

EVALUACIÓN DE LA VALIDEZ DEL CAMBIO INDUCIDO
POR LOS TRATAMIENTOS PSICOLÓGICOS
Una propuesta metodológica aplicada a la fobia social

K. Salabería-D. Páez-E. Echeburúa

Karmele Salaberria es profesora en el Departamento de Personalidad, Evaluación y Tratamientos Psicológicos de la Universidad del País Vasco Enrique Echeburúa es Catedrático en el mismo Departamento (Avda. de Tolosa, 70, 20009-San Sebastián). Darío Páez es Catedrático de Psicología Social en la Universidad del País Vasco.

El problema de la validez individual de la evaluación de tratamientos

Una crítica razonable a la evaluación clínica mediante diseños experimentales y cuasi-experimentales es la invalidez individual de los resultados basados sobre medias de grupo (Gerin,

1984; Gerin, Dazord y Sali, 1991; Matalon, 1988). Al basarse en comparaciones grupales, se ocultan los cambios individuales. De hecho, un tratamiento puede aparecer como el más efectivo y, sin embargo, mejorar sólo el estado psicológico de algunas personas. Asimismo las medias pueden permanecer constantes y no revelar que un subgrupo ha mejorado y otro empeorado.

Por otra parte, si el grupo de control tiene una media inferior posttest al grupo experimental, puede deberse a distintas circunstancias: a) que una minoría sustancial haya mejorado y la mayoría siga igual en el grupo de tratamiento; b) que un sector haya mejorado y otro haya empeorado, siendo el primero predominante en el grupo de tratamiento. Este último resultado también es compatible con una ausencia de diferencias de medias, si los grupos de empeoramiento y mejora son iguales. En un caso extremo, las medias grupales pueden no corresponder a ningún sujeto ni grupo de sujetos.

Según Matalon (1988), la validez individual es un componente y una profundización de la validez interna. La validez individual es la capacidad de trasponer a nivel individual una relación que se ha observado entre grupos a nivel agregado. Se trata de una versión del clásico problema de la falacia ecológica y de la desagregación de resultados.

La validez individual es importante cuando los mecanismos de cambio estudiados (disonancia, relación entre dimensiones de actitudes, afectividad y procesos atribucionales o cognitivos en general) se han formulado teóricamente a un nivel individual. La heterogeneidad de los sujetos —que la aleatorización en el caso experimental y el apareo en el caso de los grupos no equivalentes, buscan reducir pero que no eliminan del todo—, puede hacer que las relaciones observadas entre los grupos sean compatibles con diferentes relaciones individuales.

Además, la validez individual es importante, ya que ella expande el denominado análisis interno, para contrastar directamente la acción de los mecanismos de cambio individuales postulados. El análisis interno consiste en estimar la asociación entre las variables dependientes en un grupo experimental. Recuérdese que muchos de los modelos de relaciones internas postulados para el cambio de actitudes, basados en medias, no se confirman cuando se analizan las relaciones entre las variables mediadoras y las variables dependientes. Las variables mediadoras son las que explican la influencia de la causa en el efecto. Por ejemplo, las creencias favorables/desfavorables sobre los efectos de una conducta determinan la actitud ante ella. Un cambio de actitud favorable, desde esta perspectiva, se debe acompañar de una disminución de las creencias desfavorables y de un aumento de las favorables. Sin embargo, se ha comprobado que una intervención puede modificar las conductas, sin modificar las creencias sobre los efectos de ellas (Dillard, 1991).

Otra crítica, diferente pero convergente con la anterior, se dirige al hecho que los diseños experimentales y cuasi-experimentales clásicos demuestran relaciones funcionales (no en un sentido estrictamente matemático, en el que una función es una relación predictiva sin error entre dos variables) entre variables. Los diseños se limitan a mostrar la existencia de relaciones de tipo funcional, mostrando que a mayor nivel de la variable independiente, mayor nivel de la variable dependiente. Como afirman Cook y Shadish (1994), los diseños clásicos integran una información causal descriptiva. Dicho de otra manera, muestran una relación entre X e Y, sin permitir comprender por qué X afecta a Y. Son diseños que aportan poca información sobre los procesos y los mecanismos de los que las relaciones son la manifestación y la consecuencia. La descomposición de la relación entre X e Y en procesos mediadores y la

obtención de indicadores de éstos, así como de la relación entre éstos y las variables dependientes, constituyen la forma natural de progresar en el conocimiento.

Finalmente, los diseños clásicos tienen dificultades para estimar el impacto individual de los tratamientos. Esto último es muy importante para evaluar programas de intervención. La estimación del nivel de cambio individual en la variable dependiente es fundamental para estimar tanto en qué sujetos el tratamiento ha tenido éxito como en cuáles otros hay que seguir interviniendo porque no han cambiado o han empeorado. En este texto se van a proponer respuestas metodológicas a estos problemas.

La aplicación de las propuestas metodológicas presentadas en este trabajo se lleva a cabo, a modo de ejemplo ilustrativo, sobre una investigación referida a la eficacia diferencial de tratamientos psicológicos en la terapia de la fobia social (Salaberría y Echeburúa, 1995; *cfr.* anexo).

La evaluación del cambio individual

Una forma sencilla de poder evaluar si los cambios producidos entre grupos se pueden trasponer a nivel individual es utilizar indicadores de cambio y de estabilidad entre sujetos. Las diferencias individuales antes/después son un ejemplo de lo primero y las correlaciones constituyen un ejemplo de lo segundo.

Las medidas de cambio permiten establecer tipologías de cambio, así como comprobar si las diferencias son mínimas o nulas. Además, estos indicadores dan una visión de la heterogeneidad o variabilidad de la variable dependiente evaluada. Hay cuatro métodos principales utilizados para estimar los cambios:

a) *Las puntuaciones brutas de cambio en la variable dependiente entre el tiempo 1 y el tiempo 2.* Estas medidas dependen de que las medidas sean de intervalo, aunque el uso de la tasa de cambio (cambio en relación al nivel inicial) se puede utilizar como una alternativa. Sin embargo, este indicador no toma en cuenta la influencia del nivel inicial de la variable dependiente. Además, los errores en la medición también atentan contra la fiabilidad de este indicador de cambio.

b) *Las puntuaciones de cambio residual o de cambio corregido.* Estiman el cambio de cada sujeto en comparación con el cambio medio de sujetos que estaban en el mismo nivel o estado. Operacionalmente, implica realizar un análisis de covarianza o una regresión, utilizando el nivel inicial de puntuación en la variable dependiente como predictor. Si bien es más satisfactorio, este análisis puede enmascarar fenómenos específicos de ciertos niveles de partida de la variable dependiente. Por

ejemplo, generalmente los sujetos de puntuaciones extremas tienen un ritmo de cambio diferente del de los sujetos medios.

c) *El tamaño del efecto*. Sirve para comparar los efectos del tratamiento en contextos y con instrumentos diferentes. Consiste en utilizar como baremo la dispersión del grupo de control, de un grupo normativo o la dispersión conjunta del grupo de control y del grupo de tratamiento. Operacionalmente es la diferencia entre el antes y el después, dividido por la desviación típica —habitualmente del grupo de control. Si se aplican las áreas de la distribución sobre el punto medio del grupo de tratamiento, se puede estimar dónde se sitúa este grupo en relación al grupo de control. Si el punto medio del grupo de tratamiento se encuentra a 0.93 desviaciones típicas por encima de la media del grupo de control, normativo o de la distribución común —como se encuentra en el metaanálisis de Shapiro y Shapiro, 1982— aplicando la distribución normal, se puede afirmar que el paciente medio tratado se halla en una situación superior al 82% de los pacientes del grupo no tratado.

d) *La evaluación del nivel alcanzado*. Esta evaluación se realiza en relación a un criterio establecido de significación clínica, que permite evaluar el nivel de funcionalidad o disfuncionalidad del sujeto. Esta medida es importante, en particular en áreas clínicas, educativas y de salud, en las que existen criterios adaptativos mínimos relativamente claros. Esta medida de cambio ignora el nivel de partida ya que toma en cuenta sólo adónde llega el sujeto (Gerin, 1984, Gerin *et al.*, 1991; Matalon, 1988).

