, nuestro modelo únicamente tendría 2 residuos grandes (A3-A4: -4.15) y A3-A5: 4.23).

20 Puesto que el gráfico de residuos que proporciona LISREL es de baja resolución y de muy escasa calidad, recomendamos realizar dicho gráfico en programas especializados (e.g., Excel, DeltaGraph, SigmaPlot). El modo de generar en Excel los datos para su graficación es, en síntesis, el siguiente: en la columna A escribimos los rangos, desde 1 a n, siendo n el número de residuos (en nuestro ejemplo, 45, puesto que desechamos los 10 valores 0.00 de la diagonal); en la B, la probabilidad acumulada, utilizando la fórmula  $B2=(A2-0,5)/n$  y replicando hacia abajo; en la columna C escribimos los residuos y los ordenamos de menor a mayor; en la celda D2 calculamos la puntuación normalizada con la función  $D2=DISTR.NORM.ESTAND.INV(B2)$  y replicamos hacia abajo;  $D3= DISTR.NORM.ESTAND.INV(B3)$  y así sucesivamente. Graficamos a continuación con un dispersigrama los datos de las columnas C y D. Para asemejar el nuevo gráfico al ofrecido por LISREL, sustituimos los valores menores que -3.5 por -3.5 y los mayores que 3.5 por 3.5.

Figura 8. Residuos estandarizados y cuantiles normalizados



El mejor ajuste posible sucedería si todos los puntos estuvieran ordenados en una línea paralela al eje de ordenadas; el peor, si estuvieran en una línea paralela al eje de abscisas. El ajuste aceptable queda indicado cuando los puntos se sitúan en torno a la diagonal.

El siguiente apartado de la salida de resultados en LISREL se refiere a los *índices de modificación*. En este punto, conviene plantearse dos preguntas: ¿Puede mejorarse un modelo que ajusta bien? ¿Puede modificarse un modelo que ajusta mal para mejorar ese ajuste?

En principio, ha de tenerse en cuenta que existen dos tipos de errores de especificación: a) errores de especificación interna (por omisión de parámetros relevantes o inclusión de parámetros irrelevantes) y b) errores de especificación externa (variables omitidas). Para diagnosticar el modelo seguiremos dos vías: estadísticos residuales e índices de modificación.

Un índice de modificación muestra el decremento mínimo en el valor de  $\chi^2$  del modelo si un parámetro fijo se hiciera libre y se volviera a estimar el modelo. Se consideran 'grandes' los índices de modificación superiores a 3.84 (puesto que 3.84 es el valor crítico del estadístico  $\chi^2$  con 1 grado de libertad al nivel de confianza del 5%). En nuestro modelo sólo hay un índice superior a 3.84 en la matriz LAMBDA-X (saturación de A5 en el factor COGNI).

Tabla 14 (Cont.). Índices de modificación

Modification Indices for LAMBDA-X			Expected Change for LAMBDA-X		
	AFFECT	COGNI		AFFECT	COGNI
	-----	-----		-----	-----
A1	- -	1.02	A1	- -	-0.13
A2	- -	0.16	A2	- -	-0.05
A3	- -	1.28	A3	- -	-0.10
A4	- -	0.02	A4	- -	-0.02
A5	- -	<b>4.97</b>	A5	- -	0.24
C1	0.00	- -	C1	0.00	- -
C2	0.51	- -	C2	0.03	- -
C3	0.69	- -	C3	-0.05	- -
C4	1.78	- -	C4	-0.07	- -
C5	1.74	- -	C5	0.06	- -
Standardized Expected Change for LAMBDA-X			Completely	Standardized	Expected
			Change for LAMBDA-X	Change for LAMBDA-X	
	AFFECT	COGNI		AFFECT	COGNI
	-----	-----		-----	-----
A1	- -	-2.84	A1	- -	-0.06
A2	- -	-1.10	A2	- -	-0.03
A3	- -	-2.19	A3	- -	-0.06
A4	- -	-0.39	A4	- -	-0.01
A5	- -	5.14	A5	- -	0.11
C1	-0.09	- -	C1	0.00	- -
C2	1.41	- -	C2	0.05	- -
C3	-2.04	- -	C3	-0.06	- -
C4	-2.93	- -	C4	-0.10	- -
C5	2.49	- -	C5	0.09	- -

La información de los índices de modificación sigue el mismo orden que los parámetros (LAMBDA-Y, LAMBDA-X, BETA, GAMMA, PHI, PSI, THETA-EPSILON Y THETA-DELTA). Los valores suministrados servirán para, en su caso, introducir modificaciones que mejoren el ajuste del modelo. Los índices de modificación sugieren añadir covarianzas de error entre los indicadores A4 y A2, A5 y A3, y entre A4 y A3. Si volvemos a estimar el modelo con dichos cambios, obtendríamos un decremento en el valor de  $\chi^2$ . En todo caso, tanto la introducción de nuevos parámetros como la eliminación de alguno presente en el modelo que resulte innecesario –fijándolo a cero–, deberá estar justificada

Tabla 14 (Cont.). Índices de modificación para  $\theta - \delta$

Modification Indices for THETA-DELTA										
	A1	A2	A3	A4	A5	C1	C2	C3	C4	C5
	----	----	----	----	----	----	----	----	----	----
A1	- -									
A2	0.91	- -								
A3	0.74	1.25	- -							
A4	6.49	14.55	17.23	- -						
A5	8.44	5.13	17.90	1.21	- -					
C1	4.19	2.05	0.21	1.38	0.57	- -				
C2	0.05	1.53	0.25	0.02	0.03	1.35	- -			
C3	0.01	0.41	0.56	0.34	0.16	11.58	0.01	- -		
C4	0.34	4.90	0.00	0.85	3.90	1.56	0.25	7.63	- -	
C5	2.90	0.05	0.03	1.73	1.13	9.73	0.00	- -	3.87	- -

Tabla 14 (Cont.). Solución estandarizada

Standardized Solution		
LAMBDA-X		
	AFECT	COGNI
	-----	-----
A1	40.65	- -
A2	39.66	- -
A3	34.03	- -
A4	39.18	- -
A5	43.24	- -
C1	- -	21.40
C2	- -	23.86
C3	- -	28.46
C4	- -	25.10
C5	- -	24.11
PHI		
	AFECT	COGNI
	-----	-----
AFECT	1.00	
COGNI	0.81	1.00

La diferencia entre la solución estandarizada (*Standardized Solution*) y la completamente estandarizada (*Completely Standardized Solution*) estriba en que en la primera únicamente se estandarizan las variables latentes (las manifiestas siguen conservando su métrica original), en tanto que en la segunda se estandarizan tanto las variables latentes como las manifiestas. A la vista de la solución 'completamente estandarizada' podemos obtener una serie de conclusiones con respecto al modelo: en primer lugar, hemos conseguido estimar el modelo sin problemas (debidos a errores de sintaxis, errores en el archivo de datos –e.g., matriz singular–, incompatibilidad entre el modelo y los datos, etc.); en segundo lugar, la revisión de los parámetros estimados revela que no existen estimaciones impropias o no razonables; en tercer lugar, todos los parámetros son significativamente distintos de cero como denotan los valores de *t* obtenidos; en cuarto lugar, los valores de  $R^2$  son relativamente altos en todos los casos (el rango va de .73 a .91), lo que nos sugiere que los indicadores son medidas razonablemente buenas de las variables latentes; por último, la correlación entre los factores es positiva y moderadamente elevada, en consonancia con la teoría previa.

Tabla 14 (Cont.). Solución completamente estandarizada

Completely Standardized Solution					
LAMBDA-X					
	AFFECT	COGNI			
	-----	-----	-----	-----	-----
A1	0.90	-	-		
A2	0.90	-	-		
A3	0.94	-	-		
A4	0.87	-	-		
A5	0.95	-	-		
C1	-	-	0.85		
C2	-	-	0.88		
C3	-	-	0.86		
C4	-	-	0.86		
C5	-	-	0.89		
PHI					
	AFFECT	COGNI			
	-----	-----	-----	-----	-----
AFFECT	1.00				
COGNI	0.81	1.00			
THETA-DELTA					
	A1	A2	A3	A4	A5
	-----	-----	-----	-----	-----
	0.19	0.19	0.12	0.25	0.09
	C1	C2	C3	C4	C5
	-----	-----	-----	-----	-----
	0.27	0.23	0.26	0.26	0.20

Llevaremos a cabo, como complemento a los análisis efectuados en SIMPLIS y LISREL, la evaluación de la fiabilidad y validez de los indicadores y los constructos. La *fiabilidad de los indicadores* equivale al cuadrado de la correlación entre el indicador y la variable latente. En otras palabras, la fiabilidad indica la proporción de variación del indicador que es explicada por el factor o variable latente que se supone que mide (vid. Tabla 15). Un valor de  $R^2$  elevado indica que el indicador es fiable. Estos valores son proporcionados directamente por LISREL (vid. valores  $R^2$  en Tabla 14). En el modelo que comentamos, A5 es el indicador más fiable es y C1 el menos fiable.

Tabla 15. Fiabilidad de los indicadores

A1	.900	.810
A2	.902	.814
A3	.939	.882
A4	.869	.755
A5	.952	.906
C1	.852	.726
C2	.876	.767
C3	.863	.745
C4	.858	.736
C5	.893	.797

Además de la fiabilidad de los indicadores individuales, puede calcularse la *fiabilidad compuesta* de cada variable latente ('fiabilidad de constructo'). Para ello utilizaremos la infor-

mación de los coeficientes y varianzas de error de la tabla *Completely Standardized Solution* y aplicaremos la siguiente fórmula (LISREL no ofrece el cálculo directo de esta clase de fiabilidad):

$$\rho_c = \left( \sum_{i=1}^n \lambda_i \right)^2 / \left[ \left( \sum_{i=1}^n \lambda_i \right)^2 + \left( \sum_{i=1}^n \delta_i \right) \right]$$

donde  $\rho_c$  es la fiabilidad compuesta,  $\lambda_i$  las saturaciones estandarizadas y  $\delta_i$  las varianzas de error asociadas con las variables observadas individuales (i.e., varianzas de  $\delta$  o  $\epsilon$ ). La fiabilidad compuesta es similar al coeficiente  $\alpha$  de Cronbach, e indica la consistencia interna de los indicadores dentro de un mismo factor. En el ejemplo, la fiabilidad compuesta de la subescala afectiva es igual a

$$20.182/(20.182+0.381)=0.962$$

la fiabilidad compuesta de la subescala cognitiva es igual a

$$18.853/(18.853+1.227)=0.939$$

la fiabilidad compuesta de la escala total equivale a

$$79.282/(79.281+18.853)=0.975$$

Puesto que en todos los casos  $\rho_c$  es superior a 0.60, podemos concluir que los indicadores de cada subescala, considerados en conjunto, son una medida fiable del constructo.

Una medida complementaria de la fiabilidad compuesta es la *varianza media extractada* (average variance extracted), que se calcula mediante la fórmula

$$\rho_v = \sum_{i=1}^n \lambda_i^2 / \left[ \sum_{i=1}^n \lambda_i^2 + \left( \sum_{i=1}^n \delta_i \right) \right] = \sum_{i=1}^n \lambda_i^2 / n$$

donde  $\rho_v$  es la varianza media extractada,  $\lambda_i^2$  las saturaciones estandarizadas al cuadrado y  $n$  el número de ítems de cada subescala. Una vez realizados los cálculos (lo que puede hacerse fácilmente en una hoja de cálculo a partir de los datos ofrecidos en la solución LISREL), tenemos las siguientes varianzas medias extractadas: para la subescala afectiva:  $4.167/5=0.834$ , lo que indica que el 83% de la varianza del factor afectivo ha sido capturada por el constructo, en tanto que únicamente el 17% ( $1-.83=.17$ ) es debido al error de medida; la subescala afectiva presenta una VME algo más baja:  $3.772/5=0.755$ , y la escala total:  $7.938/5=0.794$ .

Dado que  $\rho_v > 0.50$  en todos los casos, podemos concluir que una cantidad sustancial de la varianza de los indicadores es capturada por cada constructo comparada con la que es capturada por el error de medida. Esto proporciona confianza adicional en la operacionalización de las variables latentes. En resumen, la evaluación del componente de medida del modelo hasta el momento realizada, revela evidencia de fiabilidad y validez de las operacionalizaciones de las variables latentes.

La *validez convergente* se evalúa en los modelos AFC revisando los valores de  $t$  correspondientes a las saturaciones factoriales. Si todas las saturaciones de los indicadores que evalúan el mismo constructo fueran estadísticamente significativas (i.e., superiores al doble de sus errores estándar o inferiores a los umbrales críticos de los distintos niveles de significación a posteriori), dispondríamos de evidencia a favor de la validez convergente de los indicadores, en la medida en que valores significativos de  $t$  indican que, efectivamente, todos los indicadores evalúan el mismo constructo. Los valores de  $t$  obtenidos en la subescala afectiva son 21.099, 23.566, 19.213 y 24.579; en la subescala cognitiva, 17.055, 16.604, 16.438 y

17.674. Como se ve, todos ellos superan ampliamente el valor crítico de  $t=1.96$  con  $p<.05$ ,  $t=2.58$  con  $p<.01$ , o  $t=3.29$  con  $p<.001$ .

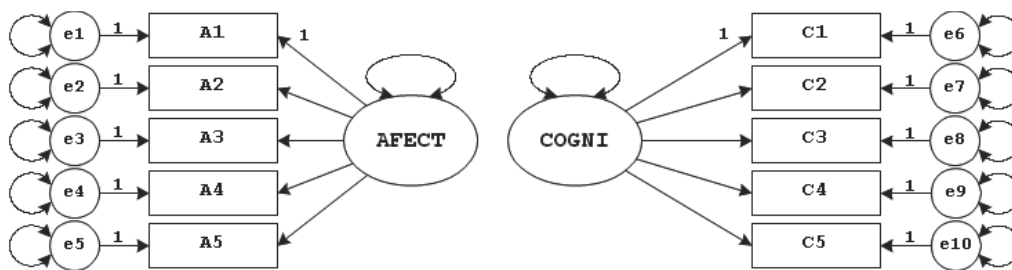
En lo que atañe a la *validez discriminante*, podría determinarse en la medida en que la varianza extractada de cada variable latente fuera superior al cuadrado de la correlación entre ellas (Hair et al., 2006). Esto es lo que sucede en el ejemplo, toda vez que  $r_{AC}^2 = .656$ , claramente inferior a  $VME_A = .834$  y  $VME_C = .755$  como más arriba se indicó.

Cabría determinar, finalmente, la *validez nomológica* (examinando si las correlaciones entre las variables latentes tienen sentido en el ámbito de la teoría en que se inscribe la investigación) y la *validez aparente* (que, en todo caso, debería establecerse con anterioridad a los propios análisis).

### C. Modelo de dos factores no correlacionados

El modelo que se somete a prueba a continuación es idéntico al comentado en párrafos anteriores, con la diferencia de que en esta ocasión los factores no están correlacionados (vid. Figura 9).

Figura 9. Modelo AFC de dos factores no correlacionados



Mostramos en la Tabla 16 la sintaxis en LISREL para el análisis de este modelo. Como se ve, es idéntica a la del modelo anterior, con la única diferencia de que se especifica el valor 0 en la matriz  $\Phi$ . Tenemos en este caso un parámetro menos a estimar (i.e., la correlación entre los dos factores) y, en consecuencia, un grado de libertad más ( $GL=35$ ) que en el modelo anterior.



Tabla 16. Sintaxis (LISREL) del modelo de dos factores no correlacionados

```

TITLE MODELO B2 WILLIAMS EAVES & COX
DA NI=10 NO=216 MA=CM
LA
A1 A2 A3 A4 A5 C1 C2 C3 C4 C5
CM
[ESCRIBIR AQUÍ LA MATRIZ DE COVARIANZAS, VID. TABLA 6]
MO NX=10 NK=2 PH=SY,FR LX=FU,FR TD=SY,FR
LK
AFECT COGNI
PA LX
0 0
1 0
1 0
1 0
1 0
0 0
0 1
0 1
0 1
0 1
0 1
VA 1.0 LX(1,1) LX(6,2)
PA TD
1
0 1
0 0 1
0 0 0 1
0 0 0 0 1
0 0 0 0 0 1
0 0 0 0 0 0 1
0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 0 1
PA PH
1
0 1
OU ME=ML RS MI SC ND=2

```

Los resultados (Tabla 17 y Figura 10) denotan claramente un mal ajuste que conduce al rechazo de la hipótesis de que el modelo se ajusta a los datos observados.

Tabla 17. Resumen de resultados del modelo de dos factores no correlacionados

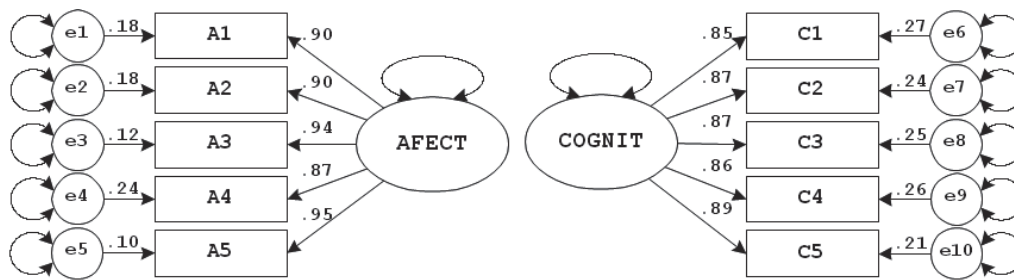
```

Goodness of Fit Statistics

Degrees of Freedom = 35
Minimum Fit Function Chi-Square = 283.34 (P = 0.0)
Normal Theory Weighted Least Squares Chi-Square = 207.99 (P = 0.0)
. . .
Root Mean Square Error of Approximation (RMSEA) = 0.15
90 Percent Confidence Interval for RMSEA = (0.13 ; 0.17)
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.00
. . .
Normed Fit Index (NFI) = 0.94
Non-Normed Fit Index (NNFI) = 0.93
Parsimony Normed Fit Index (PNFI) = 0.73
Comparative Fit Index (CFI) = 0.95
Incremental Fit Index (IFI) = 0.95
Relative Fit Index (RFI) = 0.93
. . .
Standardized RMR = 0.43
Goodness of Fit Index (GFI) = 0.84

```

Figura 10. Modelo AFC de dos factores no correlacionados



*D. Modelo jerárquico (dos factores de primer orden y un factor de segundo orden)*

El último de los modelos sometidos a prueba (Figura 11) hipotetiza que los factores afectivo y cognitivo están causados por un factor de segundo orden (arousal). Los parámetros a estimar en este caso son 21 y las observaciones 55, por lo que tenemos un modelo supraidentificado con 34 grados de libertad. El modelo consta en total de 25 variables (10 manifiestas y 15 latentes; 13 exógenas y 12 endógenas). En la Tabla 18 se ofrece la sintaxis (en SIMPLIS) para analizar el modelo.

Figura 11. Modelo AFC de dos factores no correlacionados

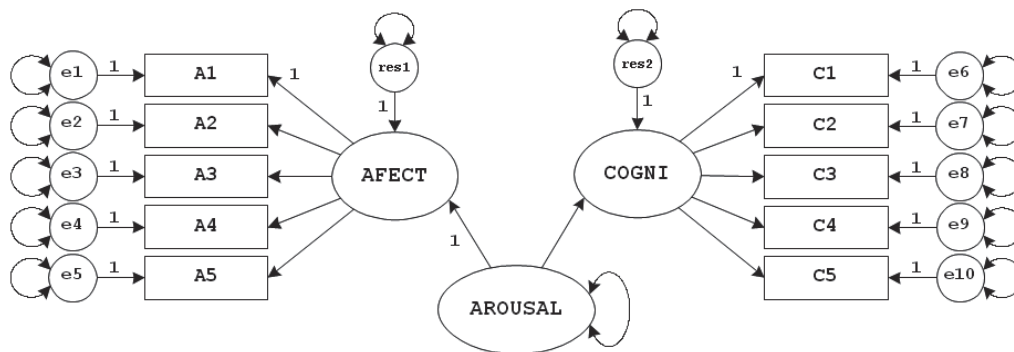


Tabla 18. Sintaxis (en SIMPLIS) del modelo de dos factores de primer orden y un factor de segundo orden

```

!Matriz de covarianzas Williams Eaves & Cox, modelo C (Jerárquico)
Observed Variables
A1 A2 A3 A4 A5 C1 C2 C3 C4 C5
Latent Variables
AFECT COGNI AROUSAL
Covariance Matrix
[ESCRIBIR AQUÍ LA MATRIZ DE COVARIANZAS, VID. TABLA 6]
Sample Size is 216
Relationships
A1 = 1*AFECT
A2 = AFECT
A3 = AFECT
A4 = AFECT
A5 = AFECT
C1 = 1*COGNI
C2 = COGNI
C3 = COGNI
C4 = COGNI
C5 = COGNI
AFECT = 1*AROUSAL
COGNI = AROUSAL
Set the error variance of AFECT to 1
OPTIONS wp rs ef ss sc mi nd=2
Path Diagram
End of Program

```

En la Tabla 19 y la Figura 12 se presentan, respectivamente, una síntesis de los resultados y la configuración final del modelo.

Tabla 19. Resumen de resultados del modelo jerárquico

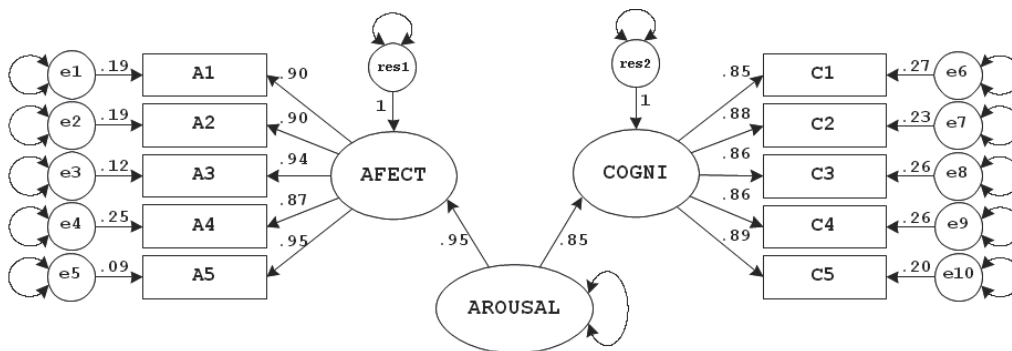
```

Goodness of Fit Statistics

Degrees of Freedom = 34
Minimum Fit Function Chi-Square = 88.47 (P = 0.00)
Normal Theory Weighted Least Squares Chi-Square = 85.11 (P = 0.00)
. . .
Root Mean Square Error of Approximation (RMSEA) = 0.084
90 Percent Confidence Interval for RMSEA = (0.062 ; 0.11)
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.0076
. . .
Normed Fit Index (NFI) = 0.98
Non-Normed Fit Index (NNFI) = 0.99
Parsimony Normed Fit Index (PNFI) = 0.74
Comparative Fit Index (CFI) = 0.99
Incremental Fit Index (IFI) = 0.99
Relative Fit Index (RFI) = 0.98
. . .
Standardized RMR = 0.022
Goodness of Fit Index (GFI) = 0.93

```

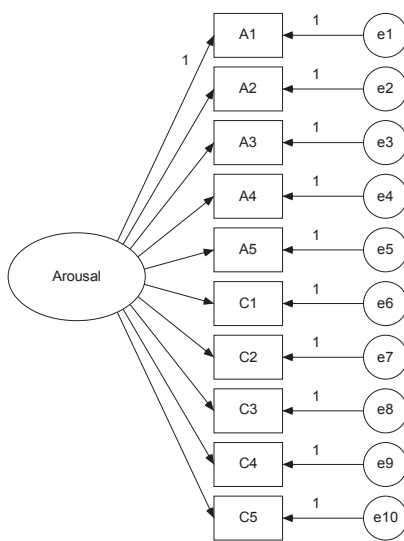
Figura 12. Modelo AFC de dos factores de primer orden y uno de segundo orden



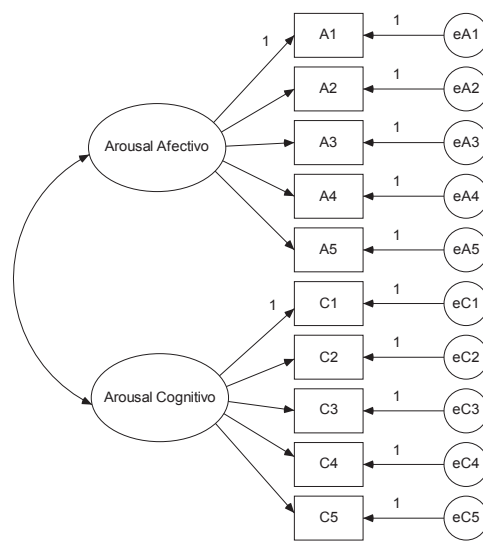
ANÁLISIS CON AMOS Y SAS

En los Apéndices A y B incluimos, respectivamente, una muestra del procedimiento en AMOS y SAS para llevar a cabo los análisis explicados en este trabajo.

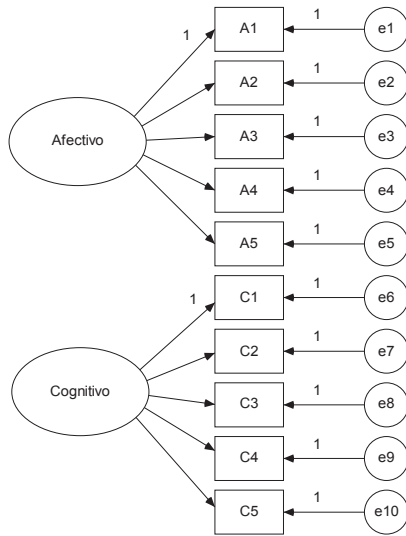
Apéndice A. Solución en AMOS



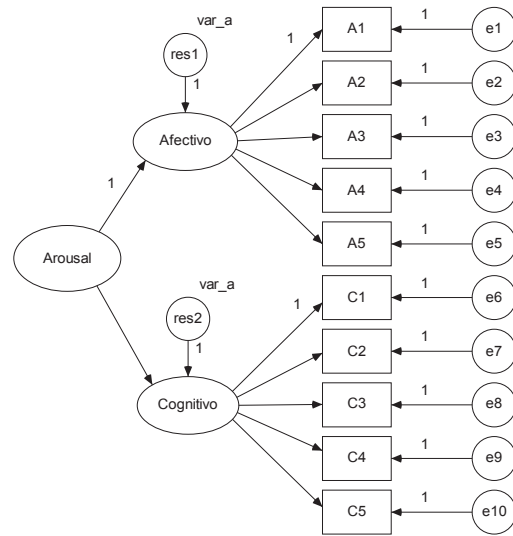
Modelo A



Modelo B1



Modelo B2



Modelo C

rowtype_	varname_	A1	A2	A3	A4	A5	C1	C2	C3	C4	C5	
1	n	216,000	216,000	216,000	216,000	216,000	216,000	216,000	216,000	216,000	216,000	
2	cov	A1	2038,035	.	.	.	.	.	.	.	.	
3	cov	A2	1631,766	1932,163	.	.	.	.	.	.	.	
4	cov	A3	1393,567	1336,871	1313,809	.	.	.	.	.	.	
5	cov	A4	1657,299	1647,164	1273,261	2034,216	.	.	.	.	.	
6	cov	A5	1721,830	1688,292	1498,401	1677,767	2061,875	.	.	.	.	
7	cov	C1	659,795	702,969	585,019	656,527	775,118	630,518	.	.	.	
8	cov	C2	779,519	790,448	653,734	764,755	871,211	500,128	741,767	.	.	
9	cov	C3	899,912	879,179	750,037	897,441	1008,287	648,935	679,970	1087,409	.	
10	cov	C4	775,235	739,157	659,867	751,860	895,616	550,476	603,950	677,865	855,272	
11	cov	C5	821,170	785,419	669,951	802,825	877,681	491,042	574,775	686,391	622,830	728,674

SINTAXIS AMOS (en VisualBasic) del modelo A (unidimensional)<sup>21</sup>

```

#Region "Header"
Imports System
Imports System.Diagnostics
Imports Microsoft.VisualBasic
Imports AmosEngineLib
Imports AmosGraphics
Imports AmosEngineLib.AmosEngine.TMatrixID
Imports PBayes
#End Region
Module MainModule
' AFC WILLIAMS EAVES & COX MODELO A
'
Sub Main ()
Dim sem As New AmosEngine

    Try
        Sem.TextOutput ()
        Sem.Standardized ()
        Sem.Smc ()
        Sem.AllImpliedMoments ()
        Sem.FactorScoreWeights ()
        Sem.TotalEffects ()

        Sem.BeginGroup (AmosEngine.AmosDir & "Examples\williams_eaves.sav")

sem.AStructure ("A1 = (1) AROUSAL + (1) e1")
sem.AStructure ("A2 = AROUSAL + (1) e2")
sem.AStructure ("A3 = AROUSAL + (1) e3")
sem.AStructure ("A4 = AROUSAL + (1) e4")
sem.AStructure ("A5 = AROUSAL + (1) e5")
sem.AStructure ("C1 = AROUSAL + (1) e6")
sem.AStructure ("C2 = AROUSAL + (1) e7")
sem.AStructure ("C3 = AROUSAL + (1) e8")
sem.AStructure ("C4 = AROUSAL + (1) e9")
sem.AStructure ("C5 = AROUSAL + (1) e10")

        Sem.FitModel ()
    Finally
        Sem.Dispose ()
    End Try
End Sub
End Module

```

21 Se presenta únicamente, a título de ejemplo, la sintaxis en AMOS-VB de no de los modelos.

## Apéndice B. Solución en SAS

```

*/ WILLIAMS EAVES MODELO B1 DOS FACTORES CORRELACIONADOS /*;
DATA D1 (TYPE=COV);
  INPUT _TYPE_ $ _NAME_ $ A1 A2 A3 A4 A5 C1 C2 C3 C4 C5;
CARDS;
N      . 216      216      216      216      216      216      216      216      216
216
COV A1  2038.035      .      .      .      .      .      .      .      .
.
COV A2  1631.766 1932.163      .      .      .      .      .      .      .
.
COV A3  1393.567 1336.871 1313.809      .      .      .      .      .      .
.
COV A4  1657.299 1647.164 1273.261 2034.216      .      .      .      .      .
.
COV A5  1721.830 1688.292 1498.401 1677.767 2061.875      .      .      .      .
.
COV C1  659.795 702.969 585.019 656.527 775.118 630.518      .      .      .
.
COV C2  779.519 790.448 653.734 764.755 871.211 500.128 741.767      .      .
.
COV C3  899.912 879.179 750.037 897.441 1008.287 648.935 679.970 1087.409      .
.
COV C4  775.235 739.157 659.867 751.860 895.616 550.476 603.950 677.865
855.272      .
COV C5  821.170 785.419 669.951 802.825 877.681 491.042 574.775 686.391
622.830 728.674
;
PROC CALIS COVARIANCE CORR RESIDUAL MODIFICATION ;
LINEQS
  A1 = LV1F1 F1 + E1,
  A2 = LV2F1 F1 + E2,
  A3 = LV3F1 F1 + E3,
  A4 = LV4F1 F1 + E4,
  A5 = LV5F1 F1 + E5,
  C1 = LV6F2 F2 + E6,
  C2 = LV7F2 F2 + E7,
  C3 = LV8F2 F2 + E8,
  C4 = LV9F2 F2 + E9,
  C5 = LV10F2 F2 + E10;
STD
  F1 = 1,
  F2 = 1,
  E1-E10 = VARE1-VARE10;
COV
  F1 F2 = CF1F2;
VAR A1 A2 A3 A4 A5 C1 C2 C3 C4 C5;
RUN;

```

## REFERENCIAS BIBLIOGRÁFICAS

- Arbuckle, J. L. (2000). *Exploratory structural equation modeling*. Fordham University. Department of Psychology colloquium series.
- Bandalos, D. L. (2002). The effects of ítem parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 9, 78-102.
- Bandalos, D. L. & Finney, S. J. (2001). Ítem parceling issues in structural equation modeling. En G. A. Marcoulides & R. E. Schumacker (Eds.), *New development and techniques in structural equation modeling* (pp. 269-296). Mahwah, NJ: Lawrence Erlbaum.
- Bentler, P. M. (1995). *Structural Equations Program Manual*. Encino, CA: Multivariate Software, Inc.
- Bentler, P. M., & Chu, C. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, 16, 78-117.
- Boomsma, A. (2000). Reporting Analyses of Covariance Structures. *Structural Equation Modeling: A Multidisciplinary Journal*, 7(3), 461-483.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: The Guilford Press.
- Breckler, S. J. (1990). Applications of covariance structure modeling in psychology: Cause for concern? *Psychological Bulletin*, 107, 260-263.
- Cribbie, R. A. (2007). Multiplicity Control in Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 14, 98-112.
- Diamantopoulos, A., & Siguaw, J. A. (2000). *Introducing LISREL*. Thousand Oaks, CA: Sage.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (6ª ed.). Upper Saddle River, NJ: Pearson-Prentice Hall.
- Hancock, G. R. (1999). A sequential Scheffé-type respecification procedure for controlling Type I error in exploratory structural equation model modification. *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 158-168.
- Hatcher, L. (2006). *A step-by-step approach to using SAS for factor analysis and structural equation modeling*. Cary, NJ: The SAS Institute.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 1-55.
- Jackson, D. L. (2007). The effect of the number of observations per parameter in misspecified confirmatory factor analytic models. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(1), 48-76.
- Jöreskog, K. G. (1993). Testing structural equation models. En K.A. Bollen & J.S. Lang (Eds.), *Testing structural equation models* (pp. 294-316). Newbury Park, CA: Sage.
- Jöreskog, K. G., & Sörbom, D. (1996a). *LISREL 8: User's Reference Guide*. Chicago, IL: Scientific Software International.
- Jöreskog, K. G., & Sörbom, D. (1996b). *LISREL 8: Structural Equation Modeling with the SIMPLIS command Language*. Chicago, IL: Scientific Software International.
- Kenny, D. A. (1979). *Correlation and causality*. New York: John Wiley and Sons.
- Kishton, J. M., & Widaman, K. F. (1994). Unidimensional versus domain representative parceling of questionnaire ítems: An empirical example. *Educational and Psychological Measurement*, 54, 757-765.



