Arce, R., Seijo, D., y Rodríguez, M. J. (2005). Falsas alarmas de los criterios de realidad en delitos de agresiones sexuales e intimidación en muestras de adultos. En J. Sobral, G. Serrano, y J. Regueiro (Comps.), *Psicología jurídica, de la violencia y de género* (pp. 31-43). Madrid: Biblioteca Nueva. **ISBN: 84-9742-452-2.**

Falsas alarmas de los criterios de realidad en delitos de agresiones sexuales e intimidación en muestras de adultos.

Autores: Ramón Arce, Dolores Seijo y María Jesús Rodríguez

Resumen

Los criterios de realidad recogidos en el CBCA y el Reality Monitoring se han mostrado efectivos para la discriminación entre declaraciones verdaderas o reales y falsas o fabricadas. En principio se aplicó a las declaraciones de menores víctimas de agresiones sexuales, pero la literatura encontró que también eran efectivos en las declaraciones de adultos y en otros tipos de delitos. El sistema de evaluación se basa en el análisis del efecto acumulativo de los criterios de realidad en la declaración lo que lo convierte en un procedimiento semi-objetivo. En este contexto nos plateamos un estudio con simuladores adultos de una agresión sexual o amenazas intimidatorias muy graves y analizar el número de falsas alarmas en cada criterio así como el efecto acumulativo de éstas. Se analizan los resultados de una muestra de 105 participantes que simularon haber sufrido una agresión sexual, el 51.4%, o amenazas muy graves a la integridad física, el 48.6%. Finalmente se discuten los resultados así como las implicaciones para la práctica forense.

Palabras clave: testimonio, credibilidad, CBCA, Reality monitoring, pericial psicológica.

Arce, R., Seijo, D., y Rodríguez, M. J. (2005). Falsas alarmas de los criterios de realidad en delitos de agresiones sexuales e intimidación en muestras de adultos. En J. Sobral, G. Serrano, y J. Regueiro (Comps.), *Psicología jurídica, de la violencia y de género* (pp. 31-43). Madrid: Biblioteca Nueva. **ISBN: 84-9742-452-2.**

Introducción

El sistema judicial asume una casi infalibilidad en sus decisiones. Es cierto que permite revisiones de las decisiones tomadas, pero también lo es el principio de interpretación única de las pruebas. Éstas son mayoritariamente testimonios en los que la estimación de su valor probatorio es la clave del mismo. Entre éstos, Jones y McGraw (1987) han cuantificado en un 8% el total de testimonios falsos de víctimas, reduciéndose tal porcentaje a un 2% en casos de menores. Si a este porcentaje de falsedad unimos las discrepancias que observamos en las vistas entre testigos, nos encontramos con que una de las claves de la justicia es discernir entre la verdad y la mentira o, lo que es más sorprendente, el testimonio honesto que no es verídico (Alonso-Quecuty, 1993).

Los esfuerzos científicos más salientes al respecto, todos ellos de orden psicológico, se han orientado casi exclusivamente al testimonio de menores víctimas y de un delito específico, la agresión sexual. Éstos han redundando en dos grandes modelos de análisis de las declaraciones, el *CBCA* (Criteria Based Content Analysis) (Steller y Köhnken, 1989/1994) y el *Reality Monitoring* (originario de Johnson y Raye, 1981, para diferenciar memorias de origen interno y externo, y aplicado inicialmente al testimonio por Alonso-Quecuty, 1990). Éstos se ciñen únicamente al análisis de la fiabilidad del contenido de las declaraciones bien mediante el análisis de trazos de memoria de lo imaginado y percibido, en el caso del *Reality Monitoring*, bien como consecuencia de las observaciones de las diferencias entre declaraciones verdaderas y falsas por parte de determinados experimentadores, como es el caso del *CBCA*. Dado que ambos procedimientos se orientan al estudio de la consistencia interna de las declaraciones, la fiabilidad, se pueden complementar con el análisis de la validez de las declaraciones. En éste, se da entrada a la evaluación de la consistencia con otros parámetros externos como las leyes de la naturaleza, otras pruebas, temporal o la validez del protocolo de obtención de la declaración. El protocolo más conocido para el análisis de la validez es el SVA (Statement Validity Analysis) (p.e., Steller, 1989).