La propuesta presentada en este texto para la evaluación del cambio individual está inspirada en Jacobson y Truax (1991). Se trata de establecer una puntuación de cambio exacto, que permite estimar la validez individual del cambio. Además, la puntuación de cambio baremada a partir de una muestra normativa y el establecimiento de puntos de corte sustantivos y basados estadísticamente permiten tener una estimación que integra los cuatro tipos previos de cambio estimado.

El problema de las puntuaciones de cambio

La utilización de medidas de cambio ha recibido también críticas importantes. La diferencia entre el pretest y el posttest es una medida de cambio intuitiva y muy utilizada. Sin embargo, durante los años 60 y 70 la conclusión había sido negativa sobre la validez y fiabilidad de estas medidas. Se llegó a aconsejar su abandono basándose en su falta de utilidad, en particular para los diseños cuasi experimentales de grupos no equivalentes (por ejemplo, Crano y Brewer, 1977). Las críticas centrales a las medidas de cambio eran las siguientes.

a) Las medidas de cambio están sujetas a efectos techo (el rango de puntuación limita la evaluación del cambio). La utilización del BDI, que es un indicador de depresión o malestar, como indicador de bienestar sería un ejemplo de esto. Ya que la puntuación más baja del BDI sólo indica que no se está deprimido, es incapaz de estimar a los sujetos que cambian y llegan a estar con alto bienestar en el postest. Esta limitación depende del rango de respuesta del instrumento y es resoluble fácilmente.

b) La regresión a la media hace que los sujetos que puntúan extremadamente alto o bajo la primera vez se desplacen hacia la media en evaluaciones posteriores. Las puntuaciones de cambio sufrirían una distorsión por este fenómeno.

c) Las puntuaciones de cambio no son fiables. Existe una relación inversa entre la fiabilidad de las puntuaciones de cambio de una medida (ansiedad, por ejemplo) y la estabilidad o fiabilidad test-retest (correlación entre las medidas pre y postest de ansiedad, por ejemplo). A mayor cambio, menor fiabilidad de las puntuaciones de cambio.

d) Las medidas de cambio suponen que hay intervalos iguales en los diferentes puntos del instrumento. En un test de 100 ítems, el cambio de 90 a 95 se presupone igual al que ocurre de 40 a 45, si es un test de capacidad. Igualmente, una mejoría en el BDI de 21 a 15 se supone igual a la de 10 a 4. Como ambos ejemplos demuestran, es más difícil subir de 90 a 95 en rendimiento y bajar de 21 (depresión fuerte) a 15 que de 10 (límite depresivo no clínico) a 4 (estado normal). En otras palabras, existe una correlación negativa entre la puntuación inicial y la puntuación de cambio de diferencia (cuanto más baja es la puntuación inicial, mayor es la puntuación de cambio positiva si es una medida de capacidad o malestar y menor es el cambio negativo si es una medida de déficits o síntomas). El cambio puede ser mayor en los sujetos con puntuaciones extremas.

A modo de reflejo de la evolución metodológica actual, la versión más reciente del texto de Crano y Brewer (1986) no reproduce críticas tan extremas a las puntuaciones de cambio y sugiere la utilización de varias medidas repetidas y de series temporales como alternativa. Además, según Crano y Brewer (1986), las limitaciones de las puntuaciones de cambio se dan de forma extrema sólo en el caso de los grupos no equivalentes, formados seleccionando sujetos con puntuaciones extremas en el pretest.

Con respecto al tercer problema —la relación inversa entre la fiabilidad de las puntuaciones de cambio y la consistencia temporal del mismo instrumento—, se trata de un aspecto preocupante, ya que se supone que la correlación pre-post es el indicador de la fiabilidad o grado en que

el instrumento mide el mismo constructo en dos tiempos diferentes. Sin embargo, esta presuposición sólo es válida si los sujetos cambian al mismo ritmo, lo que es poco probable. La fiabilidad de una puntuación de diferencia es función de varios factores: a) la fiabilidad o precisión del instrumento (correlación pre-post); b) el lapso de tiempo entre las dos mediciones (a mayor cercanía, mayor es la correlación en general); y c) la variabilidad del cambio real en los sujetos. Si los sujetos cambian al mismo ritmo o tasa relativa (no necesariamente el cambio es igual en puntuaciones brutas), al aumentar cada uno de los componentes antes descritos, aumenta la fiabilidad de las puntuaciones de cambio. Cuando los sujetos cambian a ritmos variables, una baja correlación pre-postest no indica que el instrumento mide diferentes constructos, sino que está reflejando las diferencias individuales en crecimiento (Francis, Fletcher, Stuebing, Davidson y Thompson, 1991).

Un aspecto central para estimar los cambios es evaluar la variabilidad de respuestas antes-después de los sujetos. Supongamos un caso en que el tratamiento va a mejorar de forma importante a un grupo de personas y empeorar claramente a otras. La media pre-post no cambiará, pero, sin embargo, la dispersión será mucho mayor en la segunda ocasión. Recuérdese que existen pruebas estadísticas, como la de Cochran, que permiten estimar si las varianzas difieren entre ellas. Por otro lado, una correlación débil entre el pre y el post-test es un indicador de la heterogeneidad de los sujetos, así como de una fuerte variabilidad de sus varianzas. En otras palabras, es un indicador del cambio entre sujetos. Por ejemplo, en un caso extremo, se puede obtener una media en el pre y la misma en el post, así como la misma varianza. Sin embargo, la correlación del pre con el post es baja o nula. Este resultado, obtenido en un diseño cuasi-experimental por Matalon (1988), lejos de indicar una estabilidad, como sugerirían las medias y las varianzas estables, muestra cambios importantes de los sujetos, sin que se pueda extraer una tendencia clara.

Finalmente, conviene señalar que una disminución de la varianza indica una tendencia a la homogeneización, es decir, que algunos sujetos han cambiado sistemáticamente a menos y otros a más. Los hechos antes expuestos implican que es conveniente utilizar análisis en los que el sujeto es su propio control (análisis de varianza simple o múltiple de medidas repetidas con el sujeto como fuente de variación; regresión múltiple con las puntuaciones originales).