- Kline, R. B. (2005). *Principles and practice of structural equation modeling*. (2 ed.) New York: The Guilford Press.
- Little, T. D., Cunningham, W. A., Shahar, G., & Widaman K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, 9, 151-173.
- Loehlin, J. C. (2004). *Latent variable models: An introduction to factor, path, and structural equation analysis*. (4<sup>a</sup> ed.) Mahwah, NJ: Lawrence Erlbaum Associates.
- Lomax, G. R. (1982). A guide to LISREL-type structural equation modeling. *Behavior Research Methods & Instrumentation*, 14, 1-8.
- McDonald, R. P., & Ho, M. R. (2002). Principles and Practice in Reporting Structural Equation Analyses. *Psychological Methods*, 7(1), 64–82.
- McKnight, P. E., McKnight, K. M., Sidaniy, S. y Figueredo, A. J. (2007). *Missing Data: A Gentle Introduction (Methodology In The Social Sciences)*. New York: Guilford.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3<sup>a</sup> ed.). New York: McGraw-Hill.
- SAS Institute (2006). *SAS*, v. 9.1.3. [Programa informático]. Cary, NC: SAS Institute Inc.
- Schumaker, R. E., & Marcoulides, G. A. (Eds.). (1998). *Interaction and nonlinear effects in structural equation modeling*. Mahwah, NJ: Erlbaum.
- Scientific Software International (2006). *LISREL*, v. 8.8 [Programa informático]. Lincolnwood, IL: Scientific Software International.
- SPSS (2006). *AMOS*, v. 6.0 [Programa informático]. Chicago, IL: SPSS, Inc.
- SPSS (2006). *SPSS*, V. 14.0 [Programa informático]. Chicago, IL: SPSS, Inc.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4<sup>a</sup> ed.). Boston: Allyn & Bacon.
- West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. En R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56-75). Thousand Oaks, CA: Sage.
- Williams, T. O., Eaves, R. C. y Cox, C. (2002). Confirmatory factor analysis of an instrument designed to measure affective and cognitive arousal. *Educational and Psychological Measurement*, 62(2), 264-283.

# META-ANÁLISIS DE LA INVESTIGACIÓN\*

JULIO SÁNCHEZ MECA  
*Departamento de Psicología Básica y Metodología*  
*Facultad de Psicología (Universidad de Murcia)*

## INTRODUCCIÓN

Durante la segunda mitad del siglo pasado se ha producido una explosión de información que ha provocado la necesidad de desarrollar métodos objetivos y sistemáticos para la acumulación del conocimiento científico en cualquier ámbito de investigación. El meta-análisis ha surgido como una metodología capaz de integrar cuantitativamente los resultados de las investigaciones sobre un determinado tema para poder establecer qué es lo que la evidencia empírica, hasta ese momento, ha demostrado. La revisión cuantitativa de la investigación se ha convertido en una tarea imprescindible entre el quehacer científico del pasado y del futuro, para orientar y dirigir nuevas investigaciones. Pero el meta-análisis es más que un mero método de revisión de la investigación. El meta-análisis comporta un nuevo modo de entender el significado y el análisis de los datos, con su énfasis en el tamaño del efecto frente a las pruebas de significación; es un nuevo enfoque que afecta a la interpretación habitual de qué entendemos por descubrimiento científico (Glass, McGaw y Smith, 1981; Schmidt, 1992, 1996; Schmidt y Hunter, 1995).

La realización de un meta-análisis es un proceso de investigación que requiere el cumplimiento de las normas propias del método científico: objetividad, sistematización y replicabilidad. Ésta es la principal diferencia con respecto a las revisiones tradicionales de la investigación. Frente a las revisiones narrativas, también denominadas revisiones cualitativas o subjetivas, el meta-análisis propugna el mismo rigor científico que se exige en las investigaciones primarias. En un meta-análisis no se prejuzgan los resultados de las investigaciones, sino que se cuantifican y se analizan conjuntamente. Así mismo, un meta-análisis tiene que garantizar su replicabilidad por otros investigadores, por lo que todas las decisiones en el proceso de revisión cuantitativa tienen que hacerse explícitas (Rosenthal, 1995).

La característica que mejor identifica al meta-análisis es el uso de los métodos estadísticos para integrar cuantitativamente los resultados de los estudios. Para ello, se requiere definir una medida de los resultados que sea homogénea a lo largo de todos los estudios. El índice conocido como "tamaño del efecto" (*effect size*) representa la magnitud en que se manifiesta el fenómeno en cuestión en cada estudio empírico (Glass et al., 1981). Hoy día existen numerosos textos que exponen cómo llevar a cabo estudios meta-analíticos, cómo

\* Trabajo financiado con un Proyecto del Fondo de Investigación Sanitaria en la convocatoria de Evaluación de Tecnologías Sanitarias (Expediente N° PI07/90384)

calcular el tamaño del efecto de cada estudio y cómo aplicar técnicas de análisis estadístico específicamente diseñadas para ser aplicadas en meta-análisis, y todo ello dirigido a diferentes ámbitos científicos, tales como las Ciencias Sociales y del Comportamiento (Botella y Gambara, 2002; Cooper, 1998; Cooper y Hedges, 1994; Glass et al., 1981; Gómez, 1987; Hedges y Olkin, 1985; Hunter y Schmidt, 2004; Lipsey y Wilson, 2001; Petticrew y Roberts, 2006; Rosenthal, 1991; Schulze, 2004; Wolf, 1986) y las Ciencias de la Salud (Davey Smith, 2001; Eddy, Hasselblad y Shachter, 1995; Egger, Smith y Altman, 2001; Martín, Tobías y Seoane, 2006; Sutton, Abrams, Jones et al., 2000; Whitehead, 2002).

## FASES DE UN META-ANÁLISIS

A lo largo del proceso de realización de un meta-análisis se suceden varias etapas, cada una de las cuales pretende aportar objetividad y sistematización al proceso de revisión. Las principales fases de un meta-análisis pueden resumirse en (Lipsey y Wilson, 2001; Marín-Martínez, Sánchez-Meca, Huedo y Fernández-Guzmán, 2007; Roberts, Kuncel, Viechtbauer y Bogg, 2007; Sánchez-Meca, 1999; Sánchez-Meca y Ato, 1989): (a) formulación del problema, (b) búsqueda de la literatura, (c) codificación de los estudios, (d) análisis estadístico y (e) publicación del estudio.

*1º Formulación del problema.* Como en cualquier investigación, el primer paso que debe dar el meta-analista es definir el objeto del meta-análisis, que generalmente será estudiar la magnitud y sentido de la relación entre dos (o más) variables o conceptos. En esta fase se precisa definir los conceptos no sólo de forma teórica, sino también las variables que se aceptarán como operativizaciones de ellas. Las definiciones operativas de los conceptos constituyen un importante criterio para la inclusión de los estudios en el meta-análisis. Así mismo, hay que distinguir entre las variables fundamentales de la investigación (aquéllas cuya relación se pretende estudiar) y variables potencialmente moderadoras de tal relación. En esta fase deben también plantearse las hipótesis que se pretenden poner a prueba.

*2º Búsqueda de la literatura.* Una vez formulado el problema de investigación, el paso siguiente consiste en realizar una búsqueda de la literatura lo más completa y exhaustiva posible. Para ello, es preciso especificar los criterios que deben cumplir los estudios empíricos para que sean incluidos en el meta-análisis, tales como el rango temporal en el que se realizaron o publicaron o el tipo de información estadística que deben aportar para poder realizar posteriormente los cálculos del tamaño del efecto.

En esta fase se recomienda utilizar tanto procedimientos formales como informales de búsqueda. Dentro de los procedimientos formales destacan las búsquedas por computador, que permiten acceder con gran economía de tiempo y recursos a las bases de datos informatizadas mediante el uso de palabras clave y otros descriptores. Pero estos procedimientos formales deben completarse con el uso de otros sistemas de búsqueda no sistemáticos que permiten localizar la literatura "fugitiva", es decir, trabajos no publicados o de muy reducida difusión, pero que pueden aportar datos relevantes al meta-análisis.

Por muy completa que sea la búsqueda de los estudios siempre cabe la posibilidad de que no seamos capaces de recoger todos los existentes, en especial aquéllos que no han sido publicados, con la consiguiente amenaza que el sesgo de publicación contra los resultados nulos puede ejercer en el meta-análisis. Siempre es recomendable, pues, realizar un estudio de la tolerancia a los resultados nulos para poder desechar el sesgo de publicación como una posible amenaza contra la validez de las conclusiones del meta-análisis (Begg, 1994; Burdett, Stewart y Tierney, 2003; Rosenthal, 1991; Rothstein, Sutton y Borenstein, 2005; Sutton, Duval, Tweedie et al., 2000).

3° *Codificación de las variables.* Por regla general, la realización de un meta-análisis se justifica por la existencia de resultados contradictorios en un determinado problema de investigación. Para determinar las razones de tales inconsistencias cada estudio seleccionado para el meta-análisis es revisado en profundidad y sometido a un protocolo de registro de aquellas variables que, en teoría, podrían estar afectando a la heterogeneidad encontrada en los resultados de los estudios. Este protocolo incluye variables moderadoras de tres tipos: sustantivas, metodológicas y extrínsecas.

- (a) Las variables *sustantivas* son aquéllas propias del objeto de investigación e incluyen aspectos tales como las características demográficas de las muestras de sujetos sometidas a estudio, cómo se operativizaron las variables, el contexto social, cultural, geográfico, económico, etc. en el que se hizo la investigación, etc.
- (b) Por variables *metodológicas* se entienden aquéllas que se refieren al método y diseño de la investigación y, por tanto, se repiten en todos los meta-análisis. Son metodológicas características tales como el tamaño de la muestra, la mortalidad experimental, el tipo de diseño o la calidad del diseño de investigación.
- (c) Las variables *extrínsecas* se caracterizan por ser externas al propio desarrollo de la investigación, aunque en ocasiones pueden afectar a los resultados de los estudios y poner en evidencia la existencia de deficiencias, sesgos o artefactos en un campo de estudio. Son ejemplos de este tipo de características, el estatus de publicación del estudio (publicado versus no publicado), la filiación de los autores del estudio o el género de los autores.

Además del registro de las variables moderadoras, es preciso definir y calcular el tamaño del efecto, que es un índice cuantitativo que resume la magnitud de la relación encontrada en cada estudio. Dependiendo de que el tipo de diseño habitual en los estudios integrados en el meta-análisis sea experimental o correlacional, se distinguen dos familias de índices del tamaño del efecto: las familias *d* y *r*, la primera para los estudios con asignación de los sujetos a grupos o condiciones experimentales, y la segunda para diseños de tipo correlacional (Cooper, 1998; Cortina y Nouri, 2000; Grissom y Kim, 2005; Lipsey y Wilson, 2001; Rosenthal, 1991; Rosenthal, Rosnow y Rubin, 2000).

La principal amenaza contra la validez del proceso de registro de las variables moderadoras es la falta de fiabilidad en el proceso de codificación. Es por ello que se recomienda la realización de un estudio de la fiabilidad de este proceso, para lo cual dos (o más) investigadores deben codificar de forma independiente los mismos estudios y, posteriormente, analizar la fiabilidad mediante el cálculo de índices de acuerdo inter-codificadores.

4° *Análisis estadístico.* Una vez que se dispone de una estimación del tamaño del efecto de cada estudio y de las variables moderadoras registradas en cada uno de ellos, el análisis estadístico de estos datos implica tomar el tamaño del efecto como la variable dependiente y las variables moderadoras como potenciales factores explicativos de aquél. Las preguntas a las que puede responder un meta-análisis son:

- (a) ¿Cuál es la magnitud del efecto medio?
- (b) ¿Es estadísticamente significativo el tamaño del efecto medio?
- (c) ¿Es el tamaño del efecto medio representativo de todos los estudios individuales? O lo que es lo mismo, ¿existe homogeneidad en torno al tamaño del efecto medio?
- (d) Si no existe homogeneidad, ¿qué características de los estudios pueden estar moderando el tamaño del efecto? ¿Puede proponerse un modelo explicativo?

*5º Publicación del estudio.* Para lograr el carácter replicable del meta-análisis, es preciso que la publicación del mismo siga las mismas normas que la publicación de investigaciones primarias (Botella y Gambara, 2006; Cooper, 1998; Rosenthal, 1995). La presentación de un meta-análisis debe, pues, incluir una introducción en la que se especifique el objetivo del estudio, las definiciones conceptuales y operativas de las variables y las hipótesis que se pretenden probar. En la sección método se incluye el proceso de búsqueda de los estudios incluidos en el meta-análisis, la codificación de las variables y de los tamaños del efecto y la especificación de las técnicas estadísticas que se han aplicado. En la sección de resultados se presentan y discuten los resultados de los estadísticos y contrastes aplicados, finalizando con las conclusiones, que deben incorporar directrices para la investigación futura. En la publicación debe ofrecerse al lector todas las facilidades para disponer de toda la base de datos meta-analítica, con objeto de posibilitar su replicación por otros investigadores.

## ÍNDICES DEL TAMAÑO DEL EFECTO

Para poder integrar los resultados de múltiples estudios o investigaciones primarias acerca de una misma temática, es preciso que vengan expresados en una métrica común que los haga directamente comparables. En Ciencias Sociales es muy frecuente que, para la medición de un mismo concepto, se empleen muy diversas definiciones operativas, con diferentes indicadores y escalas de medida. Desde el meta-análisis se ha propuesto el tamaño del efecto como el indicador idóneo para representar el resultado de una investigación, de forma que sea comparable para diferentes escalas y operacionalizaciones de las variables implicadas.<sup>1</sup> Son múltiples los índices del tamaño del efecto que se pueden utilizar para resumir cuantitativamente la magnitud de la relación entre las variables. En meta-análisis estos índices suelen agruparse en dos familias: Las familias *d* y *r*.

• **La familia *d*.** Cuando los diseños de los estudios implican asignación de sujetos a diferentes grupos o condiciones experimentales (generalmente, grupos experimental y control), el índice del tamaño del efecto más adecuado es la “diferencia media tipificada”, que se define como la diferencia entre las medias de los grupos experimental y control dividida por la desviación típica intra-grupo (Lipsey y Wilson, 2001; Shadish y Haddock, 1994):

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

donde  $\bar{y}_E$  y  $\bar{y}_C$  son las medias de los grupos experimental y control, respectivamente;  $S$  es la desviación típica intra-grupo, que se obtiene mediante:<sup>2</sup>

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

1 Aunque la corriente meta-analítica preconizada por Rosenthal proponía como resultado de los estudios el nivel crítico de probabilidad,  $p$ , obtenido en la prueba de significación, actualmente no se utiliza este índice debido a su dependencia del tamaño de la muestra.

2 Glass et al. (1981) propusieron dividir la diferencia entre las medias por la desviación típica del grupo de control,  $S_C$  (cuando exista), en lugar de dividir por  $S$ , debido a que, en ocasiones, la aplicación de un tratamiento puede alterar la variabilidad natural de los sujetos en la variable dependiente, como resultado de un efecto interactivo sujeto x tratamiento. Pero siempre que se cumpla el supuesto de homogeneidad de varianzas, es más eficiente utilizar  $S$  (Hedges y Olkin, 1985).

siendo  $n_E$  y  $n_C$  los tamaños muestrales de los grupos experimental y control, respectivamente, y  $s_E^2$  y  $s_C^2$  las respectivas varianzas. El factor  $c(m)$  corrige el ligero sesgo positivo para muestras pequeñas mediante:

$$c(m) = 1 - \frac{3}{4(n_E + n_C) - 9} .$$

En meta-análisis, cada estimación del tamaño del efecto se pondera por la inversa de su varianza muestral, con objeto de permitir que los estudios más precisos (que suelen ser los basados en tamaños muestrales mayores) ejerzan un mayor peso específico en los cálculos meta-analíticos. Por ello es fundamental disponer de un estimador eficiente de la varianza debida a error de muestreo (o varianza intra-estudio). Para el índice  $d$ , ésta viene dada por:

$$V(d) = \frac{n_E + n_C}{n_E n_C} + \frac{d^2}{2(n_E + n_C)} .$$

En el caso de que todos los estudios del meta-análisis hayan operativizado la variable dependiente con la misma escala, test o prueba, no sería necesario estandarizar la diferencia de medias dividiendo por la desviación típica conjunta, sino que la simple diferencia entre las dos medias sería un índice del tamaño del efecto apropiado (Bond, Wiitala y Richard, 2003; Rosenberg, Adams y Gurevitch, 2000):

$$D = \bar{y}_E - \bar{y}_C ,$$

siendo su varianza intra-estudio:

$$V(D) = \frac{(n_E - 1)S_E^2 + (n_C - 1)S_C^2}{n_E + n_C - 2} .$$

Cuando la variable dependiente es dicotómica, el resultado de cada estudio queda resumido en una tabla de contingencia 2 x 2. La literatura médica ha estudiado en profundidad estos casos, recomendando la aplicación de alguno de los tres siguientes índices del tamaño del efecto (Deeks y Altman, 2001; Fleiss, 1994; Haddock, Rindskopf y Shadish, 1998; Sánchez-Meca y Marín-Martínez, 2000, 2001): La diferencia de riesgos,  $dr$  (o de proporciones), la razón de riesgos,  $rr$  (o de proporciones) y el odds ratio,  $or$  (o razón de posibilidades, o de productos cruzados). Siendo  $n_E$  y  $n_C$  los tamaños muestrales de los dos grupos,  $O_{1E}$  y  $O_{1C}$  las frecuencias de éxito en ambos grupos, y  $O_{2E}$  y  $O_{2C}$  las frecuencias de fracaso respectivas, la diferencia de riesgos se obtiene mediante:

$$dr = p_E - p_C ,$$

donde  $p_E = O_{1E} / n_E$  y  $p_C = O_{1C} / n_C$ , y la varianza intra-estudio viene dada por:

$$V(dr) = \frac{p_E(1-p_E)}{n_E} + \frac{p_C(1-p_C)}{n_C} .$$

La razón de riesgos y su varianza intra-estudio, esta última en términos del logaritmo de la razón de riesgos, se obtienen mediante:

$$rr = \frac{p_E}{p_C} .$$

$$V(Lrr) = \frac{1-p_E}{n_E p_E} + \frac{1-p_C}{n_C p_C} .$$



Y el odds ratio, junto con su varianza intra-estudios, también en función del logaritmo del odds ratio, se obtienen mediante:

$$or = \frac{p_E(1-p_C)}{p_C(1-p_E)} = \frac{O_{1E}O_{2C}}{O_{1C}O_{2E}}$$

$$V(Lor) = \frac{1}{O_{1E}} + \frac{1}{O_{1C}} + \frac{1}{O_{2E}} + \frac{1}{O_{2C}} .$$

Por otra parte, se dan situaciones en las que la variable de respuesta es continua en naturaleza, pero ha sido dicotomizada, de forma que la única información disponible en el estudio es una tabla de contingencia  $2 \times 2$  del tipo justamente comentado. Lo más aconsejable en estos casos es estimar la diferencia media estandarizada aplicando alguna de las funciones de transformación de variable dicotomizada a variable continua. De las diversas propuestas, las que mejores propiedades poseen son una basada en la función probit,  $d_{\text{Probit}}$  (Glass et al., 1981), y otras dos basadas en la función logística propuestas por Cox,  $d_{\text{Cox}}$  (Haddock et al., 1998) y por Hasselblad y Hedges (1995),  $d_{\text{HH}}$ , respectivamente (Sánchez-Meca, Marín-Martínez y Chacón-Moscó, 2003). El índice  $d_{\text{Probit}}$  y su varianza intra-estudio se obtienen mediante:

$$d_{\text{Probit}} = (z_E - z_C)$$

$$V(d_{\text{Probit}}) = \left[ \frac{2\pi p_E(1-p_E)e^{-z_E^2}}{n_E} + \frac{2\pi p_C(1-p_C)e^{-z_C^2}}{n_C} \right],$$

siendo  $z_E$  y  $z_C$  la inversa de la función de distribución de la curva normal para  $p_E$  y  $p_C$ , respectivamente

$$[z_E = \Phi^{-1}(p_E); z_C = \Phi^{-1}(p_C)].$$

El índice  $d_{\text{Cox}}$  y su varianza intra-estudio se obtienen mediante:

$$d_{\text{Cox}} = Lor / 1.65$$

$$V(d_{\text{Cox}}) = 0.367 \left[ \frac{1}{O_{1E}} + \frac{1}{O_{2E}} + \frac{1}{O_{1C}} + \frac{1}{O_{2C}} \right]$$

siendo  $Lor$  el logaritmo del odds ratio. Y el índice  $d_{\text{HH}}$  y su varianza intra-estudio vienen dados por:

$$d_{\text{HH}} = Lor \frac{\sqrt{3}}{\pi},$$

$$V(d_{\text{HH}}) = \frac{3}{\pi^2} \left[ \frac{1}{O_{1E}} + \frac{1}{O_{2E}} + \frac{1}{O_{1C}} + \frac{1}{O_{2C}} \right].$$

Todos los índices hasta aquí comentados tienen en común que son aplicables a diseños de dos grupos independientes. Si el diseño es de un solo grupo con pretest y postest, el índice más apropiado para reflejar el efecto es la diferencia entre las puntuaciones de cambio estandarizada,  $d_{\text{Change}}$ :

$$d_{\text{Change}} = \frac{\bar{y}_{\text{Pre}} - \bar{y}_{\text{Post}}}{S},$$

donde  $S$  es una estimación de la desviación estándar, para la cual se han propuesto diferentes opciones. En concreto, Gibbons, Hedeker y Davis (1993) proponen utilizar la desviación estándar de las puntuaciones de cambio entre el pretest y el posttest (índice  $d_2$ ). Dunlap, Cortina, Vaslow y Burke (1996) y Becker (1988) consideran más apropiado utilizar la desviación estándar del pretest (índice  $d_1$ ) como un estimador no contaminado por el efecto del tratamiento:

$$d_1 = \frac{\bar{y}_{\text{Pre}} - \bar{y}_{\text{Post}}}{S_{\text{Pre}}}$$

$$d_2 = \frac{\bar{y}_{\text{Pre}} - \bar{y}_{\text{Post}}}{S_D}$$

Las varianzas intra-estudio de los índices  $d_1$  y  $d_2$  son, respectivamente:

$$V(d_1) = \frac{2(1-r)}{n} \left( \frac{n-1}{n-3} \right) \left( 1 + \frac{n}{2(1-r)} d_1^2 \right) - \frac{d_1^2}{[c(n-1)]^2}$$

$$V(d_2) = \frac{n-1}{n(n-3)} (1 + n d_2^2) - \frac{d_2^2}{[c(n-1)]^2},$$

siendo  $n$  el tamaño muestral,  $r$  el coeficiente de correlación de Pearson entre las puntuaciones del pretest y el posttest y  $c(m)$  el factor de corrección del sesgo para muestras pequeñas, que se obtiene mediante:

$$c(n-1) = 1 - \frac{3}{4(n-1)-1}$$

Otra situación común en los diseños con asignación a grupos se da cuando los dos grupos (experimental y control) se registran en el pretest y en el posttest. En estas condiciones, el índice de la familia  $d$  más apropiado debería comparar el cambio que se produce en un grupo con el cambio que se produce en el otro. Pero, dado que existen al menos dos posibles índices para definir el cambio estandarizado del pretest al posttest en cada grupo (índices  $d_1$  y  $d_2$ ), según la estimación de la desviación estándar utilizada, también cabe proponer dos índices de la diferencia entre cambios medios estandarizados (Becker, 1988; Morris, 2000; Morris y DeShon, 2002). Siendo  $d_{1E}$  y  $d_{1C}$  las diferencias de cambio estandarizadas, tomando las desviaciones estándar del pretest, para los grupos experimental y control respectivamente, el primer índice de la diferencia entre las puntuaciones de cambio lo representaremos por  $\Delta_1$ , y viene definido por:

$$\Delta_1 = d_{1E} - d_{1C} = \frac{\bar{y}_{\text{Pre}}^E - \bar{y}_{\text{Post}}^E}{S_{\text{Pre}}^E} - \frac{\bar{y}_{\text{Pre}}^C - \bar{y}_{\text{Post}}^C}{S_{\text{Pre}}^C}$$

con varianza intra-estudio

$$V(\Delta_1) = \frac{n_E + n_C}{n_E n_C} \left( \frac{N-2}{N-4} \right) \left( 1 + \frac{n_E n_C}{n_E + n_C} \Delta_1^2 \right) - \frac{\Delta_1^2}{[c(N-2)]^2}$$

y siendo el factor de corrección  $c(N-2)$ :

$$c(N-2) = c(m) = 1 - \frac{3}{4(n_E + n_C) - 9}$$



Del mismo modo, siendo  $d_{2E}$  y  $d_{2C}$  las diferencias de cambio estandarizadas, tomando las desviaciones estándar de las puntuaciones de cambio, para los grupos experimental y control respectivamente, el segundo índice de la diferencia entre las puntuaciones de cambio lo representaremos por  $\Delta_2$ , y viene definido por:

$$\Delta_2 = d_{2E} - d_{2C} = \frac{y_{\text{Pre}}^{-E} - y_{\text{Post}}^{-E}}{S_D^E} - \frac{y_{\text{Pre}}^{-C} - y_{\text{Post}}^{-C}}{S_D^C}$$

con varianza intra-estudio

$$V(\Delta_2) = \frac{n_E + n_C}{2(1-r)n_E n_C} \left( \frac{N-2}{N-4} \right) \left( 1 + 2(1-r) \frac{n_E n_C}{n_E + n_C} \Delta_2^2 \right) - \frac{\Delta_2^2}{[c(N-2)]^2} .$$

• **La familia  $r$ .** Cuando los estudios han aplicado un diseño correlacional en el que han medido y/o registrado variables, el índice del tamaño del efecto más adecuado es el coeficiente de correlación de Pearson,  $r$ , o alguna extensión de éste cuando la variable no es cuantitativa (correlación de Spearman, correlación biserial-puntual, coeficiente *phi*, etc.). Se recomienda utilizar la transformación a  $Z$  de Fisher de los coeficientes de correlación antes de integrarlos en el meta-análisis, con objeto de homogeneizar sus varianzas (Hedges y Olkin, 1985; Rosenthal, 1991, 1994):

$$Z_r = \frac{1}{2} \ln \left( \frac{1+r}{1-r} \right) .$$

En este caso, la varianza intra-estudio viene dada por:

$$V(Z_r) = \frac{1}{N-3} .$$

Cuando resulte oportuno, existen fórmulas algebraicas que permiten transformar un índice de la familia  $d$  a uno de la familia  $r$ , y viceversa. En concreto, una diferencia de medias estandarizada,  $d$ , puede transformarse a coeficiente de correlación mediante (Rosenthal, 1991; Sánchez Meca y Ato, 1989):<sup>3</sup>

$$r = \frac{d}{\sqrt{d^2 + 4}} .$$

## MODELOS ESTADÍSTICOS ENMATA-ANÁLISIS

A partir de las estimaciones del tamaño del efecto obtenidas en los estudios primarios, es posible responder a las preguntas que se mencionaron en el epígrafe 5.1 referente a las fases de un meta-análisis: ¿Cuál es la magnitud del efecto medio? ¿Cuál es el intervalo de confianza en torno a ésta? ¿Es significativa la magnitud del efecto medio? ¿Existe heterogeneidad entre las estimaciones del tamaño del efecto? ¿Pueden encontrarse variables moderadoras de tal heterogeneidad? A lo largo de la corta historia del meta-análisis se han propuesto diferentes modelos estadísticos para abordar los análisis. Pueden consultarse Chalmers, Hedges y Olkin (2002), Field (2001, 2003a, 2003b, 2005), Schmidt (2008), Schmidt y Hunter (1999), Schmidt, Oh y Hayes (en prensa), Viechtbauer (2007a) y algunos de nuestros trabajos (Ma-

3 A excepción de los índices diferencia de riesgos,  $dr$ , razón de riesgos,  $rr$ , y odds ratio,  $or$ , esta fórmula de transformación puede aplicarse al resto de índices de la familia  $d$ .

rín-Martínez y Sánchez-Meca, 1998; Sánchez-Meca y Marín-Martínez, 1997, 1998a, 1998b, 2008; Sánchez-Meca, Marín-Martínez y Huedo, 2006) para una descripción más detallada de las similitudes y diferencias entre ellos.

Existe en la actualidad un claro consenso hacia el reconocimiento de aquellos modelos meta-analíticos que en sus análisis estadísticos ponderan cada tamaño del efecto en función de su precisión. Aun así, cabe distinguir entre dos modelos estadísticos específicamente diseñados para su aplicación sobre bases de datos meta-analíticas: Los *modelos de efectos fijos* y de *efectos aleatorios* (Erez, Bloom y Wells, 1996; Hedges, 1994; Hedges y Olkin, 1985; Hedges y Vevea, 1998; Hunter y Schmidt, 2000; Overton, 1998; Raudenbush, 1994; Schmidt, 2008; Sutton y Higgins, 2008).

Los modelos de efectos fijos y aleatorios difieren en la concepción de la población de estudios de partida. En el **modelo de efectos fijos** se asume que los estudios incluidos en el meta-análisis están estimando a un mismo, y único, tamaño del efecto paramétrico,  $\theta$ , por lo que la única variabilidad asumida es la debida a error de muestreo aleatorio o varianza intra-estudio,  $V(T_i)$ , es decir, al hecho de que los estudios utilizan muestras de sujetos diferentes. Siendo  $T_i$  el *i*ésimo tamaño del efecto<sup>4</sup> de un conjunto de  $k$  tamaños del efecto independientes que están estimando a un mismo efecto poblacional,  $\theta$ , el modelo matemático de efectos fijos se formula como:

$$T_i = \theta + \varepsilon_i.$$

En el **modelo de efectos aleatorios** se asume que los estudios estiman a una distribución de tamaños del efecto paramétricos en la población, que sigue una ley normal [ $\theta_i \sim N(\mu_\theta; \sigma_\theta^2)$ ] por lo que, además de la variabilidad debida al error de muestreo o intra-estudio,  $V(T_i)$ , hay que contemplar también la variabilidad inter-estudios,  $\sigma_\theta^2$ . En consecuencia, el modelo matemático incorpora dos términos de error, uno debido a la variabilidad intra-estudio,  $\varepsilon_i$ , que coincide con el del modelo de efectos fijos, y otro que refleja la variabilidad inter-estudios,  $\xi_i$ :

$$T_i = \mu_\theta + \xi_i + \varepsilon_i.$$

Mientras que en el modelo de efectos fijos el factor de ponderación de cada índice  $T_i$  viene determinado exclusivamente por la varianza intra-estudio,  $w_i = 1/V(T_i)$ , en el de efectos aleatorios el factor de ponderación lo está en función de la varianza intra-estudio y la varianza inter-estudios,  $w_i^* = 1/[V(T_i) + \sigma_\theta^2]$ . Las consecuencias de asumir uno u otro modelo afectan al grado de generalización de los resultados del meta-análisis. En el modelo de efectos fijos, la generalización se limita a la población de estudios de características similares a los incluidos en el meta-análisis. En el modelo de efectos aleatorios, por el contrario, los resultados pueden generalizarse a una población mayor de posibles estudios.

La varianza intra-estudio de cada tamaño del efecto se estima con las ecuaciones tratadas en el epígrafe anterior, de forma que cada estudio tiene su propia varianza intra-estudio, que estará fundamentalmente en función del tamaño muestral: a menor tamaño muestral, mayor varianza intra-estudio y, por tanto, menor precisión, por lo que el peso específico que el estudio ejerce en los cálculos meta-analíticos será menor.

La varianza inter-estudios tiene que estimarse a partir de las propias estimaciones del tamaño del efecto incluidas en el meta-análisis. El procedimiento habitualmente utilizado es el propuesto por DerSimonian y Laird (1986), basado en el método de los momentos:

4 Se entiende que  $T_i$  puede representar a cualquiera de los índices del tamaño del efecto tratados en el epígrafe anterior.

$$\hat{\sigma}_{\theta}^2 = \begin{cases} \frac{Q - (k - 1)}{c} & (\text{para } Q \geq k - 1) \\ 0 & (\text{para } Q < k - 1) \end{cases},$$

donde  $c$  se obtiene mediante:

$$c = \sum_{i=1}^k w_i - \frac{\sum_{i=1}^k w_i^2}{\sum_{i=1}^k w_i},$$

siendo  $w_i = 1/V(T_i)$  el factor de ponderación en función únicamente de la varianza intra-estudio. Y  $Q$  es el estadístico de heterogeneidad, que se obtiene mediante:

$$Q = \sum_{i=1}^k w_i (T_i - \bar{T})^2,$$

donde  $\bar{T}$  es el tamaño del efecto medio ponderado por la inversa de la varianza intra-estudio:

$$\bar{T} = \frac{\sum_{i=1}^k w_i T_i}{\sum_{i=1}^k w_i}.$$

No obstante, se han propuesto otros estimadores de la varianza inter-estudios que pueden funcionar mejor que el aquí descrito. Pueden consultarse a este respecto los trabajos de Viechtbauer (2005, 2007b, 2007c) y de Sánchez-Meca y Marín-Martínez (2008).