Arce, R., Seijo, D., y Rodríguez, M. J. (2005). Falsas alarmas de los criterios de realidad en delitos de agresiones sexuales e intimidación en muestras de adultos. En J. Sobral, G. Serrano, y J. Regueiro (Comps.), *Psicología jurídica, de la violencia y de género* (pp. 31-43). Madrid: Biblioteca Nueva. **ISBN: 84-9742-452-2.**

Johnson y Raye (1981) propusieron un marco de referencia para comprender cómo pueden discriminarse los sucesos percibidos o externos de los imaginados o internos. Para estos autores, las memorias varían en una serie de rasgos, de tal manera que aquellas cuyos orígenes se encuentran en sucesos percibidos contienen más información sensorial, mayor número de detalles contextuales y menos referencias a procesos cognitivos que las memorias con base interna o imaginada. Al proceso de discriminar entre recuerdos de origen interno y recuerdos de origen externo lo denominaron *Reality Monitoring* (Control de la Realidad). En el siguiente esquema se resume el modelo de Control de la Realidad propuesto por Johnson y Raye (1981):

a) Tipos de atributos que pueden formar parte de los recuerdos:
   - Contextuales.
   - Sensoriales.
   - Operaciones cognitivas.
b) Dimensiones que generalmente diferencian los recuerdos según su origen:
   - Origen externo: más atributos contextuales (espacio-temporales) y sensoriales (v. gr., sonidos, olores).
   - Origen interno: más información sobre operaciones cognitivas, esto es, información idiosincrásica (por ejemplo, yo pensé, recuerdo ver, me sentía nervioso).

A partir de este desarrollo se han formulado sistemas categoriales para discriminar los testimonios verdaderos de los falsos (p.e., Alonso Quecuty, 1990; Porter y Yuille, 1996; Sporer y Küpper, 1995; Vrij, Edward, Roberts y Bull, 2000). En todo caso, el más manido y que responde a un mínimo común denominador es el propuesto por Alonso-Quecuty que engloba las categorías información sensorial, información contextual, información semántica, operaciones cognitivas y longitud (aunque ésta no es propia del modelo original pero también ha sido asumida por otros autores). Los estudios de contraste sobre capacidad predictiva del *Reality Monitoring* lo sitúan entre, aproximadamente, el 60 y el 80% a la vez que ponen de manifiesto que tiene un valor predictivo ligeramente inferior al CBCA (véase Vrij, 2000 para una revisión). En consecuencia, el margen de error promedio en la clasificación rondaría el 30%. Además,

el estado de la cuestión presenta otras limitaciones a tener en cuenta: dichas propuestas no son concordantes ni en los criterios a analizar ni en la definición de los mismos y, para la práctica forense, adolece de criterios categóricos (las evaluaciones se realizan bien en recuentos de frecuencias, bien en escalas de repuesta múltiple, bien en ecuaciones de regresión o funciones discriminantes). En suma, los estudios sobre el *Reality Monitoring* aplicados al análisis de la verdad y mentira en el contexto judicial se han adaptado a protocolos de medida muy robustos científicamente, pero que no son adaptables a la práctica forense. De hecho, no se han formulado cuántos atributos de una determinada categoría serían, en su caso, precisos para establecerla. Por ello, en el contexto forense se ha de tomar como criterio la presencia vs. ausencia.

Steller y Köhnken (1989/1994) propusieron un conjunto de criterios (véanse en el apartado de instrumentos de medida) para la detección de las declaraciones verdaderas (la detección de una categoría ha de interpretarse como un indicio de veracidad, pero de la ausencia no se puede extraer que sea falsa) de niños víctimas de abusos sexuales. Éstos venían a ser un compendio más sistemático y actualizado de anteriores propuestas de Undeutsch, Arntzen y Szewczyk. No obstante, no deja de ser un sistema semi-objetivo. Los criterios de análisis se pueden evaluar como presentes o ausentes, o en función del grado o intensidad de los mismos. La evaluación categórica se ajusta perfectamente al contexto forense. Además, se ha observado (Steller, 1989) que en las declaraciones verdaderas aparecen, al menos, 7 criterios de veracidad. Una revisión de Vrij y Akehurst (1998) sobre la puesta a prueba de la validez predictiva de este sistema ha informado de un valor predictivo en los estudios de simulación y laboratorio que oscilaba entre el 65% y el 85%, a la vez que una mayor efectividad en la detección de declaraciones verdaderas que falsas. No obstante, en la misma revisión Vrij y Akehurst hallaron un valor de clasificación correcto del 100% de los casos en estudios de campo.