En lo referente al cuarto problema —la presunción de intervalos iguales en los diferentes puntos del cuestionario—, es importante cuando se trata de estimar cuáles son los mediadores o variables micro que explican la acción causal del tratamiento. Recuérdese que los mismos

cambios en las puntuaciones tienen un significado diferente cuando están en la zona media o extrema. Hay que distinguir la correlación entre la puntuación inicial del sujeto y el cambio observado, de la correlación entre la puntuación real o verdadera inicial y la puntuación real de cambio. El valor de la segunda correlación, que es el indicador válido para estimar los mediadores y correlatos del cambio, sólo está indicado parcialmente por la primera. El error de medida es el factor fundamental de distorsión de esta relación. El efecto del error medida es generalmente el de subestimar la relación entre el status inicial y la puntuación de cambio real. La puntuación de cambio directa no es un buen estimador del cambio real, ya que está distorsionada por la puntuación de error. Debido a esto, no es una medida adecuada para analizar la acción de los mediadores o las correlaciones entre el cambio y las variables procesuales que explican la acción de la causa.

Para precisar los puntos anteriores, vamos a definir y analizar más exactamente la puntuación de cambio y los procedimientos estadísticos que nos permiten operacionalizar el estudio del cambio.

La puntuación de cambio o C es igual a la diferencia entre el pre y el post-test ($C = A \text{ o pre} - B \text{ o post}$).

La fiabilidad de la puntuación de cambio o r_{CC} es igual a una combinación de la consistencia interna y de la estabilidad del instrumento:

$$r_{CC} = (r_{vv} - r_{ab}) / (1 - r_{ab})$$

En esta ecuación r_{CC} es la fiabilidad de la puntuación de cambio, r_{vv} la fiabilidad o consistencia del instrumento en las dos ocasiones y r_{ab} la estabilidad o fiabilidad test-retest.

Recuérdese que este último indicador (r_{ab}) contiene tanto la fiabilidad test-retest como el efecto de las diferencias individuales de cambio. Supongamos que el coeficiente alfa medio es alto (0.81) y el test-retest medianamente alto (0.64); en este caso, la fiabilidad es sólo de 0.47 (Cohen y Cohen, 1983). La solución a esta dificultad es utilizar la correlación atenuada por la fiabilidad (Cohen y Cohen, 1983).

La sugerencias para afrontar los problemas anteriores de estudio del cambio proponen utilizar puntuaciones de cambio residual o estandarizado. Operacionalmente, se utiliza la regresión de la puntuación inicial en la puntuación de cambio, o la regresión simultánea del tratamiento y de la puntuación inicial en el posttest (Cronbach y Furby, 1970; Reichardt, 1979; Cohen y Cohen, 1983; Francis, Fletcher, Stuebing, Davidson y Thompson, 1991).

La regresión de B o posttest en A y en el tratamiento (1=grupo experimental; 0= grupo control) permite una estimación del cambio en B, manteniendo constante A, ajustando B para A o residualizando B por A. En el caso de B residualizado por A es una variable cuya relación con el tratamiento es atribuible únicamente a éste. El efecto de la pertenencia de grupo es ajustado o igualado por el grupo de covariables en el análisis de varianza —en nuestro caso, por el nivel inicial de la variable dependiente o A. Algunos autores han criticado, sin embargo, la lógica y validez estadística de este análisis (Willet, Ayoub y Robinson, 1991). En el caso de la psicoterapia, los resultados obtenidos con la puntuación de cambio bruta y la puntuación de cambio residual son similares (Gerin et al., 1991). Las revisiones actuales llegan a las siguientes conclusiones:

a) Las puntuaciones de cambio no tienen necesariamente baja fiabilidad en condiciones normales, e incluso, cuando las tasas de cambio de los sujetos son diferentes entre los sujetos, la fiabilidad de las puntuaciones de cambio puede ser más alta que la de sus componentes pre y posttest.

b) La correlación entre la puntuación inicial y la puntuación de cambio no es siempre negativa. Esto tiende a ocurrir cuando no se corrige o ajusta la correlación del error de medida, utilizando la corrección de fiabilidad, o cuando se tipifican simultáneamente el pre y el posttest, provocando la desaparición de información importante para el cambio (Willet, Ayoub y Robinson, 1991).

En síntesis, las puntuaciones de cambio son fiables y útiles y resultan válidas para estimar el cambio antes/después de un tratamiento, sobre todo cuando se ajustan o corrigen del error de medida y se toma en cuenta el nivel de partida.

Recuérdese que una medida de cambio integrativa debería reunir los siguientes requisitos:

- a) Otorgar una estimación del tamaño del efecto generalizable a otros contextos.
- b) Tomar en cuenta criterios de funcionalidad.
- c) Atenuar la estimación del cambio por el error de medida; tomar en cuenta el nivel de partida; y ofrecer una estimación de cambio individual, que nos permita establecer la validez del cambio

Además, es importante contrastar la validez individual de los mecanismos mediadores del cambio (d). En el caso de la fobia social se supone que los cambios cognitivos son los que explican la mejoría conductual. Que la media posttest del grupo de tratamiento sea menor en cogniciones ansiógenas y muestre un menor nivel de conductas de evitación social no es suficiente para confirmar la validez de un modelo causal de tipo cognitivo (la cognición disfuncional provoca o media la

conducta evitativa). Los siguientes procedimientos responden a estas exigencias.

a) *Efecto del tratamiento generalizable a otros contextos. Evaluación de tratamientos y diseño de línea base no-causal construida*

El diseño de línea base no-causal construida, que no es sino una ampliación de la vieja técnica psicométrica de la baremación, permite obtener una estimación del efecto de cambio generalizable (Perachio y Cook, 1988).

En este diseño, el grupo de control no equivalente es reemplazado por la comparación con una puntuación baremada de una muestra representativa. Si se construyera un baremo de percentiles (a partir, por ejemplo, de una muestra representativa de 800 sujetos) sobre ansiedad, depresión, competencia social, apoyo social, etc., se podría evaluar la eficacia de los tratamientos psicológicos y psiquiátricos en un país determinado mediante la comparación con este baremo. En lugar de comparar la evaluación pretratamiento con la evaluación postratamiento, con las limitaciones de validez interna clásicas, se puede comparar el percentil medio en el pre-test con el percentil medio en el post-test. La diferencia entre ambas medias permite estimar la fuerza de la intervención y facilita los procedimientos de acumulación evaluativa mediante el metaanálisis. Asimismo, si se igualan los grupos en, por ejemplo, edad, sexo y nivel de ingresos, se pueden comparar el pre-test y el post-test del grupo de tratamiento con un grupo apareado de la muestra representativa (cfr. Perachio y Cook, 1988; Páez, Echeburúa y Borda, 1993).

Los problemas de este diseño se pueden superar con una sola investigación en una muestra normativa amplia con medidas repetidas.

El diseño de línea base no-causal construida debe adecuarse a ciertas exigencias. En primer lugar, se puede suponer que los tratamientos psicológicos afectan a sólo una pequeña parte de la población, por lo que la presencia del factor de protección en la población general será limitado.