Desde el modelo de efectos fijos, el tamaño del efecto medio,  $\bar{T}$ , se estima según acabamos de exponer; mientras que asumiendo un modelo de efectos aleatorios, los factores de ponderación incluyen no sólo varianza intra-estudio, sino también una estimación de la varianza inter-estudios:

$$\bar{T}^* = \frac{\sum_{i=1}^k w_i^* T_i}{\sum_{i=1}^k w_i^*}.$$

Puede, así mismo, obtenerse un intervalo de confianza en torno al tamaño del efecto medio en ambos modelos, asumiendo una distribución normal:

$$\text{Modelo de efectos fijos: } P\left(\bar{T} - |z_{\alpha/2}| \sqrt{V(\bar{T})} \leq \theta \leq \bar{T} + |z_{\alpha/2}| \sqrt{V(\bar{T})}\right) = 1 - \alpha$$

$$\text{Modelo de efectos aleatorios: } P\left(\bar{T}^* - |z_{\alpha/2}| \sqrt{V(\bar{T}^*)} \leq \theta \leq \bar{T}^* + |z_{\alpha/2}| \sqrt{V(\bar{T}^*)}\right) = 1 - \alpha$$

Para cada caso, la varianza del tamaño del efecto medio viene dada por:

$$V(\bar{T}) = 1 / \sum_{i=1}^k w_i$$

$$V(\bar{T}^*) = 1 / \sum_{i=1}^k w_i^*.$$

En el modelo de efectos fijos, la hipótesis de homogeneidad entre los tamaños del efecto, se contrasta planteando ésta como hipótesis nula:  $H_0: \theta_1 = \theta_2 = \dots = \theta_i = \dots = \theta_k$ , y aplicando el estadístico  $Q$  arriba descrito, bajo cuya hipótesis se distribuye según  $\chi^2$  de Pearson con  $k - 1$  grados de libertad.

En el modelo de efectos aleatorios, al constituir un modelo de componentes de varianza, la principal hipótesis a contrastar es la de que no existe variabilidad inter-estudios:  $H_0: \hat{\sigma}_\theta^2 = 0$ , que se contrasta con el mismo estadístico de heterogeneidad,  $Q$ .

Si existe heterogeneidad entre los tamaños del efecto estimados a través de los estudios, el paso siguiente deberá ser poner a prueba el influjo, o la asociación, de variables moderadoras sobre los tamaños del efecto, con objeto de poder explicar tal heterogeneidad. Cuando la variable moderadora es cualitativa el análisis estadístico es el equivalente a un modelo de ANOVA, pero estimando los parámetros por mínimos cuadrados ponderados; mientras que si la variable moderadora es continua, se aplica un modelo de regresión, nuevamente por el método de mínimos cuadrados ponderados. En general, el modelo de análisis cuando ponemos a prueba el influjo conjunto de un grupo de variables moderadoras (cualitativas y/o continuas), es un modelo lineal en el que el vector de coeficientes de regresión,  $\beta$ , se estima mediante:

$$\beta = (X'WX)^{-1}X'T$$

siendo  $X$  la matriz de diseño, que contiene las variables moderadoras,  $T$  el vector de tamaños del efecto, que actúa como variable dependiente, y  $W$  es una matriz de ponderación diagonal  $n \times n$ , que contiene los factores de ponderación para cada tamaño del efecto. Esta matriz de ponderaciones se obtiene como la inversa de la matriz diagonal de varianzas de los tamaños del efecto estimados,  $W = \Sigma^{-1}$ . Dependiendo de que el modelo asumido sea de efectos fijos o de efectos aleatorios, los elementos de la diagonal de  $W$  se obtienen en función de la varianza intra-estudio, para el primer modelo [ $w_i = 1/V(T_i)$ ], o bien en función de la suma de la varianza intra-estudio y de una estimación de la varianza inter-estudios, para el modelo de efectos aleatorios [ $w_i^* = 1/(V(T_i) + \hat{\sigma}_\theta^2)$ ]. En este último caso, la varianza inter-estudios tiene que estimarse teniendo en cuenta la matriz de diseño mediante (Rosenberg et al., 2000):

$$\hat{\sigma}_\theta^2 = \begin{cases} \frac{Q_E - (k - h)}{trW - tr[\mathbf{WX}(\mathbf{X}'\mathbf{WX})^{-1}\mathbf{X}'\mathbf{W}]} & (\text{para } Q_E \geq (k - h)) \\ 0 & (\text{para } Q_E < (k - h)) \end{cases}$$

donde  $k$  es el número de estudios (o tamaños del efecto),  $h$  es el número de parámetros del modelo,  $tr$  es la traza de una matriz y  $Q_E$  es la suma de cuadrados de error por mínimos cuadrados ponderados, que se obtiene mediante:

$$Q_E = T'WT$$

A partir del vector de coeficientes de regresión y de la matriz de varianzas y covarianzas estimadas de los coeficientes de regresión del modelo,  $S_\beta$ , es posible poner a prueba el influjo de variables moderadoras. Así, el contraste de la hipótesis nula de ausencia de efecto de las variables moderadoras,  $H_0: \beta = \mathbf{0}$ , se efectúa mediante la suma de cuadrados ponderada del modelo,  $Q_R$ :

$$Q_R = \beta' S_\beta^{-1} \beta,$$

que bajo dicha hipótesis se distribuye según  $\chi^2$  de Pearson con  $h$  grados de libertad.

Así mismo, es posible poner a prueba la especificación del modelo contrastando la hipótesis nula de una correcta especificación,  $H_0: \delta = X\beta$ , siendo  $\delta$  el vector de tamaños del

efecto poblacional, mediante la suma de cuadrados de error ponderada,  $Q_E$ , que bajo dicha hipótesis se distribuye según  $\chi^2$  de Pearson con  $k - h$  grados de libertad.

El modo de operar analíticamente en un meta-análisis es similar para los dos modelos asumibles: efectos fijos o efectos aleatorios. Pero la capacidad de generalización de los resultados es mayor en efectos aleatorios que en efectos fijos. Además, se considera actualmente más plausible el modelo de efectos aleatorios, ya que es más difícil imaginar que en la realidad toda una población de estudios realizados sobre un mismo problema estén estimando exactamente a un mismo efecto poblacional. De hecho, puede considerarse el modelo de efectos fijos como un caso particular del modelo, más general, de efectos aleatorios, cuando no existe varianza inter-estudios. Pero ante una base de datos meta-analítica concreta, nunca puede saberse con certeza si el modelo más apropiado es uno u otro, ya que desconocemos los parámetros poblacionales.

Una estrategia que suele recomendarse consiste en iniciar los análisis poniendo a prueba la hipótesis de homogeneidad de los tamaños del efecto mediante la aplicación de la prueba  $Q$ . Si ésta resulta significativa, ello sería indicativo de que existe heterogeneidad entre los tamaños del efecto, por lo que debería asumirse un modelo de efectos aleatorios en los análisis subsiguientes. Si, por el contrario, el estadístico  $Q$  no alcanza la significación estadística, podría adoptarse el modelo de efectos fijos. Esta estrategia sería factible si no fuera porque la prueba  $Q$  de heterogeneidad se muestra poco potente cuando el número de estudios,  $k$ , no es elevado (Harwell, 1997; Sánchez-Meca y Marín-Martínez, 1997). Por tanto, un resultado no significativo de la prueba  $Q$  no es garantía suficiente para asumir un modelo de efectos fijos, pero un resultado significativo de dicha prueba sí lo es para asumir con fiabilidad un modelo de efectos aleatorios.

Otra opción, compatible en general con la acabada de exponer, se basa en la estimación de la varianza inter-estudios. Dado que el estimador de dicha varianza se trunca en el valor 0 para evitar estimaciones negativas de la varianza inter-estudios, se puede asumir un modelo de efectos fijos cuando se dé tal situación, y adoptar un modelo de efectos aleatorios en caso contrario. Por último, se ha propuesto recientemente un nuevo índice, denominado índice  $I^2$ , para estimar el grado de heterogeneidad de los tamaños del efecto en un meta-análisis que, aunque está en función del estadístico  $Q$ , no se deja afectar por el número de estudios,  $k$ , y puede interpretarse como un porcentaje de heterogeneidad exhibido por los tamaños del efecto (Higgins y Thompson, 2002; Huedo-Medina, Sánchez-Meca, Marín-Martínez y Bote-lla, 2006).

## VENTAJAS Y LIMITACIONES DEL META-ANÁLISIS

No cabe duda de que el meta-análisis se ha convertido en una herramienta metodológica de gran utilidad para ayudar a acumular el conocimiento científico en un determinado campo de investigación de una forma objetiva, sistemática y rigurosa. Entre sus ventajas cabe destacar las siguientes (Cooper, 1998; Hedges y Olkin, 1985; Rosenthal, 1991; Sánchez-Meca, 2003; Sánchez-Meca y Ato, 1989):

- (a) *Eficiencia*. El meta-análisis tiene una mayor capacidad para tratar grandes cantidades de información que las revisiones narrativas, gracias a sus posibilidades de cuantificar y de codificar objetivamente las variables implicadas en los resultados de los estudios.
- (b) *Rigor científico*. El meta-análisis cumple con las normas de rigor científico que se exigen en las investigaciones primarias, posibilitando la replicación del proceso. Se

trata, en definitiva, de exigir en el proceso de revisión de la investigación las mismas normas de rigor científico que se exigen en la realización de estudios primarios.

- (c) *Detección de efectos pequeños.* Al centrarse en la magnitud de los efectos en lugar de en los resultados de las pruebas de significación estadística, el meta-análisis tiene una mayor capacidad para detectar efectos pequeños que, sin embargo, pueden tener relevancia práctica o real. Las revisiones narrativas tienen más dificultades para detectar estos efectos que, precisamente, son los más habituales en las ciencias empíricas en general y en las Ciencias Sociales y del Comportamiento, en particular.
- (d) *Potencia estadística.* Al acumular los tamaños muestrales de los diferentes estudios integrados, los procedimientos de análisis estadístico propios del meta-análisis tienen mayor potencia estadística para detectar los efectos y las relaciones entre las variables implicadas.
- (e) *Énfasis en el tamaño del efecto.* Otra de las grandes aportaciones del meta-análisis al proceso de investigación es el gran énfasis que pone en el tamaño del efecto, relegando las pruebas de significación estadística a un segundo plano. Este énfasis ha contribuido a modificar las recomendaciones que actualmente se están proponiendo contra el uso abusivo de las pruebas de significación.
- (f) *Aprovechamiento de resultados contradictorios.* El meta-análisis dispone actualmente de procedimientos estadísticos que permiten detectar variables moderadoras responsables de los resultados heterogéneos y contradictorios que pueden encontrarse en un determinado campo de investigación.
- (g) *Seguridad.* Al basarse en una metodología sistemática, objetiva y rigurosa, las conclusiones a las que se llegan con los meta-análisis son más fiables y seguras que las alcanzadas en revisiones cualitativas o narrativas de la investigación.

No obstante, el meta-análisis tiene también limitaciones que es preciso conocer para aplicarlo correctamente. Éstas son las principales junto con las recomendaciones para tratar de paliar sus efectos adversos:

(a) *Propensión a cometer errores.* Datos e información defectuosa en los estudios primarios pueden afectar a la calidad de los datos meta-analíticos y, en consecuencia, a la fiabilidad de sus conclusiones. El meta-análisis está, pues, limitado por las propias deficiencias de los estudios primarios. Si no es posible obtener datos exactos de sus resultados, es preferible no hacer el meta-análisis.

(b) *Heterogeneidad.* Estudios muy heterogéneos entre sí no deberían ser integrados, ya que los resultados globales serían poco informativos. Es por ello que se recomienda actualmente realizar estudios meta-analíticos con unos objetivos claramente focalizados, para garantizar que las conclusiones tendrán relevancia práctica. Precisamente, una de las críticas más severas que ha recibido el meta-análisis es el denominado “problema de las manzanas y las naranjas” (‘the apples and oranges problem’), dirigido por Eysenck (1978, 1994) contra el ‘macro’ meta-análisis realizado originalmente por Smith y Glass (1977) sobre la eficacia de la psicoterapia. No exento de razón, estas críticas contra la mezcla de estudios muy heterogéneos, ha provocado que actualmente se lleven a cabo meta-análisis de tamaño mucho más reducido, con estudios cuyos diseños sean similares (por ejemplo, todos los estudios deben tener dos grupos, uno de control y otro experimental). En el límite de la homogeneización se encuentra la línea de investigación meta-analítica médica, que recomienda la integración mediante meta-análisis únicamente de ensayos controlados aleatorizados (Whitehead, 2002).



(c) *Calidad de los datos*. La mezcla de estudios de buena calidad metodológica con estudios de baja calidad puede dar lugar a estimaciones de los efectos sesgadas. Se recomienda codificar como una variable moderadora más la calidad del diseño del estudio y analizar su posible relación con los tamaños del efecto; o bien, fijar desde el principio normas estrictas de calidad metodológica que deben cumplir los estudios para ser incluidos en el meta-análisis. Existe un buen número de escalas para valorar la calidad metodológica de los estudios primarios, buena parte de las cuales se han desarrollado en la literatura médica (Jüni, Altman y Egger, 2001; Moher, Jones y Lepage, 2001; Tritchler, 1999; Wortman, 1994).

(d) *Representatividad*. Por muy exhaustiva que sea la búsqueda de la literatura, nunca será posible localizar todos los estudios seleccionables. Si, además, no se dispone de estudios no publicados, el sesgo de publicación puede ser una seria amenaza contra la validez de los resultados del meta-análisis. La inclusión de estudios no publicados es muy recomendable y, en cualquier caso, realizar un análisis del sesgo de publicación mediante el cálculo de índices de tolerancia a los resultados nulos (fail-safe  $N$ ), para valorar la robustez del meta-análisis frente a esta amenaza. El problema del sesgo de publicación y sus efectos sobre los resultados de los meta-análisis es una línea de trabajo actualmente muy activa (cf. por ejemplo, Rothstein, Sutton y Borenstein, 2005).

(e) *Dependencia*. La inclusión de más de un índice del tamaño del efecto calculado sobre la misma muestra de sujetos atenta contra el supuesto de independencia de los datos que asumen las técnicas meta-analíticas de análisis estadístico. Este problema surge cuando un mismo estudio primario presenta resultados sobre diferentes variables dependientes. Si se calcula un tamaño del efecto para cada una de ellas y se incorporan todas las estimaciones al meta-análisis, se incurre en un problema de dependencia que afecta a la validez de la conclusión estadística. Se han propuesto varias soluciones para resolver este problema. Una de ellas es obtener un promedio de los tamaños del efecto correspondientes a un mismo estudio (Gleser y Olkin, 1994; Hedges y Olkin, 1985; Marín-Martínez y Sánchez-Meca, 1999; Rosenthal y Rubin, 1986). Otras soluciones implican, bien realizar meta-análisis diferentes para cada variable dependiente, bien modelar la estructura correlacional entre ellas (Becker, 2000; Kalaian y Raudenbush, 1996; Raudenbush, Becker y Kalaian, 1988).

Como conclusión, podemos decir que el meta-análisis es hoy día una metodología de investigación que se ha convertido en un elemento indispensable como puente entre la investigación pasada y la futura: Aunque se basa en la integración cuantitativa de las evidencias acumuladas sobre un determinado tópico, su principal función es orientar la investigación futura, detectando vacíos conceptuales y/o metodológicos, denunciando deficiencias en la literatura y revelando relaciones no anticipadas. Su gran ventaja está en la “suma de esfuerzos” y en su empeño por explicar las desviaciones que se observan entre los resultados de los estudios: “Estas pequeñas desviaciones, que por sí mismas pueden no llegar a ser significativas, cuando se combinan pueden revelar importantes conexiones científicas desde las que se puede extraer nuevas evidencias científicas” (Garfield, 1991, p. 5).

El meta-análisis no debe entenderse como un mero método para hacer revisiones de literatura, sino como un modo de afrontar el análisis de los datos que implica cambios en la forma de entender el progreso de una ciencia. Su énfasis en el tamaño del efecto ha ejercido una importante influencia a favor de las críticas contra la excesiva atención prestada a las pruebas de significación a lo largo de la historia de la psicología científica. Al mismo tiempo, el meta-análisis ha contribuido a relativizar el papel que los estudios primarios juegan en la demostración de hechos científicos y en la importancia de acumular de forma sistemática la investigación para el desarrollo de teorías cada vez más sólidas. En palabras de Schmidt (1992, p. 1179):

“Los datos nos llegan encriptados, y para entender su significado primero tenemos que descifrar el código. Para hacer esto se necesita del meta-análisis. Por tanto, los estudios individuales deben ser considerados como meros datos puntuales que contribuirán a un meta-análisis futuro. De esta forma, el estatus y el valor científico del estudio individual se ve necesariamente reducido”.

Siempre que se utilice con buen juicio y siendo consciente de sus limitaciones, el meta-análisis puede aportar una importante contribución a la acumulación del conocimiento científico y, por ende, al progreso de la ciencia.

## REFERENCIAS BIBLIOGRÁFICAS

- Becker, B.J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical & Statistical Psychology*, 41, 257-278.
- Becker, B.J. (2000). Multivariate meta-analysis. En H.E.A. Tinsley y S. S.D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 499-525). San Diego, CA: Academic Press.
- Begg, C.B. (1994). Publication bias. En H. Cooper y L.V. Hedges (Eds.), *The Handbook of Research Synthesis* (pp. 399-409). Nueva York: Russell Sage Foundation.
- Bond, C.F.Jr., Wiitala, W.L. y Richard, F.D. (2003). Meta-analysis of raw mean differences. *Psychological Methods*, 8, 406-418.
- Botella, J. y Gambara, H. (2002). *¿Qué es el meta-análisis?* Madrid: Siglo XXI.
- Botella, J. y Gambara, H. (2006). Doing and reporting a meta-analysis. *International Journal of Clinical & Health Psychology*, 6, 425-440.
- Burdett, S., Stewart, L.A. y Tierney, J.F. (2003). Publication bias and meta-analyses. *International Journal of Technology Assessment in Health Care*, 19, 129-134.
- Chalmers, T.C., Hedges, L.V. y Cooper, H. (2002). A brief history of research synthesis. *Evaluation & the Health Professions*, 25, 12-37.
- Cooper, H. (1998). *Integrating research: A guide for literature reviews* (3ª ed.). Newbury Park, CA: Sage.
- Cooper, H. y Hedges, L.V. (Eds.) (1994). *The handbook of research synthesis*. Nueva York: Russell Sage.
- Cortina, J.M. y Nouri, H. (2000). *Effect size for ANOVA designs*. Thousand Oaks, CA: Sage.
- Davey Smith (Ed.) (2001). *Clinical Meta-analysis*. Nueva York: Wiley.
- Deeks, J.J. y Altman, D.G. (2001). Effect measures for meta-analysis of trials with binary outcomes. En M. Egger, G.D. Smith y D.G. Altman (Eds.), *Systematic reviews in health care: Meta-analysis in context* (2ª ed.) (pp. 313-335). Londres: BMJ Books.
- DerSimonian, R. y Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7, 177-188.
- Dunlap, W.P., Cortina, J.M., Vaslow, J.B. y Burke, M.J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, 1, 170-177.
- Eddy, D.M., Hasselblad, V. y Shachter, R. (1995). *Meta-analysis by the confidence profile methods: The statistical synthesis of evidence*. Boston, MA: Academic Press.
- Egger, M., Smith, G.D. y Altman, D.G. (Eds.) (2001). *Systematic reviews in health care: Meta-analysis in context* (2ª ed.). Londres: BMJ Books.



- Erez, A., Bloom, M.C. y Wells, M.T. (1996). Using random rather than fixed effects models in meta-analysis: Implications for situational specificity and validity generalization. *Personnel Psychology*, 49, 275-306.
- Eysenck, H.J. (1978). An exercise in mega-silliness. *American Psychologist*, 33, 517-517.
- Eysenck, H.J. (1994). Meta-analysis and its problems. *British Medical Journal*, 309, 789-792.
- Field, A.P. (2001). Meta-analysis of correlation coefficients: A Monte Carlo comparison of fixed- and random-effects methods. *Psychological Methods*, 6, 161-180.
- Field, A.P. (2003). Can meta-analysis be trusted? *The Psychologist*, 16, 642-645.
- Field, A.P. (2003). The problems of using fixed-effects models of meta-analysis on real-world data. *Understanding Statistics*, 2, 77-96.
- Field, A.P. (2005). Is the meta-analysis of correlation coefficients accurate when population correlations vary? *Psychological Methods*, 10, 444-467.
- Fleiss, J.L. (1994). Measures of effect size for categorical data. En H. Cooper y L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 245-260). Nueva York: Russell Sage Foundation.
- Garfield, E. (1991). Meta-analysis and the metamorphosis of the scientific literature review. *Current Contents: Social & Behavioral Sciences*, 43, 5-8.
- Gibbons, R.D., Hedeker, D.R. y Davis, J.M. (1993). Estimation of effect size from a series of experiments involving paired comparisons. *Journal of Educational Statistics*, 18, 271-279.
- Glass, G.V., McGaw, B. y Smith, M.L. (1981). *Meta-analysis in social research*. Newbury Park, CA: Sage.
- Gleser, L.J. y Olkin, I. (1994). Stochastically dependent effect sizes. En H. Cooper y L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 339-355). Nueva York: Russell Sage Foundation.
- Gómez, J. (1987). *Meta-análisis*. Barcelona: PPU.
- Grissom, R.J. y Kim, J.J. (2005). *Effect sizes for research: A broad practical approach*. Mahwah, NJ: Erlbaum.
- Haddock, C.K., Rindskopf, D. y Shadish, W.R. (1998). Using odds ratios as effect sizes for meta-analysis of dichotomous data: A primer on methods and issues. *Psychological Methods*, 3, 339-353.
- Harwell, M. (1997). An empirical study of Hedges's homogeneity test. *Psychological Methods*, 2, 219-231.
- Hasselblad, V. y Hedges, L.V. (1995). Meta-analysis of screening and diagnostic tests. *Psychological Bulletin*, 117, 167-178.
- Hedges, L.V. (1994). Fixed effects models. En H. Cooper y L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 285-299). Nueva York: Russell Sage Foundation.
- Hedges, L.V. y Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Hedges, L.V. y Vevea, J.L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, 3, 486-504.
- Higgins, J.P.T. y Thompson, S.G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21, 1539-1558.
- Huedo-Medina, T.B., Sanchez-Meca, J., Marin-Martinez, F. y Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or  $I^2$  index? *Psychological Methods*, 11, 193-206.
- Hunter, J.E. y Schmidt, F.L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection & Assessment*, 8, 275-292.

- Hunter, J.E. y Schmidt, F.L. (2004). *Methods of meta-analysis: Correcting error and bias in research synthesis* (2ª ed.). Sage.
- Jüni, P., Altman, D.G. y Egger, M. (2001). Assessing the quality of randomised controlled trials. En M. Egger, G.D. Smith y D.G. Altman (Eds.), *Systematic reviews in health care: Meta-analysis in context* (2ª ed.) (pp. 87-108). Londres: BMJ Books.
- Kalaian, H.A. y Raudenbush, S.W. (1996). A multivariate mixed linear model for meta-analysis. *Psychological methods*, 1, 227-235.
- Lipsey, M.W. y Wilson, D.B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- Marín Martínez, F. y Sánchez Meca, J. (1998). Testing for dichotomous moderators in meta-analysis. *Journal of Experimental Education*, 67, 69-81.
- Marín Martínez, F. y Sánchez Meca, J. (1999). Averaging dependent effect sizes in meta-analysis: A cautionary note about procedures. *Spanish Journal of Psychology*, 2, 32-38.
- Marín-Martínez, F., Sánchez-Meca, J., Huedo, T.B. y Fernández-Guzmán, I. (2007). *Meta-análisis: ¿Dónde estamos y hacia dónde vamos?* En A. Borges y P. Prieto (Eds.), *Psicología y ciencias afines en los albores del siglo XXI (Homenaje al profesor Alfonso Sánchez Bruno)* (pp. 87-102). Grupo Editorial Universitario.
- Martín, J.L.R., Tobías, A. y Seoane, T. (Coords.) (2006). *Revisiones Sistemáticas en las Ciencias de la Vida*. Toledo: FISCAM.
- Moher, D., Jones, A. y Lepage, L. for the CONSORT Group (2001). Use of the CONSORT statement and quality of reports for randomized trials: A comparative before-and-after evaluation. *Journal of the American Medical Association*, 285, 1992-1995.
- Morris, S.B. (2000). Distribution of the standardized mean change effect size for meta-analysis on repeated measures. *British Journal of Mathematical & Statistical Psychology*, 53, 17-29.
- Morris, S.B. y DeShon, R.P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, 7, 105-125.
- Overton, R.C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods*, 3, 354-379.
- Petticrew, M. y Roberts, H. (2006). *Systematic Reviews in the Social Sciences: A practical guide*. Malden, MA: Blackwell.
- Raudenbush, S.V. (1994). Random effects models. En H. Cooper y L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301-321). Nueva York: Russell Sage Foundation.
- Raudenbush, S.W., Becker, B.J. y Kalaian, H.A. (1988). Modeling multivariate effect sizes. *Psychological Bulletin*, 103, 111-120.
- Roberts, B.W., Kuncel, N.R., Viechtbauer, W. y Bogg, T. (2007). Meta-analysis in personality psychology: A primer. En R.W. Robins, R.C. Fraley y R.F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 652-672). Nueva York: Guilford Press.
- Rosenberg, M.S., Adams, D.C. y Gurevitch, J. (2000). *MetaWin: Statistical software for meta-analysis* (Vers. 2.0). Sunderland, MA: Sinauer Assoc.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research* (ed. Rev.). Newbury Park, CA: Sage.
- Rosenthal, R. (1994). Parametric measures of effect size. En H. Cooper y L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231-244). Nueva York: Russell Sage Foundation.
- Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin*, 118, 183-192.
- Rosenthal, R., Rosnow, R.L. y Rubin, D.B. (2000). *Contrasts and effect sizes in behavioral research: A correlational approach*. Cambridge, UK: Cambridge University Press.

- Rosenthal, R. y Rubin, D.B. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin*, 99, 400-406.
- Rothstein, H.R., Sutton, A.J. y Borenstein, M. (Eds.) (2005). *Publication bias in meta-analysis: Prevention, assessment, and adjustments*. Chichester, UK: Wiley.
- Sánchez-Meca, J. (1999). Meta-análisis para la investigación científica. En F.J. Sarabia (Coord.), *Metodología para la Investigación en Marketing y Dirección de Empresa* (pp. 173-201). Madrid: Pirámide.
- Sánchez-Meca, J. (2003). La revisión del estado de la cuestión: El meta-análisis. En C. Camisón, M.J. Oltra y M.L. Flor (Eds.), *Enfoques, problemas y métodos de investigación en economía y dirección de empresas. Tomo I* (pp. 101-110). Castellón: ACEDE / Fundación Universitat Jaume I – Empresa.
- Sánchez-Meca, J. y Ato, M. (1989). Meta-análisis: Una alternativa metodológica a las revisiones tradicionales de la investigación. En J. Arnau y H. Carpintero (Eds.), *Tratado de psicología general I: Historia, teoría y método* (pp. 617-669). Madrid: Alhambra.
- Sánchez-Meca, J. y Marín-Martínez, F. (1997). Homogeneity tests in meta-analysis: A Monte Carlo comparison of statistical power and Type I error. *Quality & Quantity*, 31, 385-399.
- Sánchez-Meca, J. y Marín-Martínez, F. (1998a). Testing continuous moderators in meta-analysis: A comparison of procedures. *British Journal of Mathematical & Statistical Psychology*, 51, 311-326.
- Sánchez-Meca, J. y Marín-Martínez, F. (1998b). Weighting by inverse-variance or by sample size in meta-analysis: A simulation study. *Educational & Psychological Measurement*, 58, 211-220.
- Sánchez-Meca, J. y Marín-Martínez, F. (2000). Testing the significance of a common risk difference in meta-analysis. *Computational Statistics & Data Analysis*, 33, 299-313.
- Sánchez-Meca, J. y Marín-Martínez, F. (2001). Meta-analysis of 2x2 tables: Estimating a common risk difference. *Educational & Psychological Measurement*, 61, 249-276.
- Sánchez-Meca, J. y Marín-Martínez, F. (2008). Confidence intervals for the overall effect size in random-effects meta-analysis. *Psychological Methods*, 13, 31-48.
- Sánchez-Meca, J., Marín-Martínez, F. y Chacón-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods*, 8, 448-467.
- Sánchez-Meca, J., Marín-Martínez, F. y Huedo, T. (2006). Modelo de efectos fijos y modelo de efectos aleatorios. En J.L.R. Martín, A. Tobías y T. Seoane (Coords.), *Revisiones Sistemáticas en Ciencias de la Vida* (pp. 189-204). Toledo: FISCAM.
- Schmidt, F.L. (1992). What do data really mean? *American Psychologist*, 47, 1173-1181.
- Schmidt, F.L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods*, 1, 115-129.
- Schmidt, F.L. (2008). Meta-analysis: A constantly evolving research integration tool. *Organizational Research Methods*, 11, 96-113.
- Schmidt, F.L. y Hunter, J.E. (1995). The impact of data analysis method on cumulative knowledge: Statistical significance testing, confidence intervals, and meta-analysis. *Evaluation & the Health Professions*, 18, 408-427.
- Schmidt, F.L. y Hunter, J.E. (1999). Comparison of three meta-analysis methods revisited: An analysis of Johnson, Mullen, and Salas (1995). *Journal of Applied Psychology*, 84, 144-148.
- Schmidt, F.L., Oh, I.-S. y Hayes, T.L. (en prensa). Fixed versus random effects models in meta-analysis: Model properties and an empirical comparison of difference in results. *British Journal of Mathematical and Statistical Psychology*.
- Schulze, R. (2004). *Meta-analysis: A comparison of approaches*. Hogrefe & Huber Pub.

- Shadish, W.R. y Haddock, C.K. (1994). Combining estimates of effect size. En H. Cooper y L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 261-281). Nueva York: Russell Sage Foundation.
- Smith, M.L. y Glass, G.V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32, 752-760.
- Sutton, A.J., Abrams, K.R., Jones, D.R., Sheldon, T.A. y Song, F. (2000). *Methods for meta-analysis in medical research*. Chichester, UK: Wiley.
- Sutton, A.J., Duval, S.J., Tweedie, R.L., Abrams, K.R. y Jones, D.R. (2000). Empirical assessment of effect of publication bias on meta-analyses. *British Medical Journal*, 320, 1574-1577.
- Sutton, A.J. y Higgins, J.P.T. (2008). Recent developments in meta-analysis. *Statistics in Medicine*, 27, 625-650.
- Titchler, D. (1999). Modelling study quality in meta-analysis. *Statistics in Medicine*, 18, 2135-2145.
- Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational & Behavioral Statistics*, 30, 261-293.
- Viechtbauer, W. (2007a). Accounting for heterogeneity via random-effects models and moderator analyses in meta-analysis. *Journal of Psychology*, 215, 104-121. Viechtbauer, W. (2007b). Confidence intervals for the amount of heterogeneity in meta-analysis. *Statistics in Medicine*, 26, 37-52.
- Viechtbauer, W. (2007c). Hypothesis tests for population heterogeneity in meta-analysis. *British Journal of Mathematical & Statistical Psychology*, 60, 29-60.
- Whitehead, A. (2002). *Meta-analysis of controlled clinical trials*. Chichester, UK: Wiley.
- Wolf, F.M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. Beverly Hills, CA: Sage.
- Wortman, P.M. (1994). Judging research quality. En H. Cooper y L.V. Hedges (Eds.), *The Handbook of Research Synthesis* (pp. 97-109). Nueva York: Russell Sage Foundation.



# METODOLOGÍAS CUALITATIVAS: CARACTERÍSTICAS, PROCESOS Y APLICACIONES

M<sup>a</sup> TERESA ANGUERA ARGILAGA

*Departament de Metodologia de les Ciències del Comportament  
Facultat de Psicologia (Universitat de Barcelona)*

## RELEVANCIA DE LA METODOLOGÍA CUALITATIVA EN EL MOMENTO ACTUAL

Hasta hace unos años se trataba de una opción metodológica claramente marginal y con escaso poder de convocatoria. En la actualidad se halla en la cresta de la ola, en el epicentro de un movimiento que en algo se diferencia 'cualitativamente' de lo que venía siendo habitual en pasadas décadas, en donde se daba un movimiento pendular de alternancia.

Szent-Györgyi (1980), en una frase epítomica, se refería al carácter desconocido de gran número de episodios de la naturaleza, en donde aparece una extraña mezcla de lo contradictorio, ilógico e incoherente, y en este enclave es donde la metodología cualitativa debería "suavizar" estas contradicciones y dar una imagen comprensiva de dicha realidad (Morse, 1994). El *laboratorio* de quien se dedique a investigación cualitativa es la vida diaria, y las variables, que no se controlan, no se conocen hasta el fin del estudio.

Benoliel (1984) describió la investigación cualitativa como "modos de cuestionamiento sistemático enfocados a entender a los seres humanos y a la naturaleza de sus interacciones con ellos mismos y con su entorno" (pág. 3). Con frecuencia, la investigación cualitativa se describe como holística; es decir, que se preocupa por los seres humanos y su ambiente en toda su complejidad (Valles & Baer, 2005). A menudo se basa en la premisa de que el obtener conocimiento de los seres humanos es imposible sin describir cómo se vive y cómo se define la experiencia humana por los propios actores.

Al tratar de conceptualizar la metodología cualitativa, se hace patente la tentación fácil de apoyarnos en el concepto aparentemente mejor delimitado de metodología cuantitativa, pero deberá obviarse esta vía. En la actualidad, en efecto, buen número de discusiones se plantean en base a dicotomías diversas: investigación nomotética/idiográfica, investigación psicométrica/etnometodología, investigación de laboratorio/de campo, investigación experimental/naturalista (Alvarez, 1986). Con ellas se abordan aspectos complementarios de los que no podrían negarse, además, sus posibilidades de colisión ni su atributo opuesto de mutua exclusividad.