En este contexto nos hemos planteado un estudio con el objetivo general de conocer el potencial de falsas alarmas informadas por evaluadores entrenados en la detección de las categorías de los dos grandes protocolos de análisis de la credibilidad,

Arce, R., Seijo, D., y Rodríguez, M. J. (2005). Falsas alarmas de los criterios de realidad en delitos de agresiones sexuales e intimidación en muestras de adultos. En J. Sobral, G. Serrano, y J. Regueiro (Comps.), *Psicología jurídica, de la violencia y de género* (pp. 31-43). Madrid: Biblioteca Nueva. **ISBN: 84-9742-452-2.**

el *CBCA* y el *Reality Monitoring* en declaraciones imaginadas prestadas por adultos a la vez que la consistencia de las predicciones de los modelos inter-contextos.

Método

Muestra.

Para este estudio contamos con la colaboración voluntaria de 105 sujetos, en su mayoría estudiantes universitarios, sin conocimientos técnicos de Psicología ni de aplicación forense. El 48.6% de ellos eran hombres mientras que las mujeres constituían el 51.4% de la muestra. El rango de edad de los sujetos abarcaba de los 20 a los 27 años, con un promedio de 24.32 años.

Procedimiento y diseño.

Los sujetos que participaron en este experimento debían relatar ante una cámara de vídeo que habían sido víctimas de un delito, narrando los hechos supuestamente ocurridos. Para ello dos entrevistadores, previamente entrenados, procedieron con los 5 primeros pasos de la entrevista cognitiva mejorada (Fisher y Geiselman, 1992): presentación y personalización de la entrevista; establecimiento de la comunicación; explicación del objeto de la entrevista, reinstauración de contexto y recuerdo libre.

Una semana antes de que realizasen esta declaración se les administraba de forma oral las siguientes instrucciones:

“Tienes que imaginarte que vas a declarar en un juicio en el que acusas a alguien sin ser cierto:

a) de haberte violado,

   o

b) de haberte amenazado, pero amenazas muy graves (amenazas intimidatorias), que han supuesto una seria amenaza para tu integridad física.

Arce, R., Seijo, D., y Rodríguez, M. J. (2005). Falsas alarmas de los criterios de realidad en delitos de agresiones sexuales e intimidación en muestras de adultos. En J. Sobral, G. Serrano, y J. Regueiro (Comps.), *Psicología jurídica, de la violencia y de género* (pp. 31-43). Madrid: Biblioteca Nueva. **ISBN: 84-9742-452-2.**

En realidad no es cierto, se trata de una acusación falsa con la que pretendes vengarte de esa/s persona/s".

A fin de amortiguar una posible falta de implicación se les advirtió de la importancia de la misma para el buen fin del experimento a la vez que creamos un motivador externo de tipo económico mediante un premio de 150 euros a las dos mejores simulaciones.

Del total de sujetos, aproximadamente la mitad (51.4%) simularon que habían sido víctimas de agresión sexual y el resto (48.6%) fingieron haber sufrido amenazas intimidatorias muy graves. Cada uno de estos grupos estaba formado por un número aproximadamente igual de varones y mujeres. Pretendíamos contrabalancear, así, un posible efecto del género.

Análisis de los protocolos.

Dos codificadores expertos y entrenados en el análisis de contenido, pero no en análisis de credibilidad, evaluaron los protocolos independientemente en términos categoriales (v.gr., presencia vs. ausencia) en cada criterio del *CBCA* y *Reality Monitoring*. Las categorías de análisis fueron:

1) Del *Reality-Monitoring* de Johnson y Raye (1981), se tomaron los atributos de análisis utilizados por Alonso-Quecuty (1995 para una revisión) a los que añadió la longitud de la declaración así como su operativización (téngase presente que no todos los autores toman los mismos criterios ni los definen igual) complementada para los atributos longitud e información semántica para una codificación como una variable categórica con el criterio propuesto por Undeutsch (1967) “si exceden o no las capacidades cognitivas del testigo”:

- Información contextual (p.e., espacial, temporal).
- Información sensorial (i.e., colores, ruidos,....).

- Información semántica, es decir, información específica, con significado (propia de una memoria de lo imaginado vs. propia de una memoria de lo percibido). La memoria de lo imaginado contiene menos información semántica que la de lo percibido.
- Información idiosincrásica (ej., yo estaba asustado, pienso que no debió ocurrir,...).
- Longitud del relato (propia de una memoria de lo imaginado vs. propia de una memoria de lo percibido). La memoria de lo imaginado es más corta que la de lo percibido.