En segundo lugar, no siempre se dispone de grupos de tratamiento que respondan a los criterios sociodemográficos de la muestra normativa (por ejemplo, en lo relativo a la edad, el sexo, la ocupación, etc.), cuando el modelo de comparación baremada supone que el grupo de tratamiento y el grupo de control provienen de poblaciones equivalentes en cuanto a historia, tiempo, lugar y características demográficas. Si esto no es así —y una limitación de los tests baremados disponibles es la baremación sobre muestras no siempre representativas— se cuestiona un supuesto básico: que los intervalos de cambio son estables en el tiempo y que son

los mismos para la muestra normativa y para los grupos en tratamiento. Una forma de resolver este problema es utilizar baremos extraídos de muestras del mismo ámbito geográfico al que pertenecen los pacientes y utilizar como grupo de comparación a sujetos "normales" dentro del rango de edad, sexo, profesión, etc., predominante entre los pacientes. En nuestro caso tenemos una muestra de sujetos de Canadá.

En tercer lugar, el modelo de comparación baremada supone que un cambio en los puntos brutos es igual al número de cambios en el percentil de la población, pero esto no es así ya que los resultados son distintos según sea el nivel de partida. De hecho, si se parte de percentiles medios con menos puntos, se consiguen más cambios (Perachio y Cook, 1988).

Y en cuarto lugar, las limitaciones genéricas de los diseños cuantitativos (prestar atención a la evaluación del resultado de las terapias psicológicas, pero no así a la evaluación del proceso de las mismas, es decir, a cómo funcionan y a qué tipo de factores se atribuye el cambio del comportamiento) se mantienen también en este caso (Clemente, 1989; Hersen y Michelson, 1986).

Estas críticas han sido mencionadas previamente y su solución se asocia, por un lado, a estimaciones más refinadas del cambio, tomando en cuenta el nivel de partida y el error de medición, y por otro, a la necesidad de tener indicadores de los mecanismos procesuales de la acción causal. En nuestro caso, podemos suponer que los cambios en el miedo a la evaluación negativa serán mediadores del cambio de conducta. Más adelante desarrollaremos nuestras respuestas a estas limitaciones. Previamente vamos a mostrar la aplicación del diseño de línea base no causal construida en los datos del ejemplo que estamos utilizando.

Para ello utilizaremos dos instrumentos de evaluación que han sido utilizados en las principales investigaciones sobre la fobia social: la *Escala de Ansiedad y Evitación Social* (SAD) (Watson y Friend, 1969) y la *Escala de Temor a la Evaluación Negativa* (FNE) (Watson y Friend, 1969).

La *Escala de Ansiedad y Evitación Social* (SAD) (Watson y Friend, 1969) consta de 20 ítems, con dos posibilidades de respuesta (verdadero o falso), referidos al malestar subjetivo en situaciones sociales y a la evitación activa o deseo de evitación de dichas situaciones. La fiabilidad test-retest es de 0.68 y cuenta con un coeficiente alfa de consistencia interna de 0.90. Las puntuaciones oscilan entre 0 y 28: a mayor puntuación, mayor es la cantidad de situaciones en las que el paciente experimenta ansiedad social y adopta conductas de evitación.

La *Escala de Temor a la Evaluación Negativa* (FNE) (Watson y Friend, 1969) consta de 30 ítems de tipo dicotómico relacionados con el temor

a recibir críticas o apreciaciones negativas por parte de los demás. La fiabilidad test-retest es de 0.78 y cuenta con un coeficiente alfa de 0.94.

Las dos escalas se utilizan conjuntamente, la primera para evaluar aspectos conductuales y la segunda para evaluar aspectos cognitivos.

Existen dos formas de aplicación del modelo de comparación baremada:

1) Comparar las puntuaciones medias de antes y de después del tratamiento en los grupos terapéuticos con los percentiles de la muestra normativa.

La distribución en percentiles de las puntuaciones del SAD y del FNE no están disponibles para una muestra normativa de nuestro contexto. Las medias y desviaciones típicas del estudio de fiabilidad y validez, que fue realizado con una muestra de 205 estudiantes universitarios normales de Toronto (Canadá) en 1969, son las siguientes: SAD $X=9.11$ ($DT=8.01$) y FNE $X=15.47$ ($DT=8.62$).

La distribución de una muestra disfuncional de nuestro contexto está disponible a partir de los pretests de los 60 sujetos de nuestra investigación (Tabla 1).

TABLA 1
PERCENTILES DE LA MUESTRA EXPERIMENTAL ($N=60$)

	<i>SAD pre</i>	<i>FNE pre</i>
90	28	29
80	27	28.5
70	26.5	28
60	26	27
50	24.5	27
40	23.5	25.5
30	22	25
20	21	23
10	20	21

Las medias posttest del grupo de tratamiento se sitúan por debajo del percentil 10 de la muestra disfuncional unificada, tanto para el SAD ($X=13.25$; $DT=7.42$) como para el FNE ($X=16.52$; $DT=8.19$). Esto quiere decir que los sujetos del grupo de tratamiento están mejor que el 90% de la muestra disfuncional de referencia (sus propias puntuaciones y las del grupo de control en el pretest). El cambio es de 3 deciles para la escala FNE (de media 40 a 10) y de 4 deciles para la escala SAD (de media 50

a 10), es decir, el cambio es similar al efecto medio de Shapiro y Shapiro (1982), aunque más marcado en la Escala de Estrés y Evitación Social.

2) Comparar los grupos terapéuticos y la muestra normativa antes y después del tratamiento por contraste de hipótesis (t). Si la terapia es eficaz, no deben existir diferencias entre las medias del grupo que ha recibido el tratamiento en el post y las medias de los sujetos normales o, al menos, estas diferencias deben ser menos significativas que en el pretratamiento.

TABLA 2
MEDIAS Y DESVIACIONES TÍPICAS EN EL "SAD"

	PRE	6 MESES
Baremo (N=205)	9.11 (8.01)	-----
Tratamiento (N=40)	24.02 (3.22)	13.25 (7.42)
Control (N=20)	23.91 (3.71)	23 (4)

Antes del tratamiento $t = -20.14$ $p < 0.001$ (Baremo y Tratamiento)		
Después del tratamiento $t = -3.20$ $p < 0.01$ (Baremo y Tratamiento)		
Antes del tratamiento $t = 0.22$ (n.s.) (control y tratamiento)		
Después del tratamiento $t = -9.11$ $p < 0.001$ (control y tratamiento)		

El efecto del tratamiento comparado con el baremo es importante pero las diferencias siguen siendo significativas a los 6 meses. Si se compara con el grupo de control en el pre, no existen diferencias y sí en cambio a los 6 meses a favor del grupo tratado (Tabla 2).

TABLA 3
MEDIAS Y DESVIACIONES TÍPICAS EN EL "FNE"

	PRE	6 MESES
Baremo (N=205)	15.47 (8.62)	-----
Tratamiento (N=40)	25.58 (3.67)	16.52 (8.19)
Control (N=20)	25.52 (3.39)	24.45 (5.36)

Antes del tratamiento $t = -12.18$ $p < 0.001$ (Baremo y tratamiento)		
Después del tratamiento $t = -0.73$ (n.s.) (Baremo y tratamiento)		
Antes del tratamiento $t = 0.10$ (n.s.) (control y tratamiento)		
Después del tratamiento $t = -5.38$ $p < 0.001$ (control y tratamiento)		

Al comparar las puntuaciones del baremo y de los grupos tratados las diferencias desaparecen por efecto del tratamiento y, en cambio, aparecen con el grupo de control (Tabla 3).