Nos hallamos, pues, en una verdadera encrucijada, en la cual retomamos la consideración de la **metodología cualitativa** como "*una estrategia de investigación fundamentada en una depurada y rigurosa descripción contextual del evento, conducta o situación que garantice la máxima objetividad en la captación de la realidad, siempre compleja, y preserve la espontánea*



*continuidad temporal que le es inherente, con el fin de que la correspondiente recogida sistemática de datos, categóricos por naturaleza, y con independencia de su orientación preferentemente idiográfica y procesual, posibilite un análisis (exploratorio, de reducción de datos, de toma de decisiones, evaluativo, etc.) que dé lugar a la obtención de conocimiento válido con suficiente potencia explicativa, acorde, en cualquier caso, con el objetivo planteado y los descriptores e indicadores a los que se tuviera acceso” (Anguera, 1986, p. 24).*

Se imponen varias matizaciones a esta conceptualización, que pueden estructurarse en torno a criterios epistemológicos, metodológicos, técnicos y de contenido, los cuales facilitan su justificación (Anguera, 1995a, 1999):

- a) La vía hipotético-deductiva, implantada tradicionalmente en buen número de líneas de investigación, ofrece graves dificultades en la investigación cualitativa, ya que habitualmente conduce a una pérdida de flexibilidad en la descripción y a su descontextualización. Por el contrario, y sin asumir de forma generalizada la vía inductiva, ésta se revela con mayor viabilidad, lo cual se halla en consonancia con el carácter previamente no consolidado de los estudios temáticos a los que se dirige (Miles & Huberman, 1984).
- b) Metodológicamente, los datos cualitativos deben ofrecer la necesaria consistencia, y emergen de una descripción y registro cuidadosos, aunque su contenido es variable y su análisis difícil debido a la nula estandarización de las respuestas y su compleja sistematización. El papel fundamental lo juega el proceso de categorización (Lofland, 1971), ya que no se puede aspirar a una adecuada “captación de la realidad en sus propios términos” si no se logran elaborar las categorías que la hacen explicable y dan coherencia al flujo de eventos y/o conductas necesariamente contextualizados. En este mismo sentido, Denzin (1970) y Patton (1980) están de acuerdo en que la tarea de un metodólogo cualitativo es la de suministrar un marco dentro del cual los sujetos respondan de forma que se representen fielmente sus puntos de vista respecto al mundo y su experiencia.
- c) A nivel genérico, las descripciones detalladas procedentes de registros directos y datos documentales (Patton, 1980) constituyen el grueso de los datos cualitativos que, por otra parte, requerirán diversos instrumentos de medida, a pesar del importante papel jugado por lo interpretativo (Smith, 1983). De forma particularizada, las técnicas que más propiamente abarca la metodología cualitativa son el registro de conducta en observación –y de forma especial en observación participante, entrevista y material documental, entendiendo que la fase definitoria es la de recogida de datos (Blaxter, 1979), ya que nada impide que el análisis al cual se sometan los datos implique el adentrarnos en lo cuantitativo (Blanco, 1983).
- d) Desde un criterio de contenido, existe una primera gran restricción relativa al nivel de observabilidad (Norris, 1984), y aunque la posición oficial del cognitivismo rechaza los procedimientos introspectivos (Nisbett & Wilson, 1977), se progresa en el camino que pretende el acceso a los fenómenos mentales, reconociéndose que los individuos tienen acceso directo a una gran cantidad de hechos privados. La necesaria contextualización inherente a la metodología cualitativa y sus implicaciones a nivel de los términos acuñados “ad hoc” en el proceso de categorización (Bulmer, 1979), así como la indudable tradición de interaccionismo simbólico (Schwartz & Jacobs, 1984) que lo ha propiciado, muestran una inclinación o mejor predisposición temática con dicha orientación, sin dejar de relacionarse con la teoría de la titulación de Howard Becker, la inducción analítica de Florian Znaniecki, la teoría fundamentada de Barney Glaser, o la etnografía básica de John Lofland, por citar algunos de ellos.

En la actualidad, la investigación cualitativa se entronca, desarrolla y aplica principalmente en psicología, educación, sociología, antropología, relaciones humanas, medicina y justicia; y la discapacidad, en cuyo ámbito nos interesa situar la metodología cualitativa, se entronca con todas estas áreas del conocimiento.

Y, finalmente, la discapacidad puede abordarse, a su vez, desde distintas vertientes en el marco de la metodología cualitativa. Los más radicales lo plantearían desde propuestas teóricas más o menos críticas al efectuar una construcción social de la discapacidad (Ferreira y Rodríguez Caamaño, 2006; Ferreira, 2007; Urraca, 2007), pero aquí se desarrollará desde una perspectiva empírica.

## CARACTERÍSTICAS DE LA METODOLOGÍA CUALITATIVA

Como consecuencia de los principios inspiradores indicados se derivan una serie de características que concretan los rasgos diferenciales que los identifican y que, según sea la técnica concreta de recogida de información, se manifiestan de forma más o menos acusada.

Las principales características son las siguientes:

- a) La fuente principal y directa de los datos son las situaciones naturales, ubicadas especialmente en el contexto de la cotidianidad (Anguera, 1999). Ningún fenómeno puede ser entendido fuera de sus referencias espacio-temporales y de su contexto.
- b) El investigador se convierte en el principal “instrumento” de recogida de datos, en el sentido de actor del proceso que implica la captación de la realidad, y con la capacidad para aportar datos tan fiables como los generados por medios más objetivos (Anguera, 2000b). Entre las principales ventajas que ello reporta, destacan: 1) Su adaptabilidad para registrar información simultánea sobre múltiples factores y a varios niveles. 2) Visión holística, es decir, capacidad para captar el contexto de forma global. 3) Mayor amplitud de conocimientos. 4) Posibilidad de explorar respuestas atípicas e idiosincráticas, las cuales son difíciles de captar por medios ordinarios, y tienen una enorme relevancia para lograr una más óptima comprensión.
- c) Incorporación del conocimiento tácito, es decir, el correspondiente a intuiciones, aprehensiones o sentimientos que no se expresan de forma lingüística pero que se refieren a aspectos conocidos de algún modo. Muchos de los matices de la realidad sólo pueden ser captados por esta vía, a la vez que muchas de las interacciones entre investigador e investigado ocurren en este nivel.
- d) Aplicación de técnicas de recogida de datos abiertas, por adaptarse mejor a las influencias mutuas y ser más sensibles para detectar patrones de comportamiento (Anguera, 1995d).
- e) Muestreo intencional. La selección de la muestra no pretende representar a una población con el objeto de generalizar los resultados, sino que se propone ampliar el abanico y rango de los datos tanto como sea posible, a fin de obtener la máxima información de las múltiples realidades que pueden ser descubiertas.
- f) Análisis inductivo de los datos. Ello implica una primera descripción de las situaciones de cada uno de los casos o eventos estudiados, con el fin de detectar progresivamente la existencia de unas regularidades entre ellos que constituyen la base o germen de una futura teoría adecuada a las condiciones y valores locales.



- g) La teoría se genera a partir de los datos de una realidad concreta, no partiendo de generalizaciones “a priori”. Autores relevantes como Goetz y LeCompte (1988) definen la teoría como: 1) generativa, por preocuparse por el descubrimiento de constructos y proposiciones, 2) inductiva –aunque pueda parecer una incongruencia–, pues las teorías se desarrollan desde abajo, a través de la interconexión de evidencias y datos recogidos, 3) constructiva, dado que las unidades de análisis comienzan a aparecer en el curso de la observación y descripción, 4) subjetiva, entendido como el propósito de reconstruir categorías específicas que los participantes utilizan para conceptualizar sus propias experiencias y su visión de la realidad.
- h) El diseño de la investigación es emergente y en cascada, ya que se va elaborando a medida que avanza la investigación. La situación generadora del problema da lugar a un cuestionamiento continuado y a una reformulación constante (Figura 1), en función de la incorporación de nuevos datos. Esta filosofía de “diseños no estandar” flexibiliza el estudio de forma acorde con la propia realidad y los datos que se obtienen, lo cual le aporta infinito número de posibilidades. Además, este tipo de diseños permite una adecuación a las múltiples realidades, a los contextos específicos y a las interacciones entre investigador y contexto. Un diseño previo prefijado relegaría la realidad vivencial.
- i) La metodología cualitativa se plantea criterios de validez específicos, utilizando técnicas propias que garantizan la credibilidad de los resultados (Sandín, 2003).

## EXIGENCIAS METODOLÓGICAS

De acuerdo con los apartados anteriores, puede surgir una seria duda sobre la calidad que se requiere a la metodología cualitativa, y, en consecuencia, a los datos que van a obtenerse mediante diferentes técnicas. Patton (1980, p. 22) ha señalado que “los datos cualitativos consisten en descripciones detalladas de situaciones, eventos, sujetos, interacciones y conductas observadas; citas directas de sujetos acerca de sus experiencias, actitudes, creencias y pensamientos; y fragmentos o pasajes enteros de documentos, correspondencia, registros e historias de casos”. Pero, ¿qué garantía ofrece este tipo de material?

Se pueden señalar diversos marcos de exigencias (Erickson, 1986; Zabalza, 1991) respecto a las condiciones de legitimidad metodológica:

**A) Representatividad, relevancia y plausibilidad de los datos (validez semántica de la investigación).** Este requisito implica identificar las diversas perspectivas de significación, proceder a una correcta contextualización, y constatar con objetividad los sucesos, o eventos, o conductas que se presentan. Ello significa –y en este sentido estamos de acuerdo con Zabalza (1991)– cumplir determinadas condiciones:

- a. Ampliar al máximo el contexto de análisis, de manera que puedan incorporarse a la situación analizada todas las variables, sujetos o elementos que puedan aportar un papel clarificador respecto al ámbito estudiado. De aquí que se recomiende la *triangulación*, o uso complementario de varias –al menos tres– modalidades distintas de recogida de datos, lo cual minimizará las carencias de alguna de ellas al complementarse con las virtudes de otras.
- b. Describir el propio proceso seguido en la obtención y análisis de la información. La investigación cualitativa tiene una naturaleza deliberativa, apenas existe estandarización, no existe uniformidad en los procesos seguidos, y algunos de los datos obtenidos pueden ser considerados confusos o parciales; la debilidad que ello comporta

exige en contrapartida que el investigador dé cuenta de cómo planteó y desarrolló cada uno de los pasos de su investigación, de manera que ésta pueda ser valorada, y, en su caso, replicada en otros contextos.

- c. Configurar la investigación como un auténtico proceso de búsqueda deliberativa, en el que se van presentando –y deben resolverse– cadenas de dilemas de muy diversos órdenes, y sobre los que hay que tomar decisiones (por ejemplo, sobre si anotar un determinado aspecto observado, si conceder o no relevancia a una determinada dimensión de un texto, etc.). Además, es frecuente la obtención de una inmensa cantidad de datos (sobre todo a nivel de descripciones casuísticas), que colocan frecuentemente al investigador ante una perplejidad decodificadora, y con verdaderos problemas al tratar de simplificar la información sin perder datos que pudiesen ser relevantes.

**B) Fundamentación teórica de la investigación y de los análisis e interpretaciones que incluye (validez hermenéutica de la investigación).** La dialéctica entre inductivismo y deductivismo adquiere un especial sentido en metodología cualitativa, pues no en vano se basa fundamentalmente en interpretaciones, o al menos acude frecuentemente a ellas para dar sentido a los datos e interpretaciones. Consideramos obvio que no se puede comprender sin un bagaje previo de preconcepciones, ni tampoco al margen del equipamiento interpretativo del que actúa de intérprete.

Pero hay que matizar que esta estructura o marco general de significaciones previas (en cuyo seno se puede entender la nueva situación o la nueva información) no constituye una barrera inflexible, sino que se trata de un conjunto de componentes cognitivos y experienciales que van alterando sus contenidos y filtros en base a la interacción con cada nuevo proceso de conocimiento e interpretación.

**C) Dinámica relacional de la investigación (validez pragmática).** Se incluyen esencialmente cuestiones relativas a condiciones metodológicas, optimización de los procesos que se estudian, y se salvaguardan los derechos de las personas que participan en la investigación. Se pueden concretar en los siguientes puntos:

- a. Indicación a los participantes de los objetivos del trabajo (lo cual no se cumple en los casos en que se obtiene un registro de conducta mediante la observación directa), y, en su caso, información sobre el proceso de negociación previa.
- b. Audiencias a las que se debe acudir, y que son los distintos colectivos, o comunidades, o individuos, relacionados con la cuestión que interesa. Resulta obvio considerar la incidencia del colectivo, sea cual sea, así como el papel de los propios participantes en la investigación como audiencia preferente.

## DIVERSIFICACIÓN

Existen unas características comunes entre sí, que tienen su razón de ser en la forma de obtener información de la situación estudiada y de su contexto, y todas ellas participan en común de las estrategias de investigación cualitativa (Riba, 2007).

Resulta obvio que en estudios empíricos realizados en el ámbito de la discapacidad se producen datos al traducir la realidad a sistemas de notación escrita. Pero surge una dicotomización (no una dicotomía real) en función del tipo de datos que se obtienen (cuantitativos vs. cualitativos), la cual, a su vez, se halla supeditada en buena medida a la propia naturaleza del problema.

Si se trata de un estudio de tiempos de reacción en la respuesta a un estímulo visual que se presenta en un monitor a una persona con discapacidad sensorial leve en visión, lógicamente la recogida de datos implicará determinado tipo de datos, seguramente expresados en unidades convencionales de tiempo (segundos, décimas de segundo, milisegundos, ...). Pero son muchas las temáticas de estudio en la discapacidad en las cuales se producen igualmente "datos", pero de los que no es posible su operativización, o no resulta factible sin incurrir en un grave reduccionismo; así, en programas de intervención en familias de discapacitados en los cuales se deba recabar la opinión de los allegados y la descripción de episodios diversos, ¿sería factible realizar el recuento de algún tipo de cantidad? No, dadas las múltiples manifestaciones del apoyo necesario, la borrosidad de algunas de ellas, la necesaria contextualización, y seguiría un largo etcétera.

Las diversas modalidades se caracterizan en común por las características de los datos, pero cabe establecer una gradación en función de su naturaleza, que a su vez da lugar a la propuesta taxonómica indicada en la Tabla 1, donde se evidencia la gradación que recoge el abanico de posibilidades que se disponen ordenadas entre los datos procedentes del registro de conductas perceptibles –que serán los más fáciles de codificar y cuantificar posteriormente (ver apartado 5)– y los propios de informes o "dossiers" documentales, en los que está más distante esta posibilidad, aunque se haya avanzado enormemente en la misma dirección en los últimos años. Y no obviamos la posición central que ocupa la conducta verbal transformable en material textual, proceso de transcripción que ha tenido una gran relevancia, pero que en la actualidad cada vez es menos importante, dadas las prestaciones de los actuales programas informáticos.

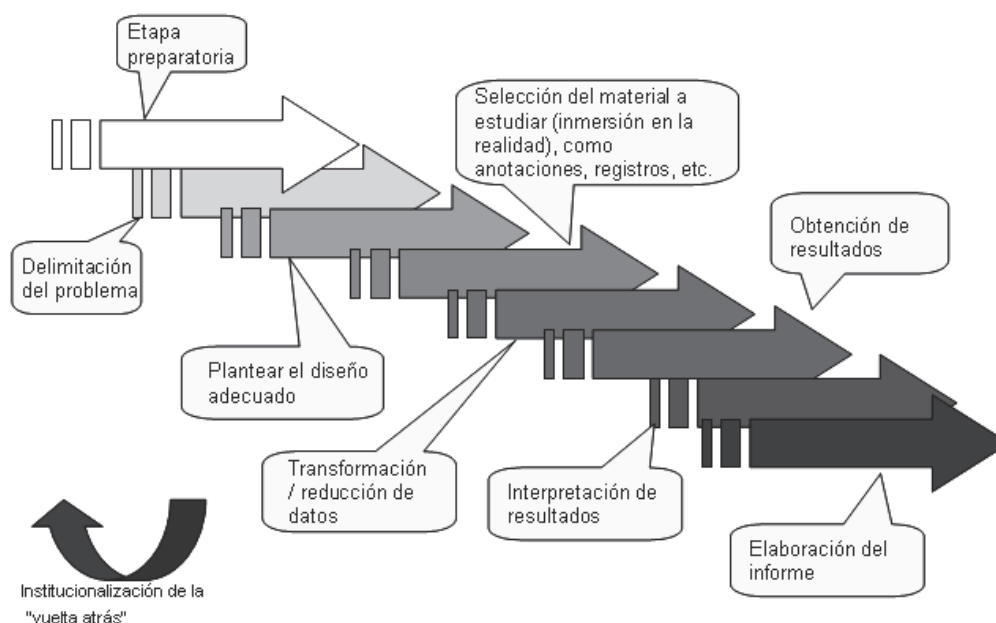
Tabla 1. Criterios taxonómicos en metodología cualitativa [Adaptado de Anguera (1995a, 1995b)]

OBSERVACION DIRECTA	Registros de conducta: Descriptivos, semisistematizados y sistematizados [Todos ellos admiten cualquier nivel de participación: observación no participante, observación participante, participación-observación, autoobservación de conductas heteroobservables].
CONDUCTA VERBAL TRANSFORMABLE EN MATERIAL DOCUMENTAL	Material grabable en audio (entrevistas en profundidad, grupos focales, conversaciones informales, monólogos, etc.)
MATERIAL DOCUMENTAL (OBSERVACIÓN INDIRECTA)	Textos originalmente producidos en forma documental (documentos personales, diarios de campo, registros biográficos, informes monográficos, etc.), que pueden incluir también información gráfica (fotografías, dibujos, animaciones, etc.).

## PROCESO METODOLÓGICO

La metodología cualitativa, como se indica en la Figura 1, consta de una serie de etapas en las cuales se halla institucionalizada la vuelta atrás. Ninguna de ellas se puede obviar y el procedimiento debe guiarse por un proceso continuado de decisiones del investigador (Rodríguez Gómez, Gil Flores y García Jiménez, 1996).

Figura 1. Proceso que implica la metodología cualitativa

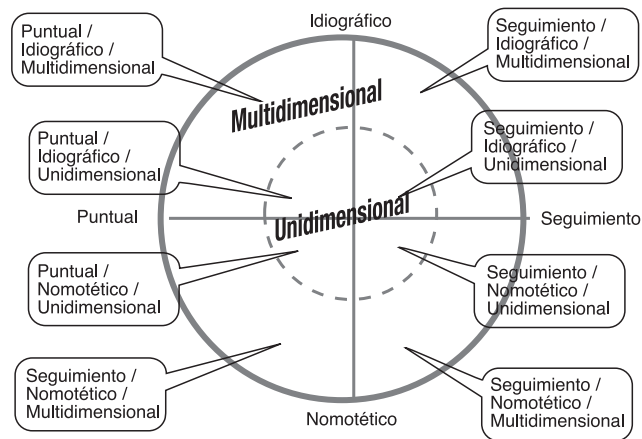


La esencia del proceso, una vez se ha planteado el diseño adecuado, consiste en seleccionar el material adecuado, plantear los criterios de segmentación de unidades, recoger la información de forma adecuada, transformarla, gestionarla, y analizarla, de forma que los resultados revierten de nuevo en el proceso (Anguera, 2000b).

Esquemáticamente se comentan los respectivos apartados:

- a) *Etapa preparatoria*. Establecimiento del marco teórico-conceptual desde el que parte la investigación, y dirigido especialmente a determinar el tópico de interés en función de las razones por las que se elige el tema (motivación personal, la propia vida cotidiana, experiencias concretas, práctica profesional, contraste con otros especialistas, lectura de trabajos de otros investigadores, etc.).
- b) *Delimitación del problema*. La acotación del problema implicará centrarse en algún aspecto específico del tema seleccionado. Es esencial que se efectúe de forma clara, sin ambigüedades, y dejando claras las tres vertientes que abarca su cobertura: Temática (problema de estudio acotado), espacial (ámbito geográfico o administrativo en el que se ubica) y temporal (referente relativo al período de tiempo que interesa). Por ejemplo, resolución de conflictos relativos al manejo de la vida cotidiana en padres de personas con discapacidad intelectual severa de una determinada ciudad a lo largo de un año.
- c) *Plantear el diseño adecuado*. El diseño es una pauta o guía que debe establecerse una vez está delimitado el problema de estudio, y que nos conducirá a lo largo del proceso empírico, especialmente en la obtención, gestión y análisis de datos. En metodología cualitativa los diseños se plantean con carácter flexible, y existen diversas taxonomías, en función de que se siga un planteamiento paralelo al de la metodología observacional (Anguera, Blanco y Losada, 2001), considerando los ocho diseños posibles (Figura 2), o bien se adopten posiciones más abiertas (Rodríguez Gómez, Gil Flores y García Jiménez, 1996; Flick, 2007).

Figura 2. Diseños observacionales (Extraído de Anguera, Blanco y Losada, 2001)



- d) *Selección del material a estudiar.* Se trata de llevar a cabo muestreos diversos –que habitualmente son de carácter no probabilístico, con las connotaciones negativas que ello supone respecto a la representatividad de la muestra y posterior generalización de los resultados– para los cuales se habrán debido decidir los criterios a tener en cuenta. Las decisiones se realizan a distintos niveles, que básicamente se refieren a los casos pero, a su vez, para cada uno de ellos se supeditarán otras decisiones, como en qué contexto, qué episodios, que fragmentos del texto en su caso, etc. La selección de material se efectúa de acuerdo con algún tipo de triangulación, que habitualmente es la triangulación de datos, pero que no quedan excluidos otros tipos (triangulación del investigador, teórica, metodológica o disciplinar). El proceso de muestreo se continúa habitualmente hasta que se llega a la saturación, cuando sucesivas muestras ya no aportan más información.
- e) *Transformación/reducción de datos.* Resulta básico, en primer lugar, el establecimiento de unidades, que se corresponde directamente con el planteamiento del muestreo, y, en este sentido, deberá diferenciarse entre unidad de análisis y unidad de codificación (Boyatzis, 1998), partiendo de la base de que la primera consiste en el mínimo segmento significativo de la realidad, mientras que la segunda se basa en el mínimo dato o información que puede ser evaluado y, por supuesto, nunca podrá tener mayor magnitud que la unidad de análisis. Una vez se especifican las unidades, se deberá proceder a la sistematización del registro, en aras a facilitar el proceso de gestión de la información. Si se trabaja con notas de campo, transcripciones provenientes de entrevistas no estructuradas, documentos históricos o algún otro material cualitativo, una tarea determinante es la preparación cuidadosa de la codificación (Miles & Huberman, 1984; Strauss & Corbin, 1990; Edwards & Lampert, 1993) mediante la imposición de alguna estructura en la mayor parte de la información. En primer lugar, al igual que en los estudios cuantitativos, es importante revisar que los datos estén completos, que tengan buena calidad y que estén en un formato que facilite su organización. Se debe confirmar que las transcripciones *textuales* en realidad lo sean, y que se hallen completas. La principal tarea en la organización de los datos cualitativos es desarrollar un método para indizar el material; por ejemplo, listados que relacionan los números de identificación de materia con otros tipos de información, como fechas y lugares de la recogida de datos. Todo registro, por ajustarse al objetivo previamente delimitado, implica una selección de las conductas consideradas relevantes, y en base



a sus características, a la técnica de registro elegida y a los recursos de que se dispone, deberá escogerse un sistema (escrito, oral, mecánico, automático, icónico, etc.) que facilite su simplificación y almacenamiento. Ahora bien, el plano en que se sitúa el registro es pobre e insuficiente si no se construye un instrumento de carácter no estándar (sistema de categorías, formato de campo, o la combinación entre ambos, al margen del carácter residual de las *rating scales*) que permita canalizar adecuadamente dicha información y codificarla adecuadamente, y en donde jugará un papel primordial la dimensionalidad del diseño (unidimensional vs. multidimensional) para elegir el tipo de instrumento más adecuado (Anguera, Magnusson & Jonsson, 2007).

- f) *Obtención de resultados.* La metodología cualitativa muestra su especial carencia en esta etapa del proceso.
- g) *Interpretación de resultados.* Los resultados deberán comentarse atendiendo al problema planteado, con referencia al marco teórico que se toma como referencia, y
- h) *Elaboración del informe.* Deberá realizarse de forma estándar, y siguiendo las directrices de la American Psychological Association (<http://www.apastyle.org/>).

## COMPLEMENTARIEDAD ENTRE LAS METODOLOGÍAS CUALITATIVA Y CUANTITATIVA

Mientras que tradicionalmente las dos grandes vertientes metodológicas han mantenido una pugna que dicotomizaba a sus seguidores, en los últimos años, por fortuna, se ha ido instaurando la cordura metodológica necesaria para iniciar una vía fructífera de complementariedad. Ha resultado un proceso duro y difícil, no exento de incomprendimientos desde ambas orillas respecto a intentos iniciales de colaboración (Anguera, 1985), pero que en la actualidad se halla en un proceso imparable que resulta sumamente fértil y prometedor.

El nudo gordiano de la cuestión se hallaba, por un lado, en que por parte de los seguidores de la *tradicional metodología cualitativa*, una vez se había recogido la información, sea en forma de notas de campo, diarios, transcripción de entrevistas en profundidad, etc., material valorado como de gran riqueza informativa, todo el esfuerzo se dedicaba unidireccionalmente a 'enlazar' datos, elaborar memorandos (comentarios reflexivos), analizar el contenido contando frecuencias (solamente parte de los investigadores cualitativos acceden a informaciones de frecuencia), generar mapas conceptuales, organizar matrices o redes para condensar la información recogida, proponer nuevas teorías a partir de las relaciones detectadas por inspección visual y la posterior reflexión, etc. (Huberman & Miles, 1991), pero siempre había una carencia metodológica latente en todas estas actividades, y era la subjetividad, o, dicho de forma edulcorada, la carencia de objetividad demostrada, debilidad que ni siquiera se situaba en el epicentro de las preocupaciones de los que realizaban investigaciones cualitativas, mucho más preocupados habitualmente por plasmar sus discursos reflexivos o por construir teoría. Pero, por otro lado, los seguidores de la *tradicional metodología cuantitativa* operaban exclusivamente con valores numéricos procedentes de medida mediante instrumentos estándar que cumplían los requisitos métricos, o de constatación del *iter* u orden, permitiendo descubrir, verificar o identificar relaciones simétricas o no entre conceptos que derivaban de un esquema teórico elaborado de acuerdo con los criterios que regían cada una de las situaciones que interesaba estudiar exigiendo, para llevar a cabo el contraste de la hipótesis, los requisitos de representatividad y aleatorización, lo cual comportaba unas adecuadas técnicas de muestreo, y después se llevaban a cabo técnicas cuantitativas de análisis que podían alcanzar elevada sofisticación; no obstante, siempre había una carencia metodológica latente en todas estas actividades, que era la omisión voluntaria de cualquier información procedente de la descripción de la realidad, fuesen episodios, documentos personales,

registros descriptivos, autoinformes, etc., a los cuales se los consideraba como 'material de chatarra' para el estudio del comportamiento y, por consiguiente, se despreciaba la riqueza de información que podían aportar.

Por supuesto, los planteamientos son divergentes, y mucho se ha discutido acerca de si comportan un diferente cuño epistemológico, así como de la necesaria elección del paradigma (Cook y Reichardt, 1986). Pero la cuestión que aquí nos interesa es hasta qué punto es factible aunar sus ventajas o puntos fuertes a la vez que se compensen sus carencias o puntos débiles.

La tarea no es fácil, pero si es factible, a nuestro humilde entender, y la clave se reduce a dos cuestiones fundamentales: Por una parte, cubrir las respectivas deficiencias de ambos colectivos de investigadores, considerando la extraordinaria riqueza que aportan los registros de conducta, la conducta verbal transformable a material documental, y los materiales documentales (Tabla 1), con lo cual alcanzarán el rango de 'datos' (Anguera, 2000a, 2004), a la vez que se fortalecen todas las cautelas y acciones metodológicas tendentes a controlar la calidad del dato (Blanco, 1989, 1991, 1992, 1993, 2001; Blanco y Anguera, 2000, 2003; Blanco, Castellano y Hernández Mendo, 2000; Blanco y Hernández Mendo, 1998; Blanco y Losada, 2002; Blanco, Losada y Anguera, 1991) con el fin de garantizar la necesaria objetividad propia de cualquier estudio científico. Y, por otra parte, aprovechar la pertinencia de los instrumentos no estándar (Bakeman y Gottman, 1989; Anguera y Blanco, 2006; Anguera, Magnusson y Jonsson, 2007) que, mediante los sistemas de categorías (en los diseños unidimensionales) y los formatos de campo (en los diseños multidimensionales), así como la combinación entre ambos (en los diseños multidimensionales), dotan a los investigadores de importantes herramientas que posibilitan la transformación de datos cualitativos a datos tratables analíticamente de forma cuantitativa.

En este sentido, la metodología observacional ha sido pionera en lograr esta complementariedad (Anguera, 2004, 2005; Anguera & Izquierdo, 2006), desde su origen –considerado tradicionalmente como una metodología cualitativa (Blanco, 1997)– hasta los desarrollos analíticos posteriores (Blanco, Losada y Anguera, 2003), que prácticamente no presentan ninguna limitación respecto a los tradicionales estudios cuantitativos.

Con estas nuevas consideraciones metodológicas, se supera la crítica de la endémica debilidad metodológica de los estudios cualitativos, a la vez que se consiguen importantes avances consistentes en el uso de recursos metodológicos sofisticados que permiten un rigor mucho más elevado. Una idea básica y esencial que forma parte de esta nueva concepción de la metodología cualitativa en su complementariedad con la cuantitativa consiste en que, en función del diseño planteado y de la naturaleza de los datos, procederá una u otra técnica analítica (Anguera, Blanco y Losada, 2001), aunque originariamente hubiese estado reservada únicamente a estudios cuantitativos (Anguera, Blanco, Losada y Sánchez-Algarra, 1999).

En cualquier caso, si la metodología cualitativa nos ayudó en la obtención de datos que aportan una gran riqueza informativa, la cuantitativa nos suministra los recursos para su análisis más conveniente. La lógica sucesión de etapas de forma organizada debe permitir este cambio de perspectiva (de *lo* cualitativo a *lo* cuantitativo) de forma pacífica, sin tensión interna en el seno del procedimiento a seguir.

Ambas vertientes metodológicas pueden beneficiarse mutuamente entre sí, y son muchas las ocasiones en que se utilizan de manera conjunta, dando garantía de su complementariedad. Es cierto que en ocasiones esta opción presenta graves problemas por su costo en tiempo y dinero, o por falta de personal preparado el efecto, pero en cualquier caso se trata de superar la posición enfrentada de ambas perspectivas.

El empleo conjunto de la metodología cualitativa y de la cuantitativa, dado que se interesa por el proceso y el resultado, potencia la vigorización mutua de los dos tipos de procedimientos, y facilita la triangulación a través de operaciones convergentes (Cook y Reichardt, 1979). Sin embargo, no podemos eludir la coincidencia de diversos autores estudiosos de la cuestión al considerar que buena parte de las técnicas de recogida de datos son propias de una determinada metodología, o, lo que es lo mismo, que existen instrumentos tanto cualitativos como cuantitativos (Hernández López, 1995; Anguera, 1995c). Así, y en una primera aproximación, las entrevistas en profundidad, técnicas etnográficas, análisis histórico o historias de vida, son propias de la metodología cualitativa, mientras que indicadores estadísticos, observación sistemática, escalas de apreciación o cuestionarios, lo son de la cuantitativa. No obstante, se puede elaborar una sistemática más completa acerca de la naturaleza de las diversas técnicas de recogida de datos, así como de las posibilidades de ser utilizadas desde una u otra metodología (Cook y Reichardt, 1979; Marshall & Rossman, 1989).

Es innegable que habrá estudiosos y profesionales que, genéricamente, manifiestan una preferencia marcada y casi excluyente por la metodología cuantitativa, mientras que otros prefieren la cualitativa. *Pero cada vez es mayor el número de profesionales e investigadores que optan por la combinación de ambos planteamientos, contemplando la utilización de técnicas propias de uno y otro conforme a las características del estudio a realizar.* Nosotros igualmente nos pronunciamos en este sentido, considerando que debería intentarse una redefinición del debate, y eliminando lo que durante décadas se ha propuesto como necesaria elección del paradigma. Un investigador no tiene por qué adherirse ciegamente a uno de ambos paradigmas, sino que puede elegir libremente una relación de atributos que indistintamente provengan de uno u otro si así se logra una adaptación flexible a su problemática.

A ello ayuda también el hecho de que cada vez sea mayor el número de situaciones en que un equipo multidisciplinar, a partir de una pluralidad de técnicas, trata de aunar esfuerzos en aras a una mayor rigurosidad de la evaluación realizada. El camino está cada vez más despejado, pero todavía requerirá considerables esfuerzos en el futuro para consolidar nuevas posibilidades de colaboración, e incluso de integración, si atendemos a la propuesta de Bericat (1998), que considera la actitud positiva y recíproca entre ambas metodologías como un paso más allá de la “legítima y reconocida convivencia” (p. 31).

Finalmente, es de justicia señalar el importante papel que han desempeñado las nuevas tecnologías (Sulé Duesa, 2004) y los avances informáticos, no solamente en el avance sustantivo, sino también en esta tarea de acercamiento entre las posiciones metodológicas cualitativa y cuantitativa, si bien es cierto que existen programas que siguen los planteamientos conceptuales más radicales (Weitzman & Miles, 1995; Lewins & Silver, 2007). Es muy amplio el elenco de programas de que se dispone, así como la inusitada velocidad con que aparece nuevo *software* en el mercado y en la comunidad científica. En la actualidad ya no nos sorprende ni que se realicen determinados análisis estadísticos a partir de textos (Woods, Flechter & Hughes, 1996; Oakes, 1998; Lébart, Salem y Bécue Bertaut, 2000), ni que determinados programas, como el Hyperbase (Brunet, 2006), puedan llevar a cabo análisis de conglomerados o análisis factorial, o que otros que tradicionalmente se han aplicado a registros sistemáticos de conducta, como el Thème (Magnusson, 1996, 2000), puedan extraer T-Patterns a partir de la transcripción de sesiones de psicoterapia (Blanchet, Batt, Trognon & Masse, 2005).