2) Del C.B.C.A. (Análisis de Contenido Basado en Criterios) los criterios propuestos por Steller y Köhnken (1989/1994) que los codificadores evaluaban en una variable categórica (v.gr., presencia vs. ausencia). Como quiera que el criterio "cantidad de detalles" no se puede codificar directamente como una variable categórica, se les proporcionaba un criterio adicional "si exceden o no las capacidades cognitivas del testigo":
   - Estructura lógica (coherencia y consistencia interna).
   - Elaboración inestructurada (presentación desorganizada).
   - Cantidad de detalles (abundancia de detalles o hechos distintos). La memoria de lo imaginado contiene menos información semántica que la de lo percibido.
   - Engranaje contextual (ubicación de la narración en un espacio y tiempo).
   - Descripción de interacciones (cadena de acciones entre el testigo y otros actores).
   - Reproducción de conversación (réplica de conversaciones).
   - Complicaciones inesperadas durante el incidente (por ejemplo, interrupción imprevista).
   - Detalles inusuales (detalles con baja probabilidad de ocurrencia).
   - Detalles superfluos (detalles irrelevantes que no contribuyen significativamente a los hechos).
   - Incomprensión de detalles relatados con precisión (explicitación de detalles que el menor no comprende pero que realmente sí tienen sentido).

- Asociaciones externas relacionadas (inclusión de información externa a los hechos en sí pero relacionada con ellos, tales como en una agresión sexual recordar conversaciones anteriores sobre este tema).
- Relatos del estado mental subjetivo (referencias a sentimientos, emociones o cogniciones propias).
- Atribución del estado mental del autor del delito (referencias al estado mental del agresor y atribución de motivos).
- Correcciones espontáneas (correcciones espontáneas o mejoras de la declaración).
- Admisión de falta de memoria (reconocimiento de lagunas de memoria).
- Plantear dudas sobre el propio testimonio.
- Auto-desaprobación (actitud crítica sobre su propia conducta).
- Perdón al autor del delito (la declaración de la víctima favorece al acusado, o evitación de más acusaciones).
- Detalles característicos de la ofensa (descripciones que contradicen las creencias habituales sobre el delito).

Entrenamiento de Codificadores.

Los dos codificadores, que no coincidían con los entrevistadores, habían sido previa y exhaustivamente entrenados en este tipo de codificación y contaban con experiencia en la codificación con sistemas categoriales metódicos con fines científicos (p.e., Fariña, Arce y Novo, 2002;). El entrenamiento se llevó a cabo para la detección de las categorías, no para que discriminaran en función de éstas entre declaraciones verdaderas o falsas. Se pretendía, con este procedimiento, controlar un posible efecto confirmatorio o de negación sistemático por parte del codificador. Aún es más, no se les informaba que éste fuera el cometido final de estos análisis de contenido. En el entrenamiento de estos codificadores participaron otros dos experimentadores con experiencia en codificación con los criterios del *CBCA* y los atributos del *Reality Monitoring* tanto con fines científicos (Arce, Fariña y Freire, 2002) como forenses. En

Arce, R., Seijo, D., y Rodríguez, M. J. (2005). Falsas alarmas de los criterios de realidad en delitos de agresiones sexuales e intimidación en muestras de adultos. En J. Sobral, G. Serrano, y J. Regueiro (Comps.), *Psicología jurídica, de la violencia y de género* (pp. 31-43). Madrid: Biblioteca Nueva. **ISBN: 84-9742-452-2.**

resumen, los codificadores fueron entrenados para la búsqueda de las categorías de análisis, pero ciegos en la discriminación entre verdad y mentira.

Fiabilidad.

Los dos entrevistadores del estudio, entrenados y con experiencia en la entrevista cognitiva, obtuvieron unos protocolos similares, $F_{multivariada}(2,102)=.59$; ns, en términos de la actividad cognitiva general (palabras, proposiciones concretas y proposiciones abstractas, esto es, relacionadas o no con el objeto de la declaración) desplegada por los participantes en la declaración. En suma, las declaraciones obtenidas por los entrevistadores no están contaminadas por el factor entrevistador.

La consistencia en las codificaciones fue evaluada a través del índice de concordancia por ser las variables categóricas. Una concordancia mayor de .80 se considera consistente (Tversky, 1977). Los resultados informaron de consistencia en las codificaciones tanto intra- como inter-codificadores (IC>.80).