En resumen, a pesar de que el tratamiento es válido, los sujetos tratados no llegan a alcanzar los niveles de malestar y evitación de la población normal, pero sí se asemejan en las puntuaciones obtenidas en el miedo a la evaluación negativa. Recuérdese que la media del grupo de tratamiento indica que el sujeto medio de este grupo está mejor que el 90% de los sujetos de un grupo disfuncional, así como que el cambio es más marcado en el SAD que en el FNE.

b) Criterios de funcionalidad estadísticos

Una medida de cambio integral no sólo debería ser generalizable a partir de un baremo de población funcional y disfuncional, sino que la información sobre estas poblaciones debería concretarse en criterios de funcionalidad basados estadísticamente.

Esto implica establecer puntos de corte para obtener un indicador de recuperación significativo que tenga base estadística.

Según Jacobson y Truax (1991), hay cuatro fórmulas para establecer estadísticamente un punto de corte que diferencie a poblaciones funcionales de disfuncionales. Estas fórmulas aplican parcial o simultáneamente tres patrones: a) que el sujeto después del tratamiento esté fuera del rango de una población disfuncional (por ejemplo, dos desviaciones típicas por encima de la media de un grupo deficitario); b) que el sujeto esté en el posttest dentro del rango de una población funcional (por ejemplo, dentro de menos dos desviaciones de la media de una población "normal"); y c) que el sujeto esté en el post-test más distante de la media del grupo disfuncional y más próximo a la media del grupo normal.

1. *Coeficiente C1*: cuando las desviaciones típicas del grupo de tratamiento y de la muestra de la población general son similares (circunstancia poco habitual). Para nuestros datos, este coeficiente implica sumar la media del grupo normal (9.11) con la del grupo disfuncional (24.02) y dividir por dos $(9.11 + 24.02)/2 = 16.56$ en el SAD.

$$C1 = (Mo + M1)/2 \quad \text{SAD } C1 = 16.56 \quad \text{FNE } C1 = 20.52$$

2. *Coeficiente C2*: cuando las desviaciones típicas entre el grupo disfuncional y la muestra normal son diferentes. El procedimiento implica, para el caso del SAD, multiplicar la media del grupo disfuncional por la desviación típica del grupo funcional (24.02×8.01) y sumar esta cantidad a la media del grupo funcional por la desviación típica del grupo

disfuncional (9.11×3.22), todo ello dividido por la suma de ambas desviaciones típicas ($8.01 + 3.22 = 11.23$).

$$C2 = (DT_{oxM1} + DT_{1xMo}) / (DT_o + DT_1)$$

$$SAD\ C2 = 19.74 \quad FNE\ C2 = 22.55$$

3. *Coefficiente C3*: cuando no hay una muestra normal de comparación. Si bien esta fórmula es más exigente que C1 y C2, cuando las distribuciones entre la muestra normal y la disfuncional no se solapan esta fórmula es muy "liberal". En el caso del SAD sería así: $24.02 - (2 \times 3.22) = 17.58$.

$$C3 = M1 - (2\ DT_1) \quad SAD\ C3 = 17.58$$
$$FNE\ C3 = 18.24$$

4. *Coefficiente C4*: cuando hay una muestra normal. Este último coeficiente se puede utilizar de forma más exigente aplicando la misma fórmula con una desviación típica.

$$C4 = Mo + (2\ DT_o) \quad SAD\ C4 = 25.13$$
$$FNE\ C4 = 32.71$$

(este último resultado es imposible ya que el rango del cuestionario es de 0 a 30)

$$\text{Coeficiente } C4 = Mo + (1\ DT_o) \quad SAD\ C4 = 17.12$$
$$FNE\ C4 = 24.09$$

Si se utiliza el coeficiente C3, la media de los sujetos tratados está por debajo del punto de corte ($SAD\ \bar{X} = 13.25$; $FNE\ \bar{X} = 16.52$) a los 6 meses, pero no así la media de los sujetos del grupo de control ($SAD\ \bar{X} = 23$; $FNE\ \bar{X} = 24.45$).

c) *Estudio de los cambios individuales corregido por el error de medida y tomando en cuenta la puntuación inicial*

Como hemos dicho, una medida de cambio integral debe ser generalizable a partir de un baremo de población funcional y disfuncional. Para ello se deben postular criterios de funcionalidad basados estadísticamente, así como tomar en cuenta la puntuación de error y el

punto de partida de cada sujeto, lo que proporciona una estimación del cambio individual y permite establecer la validez individual del cambio.

Como respuesta al problema de la fiabilidad, varios autores han sugerido la corrección del error de medición como forma de mejorar las puntuaciones de cambio. Jacobson y Truax (1991) han elaborado una puntuación de cambio precisa o fiable según la cual se puede establecer en qué medida ha cambiado cada sujeto según la variabilidad error, además de tomar en consideración el punto de partida.

Una puntuación de cambio precisa superior a 1.96 es poco probable (con una probabilidad de ocurrencia menor de 0.05) en el caso de que no haya un cambio real. Todo cambio superior a 1.96 se puede considerar una recuperación real.

Este procedimiento permite además obtener una puntuación individualizada de cambio. Este hecho es de fundamental interés para el clínico, ya que permite decidir cuáles son los sujetos que deben recibir aún tratamiento, así como diferenciar los efectos de la terapia (Tabla 4).

TABLA 4
Procedimiento para calcular la puntuación de cambio precisa (PCP)

- 1) Fórmula para calcular PCP = $(M1-M2)/D_{diff}$
 $D_{diff} = \sqrt{2} (DE)^2$ $DE = DT1\sqrt{(1-r12)}$
- 2) Clasificación a los sujetos en categorías:
 I= Igual, sin cambios positivos ni negativos
 E= empeoramiento del sujeto, cambio negativo (N=no mejorado).
 M= cambio positivo, pero inferior a 1 PCP. Mejoría sin recuperación total ($M < 1$).
 RL= cambio positivo importante, superior a 1, pero inferior a 1.96 ($1.96 > RL > 1$).
 R= cambio positivo importante y muy poco probable que sea debido al azar ($R > 1.96$).
- 3) Aplicar la t de Student como estadístico de contraste para averiguar las características de los sujetos.
 SAD $r = 0.21+$ FNE $r = 0.35^{**}$ $p < 0.01$
 $\bar{X} = 23.98(3.36)$ $\bar{X} = 25.56(3.56)$
 PCP en el SAD: $DE = 3.36\sqrt{(1-0.21)} = 2.98$
 $D_{diff} = \sqrt{2} (2.98)^2 = 4.21$
 PCP en el FNE: $DE = 3.56\sqrt{(1-0.35)} = 2.87$
 $D_{diff} = \sqrt{2} (2.87)^2 = 4.05$