## REFERENCIAS BIBLIOGRÁFICAS

- Álvarez, J.J. (1986). Investigación cuantitativa / Investigación cualitativa: ¿Una falsa disyuntiva? En T.D. Cook y CH.S. Reichardt (Eds.) *Métodos cualitativos y cuantitativos en investigación evaluativa* (pp. 9-23). Madrid: Morata.
- Anguera, M.T. (1985) Posibilidades de la metodología cualitativa vs. cuantitativa. *Revista de Investigación Educativa*, 3 (6), 127-144.
- Anguera, M.T. (1986) Investigación cualitativa. *Educar*, 10, 23-50.
- Anguera, M.T. (1995a). Metodología cualitativa. En M.T. Anguera, J. Arnau, M. Ato, M.R. Martínez, J. Pascual y G. Vallejo. *Métodos de investigación en Psicología* (pp. 513-522). Madrid: Síntesis.
- Anguera, M.T. (1995b). Recogida de datos cualitativos. En M.T. Anguera, J. Arnau, M. Ato, M.R. Martínez, J. Pascual y G. Vallejo. *Métodos de investigación en Psicología* (pp. 523-547). Madrid: Síntesis.
- Anguera, M.T. (1995c). Tratamiento cualitativo de datos. En M.T. Anguera, J. Arnau, M. Ato, M.R. Martínez, J. Pascual y G. Vallejo. *Métodos de investigación en Psicología* (pp. 549-576). Madrid: Síntesis.
- Anguera, M.T. (1995d). Metodología de la evaluación: Evaluación cualitativa frente a evaluación cuantitativa. En Equipo del Gabinete Psicotécnico Municipal de Torrent (Recop.) *La evaluación ... ¿proceso final?* (pp. 27-36). Torrent: Ajuntament de Torrent.
- Anguera, M.T. (1999). *Hacia una evaluación de la actividad cotidiana y su contexto: ¿Presente o futuro para la metodología?* Discurso de ingreso como académica numeraria electa. Barcelona: Real Academia de Doctores. [Reimpreso (2001) en A. Bazán Ramírez y A. Arce Ferrer (Eds.), *Estrategias de Evaluación y Medición del Comportamiento en Psicología* (pp. 11-86). México: Instituto Tecnológico de Sonora y Universidad Autónoma de Yucatán].
- Anguera, M.T. (2000a). Cuantificación no reduccionista en Ciencias del Comportamiento: Anverso y reverso de una compleja realidad. En E. Oñate, F. García-Sicilia y L. Ramallo (Eds.), *Métodos numéricos en Ciencias Sociales* (pp. 1-29). Barcelona: Centro Internacional de Métodos Numéricos en Ingeniería (CIMNE).
- Anguera, M.T. (2000b). Del registro narrativo al análisis cuantitativo: Radiografía de la realidad perceptible. En *Ciencia i cultura en el segle XXI. Estudis en homenatge a Josep Casajuana* (pp. 41-71). Barcelona: Reial Academia de Doctors.
- Anguera, M.T. (2004). Posición de la metodología observacional en el debate entre las opciones metodológicas cualitativa y cuantitativa. ¿Enfrentamiento, complementariedad, integración? *Psicologia em Revista* (Belo Horizonte, Brasil), 10 (15), 13-27.
- Anguera, M.T. (2005). Microanalysis of T-patterns. Analysis of symmetry/asymmetry in social interaction. In L. Anolli, S. Duncan, M. Magnusson & G. Riva (Eds.), *The hidden structure of social interaction. From Genomics to Culture Patterns* (pp. 51-70). Amsterdam: IOS Press.
- Anguera, M.T. y Blanco, A. (2006). ¿Cómo se lleva a cabo un registro observacional? *Butlletí La Recerca*, 4 [<http://ice.d5.ub.es/recerca/butlleti/butlleti.html> / <http://ice.d5.ub.es/recerca/fitxes/fitxa4-cat.htm>].
- Anguera, M.T., Blanco, A. y Losada, J.L. (2001). Diseños observacionales, cuestión clave en el proceso de la metodología observacional. *Metodología de las Ciencias del Comportamiento*, 3 (2), 135-160.
- Anguera, M.T., Blanco, A., Losada, J.L. y Sánchez-Algarra, P. (1999). Análisis de la competencia en la selección de observadores. *Metodología de las Ciencias del Comportamiento*, 1 (1), 95-114.
- Anguera, M.T. & Izquierdo, C. (2006). Methodological approaches in human communication. From complexity of situation to data analysis. In G. Riva, M.T. Anguera, B.K. Wiederhold & F. Man-

- tovani (Coord.), *From Communication to Presence. Cognition, Emotions and Culture towards the Ultimate Communicative Experience* (pp. 203-222). Amsterdam: IOS Press.
- Anguera, M.T., Magnusson, M.S. y Jonsson, G.K. (2007). Instrumentos no estándar. *Avances en medicina*, 5 (1), 63-82.
- Bakeman, R. y Gottman, J.M. (1989). *Observación de la interacción: Introducción al análisis secuencial*. Madrid: Morata.
- Benoliel, J.Q. (1984). Advancing nursing science: Qualitative approaches. *Western Journal of Nursing Research*, 6, 1-8.
- Bericat, E. (1998). *La integración de los métodos cuantitativo y cualitativo en la investigación social. Significado y medida*. Barcelona: Ariel.
- Blanchet, A., Batt, M., Trognon, A. & Masse, L. (2005). Language and behaviour patterns in a therapeutic interaction sequente. In L. Anolli, S. Duncan, M.S. Magnusson & G. Tiva (Eds.), *The hidden structure of interactions: From neurons to culture patterns* (pp. 124-139). Amsterdam: IOS Press.
- Blanco, A. (1983). *Análisis cuantitativo de la conducta en sus contextos naturales: Desarrollo de un modelo de series de datos para el establecimiento de tendencias, patrones y secuencias*. Tesis Doctoral no publicada. Barcelona: Universidad de Barcelona.
- Blanco, A. (1989) Fiabilidad y generalización de la observación conductual. *Anuario de Psicología*, 43 (4), 5-32.
- Blanco, A. (1991) La teoría de la generalizabilidad aplicada a diseños observacionales. *Revista Mexicana de Análisis de la Conducta / Mexican Journal of Behavior Analysis (México)*, 17 (3), 23-63.
- Blanco, A. (1992). Aplicaciones de la teoría de la generalizabilidad en la selección de diseños evaluativos. *Bordón*, 43 (4), 431-459.
- Blanco, A. (1993) Fiabilidad, precisión, validez y generalización de los diseños observacionales. En M.T. Anguera (Ed.) *Metodología observacional en la investigación psicológica* (pp. 149-261). Barcelona: P.P.U., Vol. II.
- Blanco, A. (1997). *Metodologies qualitatives en la investigació psicològica*. Barcelona: Edicions de la Universitat Oberta de Catalunya.
- Blanco, A. (2001). Generalizabilidad de observaciones uni y multifaceta: Estimadores LS y ML. *Metodología de las Ciencias del Comportamiento*, 3 (2), 161-193.
- Blanco, A. y Anguera, M.T. (2000). Evaluación de la calidad en el registro del comportamiento: Aplicación a deportes de equipo. En E. Oñate, F. García-Sicilia y L. Ramallo (Eds.), *Métodos Numéricos en Ciencias Sociales* (pp. 30-48). Barcelona: Centro Internacional de Métodos Numéricos en Ingeniería.
- Blanco, A. y Anguera, M.T. (2003). Calidad de los datos registrados en el ámbito deportivo. En A. Hernández Mendo (Coord.), *Psicología del Deporte (Vol. 2). Metodología* (p. 35-73). Buenos Aires: Efdportes ([www.efdeportes.com](http://www.efdeportes.com)).
- Blanco, A., Castellano, J. y Hernández Mendo, A. (2000). Generalizabilidad de las observaciones de la acción del juego en el fútbol. *Psicothema*, 12 (Supl. 2), 81-86.
- Blanco, A. y Hernández Mendo, A. (1998). Estimación y generalización en un diseño de estructura espacial defensiva en el fútbol. En J. Sabucedo, R. García Mira, E. Ares y D. Prada (Coords.), *Medio ambiente y responsabilidad humana* (pp. 579-583). A Coruña: Tórculo.
- Blanco, A. y Losada, J.L. (2002). Estimación y generalización en diseños observacionales: Una aplicación iterativa de máxima verosimilitud. *Metodología de las Ciencias del Comportamiento, vol. especial*, 85-90.
- Blanco, A., Losada, J.L. y Anguera, M.T. (1991) Estimación de la precisión en diseños de evaluación ambiental. *Evaluación Psicológica / Psychological Assessment*, 7 (2), 223-257.

- Blanco, A., Losada, J.L. y Anguera, M.T. (2003). Analytic techniques in observational designs in environment-behavior relation. *Medio Ambiente y Comportamiento Humano*, 4 (2), 111-126.
- Blaxter, M. (1979). Symposium on the handling of qualitative data. Introduction. *Sociological Review*, 27 (4), 649-650.
- Boyatzis, R.E. (1998). *Transforming qualitative information*. Thousand Oaks : Sage.
- Brunet, E. (2006). *Hyperbase. Logiciel hypertexte pour le traitement documentaire et statistique des corpus textuels*. Document non publié.
- Bulmer, M. (1979). Concepts in the analysis of qualitative data. *Sociological Review*, 27 (4), 651-677.
- Cook, T.D. y Reichardt (Eds.) (1979). *Métodos cualitativos y cuantitativos en investigación evaluativa*. Madrid: Morata.
- Denzin, N.K. (1970). *The research act: A theoretical introduction to sociological methods*. New York: McGraw-Hill.
- Edwards, J.A. & Lampert, M.D. (1993). *Talking data. Transcription and coding in discourse research*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Erickson, F. (1986). Qualitative methods in research on teaching. In M.C. Wittrock (Edit.). *Handbook of research on teaching*. New York; McMillan.
- Ferreira, M. (2007). Prácticas sociales, identidad y estratificación: tres vértices de un hecho social, la discapacidad. *Intersticios. Revista Sociológica de Pensamiento Crítico* [Revista en línea], 1 (2). Disponible en: <http://www.intersticios.es/article/view/1084/854> [Consulta: 03/05/2008].
- Ferreira, M. y Rodríguez Caamaño, M.J. (2006). Sociología de la discapacidad: una propuesta teórica crítica. *Nómadas. Revista crítica de Ciencias Sociales y Jurídicas* [Revista en línea], 13, 1. Disponible en: <http://www.ucm.es/info/nomadas/13/ferreiracaama%F1o.html> [Consulta: 03/05/2008].
- Flick, U. (2007). *Introducción a la investigación cualitativa*. Madrid: Morata.
- Goetz, J. y LeCompte, M. (1988). *Etnografía y diseño cualitativo en investigación educativa*. Madrid: Morata.
- Hernández López, J.M. (1995). Procedimientos de recogida de información en evaluación de programas. En R. Fernández-Ballesteros (Ed.), *Evaluación de programas. Una guía práctica en ámbitos sociales, educativos y de salud* (pp. 117-147). Madrid: Síntesis.
- Huberman, A.M. & Miles, M.B. (1991). *Analyse des données qualitatives. Recueil de nouvelles méthodes*. Bruxelles: De Boeck-Wesmael.
- Lébart, L., Salem, A. y Bécue Beraut, M. (2000). *Anàlisis estadístico de textos*. Lleida: Milenio.
- Lewins, A. & Silver, C. (2007). *Using software in qualitative research*. Los Angeles: Sage.
- Lofland, J. (1971). *Analyzing social settings*. Belmont, Ca.: Wadsworth.
- Magnusson, M.S. (1996). Hidden real-time patterns in intra- and inter-individual behavior. *European Journal of Psychological Assessment*, 12 (2), 112-123.
- Magnusson, M.S. (2000). Discovering hidden time patterns in behavior: T-patterns and their detection. *Behavior Research Methods, Instruments, & Computers*, 32 (1), 93-110.
- Marshall, C. & Rossman, G.B. (1989). *Designing qualitative research*. Newbury Park: Sage.
- Miles, M.B. & Huberman, A.M. (1984). *Qualitative data analysis*. Beverly Hills: Sage.
- Morse, J.M. (1994). Qualitative research: Fact or fantasy? In J.M. Morse (Ed.) *Critical issues in qualitative research methods* (pp. 1-9). Thousand Oaks: Sage.
- Nisbett, R.E. & Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84 (3), 231-259.

- Norris, S.P. (1984). Defining observational competence. *Service Education*, 68 (2), 129-142.
- Oakes, M.P. (1998). *Statistics for corpus linguistics*. Edinburgh: Edinburgh University Texts.
- Patton, M.Q. (1980). *Qualitative Evaluation Methods*. Beverly Hills: Sage.
- Riba, C.E. (2007). *La metodología cualitativa en l'estudi del comportament*. Barcelona: UOC.
- Rodríguez Gómez, G., Gil Flores, J. y García Jiménez, E. (1996). *Metodología de la investigación cualitativa*. Archidona: Aljibe.
- Sandín, M.P. (2003). *Investigación cualitativa en Educación. Fundamentos y tradiciones*. Madrid: McGraw-Hill.
- Smith, J.K. (1983). Quantitative versus interpretative: The problem of conducting social inquiry. In E.R. House (Ed.) *Philosophy of evaluation* (pp. 27-51). San Francisco: Jossey-Bass.
- Schwartz, H. y Jacobs, J. (1984). *Sociología cualitativa. Método para la reconstrucción de la realidad*. México: Trillas.
- Strauss, A. & Corbin, J. (1990). *Basics of qualitative research. Grounded theory procedures and techniques*. Newbury Park: Sage.
- Sulé Duesa, A. (2004). Tecnologías de la informació (barates) aplicades a persones amb necessitats especials. *BiD: textos universitaris de biblioteconomia i documentació* [Revista en-línea], 12. Disponible en: [http://www2.ub.es/bid/consulta\\_articulos.php?fichero=12sule2.htm](http://www2.ub.es/bid/consulta_articulos.php?fichero=12sule2.htm). [Consulta: 03/05/2008].
- Szent-Györgyi, A. (1980). Dionysians and apollonians. In M.R. Key (Ed.) *The relationship of verbal and non-verbal communication* (pp. 317-318). New York: Mouton.
- Urraco, M. (2007). La metodología cualitativa para la investigación en Ciencias Sociales: una 'mediografía' orientativa. Un nuevo modelo teórico de la discapacidad: aplicación a la optimización de las tiflotecnologías. Disponible en: [http://www.um.es/discatif/PROYECTO\\_DISCATIF/Documentos/metodoloia\\_cualitativa.pdf](http://www.um.es/discatif/PROYECTO_DISCATIF/Documentos/metodoloia_cualitativa.pdf) [Consulta: 03/05/2008].
- Valles, M. S. & Baer, A. (2005, Septiembre). Investigación social cualitativa en España: Pasado, presente y futuro. Un retrato [62 párrafos]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [Revista en-línea], 6(3), Art. 18. Disponible en: <http://www.qualitative-research.net/fqs-texte/3-05/05-3-18-s.htm> [Consulta: 02/05/2008].
- Weitzman, E.A. & Miles, M.B. (1995). *Computer programs for qualitative data analysis. A software sourcebook*. Thousand Oaks: Sage.
- Woods, A., Flechter, P. & Hughes, A. (1996). *Statistics in language studies*. Cambridge: Cambridge University Press.
- Zabalza, M.A. (1991). *Los diarios de clase. Documento para estudiar cualitativamente los dilemas prácticos de los profesores*. Barcelona: P.P.U.



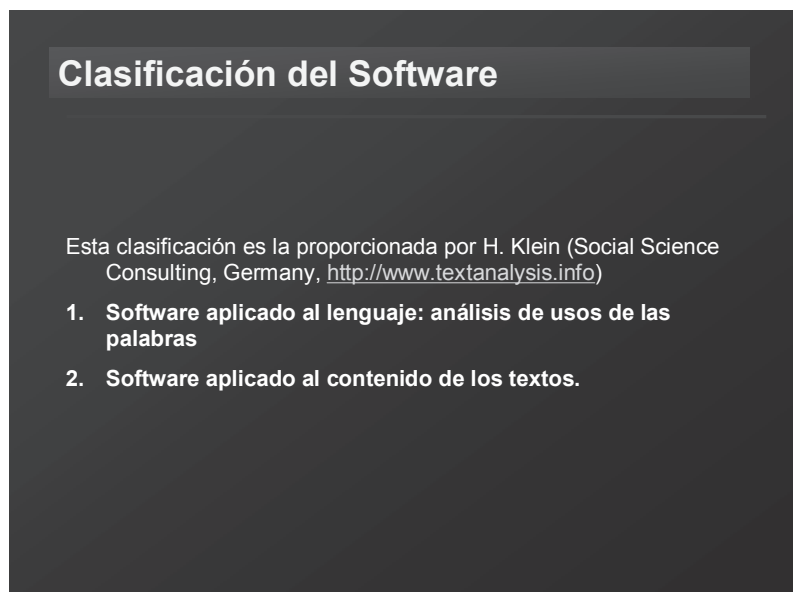
# INFORMÁTICA Y ANÁLISIS CUALITATIVO

M<sup>a</sup> CRUZ SÁNCHEZ GÓMEZ

*Departamento de Didáctica, Organización y Métodos de Investigación  
Facultad de Educación (Universidad de Salamanca)*

Como en todas las innovaciones, la utilización del ordenador en este tipo de análisis genera posiciones encontradas entre quienes lo han incorporado y pretenden ver un mundo inacabable de posibilidades, y los partidarios de continuar haciendo las cosas a mano. Los primeros llegan a plantear esta actividad como un área principal de especialización dentro del análisis, como lo es la estadística en los análisis cuantitativos.

Imagen 1. Clasificación de software de análisis textual



Desde el principio mismo del desarrollo de la investigación cualitativa ha comenzado un debate sobre las repercusiones de los programas informáticos en ella. Había diferentes preocupaciones. En primer lugar, algunos de los programas más importantes se desarrollaron sobre el trasfondo de un enfoque específico –la codificación según la teoría fundamentada– y son más difíciles de aplicar a otros enfoques. Si los programas no encajan en una interpretación más secuencial de los datos ¿simplemente ignoran ésto los investigadores que utilizan este enfoque?, ¿cambia ello la forma de analizar los datos o lleva más a menudo a algún tipo de enfoque común, una norma áurea de la investigación cualitativa? (Coffey y cols., 1996).



Por una parte, Fielding y Lee (1998) encontraron en su estudio empírico sobre el uso de programas que dos tercios de los proyectos revisados no utilizaban la teoría fundamentada pero usaban programas de uso en la investigación cualitativa. Esto muestra que la relación entre los programas informáticos y la teoría fundamentada no es tan estrecha como sospechan algunos autores.

Ten Have (1999), por otra parte, muestra cómo se puede aplicar este programa al análisis de conversaciones. Otra preocupación es que los programas informáticos fueren implícitamente su estructura lógica y de exposición sobre los datos y el análisis del investigador. El programa NUDIST, por ejemplo, apoya el desarrollo de una estructura de códigos jerárquica en forma de árbol. Entre sus usuarios, se puede encontrar una cierta inflación de sistemas de codificación estructurados en forma de árbol. Seale (2000) da una buena ilustración de este problema, cuando aplica NUDIST y ATLAS-ti a una teoría fundamentada desarrollada por Glaser y Strauss, y muestra lo diferentes que parecen la presentación y la estructura de esta teoría en ambos programas.

Por último, hay un temor a que la atención atraída por el ordenador y los programas distraigan al investigador del trabajo analítico real: leer y entender los textos, pensar de manera contemplativa sobre lo que está en los textos y lo que subyace a ellos, etc., (por ejemplo, Lee y Fielding, 1991).

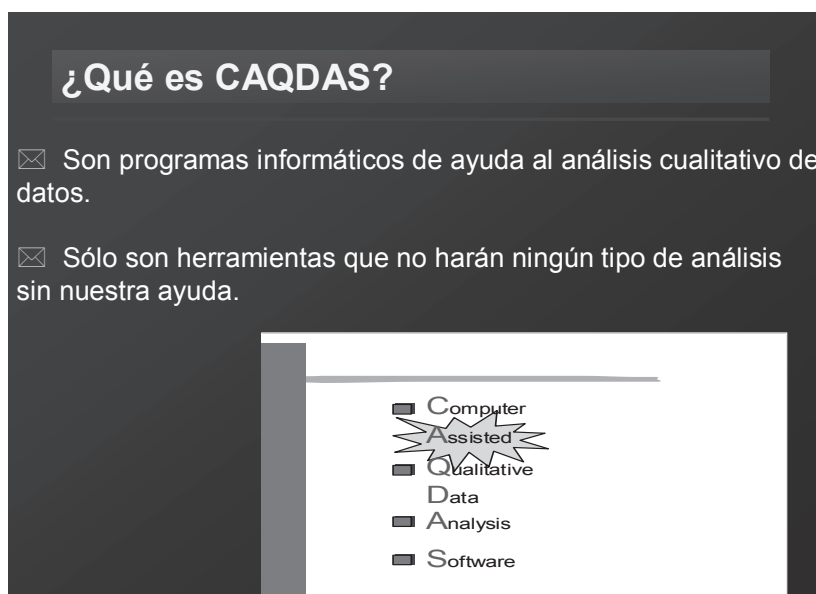
Del mismo modo, Richards y Richards (1994), que han desarrollado uno de los programas más importantes (NUDIST), afirman que *“el método informático puede tener implicaciones espectaculares para el proceso y los resultados de la investigación, desde restricciones inaceptables en los análisis a la apertura inesperada de posibilidades”*. Pero al final, depende del usuario y de su forma de hacer útil el ordenador y los programas para la investigación en curso, y de cómo reflexione sobre lo que está haciendo. Así, los ordenadores y los programas informáticos se deben ver como una herramienta pragmática de apoyo a la investigación cualitativa. Sobre sus repercusiones en la propia investigación deben reflexionar sus usuarios. No conviene sobrecargarlos de esperanzas y expectativas.

Al menos en teoría, cualquier forma de análisis cualitativo podría ser asistida por ordenador a través de programas informáticos (CAQDAS: Computer Assisted Qualitative Data Analysis Software) creados a tal efecto, que facilitan el manejo mecánico de los datos: la búsqueda, selección y organización. Programas como ETHNOGRAPH, AQUAD, NUDIST O ATLAS-Ti, resultan sumamente útiles en el análisis, para realizar operaciones como el marcado y codificación del texto, la relación de categorías y sujetos, la elaboración de tipologías o perfiles, o el recuento, búsqueda y recuperación de unidades codificadas, siendo especialmente útiles cuando se trabaja simultáneamente con grandes cantidades de información (Gil-García y otros, 2002).

Los CAQDAS nos ayudan a sintetizar, ordenar y organizar la información recogida para presentar los resultados de la investigación. El tratamiento informático de los datos ayuda al investigador a hacerse una visión de conjunto del objeto de estudio. La ingente cantidad de datos que pueden obtenerse de protocolos de observación, técnicas de registro, observaciones etnográficas, etc., deben ser tratados con herramientas potentes que permitan el aprovechamiento de esta información.

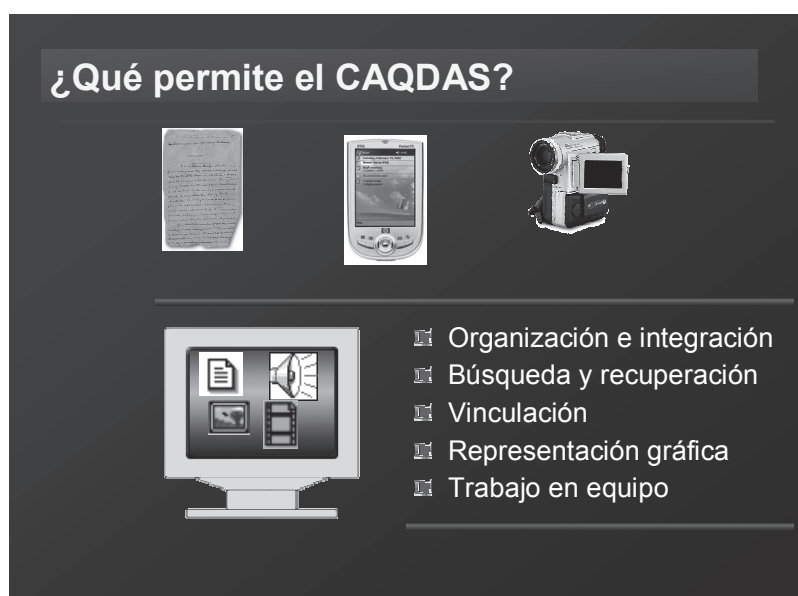
Estos instrumentos de origen multimedia, debido al diseño de los mismos, han permitido a los creadores de los CAQDAS la actualización de sus prestaciones, incluyendo el análisis de datos textuales, gráficos, sonoros y de video que los hacen especialmente interesantes para tratar la información variada obtenida de diversas fuentes en contextos multiculturales.

Imagen 2. CAQDAS



Fuente: Adaptado de Muñoz Justicia (2004)

Imagen 3. Utilidad de los CAQDAS



Fuente: Adaptado de Muñoz Justicia (2004)

Flick (2004), Miles y Huberman (1944) y Weitzman (2000) afirman que, si bien la mayor parte de los programas informáticos en la investigación cualitativa se utilizan para analizar datos, hay también otros pasos en este proceso de investigación para los que pueden utilizarse los ordenadores. En general se pueden utilizar para:

1. Tomar notas en el campo.
2. Pasar a limpio o transcribir notas de campo.



3. Editar: corregir, ampliar o revisar notas de campo.
4. Codificar: agregar palabras clave o etiquetas a segmentos del texto para permitir su recuperación posterior.
5. Almacenamiento: mantener el texto en una base de datos organizada.
6. Búsqueda y recuperación: localizar segmentos relevantes de texto y hacer que estén disponibles.
7. Enlazar datos: conectar segmentos relevantes de datos formando categorías, grupos o redes de información.
8. Elaborar memorandos: escribir comentarios reflexivos sobre algunos aspectos de los datos como base para un análisis más profundo.
9. Análisis del contenido: contar frecuencias, sucesión o localizaciones de palabras y frases.
10. Exposición de datos: colocar los datos seleccionados o reducidos en un formato condensado, organizado, como una matriz o una red.
11. Extracción de conclusiones y verificación: ayudar al analista a interpretar los datos expuestos y a someter a prueba o confirmar los hallazgos.
12. Construir teorías: desarrollar explicaciones sistemáticas coherentes conceptualmente de los hallazgos; examinar hipótesis.
13. Trazar gráficos: crear diagramas que representen los hallazgos o teorías.
14. Escribir informes: provisionales y finales.
15. La transcripción de entrevistas, la redacción de un diario de investigación, la comunicación con otros investigadores por medio de redes informáticas, el correo electrónico, el uso de Internet y la redacción de artículos sobre la propia investigación o sus métodos.

La mayor parte de estas actividades se pueden realizar con los procesadores de textos. Pero, como señala Seale (2000) el uso de procesadores de textos para estos propósitos lleva mucho más tiempo que la aplicación de programas informáticos especiales. Algunos de los CAQDAS más avanzados son capaces de gestionar, almacenar y mostrar material audio y visual –fotografías, películas, textos grabados y material de vídeo– e incluirlo en procedimientos analíticos, lo que va más allá de lo que pueden hacer los procesadores de textos.

Desde finales de los años setenta ha comenzado un debate sobre las repercusiones de los programas informáticos en la investigación cualitativa. A pesar de estas ayudas, algunos autores han alertado sobre las grandes limitaciones que entraña el análisis asistido por ordenador, como por ejemplo la segmentación de la información, la pérdida de la visión de conjunto y la descontextualización (Mercado, 2000). También existe el riesgo de quedarse atrapado por la herramienta, reduciendo el análisis a simples operaciones mecánicas y perdiendo de vista el verdadero objeto de análisis, que es la interpretación de los datos, una operación que difícilmente puede hacer un programa informático por sí solo.

El ordenador constituye una ayuda importante con un gran potencial para la realización de las tareas de rutina y como apoyo a la elaboración conceptual y teórica (Vallés, 1997), pero no puede reemplazar la capacidad del investigador.

Otra cosa que no puede hacer el ordenador es asegurar la validez de la investigación: de hecho, ningún procedimiento mecánico es capaz de distinguir si una investigación es mejor

que otra. González Martínez (2002) afirma que una investigación cualitativa podría considerarse válida sólo en función del grado de exactitud en la descripción de la perspectiva de los sujetos estudiados, la honestidad en el logro de la finalidad para la que se llevó a cabo, la adecuación de las herramientas y procedimientos utilizados durante todas las fases del estudio, y la capacidad de corroborar con evidencias las conclusiones, que deben ser creíbles para los beneficiarios de la investigación, como poníamos de manifiesto en el apartado de calidad de la investigación cualitativa. Por último decir, como Denzin (1994), que cada investigador interpreta sus datos de acuerdo con el paradigma del que proviene. Aquí volvemos a los comienzos de la investigación (proceso circular).

Como indica Richard, citado por Ruiz Olabuénaga (1999), es importante distinguir entre los programas de software informático que están orientados expresamente al análisis cualitativo de los programas no orientados a este análisis aunque con gran potencialidad para: creación y mantenimiento de códigos, búsqueda de códigos, búsqueda y establecimiento de patrones, tratamiento de textos y códigos que permiten ordenar y separar diferentes patrones, etc.

En las últimas publicaciones sobre el análisis de datos en la investigación cualitativa asistido por ordenador (Flick, 2004; Weitzman, 2000; Seale, 2000) se advierte que su utilización ayuda en el manejo de la gestión, la búsqueda y la exposición de los datos y los elementos relacionados, como códigos o memorandos en los enlaces con los datos. Dado el tiempo necesario para decidir a favor de un programa, instalarlo y aprender a utilizarlo (o incluso aprender el uso del ordenador), la ganancia real de tiempo merecerá el esfuerzo.

La segunda utilidad es el aumento de la calidad en la investigación cualitativa utilizando ordenadores, o que resulte más fácil demostrar su calidad. Aquí se menciona el aumento de coherencia en los procedimientos analíticos (Weitzman, 2000) o el rigor añadido en los análisis (Seale 2000). Kelle y Laurie (1995) vinculan el uso de estos programas a un superávit de validez en la investigación cualitativa.

Por último, afirma Flick (2004) con el uso de estos programas se incrementa la transparencia del proceso de investigación y la comunicación en un equipo de investigación. Weitzman menciona también una consolidación de la investigación, ya que el ordenador permite al investigador tener todos los documentos de ésta (desde las notas de campo iniciales hasta las exposiciones, tablas y escritos finales sobre los hallazgos) en un lugar: el disco duro del ordenador. Seale (2000) por su parte, ve con estos programas una facilitación de las decisiones de muestreo basada en el estado del análisis de los datos hasta el momento (según el muestreo teórico).

Una utilidad importante es que la gestión de los datos se hace más fácil con los ordenadores. Kelle (2000) citado por Flick (2004) enumera diversas técnicas de gestión de datos apoyada por programas de ordenador de análisis cualitativo:

- La definición de indicadores que contienen palabras clave junto con “direcciones” de pasajes del texto que se pueden utilizar para recuperar segmentos de texto indizados.
- La construcción de referencias cruzadas electrónicas con la ayuda de los llamados hipervínculos, que se pueden utilizar para saltar entre pasajes de texto que están vinculados entre sí [...].
- Funciones para el almacenamiento de comentarios de los investigadores (–memorandos–), que se pueden vincular a palabras clave o segmentos de texto.
- Rasgos para definir vínculos entre palabras clave.

- El uso de variables y filtros de manera que la búsqueda de segmentos de texto pueda estar restringida por ciertas limitaciones.
- Funciones para la recuperación de segmentos de texto con relaciones formales específicas entre ellos (por ejemplo, segmentos de texto que aparezcan con cierta distancia máxima especificada los unos de los otros).
- Funciones para la recuperación de atributos cuantitativos de la base de datos.








Las dos primeras las ofrecen todos los programas de ordenador de análisis cualitativo (pero no los procesadores de textos o sistemas de base de datos estándar), mientras que las otras cinco las aportan sólo los paquetes de programas para la investigación cualitativa más elaborados. De todos modos hay que tener en cuenta que *“no hay todavía un programa que sea el mejor”* (Flick, 2004).

## PRINCIPALES PROGRAMAS INFORMÁTICOS DE ANÁLISIS CUALITATIVO

Hasta el momento los programas disponibles se pueden resumir en varios tipos (Flick, 2004; Richards y Richards, 1998; Seale, 2000; Weitzman, 2000; Weitzman y Miles, 1995):

- Procesadores de textos, que permiten no sólo escribir sino también editar textos, buscar palabras o secuencias de palabras, al menos de forma limitada.
- Programas de recuperación de textos, que permiten específicamente investigar, resumir, enumerar, etc. ciertas secuencias de palabras.
- Gestores de bases de textos para administrar, buscar y clasificar y ordenar segmentos de texto.
- Programas de codificación y recuperación para separar el texto en segmentos, a los que se asignan códigos, y para recuperar o enumerar todos los segmentos de texto que se marcaron con cada código. Marcar, ordenar, clasificar y vincular textos y códigos reciben apoyo y ambos (el texto y el código) se presentan y administran juntos.
- Construcción de teorías basadas en códigos: además, estos programas apoyan la construcción de teorías sustentando pasos y operaciones en el nivel del texto (asignación de uno o más pasajes a un código) pero también en el nivel conceptual (relaciones entre códigos, categorías superiores y subcategorías, redes de categorías), volviendo siempre atrás a los segmentos de texto asignados. En algunos programas, se incluyen editores gráficos, y es posible integrar datos de vídeo.
- Redes conceptuales: opciones extensas para desarrollar redes conceptuales, redes de categorías y varias formas de visualizar relaciones entre las diversas partes de la red.