En resumen, considerando las consistencias inter- e intra-evaluadores, inter-métodos y de los codificadores en otros estudios y con otras muestras, podemos sostener que los datos son fiables (Wicker, 1975).

Resultados

Referidos a los atributos de realidad del Reality Monitoring.

El atributo del *Reality Monitoring* información contextual (véase la Tabla 1) presenta, en relación con las predicciones seminales del modelo, una probabilidad de falsas alarmas mayor que el simple azar (<.50), esto es, su valor práctico, en el presente estudio, en la detección de una memoria de lo imaginado es nulo. De los cuatro criterios restantes sólo el margen de error de la longitud de la declaración, $Z(n=105)=0.79$; ns, entra dentro de los límites de error estadístico y científico (≤.05), ya que tanto la

Arce, R., Seijo, D., y Rodríguez, M. J. (2005). Falsas alarmas de los criterios de realidad en delitos de agresiones sexuales e intimidación en muestras de adultos. En J. Sobral, G. Serrano, y J. Regueiro (Comps.), *Psicología jurídica, de la violencia y de género* (pp. 31-43). Madrid: Biblioteca Nueva. **ISBN: 84-9742-452-2.**

probabilidad observada en la categoría información sensorial, *Z*(n=105)=11.97; p<.001, como en la idiosincrásica, *Z*(n=105)=8.4; p<.001, como en la categoría información semántica, *Z*(n=105)=8.26; p<.001, supera dicho punto de contraste.

Tabla 1. Análisis de las falsas alarmas informadas por los atributos del *Reality Monitoring*.

Variable	p_{fa}	X^2	p
Información sensorial	.305	16.01	.000
Información contextual	.971	93.34	.000
Información idiosincrásica	.229	30.94	.000
Información semántica	.238	28.11	.000
Longitud	.067	.78.87	.000

Nota: p_{fa}= Probabilidad de falsas alarmas; G.L.(1). Los contrastes se llevaron a cabo con las predicciones generales del modelo por no haber acuerdo inter-investigadores.

El procedimiento utilizado para validar o invalidar una declaración mediante el *Reality Monitoring* se deriva de la constatación del mayor número de categorías de tal modo que si los contenidos de las declaración se ajustan a las predicciones en las 5 categorías, la probabilidad de que la declaración sea verdadera es muy alta y viceversa. Por ende, la detección de estas categorías en declaraciones falsas supone un error del modelo y la acumulación de errores un índice de ineficacia predictiva. Los resultados al respecto advierten que sólo en el 1.9% de las declaraciones no tiene cabida ningún atributo, que en el 41% de las declaraciones se observan indicios de 1 atributo, de 2 en el 36.2%, de 3 en el 17.1%, de 4 en el 2.9% y hasta de los 5 en el 1%. Si tomamos como criterio de "no validez como prueba" la falta de consistencia con el modelo en, al menos, 2 de los 5 criterios, la probabilidad de falsas alarmas (esto es, identificar como verdadera una declaración falsa) sería de .571, o sea, superior al azar.

Otro método complementario de poner a prueba la robustez del modelo es comprobar su validez inter-contextos (Wicker, 1975). Por ello contrastamos el modelo en dos contextos judiciales distintos: un delito de amenazas frente a un deleito de agresión sexual. Los resultados, que pueden verse en la Tabla 2, ponen de manifiesto

Arce, R., Seijo, D., y Rodríguez, M. J. (2005). Falsas alarmas de los criterios de realidad en delitos de agresiones sexuales e intimidación en muestras de adultos. En J. Sobral, G. Serrano, y J. Regueiro (Comps.), *Psicología jurídica, de la violencia y de género* (pp. 31-43). Madrid: Biblioteca Nueva. **ISBN: 84-9742-452-2.**

que los atributos del *Reality Monitoring* son insensibles al efecto de contexto lo que valida la consistencia inter-contextos del modelo.

Tabla 2. Contraste de la eficacia de los atributos del *Reality Monitoring* inter-contextos.

Variable	p_{faa}	p_{faas}	X^2	p	Phi
Información sensorial	.22	.39	2.94	.086	.188
Información contextual	.94	1	1.49	.222	.176
Información idiosincrásica	.22	.24	.005	.942	-.030
Información semántica	.23	.24	.000	1	.006
Longitud	.08	.06	.006	.938	-.046

Nota: p_{faa}= Probabilidad de falsa alarma en amenzas; p_{faas}= Probabilidad de falsa alarma en agresión sexual; G.L.(1).