TABLA 5
PUNTUACIONES INDIVIDUALES DE CAMBIO SIGNIFICATIVO EN EL SAD

Sujeto	Pre	6 m	Difer.	PCP	Recup.	Coef.2	Coef.3
1	28	11	17	4.03	R	S	S
2	26	8	18	4.27	R	S	S
3	21	5	16	3.80	R	S	S
4	28	17	11	2.61	R	S	S
5	21	16	5	1.18	RL	S	S
6	25	13	12	2.85	R	S	S
7	21	18	3	0.71	M	S	N
8	22	16	6	1.42	RL	S	S
9	22	23	-1	-0.23	E	N	N
10	28	27	1	0.23	M	N	N
11	27	0	27	6.41	R	S	S
12	28	18	10	2.37	R	S	N
13	27	17	10	2.37	R	S	S
14	26	0	26	6.17	R	S	S
15	28	6	22	5.22	R	S	S
16	26	9	17	4.03	R	S	S
17	28	24	4	0.95	M	N	N
18	25	7	18	4.27	R	S	S
19	15	10	5	1.18	RL	S	S
20	25	13	12	2.85	R	S	S
21	27	22	5	1.18	RL	N	N
22	20	9	11	2.61	R	S	S
23	18	2	16	3.80	R	S	S
24	23	3	20	4.75	R	S	S
25	19	18	1	0.23	M	S	N
26	22	11	11	2.61	R	S	S
27	26	23	3	0.71	M	N	N
28	24	3	21	4.98	R	S	S
29	20	19	1	0.23	M	S	N
30	23	8	15	3.56	R	S	S
31	26	14	12	2.85	R	S	S
32	22	16	6	1.42	RL	S	S
33	27	16	11	2.61	R	S	S
34	17	10	7	1.66	RL	S	S
35	27	15	12	2.85	R	S	S
36	27	17	10	2.37	R	S	S
37	24	2	22	5.22	R	S	S
38	22	24	-2	-0.47	E	N	N
39	22	13	9	2.13	R	S	S
40	24	27	-3	-0.71	E	N	N

Tipo de tratamiento

Exposición: Sujetos 1-19

Exposición más reestructuración cognitiva: Sujetos 20-40

PUNTUACIONES INDIVIDUALES DE CAMBIO SIGNIFICATIVO EN EL SAD

Sujeto	Pre	6 m	Difer.	PCP	Recup.	Coef.2	Coef.3
41	22	23	-1	-0.23	E	N	N
42	25	24	1	0.23	M	N	N
43	28	24	4	0.95	M	N	N
44	28	27	1	0.23	M	N	N
45	27	20	7	1.66	RL	N	N
46	21	18	3	0.71	M	S	N
47	21	19	2	0.47	M	S	N
48	27	26	1	0.23	M	N	N
49	25	27	-2	-0.41	E	N	N
50	21	21	0	0	IGUAL	N	N
51	27	26	1	0.23	M	N	N
52	16	13	3	0.71	M	S	S
53	24	20	4	0.95	M	N	N
54	20	2	-4	-0.95	E	N	N
55	26	26	0	0	IGUAL	N	N
56	24	26	-2	-0.47	E	N	N
57	26	26	0	0	IGUAL	N	N
58	21	17	4	0.95	M	S	S
59	24	27	-3	-0.71	E	N	N
60	27	26	1	0.23	M	N	N

Grupo de Control: Sujetos 41-60

S=Sí o N=No recuperado, según el punto de corte 2 (19.74) y el punto de corte 3 (17.58) a los 6 meses.

TABLA 6
CLASIFICACIÓN DE LOS SUJETOS

	R	RL	M	E	I
Exp. (N=19)	12 (48%)	3 (42.85%)	3 (17.64%)	1 (12.5%)	-----
Exp+R.Cog (N=21)	13 (52%)	3 (42.85%)	3 (17.64%)	2 (25%)	-----
Control (N=20)	-----	1 (14.28%)	11 (64.7%)	5 (62.5%)	3 (100%)
TOTAL	25 (41.66%)	7 (11.66%)	17 (28.66%)	8 (13.33%)	3 (5%)

Los sujetos con una mejoría importante (R+RL), situados por igual en ambas modalidades terapéuticas, son 32 (25 + 7) y constituyen el 63.32% de la muestra.

Los sujetos no mejorados (M+E+I) son 28 (17 + 8 + 3), constituyen el 46.66% de la muestra y principalmente se sitúan en el grupo de control. La prueba de Chi cuadrado es significativa, por lo que hay más sujetos recuperados en el grupo de tratamiento que en el de control (Tabla 6).

Un porcentaje importante de los sujetos del grupo de control (el 64.7%) tiende a experimentar un cambio positivo entre las dos evaluaciones. Es decir, la perspectiva de recibir tratamiento puede inducir ya cambios positivos. Los sujetos recuperados según el coeficiente PCP superan el punto de corte del coeficiente 2. Se confirma de este modo la validez convergente de ambos conceptos.

TABLA 7
PUNTUACIONES INDIVIDUALES DE CAMBIO SIGNIFICATIVO EN EL FNE

Sujeto	Pre	6 m	Difer.	PCP	Recup.	Coef.2	Coef.3
1	29	9	20	4.93	R	S	S
2	29	22	7	1.72	RL	S	N
3	21	5	16	3.95	R	S	S
4	26	12	14	3.45	R	S	S
5	25	17	8	1.97	R	S	S
6	28	26	2	0.49	M	N	N
7	16	10	6	1.48	RL	S	S
8	28	26	2	0.49	M	N	N
9	23	25	-2	-0.37	E	N	N
10	28	23	5	1.23	RL	N	N
11	28	1	27	6.66	R	S	S
12	29	26	3	0.74	M	N	N
13	23	14	9	2.22	R	S	S
14	25	2	23	5.67	R	S	S
15	29	8	21	5.18	R	S	S
16	27	9	18	4.44	R	S	S
17	29	28	1	0.24	M	N	N
18	27	17	10	2.46	R	S	S
19	27	27	0	0	IGUAL	N	N
20	25	14	11	2.71	R	S	S
21	23	25	-2	-0.17	E	N	N
22	21	10	11	2.71	R	S	S
23	25	8	17	4.19	R	S	S
24	27	7	20	4.93	R	S	S
25	28	26	2	0.49	M	N	N
26	26	13	13	3.20	R	S	S
27	27	25	2	0.49	M	N	N
28	17	11	6	1.48	RL	S	S
29	27	23	4	0.98	M	N	N
30	25	5	20	4.93	R	S	S
31	28	25	3	0.74	M	N	N
32	22	15	7	1.72	RL	S	S
33	29	23	6	1.48	RL	N	N
34	19	18	1	0.24	M	S	S
35	28	20	8	1.97	R	S	N
36	28	16	12	2.96	R	S	S
37	28	7	21	5.18	R	S	S
38	27	26	1	0.24	M	N	N
39	26	10	16	3.95	R	S	S
40	29	27	2	0.49	M	N	N

Tipo de tratamiento

Exposición: Sujetos 1-19

Exposición más reestructuración cognitiva: Sujetos 20-40

TABLA 7
PUNTUACIONES INDIVIDUALES DE CAMBIO SIGNIFICATIVO EN EL FNE

Sujeto	Pre	6 m	Difer.	PCP	Recup.	Coef.2	Coef.3
41	25	27	-2	-0.49	E	N	N
42	24	25	1	0.24	M	N	N
43	29	29	0	0	IGUAL	N	N
44	27	28	-1	-0.18	E	N	N
45	27	29	-2	-0.49	E	N	N
46	21	18	3	0.74	M	S	S
47	16	12	4	0.98	M	S	S
48	25	13	12	2.96	R	S	S
49	28	28	0	0	IGUAL	N	N
50	21	22	-1	-0.24	E	S	N
51	27	25	2	0.49	M	N	N
52	22	18	4	0.98	M	S	S
53	26	26	0	0	IGUAL	N	N
54	25	26	-1	-0.18	E	N	N
55	25	27	-2	-0.49	E	N	N
56	29	29	0	0	IGUAL	N	N
57	29	29	0	0	IGUAL	N	N
58	22	21	1	0.24	M	S	N
59	29	29	0	0	IGUAL	N	N
60	29	28	1	0.24	M	N	N

Grupo de Control: Sujetos 41-60

S=Sí o N=No recuperado, según el punto de corte 2 (22.55) y el punto de corte 3 (18.24) a los 6 meses.