Tabla 1. Programas de análisis cualitativo más utilizados en la actualidad (Sánchez Gómez y Revuelta, 2004)

LOGO	SOFTWARE	FUNCIONES
	NUDIST	➤ Organización e Integración
	NUDIST Vivo	➤ Búsqueda y Recuperación
	ATLAS.ti	➤ Enlaces
	AQUAD	➤ Representación Gráfica
	Etnograph	➤ Trabajo en grupo
	WINMAX	
	MAXQDA	

1) *NUDIST*

NUDIST son las siglas de *Non-numerical Unstructured Data \* Indexing Searching and Theorizing* (una posible traducción en castellano podría ser: Datos No Estructurados y No Numéricos \* Indexar, registrar y teorizar).

QSR NUDIST es una herramienta diseñada para ayudar a sus usuarios en el manejo de datos no numéricos y no estructurados en análisis cualitativos, para soportar procesos de codificación de datos en un sistema de catalogación, búsquedas de texto o modelos de código y teorizar sobre los datos. Hablaremos de él más adelante.

2) *NUDIST Vivo*

QSR NVivo es un programa altamente avanzado para el manejo del análisis de datos cualitativos en proyectos de investigación. Es un producto asociado a la actualización opcional del NUDIST.

Los investigadores pueden gestionar tanto datos enriquecidos, como texto enriquecido, usando negrita, cursiva, colores y otros formatos – con amplia capacidad para editar, visualizar códigos y vincular documentos tal y como son creados, codificados, filtrados, manejados y registrados.

En la última versión de este programa, las mejoras sobre la versión anterior están relacionadas con: manejo de documentos, codificación, manejo de datos, modelado (representación gráfica), informes y exportación y facilidades para el proyecto.

3) *WINMAX*

WINMAX es una herramienta para el análisis de texto, robusta y fácil de manejar, que puede codificar y recuperar información incluso para los más sofisticados análisis textuales, utilizando procedimientos cuantitativos y cualitativos de forma combinada.

El programa WINMAX ofrece: visualización de funciones básicas como códigos y memos, codificación compleja y flexible y funciones de recuperación (booleana, proximidad y recuperación semántica), búsqueda lexical y codificación automática, combinación de datos cuantitativos y cualitativos, importar y exportar matrices de datos, complejas y robustas herramientas de construcción de teoría, funciones únicas como codificación de variables según su peso significativo.

El programa tiene interfaz simple y fácil de usar, muestra cuatro ventanas: los grupos de textos, la lista de códigos, la lista de segmentos codificados y los propios textos ya importados.

#### 4) *ATLAS.ti*

ATLAS.ti es una herramienta informática cuyo objetivo es facilitar el análisis cualitativo, principalmente de grandes volúmenes de datos textuales (Muñoz Justicia, 2001).

ATLAS.ti es un potente software para el análisis visual de datos cualitativos de: textos, gráficos, audio y video. Ofrece una variedad de herramientas para llevar a cabo las tareas asociadas con una aproximación sistemática a los datos "sensibles".

Este programa puede: recopilar y organizar su texto, audio o archivos de datos visuales, junto con la codificación, memos y búsquedas, en su proyecto o "unidad hermenéutica", facilita actividades en el análisis y la interpretación, construye una red que permite conectar visualmente pasajes seleccionados, permite construir conceptos y teorías basadas en las relaciones visibles, permite descubrir significados interrelacionados.

ATLAS.ti fue desarrollado por Muhr (1991, 1994) en un proyecto de investigación en la Universidad Técnica de Berlín. El programa se basa en el enfoque de la teoría fundamentada y de la codificación teórica según Strauss (1987). Sus versiones más recientes no sólo pueden procesar textos, sino también imágenes, gráficos y sonido.

La mayoría de los autores clasifican este programa en la categoría de los "*constructores de redes conceptuales*" (Weitzman, 2000), pero principalmente en el grupo de los "*constructores de teorías basadas en códigos*". Ayuda en operaciones en el nivel de textos y el conceptual. Se forma una "unidad hermenéutica" en la pantalla que unifica el texto primario (por ejemplo, la entrevista que debe interpretarse) y las interpretaciones o las codificaciones relacionadas con ella. El programa muestra el texto primario con todos los códigos asignados y los comentarios en ventanas diferentes. Ofrece algunas funciones, que están presentes en la pantalla en forma de símbolos (recuperar, copiar, cortar, codificar, formar redes, etc.) (Flick, 2004).

Aparte de la recuperación de secuencias de palabras en el texto y la asignación de códigos, es útil la presentación de los códigos y las categorías en redes conceptuales. Se integran interfaces con el SPSS y otros programas. Además es posible que autores diferentes trabajen en el mismo texto en ordenadores distintos. Hay un apoyo aceptable del autor y una lista electrónica muy activa de usuarios: (<http://www.atlasti.de>).

#### 5) *AQUAD*

La característica especial de este programa es su capacidad, no sólo para categorizar y organizar los datos para cada categoría, sino también permitir al investigador extraer conclusiones al relacionar las categorías entre sí explorando, por ejemplo, la aparición de ciertas configuraciones típicas y repetitivas en la representación de los datos. Una vez que dichas

repeticiones son intuitivas, el investigador puede confirmar sistemáticamente los datos o bien, como señalan Miles y Huberman (1994), *comprobar las hipótesis*.

Como indica Piñuel Raigada (2003) este programa permite *“interpretación de textos mediante codificación simple del texto sin guía previa de categorías, o mediante un proceso semiautomático de codificación a partir de listados de ocurrencia de términos prefijados, que aparecen destacados sobre la pantalla siguiendo el texto. Los programas de análisis permiten desde el recuento de palabras aisladas, hasta la confección de listas de palabras (“diccionarios”), diferenciación de hablantes, codificación de las partes de textos producidos por hablantes diferentes, etc. Los segmentos de texto pueden ser recuperados por número de archivos de texto, números de línea del texto, código, palabra clave, etc. (la búsqueda de texto es completa). Permite la comparación de casos/textos mediante análisis booleanos de rasgos críticos (“minimización lógica de tablas de contingencia”).*

#### 6) MAXQDA

MAXQDA es el nuevo programa desarrollado por los creadores de WINMAX, es una potente herramienta para el análisis de texto basado en datos cualitativos.

EL MAXQDA se relaciona con los siguientes métodos y técnicas cualitativas:

- Teoría Fundamentada.
- Análisis cualitativo del contenido.
- Métodos de investigación de campo.
- Métodos etnográficos.

MAXQDA analiza los siguientes tipos de textos: transcripciones de entrevistas abiertas, entrevistas en profundidad, semiestructuradas y de expertos, observaciones, estudios sobre el terreno, debates de grupo, discursos y documentos, textos procedentes de la red, discursos políticos y charlas terapéuticas.

El programa MAXQDA se utiliza en muchas disciplinas o ámbitos, como sociología, ciencias de la educación, economía, marketing, etnología, arquitectura, planificación urbana, sanidad pública y medicina.

#### CÓMO DECIDIRSE POR UN PROGRAMA INFORMÁTICO

Autores como Weitzman (2000); Weitzman y Miles (1995); Bergmann y Meier (2004) y Flick (1991; 2004) proponen que el investigador se formule una serie de preguntas clave antes de decidirse por un programa informático o por el uso de ordenadores en general. Estas preguntas claves se detallan en cuadro siguiente.



Cuadro 1. Preguntas de guía para analizar y comparar los programas de ordenador en la investigación cualitativa

1. Preguntas relacionadas con los datos: ¿Para qué clase de datos se concibió el programa? ¿Para qué datos puede utilizarse más allá de estos datos originales? ¿Para qué datos no se debe utilizar?. ¿Se trata sólo de texto o utiliza video o fotografías, datos acústicos, imágenes en movimientos?. No todos los programas están preparados para trabajar con este tipo de datos.
2. Preguntas relacionadas con las actividades: ¿Qué actividades pueden realizarse con este programa, cuáles no deben llevarse a cabo?
3. Preguntas relacionadas con el proceso: ¿Cómo influyó el programa en el manejo de los datos y en el rol que desempeñó el investigador o intérprete de acuerdo con las experiencias tenidas hasta ahora? ¿Qué nuevas opciones abrió? ¿Qué se ha vuelto más difícil o laborioso en el proceso de interpretación debido al programa?
4. ¿Qué tipo de análisis está previsto (con un esquema de codificación predefinido con un conjunto de categorías que evolucionan)?
5. ¿Debe estar el texto que se interpreta siempre accesible (en pantalla) o sólo las categorías?
6. Preguntas técnicas ¿Cuáles son las condiciones necesarias en las opciones de red del equipo (tipo de ordenador, Ram, disco duro, tarjeta gráfica, pantalla) o de los programas (sistemas de programas, necesidad de otros programas) para otros programas (SPSS, procesadores de textos, bases de datos...)?
7. Preguntas relacionadas con la competencia: ¿Qué destrezas específicas y sobre todo técnicas requiere del usuario el programa (destrezas de programación, quizá lenguajes de programación específicos, etc.)?
8. Límites de coste ¿Puede permitirse comprar el programa y el ordenador para usarlo?

Fuente: Flick (1991: 800 y 2000: 272)

Desde nuestra experiencia personal con el manejo del software de análisis cualitativo, nos atrevemos a sugerir los programas NUDIST y ATLAS-Ti, que son los más utilizados tanto para el análisis textual como para el conceptual. Nos obstante como señala Ruiz Olabuénaga (1999) *“no conviene perder de vista la afirmación central de que el ordenador opera y ofrece sus ventajas de rapidez, capacidad y versatilidad, siempre a nivel textual, dado que, a nivel conceptual, es el investigador el que debe efectuar las operaciones oportunas en cada caso. El ordenador ejecuta las operaciones mecánicas pero el investigar debe reservarse para sí las estrictamente conceptuales”*.

## EL PROGRAMA NUDIST

De todos estos programas descritos el que se utiliza en la asignatura objeto de esta propuesta académica es el *Nudist*.

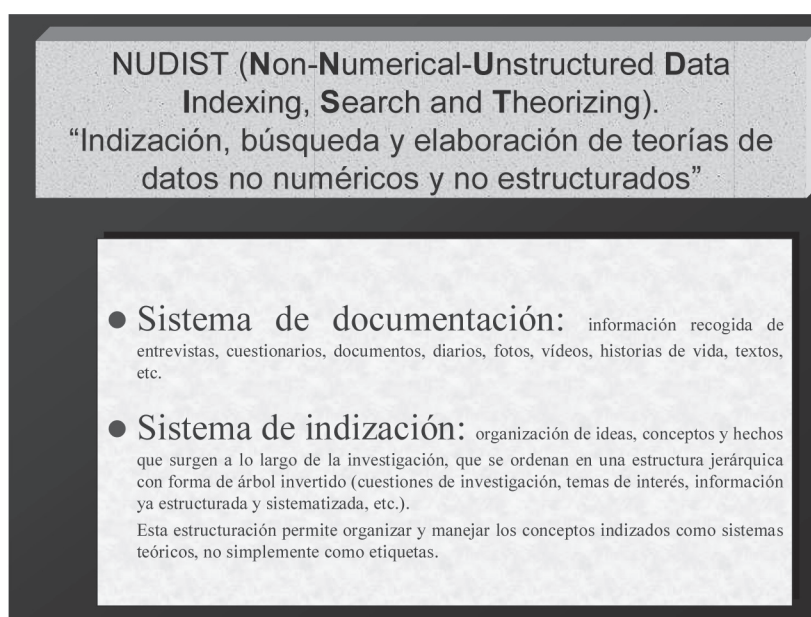
Fue desarrollado por Richards y Richards (1994). Weitzman (2000) clasifica el programa como *“constructor de teorías basadas en códigos”*. Incluye un sistema de comandos comple-



to para automatizar la codificación y la búsqueda. Permite la fusión de los archivos analíticos de dos o más proyectos de investigación ejecutados inicialmente por separado.

Este programa se distribuyó comercialmente muy pronto, y los autores lo promocionaron de una forma muy activa. Algunos rasgos como los resultados del memorando, de la búsqueda, que se puedan añadir más datos a los originales o la exposición de los códigos en la pantalla (–indización–) en una estructura jerárquica de árbol, son bastante característicos de este programa. Se puede encontrar información y ayuda en Internet (<http://www.qsr.com.au>).

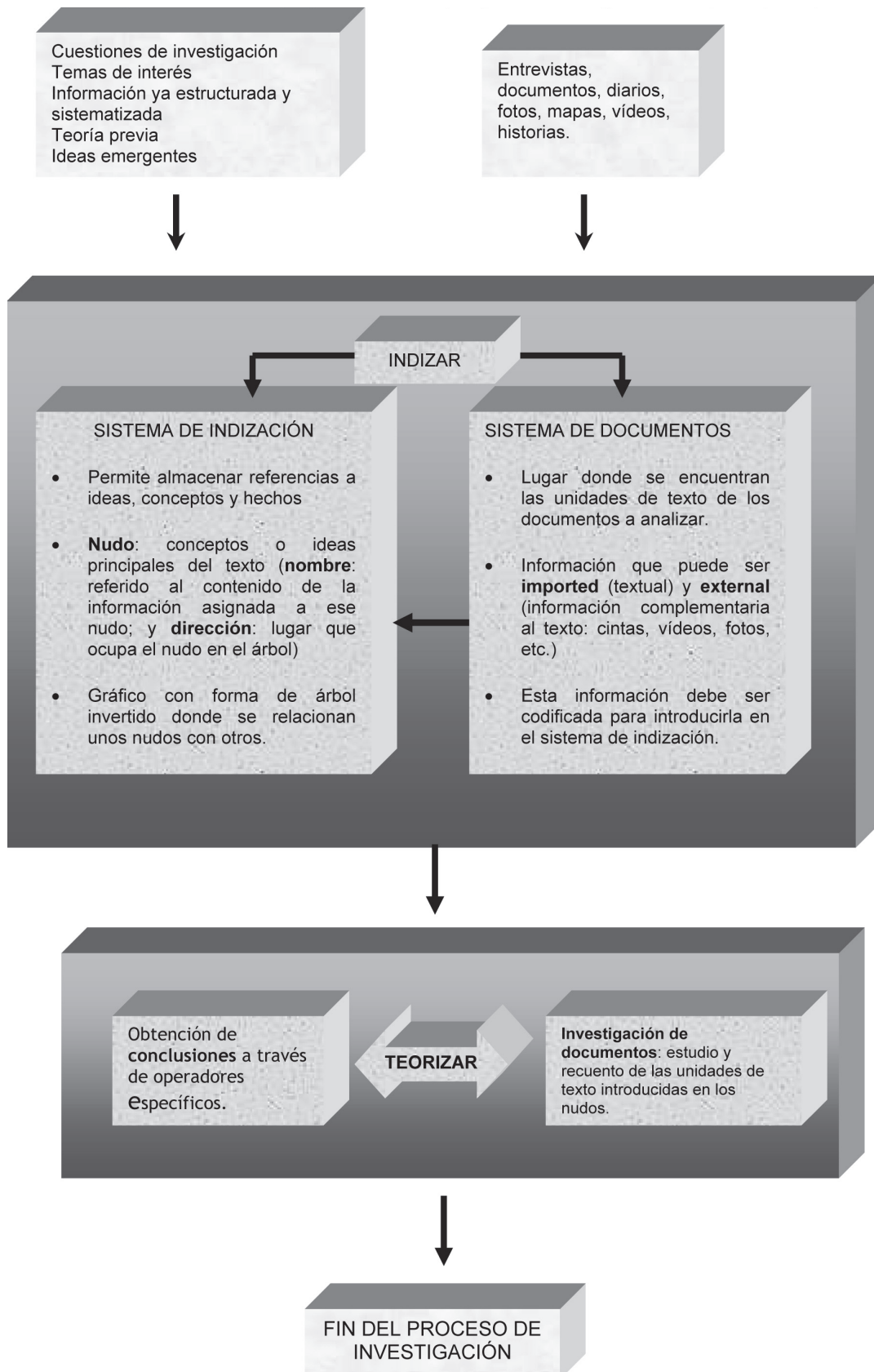
Imagen 5. Procesos del programa NUDIST



Fuente: Elaboración propia

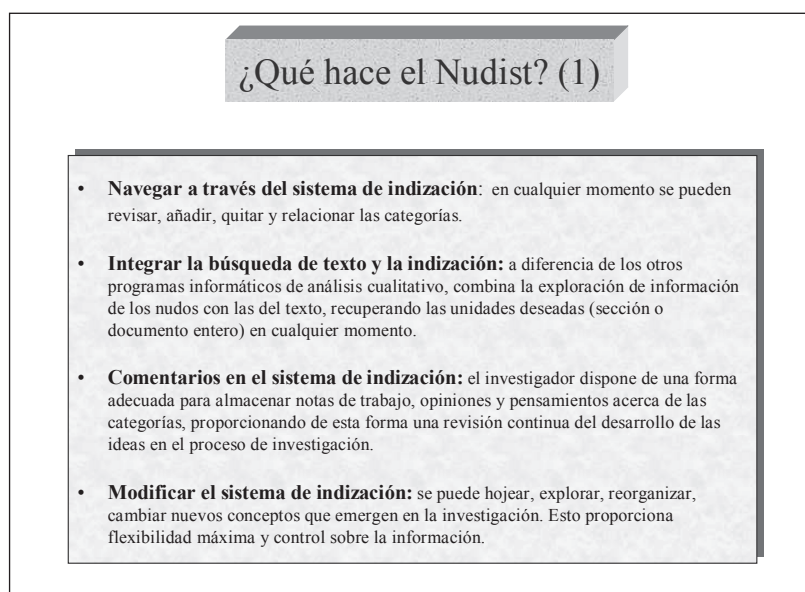
Además, este programa informático, como señalan Gil Flores y Perera Rodríguez (2001), soporta los procesos de categorización deductiva e inductiva, incluso ambas. Esta habilidad posibilita que podamos diseñar a priori un sistema de categorías desarrollado a partir de teorías existentes o establecidas en base al objeto de la investigación, o bien hipótesis o constructos que se han ido incorporando durante la recogida de la información. Una de las características de este programa es que estas categorías las podemos agrupar y organizar de forma jerárquica. Esto permite al investigador observar las relaciones que existen entre los múltiples conceptos o categorías tratadas, compararlas mediante operadores específicos (booleanos, contextuales, negativos, inclusivos, exclusivos) para esbozar conclusiones relativas a los temas de investigación.

Imagen 4. Procesos del programa NUDIST (Adaptado de Rodríguez Gómez y otros, 1996)



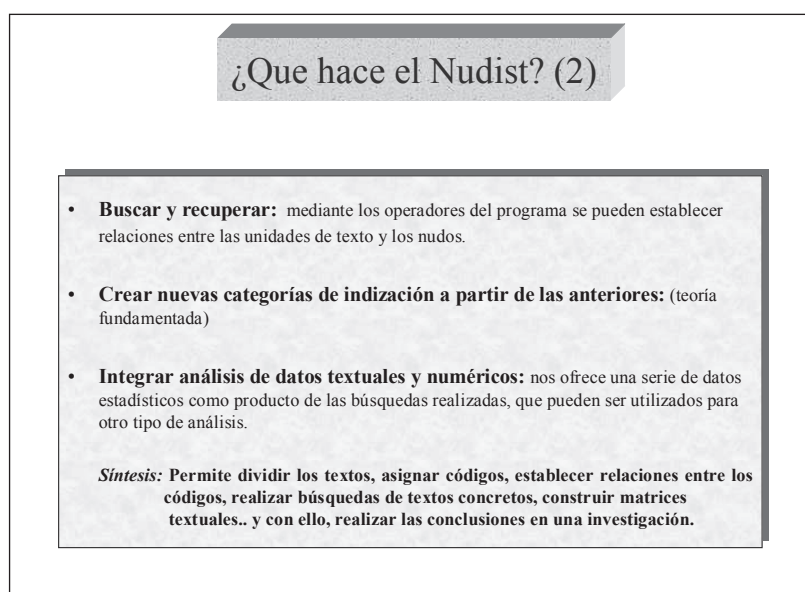
En definitiva, a través del programa NUD\*IST se puede dividir la información textual recogida en la investigación, asignar categorías, establecer relaciones entre ellas, realizar búsquedas textuales específicas, construir matrices y tablas de frecuencias con la información relevante y todo ello será lo que expondremos a continuación de forma lo más exhaustiva posible.

Imagen 6. Utilidad del programa NUDIST (1)



Fuente: Elaboración propia

Imagen 7. Utilidad del programa NUDIST (2)



Fuente: Elaboración propia

Imagen 8. Operaciones con el sistema de indización

OPERADORES				
DE BÚSQUEDA DE RELACIONES ENTRE NUDOS				
CONFRONTACION	CONTEXTUALES	NEGATIVAS	RESTRICATIVAS	JERÁRQUICAS
INTERSECT (INTERSECCIÓN)	AT. LEAST (AL MENOS)	NOT IN (NO EN)	INCLUDING- DOCS-FROM	INHERIT
UNIÓN (UNIÓN) LESS (MENOS)	IF-INSIDE (SI DENTRO) IF-OUTSIDE (SI FUERA)		EXCLUDING- DOCS - FROM	COLLECT
JUST-ONE (SOLO UNO)	NEAR (CERCA)			
OVERLAP (SOLAPAR)	FOLLOWED-BY (SEGUIDO POR)			

Fuente: Elaboración propia

## OBTENCIÓN Y VERIFICACIÓN DE CONCLUSIONES

En esta fase, utilizamos una serie de tácticas para extraer significados de los datos, como pueden ser la comparación/contraste (teoría fundamentada), la búsqueda de patrones y temas, la triangulación, la búsqueda de casos negativos (inducción analítica), etc.

Colás (1998) menciona algunas de las técnicas propuestas por Miles y Huberman (1994) para interpretar datos y extraer conclusiones. Estas son:

- Observar patrones y temas.* Supone buscar patrones o temas que aglutinen datos que, conjuntamente, adquieren cierto sentido. Estos patrones pueden hacer referencia a variables que impliquen similitudes y diferencias entre categorías o conexiones espacio-temporales. Hay que estar atentos tanto a aquellas informaciones que se puedan incluir en el mismo patrón como aquellas que no se ajusten.
- Comprobación de la posibilidad de los resultados.* Implica que los resultados obtenidos tengan sentido y se ajusten a la realidad.
- Agrupar.* Trata de aglutinar eventos concretos en unidades abstractas. Es un proceso inductivo de formación de categorías y de una reiterada inclusión de eventos dentro de categorías. Este proceso supone un nivel mayor de abstracción.
- Subsumir lo particular en lo general.* Es una actividad conceptual y teórica que conecta los datos de un primer nivel conceptual con los posteriores que representan categorías más generales. Este proceso evoluciona hasta la saturación, es decir, hasta que los nuevos datos no añaden nuevos significados a la teoría general.
- Partición de variables.* Puede darse en cualquier fase del proceso. Consiste en plantear la diferenciación que en muchas ocasiones resulta más rica e importante que la

agrupación. La partición no siempre es conveniente y debe aplicarse cuando aporte resultados coherentes, y descripciones y explicaciones integradas.

- f. *Metáforas o analogías.* Son una abstracción parcial y juegan un papel importante en el desarrollo de la teoría. Se pueden definir como la comparación de dos cosas resultando o tomando como punto de referencia sus similitudes, ignorando sus diferencias. Se utilizan para dar sentido a la experiencia.
- g. *Realizar contrastes y comparaciones.* Estos procedimientos se dirigen a generar y refinar una teoría. A través de la comparación constante con casos se identifican las propiedades de los datos, se analizan las interrelaciones y se integran en una teoría. También es importante detectar casos negativos o discrepantes para perfilar y determinar la aplicabilidad de la teoría inicial.
- h. *Factorización.* Consiste en reducir o concentrar la amplia información de los datos en un menor número de factores no observables y generalmente hipotéticos.
- i. *Establecer relaciones entre variables.* Se trata de descubrir el tipo de conexiones que se dan entre dos aspectos. Complementariamente hay que establecer tácticas para la verificación de estas relaciones, a través de su comprobaciones con explicaciones rivales, excluyendo relaciones espúreas o utilizando casos extremos.
- j. *Encontrar variables intervinientes.* Cuando se analiza la conexión entre dos variables es importante detectar la influencia de otras variables que aporten una explicación más realista, ofreciendo a su vez una explicación más compleja.
- k. *Elaborar una cadena lógica de evidencias.* Es necesario proceder de una forma gradual, primero detectando factores significativos, planteando tentativas sobre relaciones lógicas comprobándolas con nuevos datos brutos, modificando y ajustándolos a un nuevo mapa explicativo que se volverá a poner a prueba con nuevos casos.
- l. *Mantener la coherencia conceptual/teórica.* Tratar de mantener una coherencia interna entre los sucesivos estadios conceptuales que se siguen a lo largo del proceso de análisis. Por tanto hay que prestar atención a la correspondencia entre los datos y sus interrelaciones: entre metáforas y constructos y posteriormente con las teorías.

Para comprobar y verificar estos resultados es conveniente utilizar determinados procedimientos (Miles y Huberman, 1994):

- Comprobar la representatividad de la muestra.
- Comprobar los efectos del investigador.
- Control sobre los prejuicios de los distintos informantes e investigadores.
- Triangulación de datos, métodos, de diferentes teorías, e investigadores.
- Comprobación de los datos por los investigadores.

Llegar a conclusiones implica ensamblar de nuevo los elementos diferenciados en los procesos anteriores para construir un todo estructurado y significativo. En las conclusiones se tratan los resultados, los productos de la investigación y la interpretación que las personas implicadas (investigadores e implicados, o sólo investigadores) hacen de los mismos. Los resultados avanzarán en la explicación, comprensión y conocimiento de la realidad educativa y contribuirán a la teorización o intervención sobre la misma. Las conclusiones son proposiciones en las que se recogen los conocimientos adquiridos por el investigador en relación al problema estudiado (Rodríguez Gómez y otros, 1996).

## REFERENCIAS BIBLIOGRÁFICAS

- Amorós, P., Cabrera, F. y otros (1993). Una experiencia de investigación cooperativa. La metodología de intervención en medio abierto. *Revista de Investigación Educativa*, 20, 109-130.
- Anguera, M. T. (1995). Tratamiento cualitativo de datos. En M. T. Anguera y otros, *Métodos de investigación en Psicología*, (pp. 549-576). Madrid: Síntesis.
- Arnal, J., Del Rincón, D. y Latorre, A. (1992). *Investigación educativa. Fundamentos y metodología*. Barcelona: Labor.
- Ball, M. y Smith, G. (1992). *Analyzing visual data*. Newbury Park. CA: Sage.
- Bardin, L. (1986). *El análisis de contenido*. Madrid: Akal.
- Bartolomé, M. (1990). *Elaboración y análisis de datos cualitativos aplicados a la investigación-acción*. Barcelona: Universidad de Barcelona.
- Bogdan, R. y Biklen, S. (1992) (2<sup>a</sup> ed.). *Qualitative research for Education: An Introduction to theory and methods*. Boston: Houghton Mifflin.
- Bonal, X. (1995). Los límites del enfoque práctico. *Cuadernos de Pedagogía*, 236, 70-71.
- Buendía, L., Colás, P. y Hernández Pina, F. (1997). *Métodos de investigación en Psicopedagogía*. Madrid: MacGraw Hill.
- Caballero Romero, J. J. (1991). Etnometodología: una explicación de la construcción social de la realidad. *Revista de Investigaciones Sociológicas*, 56, 83-114.
- Cajide, J. (1992). La investigación cualitativa: Tradiciones y perspectivas contemporáneas. *Bordón*, 44 (4), 357-373.
- Casanova Correa, J. y Pavón Rabasco, F. *Nuevas herramientas para el procesamiento de datos cualitativos*. Disponible en: [<http://www2.uhu.es/agora/digital/numeros/03/03-articulos/miscelanea/pavon.PDF>]. Universidad de Cádiz.
- Coffey, A. y Atkinson, P. (1996). *Making sense of qualitative data*. London: Sage Publications.
- Colás, M. P. (1992). El análisis de datos en la metodología cualitativa. *Revista de Ciencias de la Educación*, 152, 521-540.
- Colás, M. P. (1997). El análisis cualitativo de datos. En L. Buendía, P. Colás y F. Hernández (Eds.), *Métodos de investigación en Psicopedagogía* (pp. 287-311). Madrid: McGraw-Hill.
- Cortés Rodríguez, L. y Camacho Adarve, M.<sup>a</sup>. M. (2003): *¿Qué es el análisis del discurso?* Barcelona: Octaedro.
- Creswell, J. W. (1998). Data analysis and representation, del libro *Qualitative inquirind research design: Choosing among five traditions* (pp. 139-165). California: Sage.
- De la Orden, R.L. (1988). Informática e investigación educativa. En I. Dendaluze (Coord.), *Aspectos metodológicos de la investigación educativa* (pp. 276-295). Madrid. Narcea.
- Delgado, J. M. y Gutiérrez, J. (Coords.) (1994). *Métodos y técnicas cualitativas de investigación en ciencias sociales*. Madrid. Síntesis.
- Dey, L. (1993). *Qualitative data analysis. A user friendly guide for social scientists*. London: Roudedge and Kegan Paul.
- Etxeberría, J. y otros (1995). *Análisis de datos y textos*. Madrid: Ra-Ma.
- Fielding, N.G. y Lee, R.M. (1991). *Using computers in qualitative research*. Newbury Park CA: Sage.
- Flick, U. (2004). *Introducción a la Investigación Cualitativa*. Madrid: Morata.



- García, E., Gil, J. y Rodríguez, G. (1994). Análisis de datos cualitativos en la investigación sobre la diferenciación educativa. *Revista de Investigación Educativa*, 23, 179-213.
- Gil, J. (1994). Aproximación interpretativa al contenido de la información textual (pp.65-107), Cap. 3 del libro *Análisis de datos cualitativos: Aplicaciones a la investigación educativa*. Barcelona: PPU.
- Gil Flores, J. (1994). *Análisis de datos cualitativos: Aplicaciones a la investigación educativa*. Barcelona: PPU.
- Gil Flores, J. y Perera Rodríguez, V. (2001): *Análisis informatizado de datos cualitativos*. Sevilla: Kronos.
- Glaser, R. y Strauss, A. (1967). *The discover of grounded theory: strategies for qualitative research*. New York: Aldine Publishing Co.
- Goetz, J., y LeCompte, M. (1988). *Etnografía y diseño cualitativo en investigación cualitativa*. Madrid: Morata.
- Green, J.C. (1994). Qualitative Program Evaluation: Practice and Promise. En N.K. Denzin y Y.S. Lincoln (Eds.), *Handbook of Qualitative Research*. (pp. 530-544). Londres: Sage Publications.
- Guba, E.G. y Lincoln, Y.S. (1994). Competing Paradigms in Qualitative Research. En N.K. Denzin y Y.S. Lincoln (eds.), *Handbook of Qualitative Research*, (pp. 105-117). Londres: Sage.
- Gustavo Rodríguez, P. (1997). *El recurso informático en el procesamiento de datos cualitativos en ciencias sociales*. Disponible en: [[http://www.naya.org.ar/analisis\\_cualitativo/vcaaspc.htm](http://www.naya.org.ar/analisis_cualitativo/vcaaspc.htm)].
- Hesse-Biber, S. y Dupuis, P. (1995). Hypothesis testing in computer-aided qualitative data analysis. En Kelle, V. (Ed.). *Computer-aided qualitative data analysis: Theory, methods and practice*. London: Sage.
- Hill, M. (1993). *Archival strategies and techniques*. Newbury Park, CA: Sage.
- Huber, G. L. (1991). *AQUAD. Análisis de datos cualitativos con ordenadores. Principios y manual del paquete de programas AQUAD 3.0*. Sevilla: Departamento de Didáctica y Organización Escolar y Métodos de Investigación y Diagnóstico en Educación.
- Huber, G. y Marcelo, C. (1990). Algo más que recuperar palabras y contar frecuencias: la ayuda del ordenador en el análisis de datos cualitativos. *Revista de Enseñanza*, 8, 69-85.
- Huber, G. (2001). *AQUAD cinco: Manual del programa para analizar datos cualitativos*. Berlín: Ingeborg Huber Verlag.
- Jacob, E. (1987). Qualitative research traditions: A review. *Review of Education Research*, 57, 1, 1-50.
- Krippendorff, K. (1990). *Metodología del análisis de contenido. Teoría y práctica*. Barcelona: Paidós.
- Latorre A., Del Rincón, D. y Arnal, J. (1996). *Bases metodológicas de la investigación educativa*. Barcelona: Hurtado.
- Llinares, S. (1992). Los mapas cognitivos como instrumento para investigar las creencias epistemológicas de los profesores. En C. Marcelo (Ed.), *La investigación sobre la formación del profesorado. Métodos de investigación y análisis de datos*. Irala, Argentina: Cincel.
- Llinares, S. y Sánchez, M. V. (1990). Las creencias epistemológicas sobre la naturaleza de las matemáticas y su enseñanza y el proceso de llegar a ser un profesor. *Enseñanza*, 8, 165-180.
- Lofland, J. (1971). *Analyzing social setting*. Belmont: CA. Wadsworth Publishing Company.
- Marcelo, C. (1992). Dar sentido a los datos: La combinación de perspectivas cualitativa y cuantitativa en el análisis de entrevistas. En C. Marcelo (Ed.), *La investigación sobre la formación del profesorado. Métodos de investigación y análisis de datos* (pp. 13-48). Irala, Argentina: Cincel.



- Miles, M. B. y Huberman, A.M. (1984). *Qualitative data analysis: A source book*. Beverly Hills, CA: Sage.
- Miller, A. (1986). *Imagery in scientific thought*. Cambridge: MIT Press.
- Morse, J. M. (1997). Considering theory derived from qualitative research. En J. M. Morse (Ed.), *Completing a qualitative project. Details and dialogue* (pp. 163-189). California: Sage.
- Muñoz Justicia, J. (2001). *Análisis cualitativo de datos textuales con ATLAS/ti*. Disponible en: [<http://antalya.uab.es/jmunoz/Cuali/ManualAtlas.pdf>].
- Murh, T. (1997). *Atlas.ti- Visual Qualitative data analysis-Management-Model Building-Release 4.1*. Berlín, Short User's Manual.
- Navarro, P. y Díaz, C. (1994). El análisis de contenido. En Delgado, J. M. y Gutiérrez, J. (1994). *Métodos y técnicas cualitativas de investigación en Ciencias Sociales*. Madrid: Síntesis.
- Noblit, G.W. y Hare, R.D. (1988). *Meta-ethnography: Synthesizing qualitative studies*. Newbury Park, CA: Sage.
- Noldus, L.P. (1991). The Observer: A software for collection and analysis of observational data. *Behavior Research Methods, Instruments & Computers*, 23 (3), 415-429.
- Pardo, A. y San Martín, P. (1994). *Análisis de datos en psicología II*. Madrid: Pirámide.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd. Ed.). Newbury Park, CA: Sage.
- Pérez Serrano, G. (1994). *Investigación cualitativa: retos e interrogantes II: Técnicas y análisis de datos*. Madrid: La Muralla.
- Peshkin, A. (1993). The goodness of qualitative research. *Educational Research*, 22 (2), 23-30.
- Piñuel Raigada, J. L. (2003). Epistemología, metodología y técnicas del análisis de contenido, *Estudios de Sociolingüística*, 4 (2).
- Pitman, M.A. y Maxwell, J.A. (1992). Qualitative Approaches to Evaluation: Models and Methods. En M. D. LeCompte, W. L. Millroy y J. Preissle (Eds.). *The Handbook of Qualitative Research in Education*, (pp. 727-770). Londres: Academic Press.
- Plume, K. (1983). *Documents of life: An introduction to the literature of humanistic method*. London: Allen and Unwin.
- Psthas, G. (1994). *Conversation Analysis*. Thousand Oaks, CA: Sage.
- Qsr (1994). *NUDIST Version 3.0 User Guide*. Victoria. Qualitative Solutions Research.
- Renom, J. (Coord) (1997). *Tratamiento informatizado de datos*. Barcelona: Masson.
- Reuelta Domínguez, F. I. y Sánchez Gómez, M. C. (2003). Programas de análisis cualitativo para la investigación en espacios virtuales de formación. *Teoría de la Educación: Educación y Cultura en la Sociedad de la información*, 4. Documento on-line [[http://www3.usal.es/~teoriaeducacion/rev\\_numero\\_04/n4\\_art\\_revuelta\\_sanchez.htm](http://www3.usal.es/~teoriaeducacion/rev_numero_04/n4_art_revuelta_sanchez.htm)].
- Richard, T. J. y Richard, L. (1994). Using computers in qualitative research, en Denzin, N. K. y Lincoln, Y. S. (Eds.) *Handbook of qualitative research*, Sage, Thousand Oaks, California.
- Richards, L. (1999): *Using Nvivo in Qualitative Research*. London: Sage.
- Richards, L. (2002): *Introducing N6 a workshop handbook*. London: Sage.
- Roberts, N. K. y otros (1993). VTLOGANL: A methodology for the qualitative analysis of video tape data. Comunicación presentada en la reunión anual de la MidWestern *Educational Research Association*, Chicago.

- Rodríguez Gómez, G., Gil Flores, J., y García Jiménez, E. (1996). El papel de la informática en el análisis de datos (pp. 237-257), Cap. 13 del libro *Metodología de la investigación cualitativa*. Granada: Aljibe.
- Rodríguez Gómez, G., Gil Flores, J., y García Jiménez, E. (1996). Aspectos básicos sobre el análisis de datos cualitativos (pp. 197-218), Cap. 11 del libro *Metodología de la investigación cualitativa*. Granada: Aljibe.
- Rodríguez, G., Gil, J. y García, E. (1996). *Metodología de la investigación cualitativa*. Málaga: Aljibe.
- Rodríguez, G. (2000). Avances en las técnicas cualitativas de investigación socioeducativa. En *XII Congreso Nacional y I Iberoamericano de Pedagogía*. Madrid. Sociedad Española de Pedagogía. Tomo I, 87-133.
- Rodríguez, G., Corrales, A., Gil, J. y García, E. (1994). Nudist: una herramienta informática para el análisis de datos cualitativos. En *Revista de Investigación Educativa* (pp.522-528).
- Rodríguez, G., Gil, J. y García, E. (1995). *Análisis de Datos cualitativos asistido por ordenador: Aquad y Nudist*. Barcelona. PPU
- Ruiz Olabuénaga, J. I. (1999). *Metodología de la investigación cualitativa*. Bilbao: Universidad de Deusto.
- Salvador Mata, F., Rodríguez Diéguez, J. L. y Bolívar Botía, A. (Dir.) (2004). *Diccionario enciclopédico de Didáctica*. Vols. I y II. Málaga: Aljibe.
- Sánchez Gómez, M.C. (2004). Investigación cualitativa. En Salvador Mata, F., Rodríguez Diéguez, J. L. y Bolívar Botía, A. (Dir.). *Diccionario enciclopédico de Didáctica*. Vol.II. Málaga: Aljibe.
- Sánchez Gómez, M. C., Revuelta Domínguez, F. I. y López Gómez, D. (2003): Estudio cualitativo de los valores como mediadores educativos en la integración del enfermo mental, *Revista Estudios*, 50. Diputación de Salamanca.
- Sandelowski, M. (1998). Writing a good read: Strategies for re-presenting qualitative data. *Research in Nursing and Health*, 21, 375-382.
- Sibert, E., y Shelly, A. (1995). Using logic programming for hypothesis generation and refinement. En Kelle, V. (Ed.). *Computer-aided qualitative data analysis: Theory, methods and practice*. London: Sage.
- Serrano Gibert, T. (1992). Una metodología cualitativa para el estudio del desarrollo conceptual en el aprendizaje de las ciencias. Análisis con redes sistémicas. *Revista de Investigación Educativa*, 20, 37-69.
- Thomas, J. (1993). *Doing critical ethnography*. Newbury Park, CA: Sage.
- Tesch, R. (1987). Comparing the most widely used methods qualitative analysis. What do they have in common? *American Educational Research Association*. Annual Convention.
- Tesch, R. (1990). *Qualitative Research: Analysis Types and Software Tools*. London: The Falmer Press.
- Tesch, R. (1991). Software for qualitative researchers: Analysis Needs and Program Capabilities. En Fielding, N., y Lee, R. *Using computers in Qualitative Research*. London: Sage
- Vallés, M. S. (1997). De las herramientas artesanales a las informáticas (pp. 391-401), del libro *Técnicas cualitativas de investigación social. Reflexión metodológica y práctica profesional*. Madrid: Síntesis.
- VV. AA. (1998): Chapter 2: *Overview*. *Qualis Research Associates*. Disponible en: [ftp://ftp.qualisresearch.com/pub/overview.exe].
- VV. AA. (1998): Chapter 4: *Quick Tour*. *Qualis Research Associates*. Disponible en: [ftp://ftp.qualisresearch.com/pub/tour.exe].

- VV. AA. (2001). *MAXqda 2001: Introducción*. Berlín. Disponible en: [<http://www.maxqda.com/maxqda-spa/download/mxintro.pdf>].
- Weaver, A. y Atkinson, P. (1996). From coding to hypertext. En R. G. Burgess (Ed.). *Using computers in qualitative research*. Greenwich, CT: JA.
- Weitzman, E. y Miles, M. (1995). *Computers programs for qualitative data analysis*. Thousand Oaks, CA: Sage.
- Wolcott, H. (1990). *Transforming qualitative data: Description, Analysis and Interpretation*. Thousand Oaks, CA: Sage Publications.

# LA ÉTICA EN LA INVESTIGACIÓN SOBRE DISCAPACIDAD

M<sup>a</sup> TERESA ANGUERA ARGILAGA  
*Departament de Metodologia de les Ciències del Comportament  
Facultat de Psicologia (Universitat de Barcelona)*

## RELEVANCIA DE LAS CUESTIONES ÉTICAS EN INVESTIGACIÓN SOBRE DISCAPACIDAD

En el momento actual podemos advertir un considerable *boomerang* acerca del interés por la ética en diversos ámbitos, así como en numerosas áreas y facetas que se refieren al ser humano y a las Ciencias del Comportamiento (França-Tarragó, 1996). Y, como es lógico esperar, las posiciones adoptadas muestran un amplio elenco de planteamientos que van desde los principios rígidos hasta un amplio relativismo (Fuentes Caballero, 1997).

La realidad de la investigación sobre discapacidad nos muestra un mundo complejo en el cual existe un conjunto de problemas y dilemas que se deben contextualizar, valorar detenidamente y resolver, y para los cuales se debe aplicar un procedimiento objetivo; es cierto que en ocasiones emergen más cuestiones por efecto del método aplicado (efecto de la demanda, de reactividad, de expectancia, etc.), ... y pueden añadirse nuevos elementos de forma incesante hasta quedar obturados por la complejidad de la situación, en una sucesión de bucles que no termina (Altarejos, Ibáñez-Martín, Jordán y Jover, 1998). Por encima de ellos, pero interactuando con todos, está la ética (Comisión Deontológica Estatal, 1999).

Los investigadores que se dedican al ámbito de la discapacidad chocan con esta realidad compleja cuando perciben que los medios, recursos y métodos de que disponen no les bastan para hallar unas soluciones éticas y adecuadas ante problemas diversos de priorización de necesidades, de uso de la información, de accesibilidad a un programa, de requerimiento de consentimiento informado, de características de ciertas intervenciones, de notificación de baja eficacia de algunos tratamientos, de incidencia del participante en casos de recidiva, etc. El quehacer cotidiano les emplaza a tomar las decisiones pertinentes para dar sentido a tales medios, recursos y métodos (Anguera, Chacón y Sanduvete, 2008), a la vez que perciben una disociación entre teoría y práctica, manifestándose un cierto desencanto que sin duda conllevará una inseguridad, como mínimo, en su respectiva competencia profesional. Y este hecho se puede ilustrar en ámbitos sumamente diversificados, como son los de la vida y la salud (Herranz, 1992; Casado, 1998; Susanne, 1998; Beauchamp & Childress, 1999), psicología forense (Hess, 1987), psicoterapia (Thompson, 1990), deficiencia mental (Allen & Allen, 1979), etc.

Los intentos para establecer una ética profesional (Callahan, 1988) han sido y son prolijos, pero los resultados a los que se ha llegado no han permitido desarrollar un modelo útil para la toma de decisiones (Newman & Brown, 1996). Existe un amplio elenco de códigos deontológicos a disposición de los profesionales en los cuales se recogen las normas, y al-

gunos de ellos de obligado cumplimiento (en España, el *Código Deontológico del Colegio Oficial de Psicólogos* (1987), pendiente en estos momentos de la aprobación del nuevo texto –*Código Deontológico de la profesión de Psicología*– elaborado durante los últimos años). Tenemos, en consecuencia, documentos consensuados sobre los principios éticos que deben regir la acción profesional, y, en mucha menor medida, la actividad investigadora.

En este momento asistimos a una creciente preocupación por la dimensión ética de la práctica profesional del investigador y del profesional en general (Brown & Newman, 1992; Bustelo Ruesta, 1998), y específicamente en el ámbito de la discapacidad. Simultáneamente, cada vez somos más conscientes de los límites existentes para hacer frente a conflictos y situaciones de alta complejidad con que nos encontramos: las carencias, precariedad y padecimientos de los individuos y de los colectivos están desencadenados, con frecuencia, por circunstancias persistentes, cuya modificación no siempre se halla bajo nuestras posibilidades de actuación; en los equipos multidisciplinares a los que acuden usuarios necesitados de apoyo existen frecuentemente múltiples tensiones larvadas o manifiestas entre los profesionales; en innumerables ocasiones se debe decidir sin tener certeza objetiva del mejor camino (Leal Rubio, 1997); no siempre se cuenta con una capacidad participativa suficiente por interferencias políticas o de estructuración administrativa; etc.

Y además un condicionante de gran relevancia es la polaridad que se establece entre la necesidad y legitimidad de la participación en la investigación, en un extremo, y un potencial usuario que exige un respeto a su autonomía individual y colectiva (Sánchez Vidal, 1999), teniendo en cuenta el especial plus de la vulnerabilidad (Arboleda-Florez, 2003) en el ámbito de la discapacidad (Bermant & Warwick, 1978). Así, por ejemplo, si un centro específico promueve un programa de mejora de calidad de vida para personas con discapacidad física que han sufrido una amputación, aunque las propias personas con discapacidad o sus familiares solicitan plaza para este programa –porque suponen que será positivo para la vida de la persona con discapacidad– será recomendable que se mejore la información sobre dicho programa, que se permita participar a las personas con discapacidad en cuanto a la toma de decisiones, que se realicen ensayos de hábitos de actuación transferibles, y que se realicen controles externos por parte de personal sanitario especialmente capacitado para esta función.

Todo ello nos lleva a la inapelable necesidad de adquirir una formación integral en investigación sobre discapacidad desde presupuestos éticos. Trabajar con las necesidades de las personas y los colectivos es sumamente difícil, así como también lo es saber incorporar en cada caso la especificidad y perfil diferencial que se adecua al caso y desde un posicionamiento ético. El investigador debe seguir creciendo en una capacidad y madurez reflexiva que sea capaz de conjugar la puesta en práctica de los planteamientos éticos, con la rigurosidad conceptual y metodológica, y el realismo de la praxis cotidiana.

## CRITERIOS ÉTICOS EN LA INVESTIGACIÓN SOBRE DISCAPACIDAD

Los diferentes códigos éticos nacionales en el ámbito de la Psicología concuerdan en diversos aspectos, y la *American Psychological Association* ha desempeñado un papel fundamental en este sentido, contribuyendo a que se estructuren más sistemáticamente los principios básicos. El *Joint Committee on Standards for Educational Evaluation* (1981) ofrece guías para los evaluadores (Brown & Newman, 1992) desde el convencimiento de que los códigos éticos son importantes pero insuficientes para orientar a los profesionales e investigadores que aspiran a una práctica ética, remite a los cinco principios defendidos por Kitchener (1984). Los cinco principios generales son los siguientes: autonomía, evitación del daño,

actuación positiva (beneficiencia), justicia y fidelidad, los cuales influyeron en buena medida en el código elaborado por la *American Psychological Association* (1982), en donde se establecieron los principios de competencia, integridad, responsabilidad profesional y científica, respeto por los derechos y dignidad de las personas, procura del bienestar de las personas, y responsabilidad social.

Ahora bien, existen aspectos éticos que son comunes a diversos contextos y ámbitos de investigación, motivo por el cual no habrá que contemplar cuestiones diferenciales entre ellos. Se podría ilustrar, en el ámbito específico de la evaluación de programas (Newman & Brown, 1996; Muñiz, 1997), como los principios éticos y deontológicos que deben regirla se articulan en torno a la utilidad de la evaluación, las posibilidades de llevarla a cabo, la adecuación y la precisión.

Una especificación de las cuestiones éticas según contextos de intervención se puede plantear a partir del Código Ético de la *American Psychological Association* (1992), que ha sido la asociación de psicología más activa en su interés por editar códigos éticos y velar por su cumplimiento, y cuyos seis principios se hallan también recogidos en el código español elaborado por el Colegio Oficial de Psicólogos. A partir de los principios generales, se ha establecido la siguiente especificación (Anguera, Chacón y Sanduvete, 2008):

**A) Grado elevado de preparación y competencia profesional.** La actualización constante de conocimientos debe constituir un compromiso para el profesional, que debe estar permanentemente cualificado para cumplir su función.

Uno de los aspectos que más repercusiones tienen en los distintos ámbitos es la adecuada instrumentación necesaria. Diferentes estudios realizados en España y en el contexto europeo muestran ciertas deficiencias relativas a esta necesaria competencia profesional del investigador. Como afirma Muñiz (1997), la potencialidad de los tests para causar daño proviene más del uso inadecuado por parte de personas sin preparación o principios, que de sus propiedades técnicas.

En los ámbitos clínico, educativo, deportivo, social, etc., se requiere una gama de diversos instrumentos, desde los de carácter estándar hasta los registros de conducta, o hasta el uso de medidas de erosión (Fang & Ellwein, 1990), y de todos ellos puede derivarse un uso inapropiado (Fremer, 1996). La forma de atacarlo, aunque con intensidad variable, consiste en combinar la formación de los usuarios con la restricción de acceso a los tests de personas no cualificadas.

Una cuestión sumamente espinosa que se plantea con determinada frecuencia en el ámbito de evaluación de programas es su ineficacia, o la baja eficacia, o especialmente la presencia de una efectividad negativa (efectos nocivos que tienen como causa la implementación de un determinado programa). Nos podemos preguntar hasta qué punto es el profesional que implementa un programa el responsable de los resultados inesperados o de los efectos nocivos imprevistos. Se plantea un reto ético fundamental y controvertido, y para cuyo análisis carecemos de criterios claros (Sánchez Vidal, 1999). Parece estar aceptado que la responsabilidad profesional (Asociación de Bioética y Derecho (ABD), 1998) del especialista que implementa un programa no le afecta en el sentido de hacerle responsable de aquello que no puede prever o controlar; pero tampoco puede ser responsable de las consecuencias de su actuación. El uso del *mejor conocimiento disponible* en el momento de decidir o de actuar puede no ser suficiente, y se podría añadir el criterio de que dispusiera de medios complementarios (información, asesoramiento especializado, etc.), pero ello no resuelve el problema ético planteado. Parece lógico que si el profesional actúa de acuerdo a los mejores principios científicos y técnicos existentes, la responsabilidad ha de recaer sobre el colectivo



profesional y científico en su conjunto, y no sobre el profesional en particular; mientras que, por el contrario, si existe negligencia personal o profesional se le debe atribuir la responsabilidad.

En cualquier caso, se requiere del profesional que implementa un programa una correcta formación científica y técnica que garantice la adquisición de los conocimientos soportados por la evidencia existente y de las técnicas conocidas como válidas.

**B) Conducta íntegra del investigador en sus relaciones con los demás.** En el marco de estas relaciones del investigador con los demás se sitúa el uso de la información (Aragón i Viñes, 1997), así como la explicación que se dé a los participantes y sus allegados o responsables acerca de los resultados del estudio.

En el **ámbito sanitario** (Metges *on line*, 2000) la información está al servicio del propósito asistencial (Health on the net (HON), 2000), y toda persona que esté en uso de suficientes capacidades psíquicas tiene derecho a ser informada (Codi de Deontologia, 1997; Roldán y Perea Pérez, 1996; Galán, 1997), respetando la forma de ser de cada usuario. En principio el principal destinatario de la información ha de ser el propio interesado –en nuestro caso, la persona con discapacidad, con la lógica diferenciación entre personas con discapacidad física o intelectual–, excepto en casos excepcionales (menores, personas a las que se puede ocasionar un grave riesgo o perjuicio –“privilegio terapéutico”–, situaciones de urgencia, pacientes que no desean recibir información), y su finalidad es la de conocer bien, consentir y rechazar –en su caso– un determinado tratamiento, así como también oponerse a recibir tratamiento. No puede olvidarse, sin embargo, que el paciente/usuario es vulnerable, y que su situación –y más cuando el pronóstico clínico es dudoso o desfavorable– le sitúan en franca dependencia de los demás, sean familiares o profesionales (Buckman, 1997).

El consentimiento dado por el usuario marca los límites o pautas de la investigación, y no exime a los profesionales de cumplir adecuadamente con sus responsabilidades en el ejercicio de su actividad. El consentimiento, además, ha de pedirse de forma continuada a lo largo de un proceso, a medida de que se utilicen diversos instrumentos, o medios de diagnóstico, y en cualquier momento es revocable por el participante o sus responsables.

A caballo entre los puntos A) y B) se halla la utilización de las nuevas tecnologías de la información en la monitorización de programas. La fuerte innovación que ha supuesto en la vida de millones de ciudadanos no puede desaprovecharse en la actuación del participante en una investigación. La comunicación presencial o por teléfono ha dado paso, en buen número de ocasiones, al E-mail o a la telecomunicación (Jerome, DeLeon, James, Folen, Earles & Gedney, 2000; Jarne y Comisión Deontológico del COPC, 2002); incluso en países anglosajones se ha acuñado el término *telehealth*, entendido como uso de tecnología electrónica y de la información que permite desarrollar una nueva forma de monitorización en la dispensación de programas de salud (Grigsby, 1997; Nickelson, 1998), como podría ser la no detección de diferencias en la alianza terapéutica entre veinte sesiones de psicoterapia por vídeo conferencia y cara a cara, o bien de forma *online*. Es cierto que la rapidez de la evolución tecnológica no permite la regulación de unos estándares, y que además en determinadas zonas rurales o de extrarradio de grandes ciudades son difícilmente accesibles. Pero es igualmente cierto que reavivan cuestiones relevantes, como son las relaciones del investigador con el participante, o del profesional con el usuario, manifestación del consentimiento informado (no hay que olvidar que las grabaciones siempre han de estar autorizadas por quien es grabado, e igualmente hay que pedir permiso si la grabación se va a difundir, por ejemplo, para fines didácticos), preservación de la privacidad y confidencialidad (con las excepciones de que el deber de protección está por encima del secreto profesional, y la salvaguarda de la integridad física del participante/paciente y de aquellos que le rodean), etc. (APA, 1997).



Los valores clave que deben preservarse son: a) salvaguarda de los participantes; b) producción de avances científicos sustantivos; y c) evitar perturbaciones innecesarias a los participantes (Allen, 1996). A éstos se les debe informar acerca de los límites de las nuevas tecnologías de la información, ya que podría derivarse una creencia errónea de que Internet, o una videoconferencia, es confidencial y privado (Maheu, Callan & Nagy, in press). En este sentido, Shapiro & Schulman (1996) han discutido tales aspectos desde un punto de vista ético y legal, y King (1996a, 1996b) ha analizado los aspectos éticos que derivan de la recogida de información a través de Internet, a la vez que ha propuesto unas guías en este sentido.

**C) Utilización de todos los medios a su alcance para asumir la responsabilidad que le corresponde en el desempeño de su actividad.** Estos medios pueden ser muy diversos, y referirse tanto a cuestiones materiales o tecnológicas, como institucionales, o colaborar con otros profesionales a los que competen aspectos relacionados.

**D) Respeto de los derechos y dignidad de las personas con discapacidad en todas sus dimensiones,** sin que sufran menoscabo en función de su raza, sexo, religión, género, nacionalidad, orientación sexual, idioma, nivel socioeconómico, etc. Las diferencias entre personas y entre grupos sociales plantean abundantes problemas éticos, que requieren un adecuado tratamiento de la diferencia (Buxarraís, Carrillo, Galcerán, López, Martín, Martínez, Payà, Puig, Trilla i Vilar, 1990), y que además no siempre plantean un claro compromiso con los valores democráticos.

**E) Responsabilidad social,** en sentido amplio, que trasciende más allá de los usuarios específicos de un determinado programa.

A caballo entre los aspectos tratados por los puntos D) y E), resulta especialmente complejo y delicado tanto el tema de la diferenciación entre individuos como el de la responsabilidad social, por sus connotaciones sociales y metodológicas. Dennis (1994), por ejemplo, aunque con cautela y restricciones, aboga a favor de los experimentos de campo aleatorizados, lo cual, a nuestro juicio, no es posible plantear en numerosas ocasiones. Considera que se requeriría una intensa monitorización, pero nos deberíamos preguntar por la propia aleatorización (¿hasta qué punto se vulnera el “derecho” a la accesibilidad al programa?). Si se plantea un diseño de grupo control, además de los problemas derivados de la habitualmente dudosa homogeneidad de los sujetos, ¿se puede admitir éticamente la adscripción a los grupos experimental y control? Jonas (1997) considera que sí sería justificable experimentar con seres humanos por el bien público, pero, a la vez, también sostiene que este bien reside tanto en el derecho de la sociedad a pedirlo como en el sacrificio del individuo en favor del bien público, de modo que el bien de la sociedad descansaría sobre la “primacía del individuo”, pues es él quien se ha de sacrificar y dar su consentimiento para hacer realidad el derecho reivindicado por la sociedad; de aquí que hay que tener conciencia del peligro que pueden suponer los ensayos clínicos con seres humanos, y se plantea qué sucede con el consentimiento y con la voluntad autónoma e informada de la persona sujeta a ensayos clínicos. Se trata de un difícil equilibrio, ya clásicamente puesto en evidencia por Suls & Rosnow (1981).

Por otra parte, Passamani (1991) sugiere que se debería determinar el momento en que se pierde el punto de equilibrio, cuando el tratamiento ya no resulta eficaz, para detenerlo, pero esta propuesta pone en evidencia la contradicción existente en determinados momentos entre la ortodoxa aplicación de la metodología –dada la perfecta sistematización que plantean los diseños experimentales– y la correcta resolución de las cuestiones éticas. Otras contradicciones entre la aplicación de la metodología y el surgimiento de problemas éticos se pueden hallar en Anguera (2000).

**F) Incremento del bienestar de las personas con discapacidad**, así como también de aquellas otras personas con las que interactúan profesionalmente.

En algunos trabajos (Guisán, 1995) se están planteando nuevas perspectivas para la ética que inciden de forma muy directa con este incremento de bienestar –la denominada *nueva ética*–, de forma que el cambio que supone consiste en “desplazar el foco de atención del mundo de los valores apriorísticos, inmutables y eternos, a un mundo de valores creados por el ser humano a partir de sus necesidades y *desiderata*, como individuo inteligente y desarrollado, imparcial e ilustrado” (p. 193). Llevándolo a nuestro terreno, se trata de un acercamiento de los comportamientos éticos a las necesidades del participante en una investigación, sin dejar de tener en cuenta una perspectiva social. Así se apunta también en Guisán (1995, p. 243) cuando afirma que no cualquier deseo es el criterio ético para el bien y el mal, “sino el mejor deseo, el que mejor contribuya a armonizar los intereses conflictivos de cada individuo en particular y los intereses de los distintos individuos en conflicto”.

Finalmente, y para resumir, se presenta un decálogo sobre la adaptación y armonización de criterios éticos al iniciar una investigación sobre discapacidad, adaptado a partir de la propuesta de Schuler (1982):

1. Que todo protocolo de investigación sobre discapacidad sea examinado por un comité ético independiente.
2. Que el criterio de selección de participantes con discapacidad sea justo y equitativo dentro de diferentes grupos sociales.
3. Que los participantes con discapacidad estén particularmente protegidos de la coerción o de las influencias indebidas.
4. Que el participante con discapacidad sepa a quién recurrir para obtener información independiente sobre la investigación antes de tomar la decisión de participar, así como para exigir la reparación de daños, en caso de que éstos se produzcan durante el transcurso de ella.
5. Que se obtenga un consentimiento válido (propio o sustituto) después de recibir la información suficiente y adecuada respecto a posibles riesgos y beneficios.
6. Que los beneficios y riesgos sean razonablemente proporcionados.
7. Que el registro de datos se haga de tal manera que asegure la privacidad de los participantes.
8. Que haya adecuadas medidas para preservar la confidencialidad de los datos obtenidos.
9. Que los participantes puedan abandonar la investigación en cualquier momento de su desarrollo.
10. Que el participante sepa el derecho que tiene a que se le explique la intención de la investigación después de haber participado en ella, y sepa dónde conocer los resultados globales que se hayan obtenido.

## ÉTICA INSTITUCIONAL EN LA INVESTIGACIÓN SOBRE DISCAPACIDAD

En el proceso de investigación aplicada se cuenta con una realidad compleja, pero tangible, en contextos, situaciones y/o programas en que se implementan acciones que a veces no se ajustan al calendario, o que no se ejecutan por igual en todos los participantes, pero en

dónde el qué, el cómo y el cuándo son registrables. En el amplio abanico institucional que se abre, la incidencia de la ética se manifiesta especialmente en los elementos humanos que constituyen una díada particular: el investigador y la institución que lo patrocina. Vamos a referirnos a ambos, y nos centraremos en el ámbito específico de la *evaluación de programas de intervención*.

Desde el punto de vista del *investigador*, algo se ha escrito sobre su deontología profesional, que acota y regula su práctica profesional. Así, en evaluación de programas, en las Guías para el proceso de evaluación (Fernández-Ballesteros, 1997) figura la ética como garantía en la evaluación psicológica. Por otra, resulta básica la orientación al evaluador acerca de los conflictos que se van a producir en cualquier situación evaluativa. Los múltiples códigos existentes se manifiestan de forma genérica al abordar las cuestiones éticas, y esta generalidad, en cierto modo inevitable e incluso deseable para evitar la rigidez, es una de las mayores debilidades desde una perspectiva de ética institucional. Una de las consecuencias de dicha generalidad es que todos los evaluadores se puedan ver reflejados desde un punto de vista positivo, de forma “políticamente correcta” (Bustelo Ruesta, 1998).

Lo “políticamente correcto” en la actualidad en el ámbito evaluativo consiste en ser flexible con las diferentes aproximaciones y modelos, teniendo en cuenta la voz de los grupos minoritarios, y aceptando las diferentes perspectivas, aunque éstas sean diametralmente opuestas entre sí. Pero hay que ser muy cauto y no dejarse seducir por esta fácil posición flexible en donde todo cabe. El evaluador deberá utilizar sus referentes teóricos y su formación metodológica y sustantiva para actuar profesionalmente y elaborar el correspondiente informe de evaluación.

En los diferentes ámbitos de investigación, y, por seguir la ilustración, también en evaluación de programas, confluyen posiciones epistemológicas plurales, metodologías más o menos interventivas, diferentes polaridades referidas a los usuarios, al programa o a la temporalidad de la implementación, en un intento máximo de integración. A nuestro juicio, ésta ha sido la causa principal de que los propios códigos hayan quedado un tanto insustanciales o desabridos, dejando las orientaciones demasiado vagas o abiertas (Bustelo Ruesta, 1998).

En este sentido, Chelimsky (1995) y Rossi (1995) concuerdan en que muchos de los principios incluidos en los códigos no se preocupan de forma especial por la integridad de los resultados, ni en hacer al evaluador lo suficientemente firme y audaz en su posición a la hora de presentar su informe valorativo, avalados por el proceso evaluativo llevado a cabo. Rossi (1995, p. 58), valientemente, llega a afirmar que la evaluación “no es una negociación para conseguir aprobación, sino la búsqueda de la mejor aproximación a la verdad”. Desde un punto de vista comparado con los de otras disciplinas se muestra pesimista, considerando que prima la debilidad y falta de firmeza: “(...) los códigos potentes y detallados existen en los campos profesionales en los que se ha alcanzado consenso sobre la sustancia y el método, mientras que los códigos débiles afloran en las asociaciones en las que los miembros están divididos en dichas materias” (Rossi, 1995, p. 55).

En efecto, el investigador, y en nuestra ilustración específica, el evaluador de un programa debe manifestarse de forma firme y objetiva –con el respaldo de una metodología rigurosa seguida durante el proceso– respecto a los diferentes aspectos que comporta el proceso evaluativo, y defenderlo de forma argumentada, en la medida en que sea preciso, ante la institución y ante el usuario, así como ante los demás profesionales que participan en la elaboración, implementación, monitorización o gestión de un programa.

En este sentido, existe abierta una polémica que conviene tener presente acerca del investigador. Mientras que, por una parte, se le considera especialista específico en el ám-

bito de la metodología y diseño, por otra, podría serlo en el tema sustantivo. Se trata de una polémica no cerrada, y que se orienta de una u otra forma según el país desde el cual se contempla. Así, en España, en donde la consolidación institucional del investigador en Psicología todavía es baja, se contratan habitualmente a especialistas en el ámbito sustantivo del estudio –en nuestro caso la discapacidad–, sin que lo sean en cuestiones metodológicas, lo cual conlleva una manifiesta debilidad en planteamiento del procedimiento, elaboración de indicadores, instrumentación, diseño, recogida y análisis de datos, proceso de generalización y elaboración del informe. Por el contrario, en países como Canadá o Estados Unidos, con una profunda consolidación institucional de la investigación, existen profesionales especialistas que, posteriormente, se pueden además especializar en el respectivo ámbito sustantivo de la discapacidad; un claro ejemplo fue el profesor G.P. Sackett (Seattle University).

El segundo elemento de la diáda antes mencionada es la *institución* en la cual opera el investigador. Esta institución se configura como uno de los niveles del contexto –el más elaborado y sofisticado–, y, como afirma Chelimsky (1995), este contexto puede llegar a generar una serie de condicionantes al investigador, sean facilitantes o dificultadores. Sin embargo, son numerosas las ocasiones en que el contexto hace que los investigadores deban ser especialmente valientes para realizar sus estudios tal y como creen que deben efectuarse (Sieber, 1980).

La institución, como contexto inmediato, guía y orienta la acción del investigador que promueve y crea nuevo conocimiento en un ámbito específico, a la vez que la limita y la coarta. Puede llegar a ser muy complicada la relación entre la institución y el investigador/evaluador si ambos son ideológicamente divergentes, y la continua o recurrente discrepancia, enfrentamiento e insatisfacción corren el riesgo de llegar a minar actuaciones inicialmente éticas por la anomia de rol creada y el potencial de apatía y conflicto que se irá acumulando entre los miembros de un equipo de trabajo.