TABLA 8
CLASIFICACIÓN DE LOS SUJETOS

	R	RL	M	E	I
Exp. (N=19)	10 (47.61%)	3 (50%)	4 (22.22%)	1 (12.5%)	1 (14.28%)
Exp+R.Cog (N=21)	10 (47.61%)	3 (50%)	7 (38.88%)	1 (12.5%)	-----
Control (N=20)	1 (4.76%)	-----	7 (38.88%)	6 (75%)	6 (85.71%)
TOTAL	21 (35%)	6 (10%)	18 (30%)	8 (13.33%)	7 (11.66%)

Los sujetos con una mejoría importante (R+RL) son 27 (21 + 6) y constituyen el 45% de la muestra, sin diferencias entre las dos modalidades terapéuticas. Es decir, la exposición por sí sola se muestra igual de eficaz que la exposición más reestructuración cognitiva para cambiar el miedo a la evaluación negativa. El 90% de los sujetos recuperados (R) según el coeficiente PCP del FNE superan el punto de corte de los coeficientes 2 y 3.

Los sujetos no mejorados (M+E+Igual) son 33 (18 + 8 + 7), constituyen el 54.99% de la muestra y principalmente se sitúan en el grupo de control. La prueba de Chi cuadrado es significativa, por lo que hay más sujetos recuperados en el grupo de tratamiento que en el de control. A diferencia de lo que ocurría con el SAD, los sujetos del grupo de control tienden a la mejoría en un porcentaje mucho menor (el 38.88%) (Tabla 8).

El porcentaje de mejoría es menor en esta variable cognitiva que en la conductual. Las explicaciones plausibles son varias: a) la preocupación o pesimismo constituyen casi un rasgo de personalidad que es difícilmente modificable; b) la intervención terapéutica no ha sido adecuada para este propósito o no se ha llevado a cabo con el formato adecuado; y c) la evaluación de las cogniciones es mucho más difícil de llevar adelante. El siguiente procedimiento nos permitirá dirimir entre estas diferentes alternativas.

d) Estudio de la validez individual de los procesos mediadores

Las puntuaciones de cambio individualizadas corregidas por el error de medida y con puntos de corte basados estadísticamente sirven para examinar la validez individual del cambio de cogniciones y conductas antes descrito. Ya sabemos que la media posttest del grupo de tratamiento es menor en evitación y en cogniciones disfuncionales. Además, debemos confirmar que las personas que más disminuyen las cogniciones disfuncionales también muestran la mayor mejoría conductual. De forma más directa, las puntuaciones de cambio de las escalas SAD y FNE deben estar asociadas. Recuérdesse que no siempre los procesos de cambio postulados teóricamente se encuentran en la práctica, aunque las medias de las puntuaciones muestren cambios aparentemente coherentes. Se pueden disminuir la conducta evitativa y las creencias negativas como media, pero, sin embargo, no existir ninguna relación -o una relación pequeña-entre el cambio cognitivo y el conductual.

Con el fin de contrastar la validez individual de los cambios evaluados en los grupos, examinamos la asociación entre el cambio cognitivo y el conductual. Este análisis entre las variables dependientes del tratamiento se ha denominado análisis interno (Tabla 9).

TABLA 9
ANÁLISIS INTERNO: CORRELACIONES ENTRE PUNTUACIONES DE CAMBIO DE CONDUCTAS Y DE CAMBIO DE COGNICIONES

PCPsad - PCP fne:	$r = 0.82$ *** (N=60)
	$r = 0.85$ *** (Exposición)
	$r = 0.73$ *** (Expo. + restruc.cognitiva)
	$r = 0.11$ n.s. (Grupo de Control)

$p < 0.001$ ***

SAD: Escala de Estrés y de Evitación Social
FNE: Escala de Miedo a la Evaluación Negativa

El cambio en el nivel de miedo a la evaluación negativa se asocia al cambio en el nivel de evitación y muestra la validez individual en este caso de la relación. Esta asociación es más clara en los grupos de tratamiento, confirmando igualmente que el tratamiento refuerza la asociación entre pensamientos de preocupación y evaluación de la conducta social. Por otro lado, en otros análisis hemos constatado que los predictores de cambio en las conductas de evitación son el menor

miedo ante las conductas- objetivo en el pre-test. Las variables predictoras de cambio en el nivel de miedo a la evaluación negativa son las variables relacionadas con la afectividad negativa (a mayores niveles de depresión y ansiedad, y menores de autoestima en el pre-test, menor cambio cognitivo). Hemos interpretado esto en el sentido que los procesos cognitivos se pueden asociar más al rasgo de alta afectividad negativa, mientras que los conductuales dependen más del nivel inicial de ansiedad de estado (véase Salaberría et al., 1995).

Conclusiones

Los sujetos tratados —en este caso, fóbicos sociales— con una terapia eficaz disminuyen en el nivel de evitación y de cogniciones disfuncionales. Este hecho se constata adecuadamente si se utiliza un estimador que coincida con un punto de corte funcional y que sea generalizable, corrigiendo sobre la puntuación de cambio la medición de error, tomando en cuenta la puntuación inicial y dando un indicador individual de cambio (Páez et al., 1993).

Las medias posttest del grupo de tratamiento se sitúan por debajo del percentil 10 de la muestra disfuncional unificada, tanto en el caso del SAD como en el del FNE. Esto quiere decir que los sujetos del grupo de tratamiento están mejor que el 90% de la muestra disfuncional de referencia (sus propias puntuaciones y las del grupo control en el pretest). El cambio es de 3 deciles en el FNE (de media 40 a 10) y de 4 deciles en el SAD (de media 50 a 10). Es decir, el cambio es similar al efecto medio de Shapiro y Shapiro (1982), aunque más marcado para las conductas motoras (SAD) que para las cognitivas (FNE).

Si se utiliza un punto de corte basado estadísticamente en la población disfuncional (C3), la media de los sujetos tratados está por debajo del punto de corte (SAD \bar{X} = 13.25; FNE \bar{X} = 16.52) a los 6 meses, pero no así la media de los sujetos del grupo de control, confirmando que el cambio es generalizable a partir de un baremo global.