En la inmensa mayoría de casos, los programas de investigación obedecen a políticas diversas, pertenecientes a la Administración central, a las autonómicas, o a las locales, aunque no podemos olvidar los promovidos desde las Administraciones institucionales, o los impulsados por asociaciones diversas de carácter no lucrativo. Genéricamente nos referiremos a la *institución* que promueve el estudio.

Los programas de investigación penden de políticas determinadas, y éstas han de ser sensibles a la realidad de cada momento. Si nos referimos a *programas sanitarios*, los nuevos avances tecnológicos, el avance de los conocimientos en programas preventivos, la obtención de nuevos fármacos, la información de que se dispone a nivel epidemiológico, etc., influirá de forma decisiva en la orientación de las actuales políticas sobre personas con discapacidad. Si nos referimos a *programas sociales*, la realidad actual en nuestro país es distinta de la de algunas décadas anteriores: las personas con discapacidad reciben mucha mayor atención; la homogeneidad de la clase trabajadora tradicional se ha visto resquebrajada por la introducción de nuevas tecnologías, por la necesidad de calificaciones crecientes y distintas para desempeñar puestos de trabajo, por la obsolescencia y casi desaparición de algunos sectores de actividad y ocupaciones tradicionales; hay que tener en cuenta la aparición de nuevos grupos, como los desempleados de distinta duración y edad, los jóvenes en busca del primer empleo o las mujeres que desean incorporarse al mercado de trabajo.

Pero es igualmente cierto que esta modernización tiene un reverso pesimista en aquellos casos en los que los ciudadanos no pueden organizar adecuadamente la defensa de sus intereses (Pérez Yruela, 1997). Y así existe población con discapacidad marginada por causa de

la pobreza, la enfermedad, u otras formas de exclusión social o económica; igualmente hay población excluida por razones étnicas o de extranjería.

En las últimas décadas, los gobiernos locales y las instituciones que de ellos dependen han irrumpido en programas que se ocupan de las personas con discapacidad (de promoción ocupacional y de formación continua, de lucha contra la exclusión, de dinamización cultural y el tiempo libre, etc.) Se ha abierto con ello todo un abanico de políticas sociales emergentes, poco relevantes en los modelos clásicos de bienestar, muy conectadas al contexto en el cual se hallan, y directamente ubicadas en el terreno de la proximidad. Estas políticas de servicios personales han permitido diseñar un conjunto de programas articulados en torno a las personas, los grupos, y la comunidad, y orientados al desarrollo sociocomunitario, sobre la base de relaciones integradoras, solidarias y participativas (Brugué y Gomá, 1998), y, aunque todavía de forma tímida, se está empezando a considerar a las personas con discapacidad como receptores de programas específicos, aunque no se trata de programas de investigación, sino de intervención, y será nuestra responsabilidad como investigadores que contribuyamos a generar nuevo conocimiento específico en este ámbito.

Existe un notable consenso en considerar que, en la actualidad, y una vez se dispone de la perspectiva que aporta la contemplación de las décadas de los ochenta y los noventa, la realidad institucional ha evolucionado, especialmente a nivel de instituciones locales, y ello tiene indudable incidencia positiva en el comportamiento ético (Rodríguez-Arana, 1993). El diseño de las políticas de servicios personales locales y de los programas que conllevan apunta hacia una recuperación de las ideas, los valores y los paradigmas como elementos centrales del proceso de toma de decisiones. Por otra parte, se enfatiza la importancia de la planificación, a la cual se aspira a dotar de un rigor técnico y de los instrumentos necesarios que aporten una garantía y estabilidad en las políticas municipales de bienestar (Caufield & Schultz, 1989), conjugando la dirección política (responsable de la institución) con una concertación pluralista y el rigor técnico a nivel legal (Anguera, 2002) e institucional. Y la tercera pata del trípode radica en la importancia de la innovación participativa (Chacón, Anguera, Pérez-Gil & Holgado, 2002), que puede dar lugar a un sinnúmero de matices, que van desde una fuerte implicación de los usuarios hasta la deliberación de alternativas y el diseño de nuevos programas, hasta las aportaciones de voluntariado.

Los tres elementos son positivos desde el punto de vista ético, y con ello se revierte la curva tradicional de relación inversamente proporcional entre participación y eficacia, generando dinámicas de potenciación recíproca entre ambos componentes, y para llevarlo a cabo se requiere (Brugué y Gomá, 1998), por una parte, transferir poder a la comunidad (*empowerment*) mediante diversos mecanismos de innovación democrática, como las conferencias deliberativas, los grupos de mediación social o los núcleos de intervención participativa; por otra, reformular el tipo y alcance de las instancias de implicación institucional del tejido asociativo; y, por último, reestructurar los diseños organizativos de los servicios personales en la línea de suavizar el peso de la burocracia y tender hacia estructuras flexibles de gestión pluralista, asociativa y personalizada. Se está apostando con ello a unos servicios personales –y por tanto programas de ellos derivados– articulados a partir de una doble dimensión (Brugué y Gomá, 1998): Nivel de cobertura e intensidad de protección. En función del nivel (de mínimo a máximo) en cada uno de ellos, se generan diversas modalidades de servicios personales (integrales, intensivo-selectivos, asistenciales y residuales). De entre ellos, la optimalidad de la ética institucional se logra con el modelo integral, caracterizado por aunar políticas contra la exclusión como resultado del aumento de la cobertura del modelo intenso-selectivo, y las de desarrollo comunitario como resultado de la intensificación del modelo asistencial.



## CONTROVERSIAS ÉTICAS EN LA INVESTIGACIÓN SOBRE DISCAPACIDAD

Una controversia es una manifestación de posiciones encontradas, en la cual se plantea la posibilidad –no siempre fácil– de pronunciamiento entre diferentes opciones, propuestas, soluciones, conductas, etc., que son disyuntivas o excluyentes entre sí (Sánchez Valverde, 1997). De forma aún más contundente, Wadeley y Blasco (1995) afirman que muchos dilemas éticos no tienen una solución única y universal, y que es probable que las soluciones puedan ser tantas como posibles los casos que se puedan dar. Aunque parezca fácil distinguir entre qué está bien y qué mal, no se llega a un acuerdo (Kane, 1992).

Sabemos que los diferentes ámbitos profesionales (psicólogos, médicos, abogados,...) cuentan con sus propios códigos deontológicos. El código deontológico constituye una de las señas de identidad que cada grupo profesional se otorga a sí mismo, pero además es un instrumento que permite, en situaciones de duda, tomar decisiones orientadas en torno a ciertos principios consensuados por el respectivo colectivo profesional, con lo cual se dulcifica un tanto la solución que se le dé a la controversia planteada.

Además, los códigos deontológicos no son estáticos, sino que van siendo modificados a medida que las nuevas realidades así lo exigen. Así, en el ámbito de la salud, mientras que hace unos años se aceptaba que un especialista de la medicina tenía capacidad casi absoluta de decisión sobre el tratamiento a un paciente, cada vez es más habitual encontrar médicos y personal sanitario que apelan al *principio de autonomía* de los pacientes a la hora de aplicar ciertos tratamientos, o de decidir sobre la salud de la población. En realidad se trata de una nueva cultura que está dejando de considerar a los usuarios de servicios como menores de edad, pero ese cambio implica una actitud distinta de los profesionales, que van a tener que adaptarse a una realidad que exige de ellos algo más que competencia técnica. En este sentido, el Colegio de Médicos de Cataluña ha modificado su código deontológico, adaptándolo a una nueva realidad tecnológica, social y cultural para la que el anterior había quedado obsoleto (Fuentes Caballero, 1997), y lo mismo ha ocurrido en la Federación de Colegios de Psicólogos de España, que se halla en la fase de aprobación de su nuevo código deontológico.

Existen además otros aspectos, probablemente más secundarios pero igualmente importantes, que igualmente se han de tener en cuenta al valorar las directrices éticas que deben guiar su actuación, como son las posibles diferencias socioculturales existentes entre los participantes y los investigadores, como ocurriría entre los usuarios y los profesionales que evalúan los programas de intervención (Olwenly, 1994; Wadeley y Blasco, 1995). Así, en la aplicación del mencionado principio de autonomía o en el respeto a la confidencialidad, se presupone que ambos conceptos se aplican exclusivamente al individuo; sin embargo, en ciertas culturas o medios sociales la autonomía y la confidencialidad pueden corresponder más a una red de relaciones familiares que al individuo. Y lo mismo puede afirmarse en la implementación y evaluación de ciertos programas de carácter sanitario, deportivo, o educativo, que si bien pueden ser no-maleficientes desde el punto de vista de quién los implementa, sí podrían ser percibidos como tal por parte del usuario que los recibe.

Fuentes Caballero (1997) define el dilema como una situación real que presenta un conflicto de valores y la necesidad de tomar una decisión en uno u otro sentido. Aunque en determinadas ocasiones la controversia puede aparecer como excesivamente simplificadora de la realidad, y no siempre las situaciones se presentan de forma dicotómica, sí es cierto que en ocasiones tienen este carácter (Reynolds, 1979). Los principios y normas de los códigos deontológicos, al aplicarlos en determinadas situaciones, pueden dar lugar a que, al cumplir con una obligación, se esté incumpliendo otra, para esa circunstancia y ese usuario concreto.

El resultado es que el investigador, igual que profesional (Corey, Corey & Callanan, 1993), se pregunta qué debe hacer en ese caso, o, dicho de otro modo, qué principio se debe aplicar, teniendo en cuenta que ambas posibilidades son igualmente relevantes.

Son muchos los interrogantes que nos podemos plantear en investigación sobre discapacidad que suponen una controversia ética: ¿Hay excepciones en disponer del consentimiento informado de los participantes? ¿Hasta qué punto prima el hecho de usar una metodología científica sobre una menos sistemática con la finalidad de adaptarse mejor a la situación real? Actuando de la primera forma se ganaría en credibilidad, pero se podría perder en humanismo ¿Habría algún grado de incomodidad o estrés que podría aceptarse a consecuencia de una intervención? ¿Dónde están los límites de la responsabilidad de quién elabora o implementa un programa? ¿Hasta qué punto se podrían utilizar nuevos avances médicos o tecnológicos que no están totalmente comprobados? ¿Cómo se distribuirían los recursos en función de la temporalidad del proceso (Sánchez-Algarra & Anguera, 2005)? ¿Qué tipo de difusión se puede hacer de los resultados de un estudio? ... Como se indica en APA (1982), todavía muchos psicólogos creen que para obtener datos válidos y generalizables es esencial que los sujetos no estén informados, o consideran que el engaño es necesario en muchas condiciones experimentales, o no están sensibilizados con la invasión de la privacidad.

Indudablemente se plantean conflictos entre la praxis de la evaluación de programas y determinados requisitos metodológicos. La evaluación de programas se desarrolla a lo largo de un proceso en el cual se plantean puntos de colisión que hasta pueden considerarse habituales entre los requisitos metodológicos y la deformación que se tiende a producir en estudios aplicados (Anguera, 2000; Anguera, Chacón y Sanduvete, 2008). Destacamos los siguientes: a) Detección y priorización de necesidades, b) evaluación de resultados vs. de proceso, c) selección de participantes; d) uso de instrumentos; y e) Elaboración de estándares de calidad.

Después de un análisis sistemático de la situación que tenga en cuenta no sólo los principios en juego, sino otros más amplios, la controversia puede diluirse, pero esto no elimina la primera vivencia, presentada como disyuntiva difícil y ante la cual el profesional se siente desorientado.

La gama de controversias éticas es muy amplia (Pryzwansky & Wendt, 1999). Entre los muchos estudios que se han llevado a cabo sobre conflictos éticos, Morris & Cohn (1993) confeccionaron, a partir de una encuesta realizada a miembros de la *American Evaluation Association*, un amplio catálogo acerca de retos o controversias, y en la mayoría de casos se presentan curiosamente en la elaboración del informe final. Pero pueden surgir controversias éticas en cualquier momento del proceso, o en un momento precedente, como en la evaluación de necesidades, o en un momento posterior, cuando se trata de dilucidar la autoría de una determinada contribución, en donde los problemas éticos que se llegan a plantear son de considerable envergadura (Fine & Kurdek, 1993).

## REFERENCIAS BIBLIOGRÁFICAS

- Allen, C. (1996). What's wrong with the "Golden Rule"? Conundrums of conducting ethical research in cyberspace. *Information Society*, 12, 175-187.
- Allen, D.F. & Allen, V.S. (1979). *Ethical issues in mental retardation*. Nashville, Tenn: Parthenon.
- Altarejos, F., Ibáñez-Martín, J.A., Jordán, J.A. y Jover, G. (1998). *Ética docente*. Barcelona: Ariel.



- American Psychological Association (APA), Ethics Committee (1997, November). *Services by telephone, teleconferencing, and Internet*. Washington, DC: Author. [Consulta hecha el 7/3/2000 en <http://www.apa.org/ethics/stmnt.01.html>].
- American Psychological Association (1982). *Ethical principles in the conduct of research with human participants*. Washington: APA.
- Anguera, M.T. (2000). Diseños de evaluación de programas: ¿Un reto para la Administración y las instituciones? En A. Blanco, D. Rojas, L. de la Corte, J.D. Delius, J.J. Miguel Tobal, R. Fernández-Ballesteros, M.T. Anguera y M.J. Díaz-Aguado, *Psicología y Sociedad* (pp. 127-158). Valencia: Real Sociedad Económica de Amigos del País.
- Anguera, M.T. (2002). Marco competencial en Servicios Sociales de Atención a la Tercera Edad en Cataluña. En J. Tornos Más (Coord.), *Los servicios sociales de atención a la tercera edad. El caso de Cataluña* (pp. 551-587). Valencia: Tirant lo Blanch.
- Anguera, M.T., Chacón, S. y Sanduete, S. (2008, en prensa). Cuestiones éticas en evaluación de programas. En M.T. Anguera, S. Chacón y A. Blanco (Coords.), *Evaluación de programas sociales y sanitarios. Un abordaje metodológico*. Madrid: Síntesis.
- Aragonés i Viñes, T. (1997, Octubre). Ús i abús de la informació (Mesa Redonda). *Actes de la 1ª Jornada sobre Serveis Socials d'Atenció Primària Dilemes ètics en la nostra pràctica: Entre el control i la inserció* (pp. 51-53). Barcelona: Col.legi Oficial de Psicòlegs de Catalunya, Col.legi Oficial de Diplomats en Treball Social i Assistents Socials de Catalunya, i Col.legi d'Educadors i Educadores Socials de Catalunya.
- Arboleda-Florez, J. (2003). La investigación en sujetos humanos: Poblaciones vulnerables. En F. Lolas y A. Quezada (Eds.), *Pautas éticas de investigación en sujetos humanos: Nuevas perspectivas* (pp. 129-142). Santiago de Chile: Programa Regional de Bioética OPS (Organización Panamericana de la Salud) / OMS (Organización Mundial de la Salud). [Consulta hecha el 2/5/2008 en <http://www.bioetica.ops-oms.org/E/docs/pautas.pdf#page=83>].
- Asociación de Bioética y Derecho (ABD) de la Universidad de Barcelona (2000). Resumen de la sesión de 25/11/1998. [Consulta hecha el 23/1/2000 en <http://www.ub.es/fildt/abdcas6e.htm>].
- Beauchamp, T.L. & Childress, J.F. (1999). *Principles of biomedical ethics*. New York: Oxford University Press.
- Bermant, G. & Warwick, D.P. (1978). The ethics of social intervention: Power, freedom and accountability. In G. Bermant, H.C. Kelman & D.P. Warwick (Comps.), *The ethics of social intervention* (pp. 379-417). New York: Halstead Press.
- Brown, R.D. & Newman, D.L. (1992). Ethical principles and evaluation standards. Do they match? *Evaluation Review*, 16 (6), 650-663.
- Brugué, Q. y Gomá, R. (1998). La dimensión local del bienestar social: El marco conceptual. En Q. Brugué y R. Gomá (Coords.), *Gobiernos locales y políticas públicas. Bienestar social, promoción económica y territorio* (pp. 39-56). Barcelona: Ariel.
- Buckman, R. (1997). *Com donar les males notícies. Una guia per a professionals de la salut*. Vic, Barcelona: Eumo.
- Bustelo Ruesta, M. (1998). Deontología de la evaluación: El modelo de los códigos éticos anglosajones. *Gestión y Análisis de Políticas Públicas*, 11-12, 141-156.
- Buxarrais, M.R., Carrillo, I., Galcerán, M. del M., López, S., Martín, M.J., Martínez, M., Payà, M., Puig, J.M., Trilla, J. i Vilar, J. (1990). *Ètica i escola: El tractament pedagògic de la diferència*. Barcelona: Edicions 62.
- Callahan, J. (Ed.) (1988). *Ethical issues in professional life*. Oxford: Oxford University Press.
- Casado, M. (Ed.) (1998). *Bioética, Derecho y Sociedad*. Madrid: Trotta.

- Caufield, I. & Schultz, J. (1989). *Planning for change*. London: Longman.
- Chacón, S., Anguera, M.T., Pérez-Gil, J.A. & Holgado, F.P. (2002). A Mutual Catalytic Model of Formative Evaluation: The Interdependent Roles of Evaluators and Local Programme Practitioners. *Evaluation. The International Journal of Theory, Research and Practice*, 8 (4), 413-432.
- Chelimsky, E. (1995). Comments on the Guiding Principle. In W.H. Shadish, D.L. Newman, M.A. Scheirer & C. Wye (Eds.), *Guiding Principles for Evaluators* (pp. 35-54). San Francisco: Jossey-Bass.
- Codi de Deontologia (1997). *Normes d'Ètica Mèdica*. Barcelona: Consell de Col.legis de Metges de Catalunya.
- Comisión Deontológica Estatal (1999). *Ética y Deontología para Psicólogos*. Madrid: Colegio Oficial de Psicólogos de España.
- Corey, G., Corey, M. & Callanan, P. (1993). *Issues and ethics in the helping professions*. Pacific Grove, CA.: Brooks Cole.
- Dennis, M.L. (1994). Ethical and practical randomized field experiments. In J.S. Wholey, H.P. Hatry & K. E. Newcomer (Eds.), *Handbook of Practical Program Evaluation* (pp. 155-197). San Francisco: Jossey-Bass.
- Fang, W.L. & Ellwein, M.C. (1990). Photography and ethics in evaluation. *Evaluation Review*, 14 (1), 100-107.
- Fernández-Ballesteros, R. (1997). Guidelines for the Assessment Process (GAP). *European Psychologist*, 2 (4), 352-355.
- Fine, M.A. & Kurdek, L.A. (1993). Reflections on determining authorship credit and authorship order on faculty-student collaborations. *American Psychologist*, 48 (11), 1141-1147.
- França-Tarragó, O. (1996). *Ética para psicólogos*. Bilbao: Desclée de Brouwer.
- Fremer, J. (1996). Promoting high standards for test use: Developments in the United States. *European Journal of Psychological Assessment*, 12, 160-168.
- Fuentes Caballero, M.T. (1997, Octubre). Dilemes ètics en la intervenció social (Mesa Redonda). *Actes de la 1ª Jornada sobre Serveis Socials d'Atenció Primària Dilemes ètics en la nostra pràctica: Entre el control i la inserció* (pp. 25-32). Barcelona: Col.legi Oficial de Psicòlegs de Catalunya, Col.legi Oficial de Diplomats en Treball Social i Assistents Socials de Catalunya, i Col.legi d'Educadors i Educadors Socials de Catalunya.
- Galán, J.C. (1997). *El consentimiento informado del usuario de los Servicios Sanitarios*. Madrid: Coalex.
- Grisgsby, B. (1997). *ATSP report on U.S. telemedicine activity*. Portland, OR: Association of Telemedicine Service Providers. [Consulta hecha el 6/1/2000 en <http://www.atsp.org>]
- Guisán, E. (1995). *Introducción a la Ética*. Madrid: Cátedra.
- Health on the net (HON) (2000). Health on the net Code of Conduct for medical and health web sites. [Consulta hecha el 23/1/2000 en <http://www.hon.ch/HONcode/>]
- Herranz, G. (1992). *Comentarios al Código de Ética y Deontología Médica*. Pamplona: Eunsa.
- Hess, A.K. (1987). The ethics of forensic psychology. In I.B. Weiner & A.K. Hess (Eds.), *Handbook of forensic Psychology* (pp. 653-681). New York: Wiley & Sons.
- Jarne, A. y Comisión Deontológica del COPC (2001). Hacia un código deontológico de la Intervención psicológica a través de Internet. *Anuario de Psicología*, 32 (2), 117-126.
- Jerome, L.W., DeLeon, P.H., James, L.C., Folen, R., Earles, J. & Gedney, J.J. (2000). The coming of age of telecommunications in psychological research and practice. *American Psychologist*, 55 (4), 407-421.

- Joint Committee on Standards for Educational Evaluation (1981). *Standards for evaluations of educational programs, projects, and materials*. New York: McGraw-Hill.
- Jonas, H. (1997). *Técnica, medicina y ética. La práctica del principio de responsabilidad*. Barcelona: Paidós.
- Kane, M.T. (1992). The assessment of professional competence. *Evaluation & The Health Professions*, 15 (2), 163-182.
- King, S.A. (1996a). Commentary on responses to the proposed guidelines. *Information Society*, 12, 199-201.
- King, S.A. (1996b). Researching Internet communities: Pproposed ethical guidelines for the reporting of results. *Information Society*, 12, 119-127.
- Kitchener, K.S. (1984). Intuition, critical thinking, and ethical principles: The foundation of ethical decisions in counseling psychology. *The Counseling Psychologist*, 12, 43-55.
- Leal Rubio, J. (1997, Octubre). Dilemes ètics en la intervenció social (Mesa Redonda). *Actes de la 1ª Jornada sobre Serveis Socials d'Atenció Primària Dilemes ètics en la nostra pràctica: Entre el control i la inserció* (pp. 32-41). Barcelona: Col.legi Oficial de Psicòlegs de Catalunya, Col.legi Oficial de Diplomats en Treball Social i Assistents Socials de Catalunya, i Col.legi d'Educadors i Educadors Socials de Catalunya.
- Metges *on line* (2000). La informació clínica facilitada al pacient. [Consulta hecha el 23/1/2000 en <http://www.comb.es/publicacions/quaderns/cat/bp8.htm>].
- Morris, M. & Cohn, R. (1993). Program evaluators and ethical challenges. A National Survey. *Evaluation Review*, 17 (6), 621-642.
- Muñiz, J. (1997). Aspectos éticos y deontológicos de la evaluación psicológica. En A. Cordero (Coord.), *La evaluación psicológica en el año 2000* (pp. 307-345). Madrid: TEA Ediciones.
- Newman, D.L. & Brown, R.D. (1996). *Applied ethics for program evaluation*. Thousand Oaks, CA.: Sage.
- Nickelson, D.W. (1998). Telehealth and the evolving health care system: Strategic opportunities for professional psychology. *Professional Psychology: Research and Practice*, 29, 527-535.
- Olwenly, C. (1994). The ethics and conduct of cross-cultural research in developing countries. *Psycho-Oncology*, 3, 11-20.
- Passamani, E. (1991). Clinical trials – Are they ethical? *New England Journal of Medicine*, 324 (22), 1589-1592.
- Pérez Yruela, M. (1997). Corporativismo y Estado del bienestar. En S. Giner y S. Sarasa (Eds.), *Buen gobierno y política social* (pp. 133-150). Barcelona: Ariel.
- Pryzwansky, W.B. & Wendt, R.N. (1999). *Professional and ethical issues in Psychology. Foundations of practice*. New York: Norton & Company.
- Reynolds, P.D. (1979). *Ethical dilemmas and social science research*. San Francisco: Jossey-Bass.
- Rodríguez-Arana, J. (1996). *Ética institucional*. Madrid: Dykinson.
- Roldán, B. y Perea Pérez, B. (1996). *El consentimiento informado en la práctica médica*. Madrid: Smi-hkline Beecham.
- Rossi, P.H. (1995). Doing it good and getting it right. In W.H. Shadish, D.L. Newman, M.A. Scheirer & C. Wye (Eds.), *Guiding Principles for Evaluators* (pp. 55-60). San Francisco: Jossey-Bass.
- Sánchez-Algarra, P. & Anguera, M.T. (2005). Time management in the cost evaluation of limited resource programs. *Quality & Quantity. International Journal of Methodology*, 39 (4), 391-411.

- Sánchez Valverde, C. (1997, Octubre). Dilemes ètics en la intervenció social (Mesa Redonda). *Actes de la 1ª Jornada sobre Serveis Socials d'Atenció Primària Dilemes ètics en la nostra pràctica: Entre el control i la inserció* (pp. 41-50). Barcelona: Col.legi Oficial de Psicòlegs de Catalunya, Col.legi Oficial de Diplomats en Treball Social i Assistents Socials de Catalunya, i Col.legi d'Educadors i Educadors Socials de Catalunya.
- Sánchez Vidal, A. (1999). *Ética de la intervenció social*. Barcelona: Paidós.
- Schuler, H. (1982). *Ethical problems of psychological research*. New York: Academic Press.
- Sieber, J.E. (1980). Being ethical: Professional and personal decisions in program evaluation. In R. Perloff & E. Perloff (Eds.) *Values, ethics, and standards* (pp. 51-63). San Francisco: Jossey-Bass.
- Shapiro, D.E. & Schulman, C.E. (1996). Ethical and legal issues in e-mail therapy. *Ethics & Behaviour*, 6, 107-125.
- Suls, J.M. & Rosnow, R.L. (1981). The delicate balance between ethics and artifacts in behavioral research. In A.J. Kimmel (Ed.), *Ethics of human subject research* (pp. 55-65). San Francisco: Jossey-Bass.
- Susanne, Ch. (Ed.) (1998). *Reading book in bioethics: A multidisciplinary approach*. Florencia-Bruxelles: European Association of Global Bioethics.
- Thompson, A. (1990). *Guide to ethical practice in psychotherapy*. New York: Wiley.
- Wadeley, A. y Blasco, T. (1995). *La ética en la investigación y la práctica psicológicas*. Barcelona: Ariel.



## ANEXO

### PROGRAMAS DE LOS SEMINARIOS DE ACTUALIZACIÓN EN INVESTIGACIÓN SOBRE DISCAPACIDAD (SAID)

#### I SAID: PRESENTE Y FUTURO DE LA INVESTIGACIÓN EN DISCAPACIDAD

El I SAID, dirigido por Miguel Ángel Verdugo Alonso, tuvo lugar los días 16 y 17 de octubre de 1998, y el programa fue el siguiente:

- Desarrollo psicológico y ciclo vital en la discapacidad. *Angel Rivière Gómez (Universidad Autónoma de Madrid)*.
- Investigación neurobiológica en discapacidad. *Julio Feroso García (Universidad de Salamanca)*.
- El estudio de la discapacidad desde el punto de vista psicosocial. *Vicente Pelechano Barberá (Universidad de La Laguna. Tenerife)*.
- Atención a las personas con discapacidad en el sistema educativo. *Climent Giné Giné (Universitat Ramón Llull. Barcelona)*.
- La medicina conductual y sus aplicaciones en el ámbito de la discapacidad. *José Antonio Carrobes (Universidad Autónoma de Madrid)*.
- Conclusiones del I SAID. *Miguel Ángel Verdugo Alonso (Universidad de Salamanca)*.

#### II SAID: INCLUSIÓN, CAMBIO ORGANIZACIONAL, PLANIFICACIÓN CENTRADA EN LA PERSONA Y APOYOS NATURALES

El II SAID, dirigido por Miguel Ángel Verdugo Alonso y Cristina Jenaro Río, tuvo lugar los días 15, 16 y 17 de marzo de 2000. Excepcionalmente, en esta ocasión, se contó con la presencia de un solo ponente, David Mank (Indiana University), que participó en las siguientes actividades:

- Conferencia: Cambio de Sistemas y Planificación Centrada en la Persona.
- Mesa Redonda: Modelos de Prestación de Servicios: de las Instituciones a la Comunidad.

### III SAID: CUESTIONES METODOLÓGICAS Y PROCEDIMENTALES

El III SAID, dirigido por Miguel Ángel Verdugo Alonso y Benito Arias Martínez y coordinado por Manuela Crespo Cuadrado, tuvo lugar los días 13 y 14 de diciembre de 2001, y el programa fue el siguiente:

- Bases metodológicas y teóricas de análisis científico. Modelos jerárquicos y lineales: Alternativas y ejemplos. *Vicente Pelechano Barberá (Universidad de La Laguna. Tenerife).*
- ¿Investigación en discapacidad y envejecimiento? Perspectivas y problemas metodológicos. *Antonio León Aguado Díaz (Universidad de Oviedo).*
- Metodologías cualitativas: Características, proceso y aplicaciones. *María Teresa Anguera Argilaga (Universidad de Barcelona).*
- Aportaciones recientes en metodología observacional: Estimación y Generalización. *Ángel Blanco Villaseñor (Universidad de Barcelona).*
- Metodología para la construcción y validación de instrumentos de evaluación: El ICAP. *Delfín Montero Centeno (Universidad de Deusto).*
- Resumen del Seminario y perspectivas. *Benito Arias Martínez (Universidad de Valladolid).*

### IV SAID: ACTUALIZACIÓN METODOLÓGICA EN INVESTIGACIÓN SOBRE DISCAPACIDAD

El IV SAID, dirigido por Miguel Ángel Verdugo Alonso y Gerardo Prieto Adánez, tuvo lugar los días 27 y 28 de mayo de 2004, y el programa fue el siguiente:

- Meta-análisis en investigación sobre discapacidad. *Julio Sánchez Meca (Universidad de Murcia).*
- Aplicaciones de los modelos tipo Rasch en la medición psicológica. *Gerardo Prieto Adánez (Universidad de Salamanca).*
- Posibilidades de los diseños observacionales en la investigación sobre discapacidad. *M<sup>a</sup> Teresa Anguera Argilaga (Universidad de Barcelona).*
- Los modelos de ecuaciones estructurales: posibilidades y aplicaciones. *Juana Gómez Benito (Universidad de Barcelona).*
- Aplicaciones del análisis factorial en el desarrollo, evaluación y validación de instrumentos psicométricos. *Pere Joan Ferrando i Piera (Universitat Rovira i Virgili).*
- La evidencia empírica en Psicología: crítica metodológica. *Juan Delgado Sánchez-Mateos (Universidad de Salamanca).*
- La investigación cualitativa: su aplicación a la investigación y análisis de la realidad. *M<sup>a</sup> Cruz Sánchez Gómez (Universidad de Salamanca).*

### V SAID: ACTUALIZACIÓN METODOLÓGICA EN INVESTIGACIÓN SOBRE DISCAPACIDAD

El V SAID, dirigido por Miguel Ángel Verdugo Alonso y Pilar Sarto Martín y coordinado por M<sup>a</sup> Cruz Sánchez Gómez, Marta Badia Corbella y Manuela Crespo Cuadrado, tuvo lugar los días 25 y 26 de mayo de 2006, y el programa fue el siguiente:



- Medición con escalas Tipo Likert. *Gerardo Prieto Adánez (Universidad de Salamanca)*.
- Técnica Delphi: el consenso entre expertos. *José Ignacio Ruiz Olabuénaga (Universidad de Deusto)*.
- Técnicas de análisis cuantitativo aplicadas en investigaciones cualitativas: análisis cluster/conglomerados. *José Serrano Angulo (Universidad de Málaga)*.
- Evaluación de Programas: práctica en ámbitos educativos. *M<sup>a</sup> José Rodríguez Conde (Universidad de Salamanca)*.
- La ética en la investigación sobre discapacidad. *M<sup>a</sup> Teresa Anguera Argilaga (Universidad de Barcelona)*.
- El análisis discriminante: posibilidades y aplicaciones. *Gregorio Rodríguez Gómez (Universidad de Cádiz)*.
- El análisis cualitativo. *M<sup>a</sup> Cruz Sánchez Gómez (Universidad de Salamanca)*.
- El proceso de análisis con ayuda del Programa NVivo. *Gregorio Rodríguez Gómez (Universidad de Cádiz)*.
- El proceso de análisis con ayuda del Programa Atlas-Ti. *Miguel Vicente Mariño (Universidad Autónoma de Barcelona)*.

## VI SAID: INTRODUCCIÓN A LA METODOLOGÍA DE ECUACIONES ESTRUCTURALES APLICADA A LA INVESTIGACIÓN EN DISCAPACIDAD

El VI SAID, dirigido por Marta Badia Corbella, Benito Arias Martínez y Miguel Ángel Verdugo Alonso y coordinado por Manuela Crespo Cuadrado, tuvo lugar los días 5 y 6 de junio de 2008, y el programa fue el siguiente:

- Introducción a la metodología SEM: concepto y propósitos fundamentales. *Begoña Orgaz Baz (Universidad de Salamanca)*.
- Modelos de medida y análisis factorial confirmatorio. *Ramón Fernández Pulido (Universidad de Salamanca)*.
- Estrategias para el manejo de "missing data". *Ángel Blanco Villaseñor (Universidad de Barcelona)*.
- Desarrollo de un ejemplo de "path analysis" en LISREL. *Begoña Orgaz Baz (Universidad de Salamanca)*.
- Modelos con componentes de medida y estructurales. *Miguel Ruiz Díaz (Universidad Autónoma de Madrid)*.
- Generalizabilidad y 'missing data' en investigación sobre discapacidad. *Ángel Blanco Villaseñor (Universidad de Barcelona)*.
- Desarrollo de un ejemplo de análisis factorial confirmatorio en LISREL, AMOS y SAS. *Benito Arias Martínez (Universidad de Valladolid)*.
- Desarrollo de un problema de ecuaciones estructurales. *Miguel Ruiz Díaz (Universidad Autónoma de Madrid)*.