Hay más sujetos recuperados individualmente en los grupos de tratamiento que en el grupo de control. Además, los sujetos recuperados según el coeficiente individual de cambio preciso (PCP) tanto del aspecto conductual (SAD) como cognitivo (FNE) superan los puntos de corte basados estadísticamente. Por ejemplo, el 90% de los sujetos recuperados (R) según el coeficiente PCP del FNE superan el punto de corte del coeficiente 3. Se confirma de este modo la validez convergente de ambos conceptos.

Por último, existe una correlación válida significativamente para los grupos de tratamiento entre el nivel de cambio cognitivo y el conductual. De este modo, se confirma la validez individual del modelo cognitivo-conductual de la fobia social (Echeburúa, 1995; Echeburúa y Salaberría, 1994). Sin embargo, aunque los cambios en los pensamientos y en las conductas están asociados, no se puede afirmar el carácter causal de las cogniciones disfuncionales sobre las conductas de evitación.

ANEXO

El objetivo de este anexo es simplemente presentar de una forma esquemática algunos de los datos de la investigación comentada en el texto que se exponen con detalle en otro lugar (Salaberría y Echeburúa, 1995) y que pueden facilitar al lector la comprensión del marco teórico y clínico de la misma.

Tema principal de investigación

Efecto de la reestructuración cognitiva en el tratamiento de la fobia social de tipo generalizado.

Número de pacientes

71 pacientes seleccionados según los criterios diagnósticos del DSM-III-R por medio de la entrevista clínica estructurada ADIS-R (Di Nardo y Barlow, 1988) y según una puntuación de 15 o más en la Escala de Estrés y Evitación Social (SAD) (Watson y Friend, 1969) o de 21 o más en la Escala de Miedo a la Evaluación Negativa (FNE) (Watson y Friend, 1969). Los pacientes fueron 37 hombres y 34 mujeres, con una media de edad de 31 años.

Diseño experimental

Diseño experimental multigrupo con medidas independientes en el factor de tratamiento y medidas múltiples y repetidas en el factor de evaluación. La asignación a los grupos se realizó al azar tras una estratificación previa en sexo y edad.

Modalidades terapéuticas

Exposición (N=24)

Exposición más reestructuración cognitiva (N=24)

Grupo de control de lista de espera de 6 meses (N=23)

Evaluación

Instrumentos principales de evaluación:

- Escala de Estrés y Evitación social (SAD) (Watson y Friend, 1969)
- Escala de Miedo a la Evaluación Negativa (FNE) (Watson y Friend, 1969)

Momentos de evaluación:

- Evaluación pretratamiento
- Evaluación postratamiento
- Evaluación en los seguimientos de 1.3 y 6 meses.

Tratamiento

Exposición a situaciones evitadas por medio de ensayos conductuales y tareas para casa. 8 sesiones grupales (de 4 a 8 pacientes), con dos terapeutas, con una duración de dos horas y media cada sesión y una periodicidad semanal.

Exposición más reestructuración cognitiva: identificación y discusión de pensamientos no adaptativos junto a exposición a situaciones evitadas por medio de ensayos conductuales y cognitivos y tareas para casa. 8 sesiones grupales (de 4 a 8 pacientes), con dos terapeutas, con una duración de dos horas y media cada sesión y una periodicidad semanal.

Referencias

- Beck, A.T.-Ward, C.H.-Mendelson, M.-Mock, J.-Erbaugh, J. (1961): An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561-571.
- Clemente, M. (1989): Metodología de investigación de los problemas psicosociales: la investigación sobre la evaluación de las intervenciones psicosociales. *Revista de Psicología Social*, 4, 85-109.
- Cohen, J.-Cohen, P. (1983): *Applied multiple regression/ correlation analysis for the behavioral sciences*. Hillsdale, N.Jersey. Lawrence Erlbaum.
- Cook, T.D.-Shadish, W.R. (1994): Social experiments: Some developments of the past fifteen years. *Annual Review of Psychology*, 45, 545-580.
- Crano, W.-Brewer, M. (1977): *Fundamentos de la investigación en psicología social*. México. El Manual Moderno (edición original, 1973).
- Cronbach, L.J.-Furby, L. (1970): How should we measure Change -or should we?. *Psychological Bulletin*, 74, 68-80.
- Dillard, J.P. (1991) the current status status of research on sequential request compliance technique. *Personality and Social psychology Bulletin*, 3, 283-288.
- Echeburúa, E. (1995): *Evaluación y tratamiento de la fobia social*. Barcelona. Martínez Roca.

- Echeburúa, E.-Salaberría, K. (1994): Técnicas de exposición y de reestructuración cognitiva en el tratamiento de la fobia social. En M. Garaigordóbil y C. Maganto (Eds.). *Socialización y conducta prosocial en la infancia y la adolescencia*. San Sebastián. Universidad del País Vasco.
- Francis, D.-Fletcher, J.-Stuebing, K.-Davidson, K.-Thompson, N. (1991): Analysis of change: Modeling Individual Growth. *Journal of Consulting and Clinical Psychology*, 59, 27-37.
- Gerin, P. (1984): *L'évaluation des psychothérapies*. París. Presses Universitaires Françaises.
- Gerin, P.-Dazord, A.-Sali, A. (1991): *Psychothérapies et changements*. París. Presses Universitaires Françaises.
- Hersen, M.-Michelson, L. (1986): *Issues in Psychotherapy Research*. New York. Plenum Press.
- Jacobson, N.S.-Truax, P. (1991): Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12-19.
- Matalon, B. (1988): *Décrire, Expliquer, Prévoir*. París. Armand Colin.
- Páez, D.-Echeburúa, E.-Borda, M. (1993): Evaluación de la eficacia de los tratamientos psicológicos: una propuesta metodológica. *Revista de Psicología General y Aplicada*, 46, 187-198.
- Perachio, L.-Cook, T. (1988): Avances en el diseño cuasi experimental. En I. Dendaluce (Ed.). *Aspectos metodológicos de la investigación educativa*. Madrid. Narcea (II Congreso Mundial Vasco).
- Reichardt, C.S. (1979): The statistical analysis of data from non equivalent group designs. En T.D. Cook y D.T. Campbell (Eds.) *Quasi-experimentation*. Chicago. Rand McNally.
- Rosenberg, M. (1965): *Society and the adolescent self-image*. New Jersey. Princeton.
- Salaberría, K.-Echeburúa, E. (1995): Tratamiento psicológico de la fobia social: un estudio experimental. *Análisis y Modificación de Conducta*, 21, 151-179.
- Shapiro, D.A.-Shapiro, D. (1982) Comparative therapy outcome research: methodological implications of meta-analysis. *Journal of Consulting and Clinical Psychology*, 51, 42-53.
- Spielberger, C.D.-Gorsuch, R.L.-y Lushene, R.E. (1970): *The state-trait anxiety inventory*. Palo Alto, California. Consulting Psychologist Press.
- Watson, D.-Friend, R. (1969): Measurement of social-evaluative anxiety. *Journal of Consulting and Clinical Psychology*, 33, 448-457.
- Willet, J.B.-Ayoub, C.C.-Robinson, D. (1991): Using growth modeling to examine systematic differences in growth. *Journal of Consulting and Clinical Psychology*, 59, 38-47.