Referidos a los criterios de realidad del CBCA.

Con relación al análisis de contenido mediante los criterios del *CBCA* el primer dato a reseñar es que las categorías “incomprensión de detalles relatados con precisión" y "detalles característicos de la ofensa" no fueron productivas. Una primera lectura de estos resultados nos diría que su funcionamiento fue perfecto en relación con las predicciones del modelo pero la literatura precedente ha advertido que estas dos categorías sólo son propias de las memorias de menores con lo que no pueden generalizarse a adultos (Landry y Brigham, 1992).

Tabla 3. Análisis de las falsas alarmas informadas por los criterios del *CBCA*.

Variable	p_{fa}	X^2	p
Estructura lógica	.952	85.95	.000
Elaboración inestructurada	.135	55.54	.000
Cantidad de detalles	.067	78.87	.000
Engranaje contextual	.495	.01	.922
Descripción de interacciones	.705	17.61	.000
Complicaciones inesperadas	.067	77.89	.000
Asociaciones externas relacionadas	.143	53.57	.000

Arce, R., Seijo, D., y Rodríguez, M. J. (2005). Falsas alarmas de los criterios de realidad en delitos de agresiones sexuales e intimidación en muestras de adultos. En J. Sobral, G. Serrano, y J. Regueiro (Comps.), *Psicología jurídica, de la violencia y de género* (pp. 31-43). Madrid: Biblioteca Nueva. **ISBN: 84-9742-452-2.**

Estado mental subjetivo	.619	5.95	.015
Estado mental autor delito	.333	11.67	.001
Reproducción de conversaciones	.429	2.14	.143
Detalles inusuales	.001	101.03	.000
Superfluos	.762	28.81	.000
Correcciones espontáneas	.238	28.81	.000
Admisión de falta de memoria	.276	21.04	.000
Dudas sobre el testimonio	.001	101.04	.000
Autodesaprobación	.095	68.81	.000
Perdón al autor del delito	.048	85.95	.000

Nota: p_{fa}= Probabilidad de falsa alarma; G.L.(1).

Sobre las falsas alarmas en relación con las predicciones del modelo (ver Tabla 3) observamos que los criterios "estructura lógica", "descripción de interacciones", "relatos del estado mental subjetivo" y "detalles superfluos" son peores indicadores que el azar, esto es, aportan más falsas alarmas que aciertos. Aún es más, en la categoría "estructura lógica", la probabilidad observada supera .95 (o sea, el margen de no falsas alarmas es inferior al margen de error estadístico .05) con lo que entra dentro de los márgenes científicos que permiten sostener que este criterio sólo aporta falsas alarmas al sistema. A nuestro entender, esta categoría de análisis no es propiamente de fiabilidad sino de validez; de ahí, ese margen de falsas alarmas. En otras palabras, la carencia de lógica en una declaración, a menos que se trate de la emitida por una persona con una discapacidad o un menor con un desarrollo cognitivo muy limitado (véase Arce, Novo y Alfaro, 2000 para una discusión), no atenta a la consistencia interna de la misma sino que la invalida Por tanto, debería pasarse del *CBCA* al *SVA*. Por su parte, los criterios "engranaje contextual" y "reproducción de conversaciones" presentan un poder discriminativo igual al azar. En lo referente a los criterios "elaboración inestructurada", "asociaciones externas relacionadas", "atribución al estado mental del autor del delito" "correcciones espontáneas", "admisión de falta de memoria" y "autodesaprobación" encontramos una ejecución mejor que el azar. Finalmente, los criterios "cantidad de detalles" [Z(n=105)=0.79; ns], "complicaciones inesperadas" [Z(n=105)=0.79; ns], "detalles inusuales" [Z(n=105)=1.87; p<.05], "dudas sobre el propio testimonio"

Arce, R., Seijo, D., y Rodríguez, M. J. (2005). Falsas alarmas de los criterios de realidad en delitos de agresiones sexuales e intimidación en muestras de adultos. En J. Sobral, G. Serrano, y J. Regueiro (Comps.), *Psicología jurídica, de la violencia y de género* (pp. 31-43). Madrid: Biblioteca Nueva. **ISBN: 84-9742-452-2.**

[Z(n=105)=1.87; p<.05], y "perdón al autor del delito" [Z(n=105)=0.09; ns], admiten una probabilidad de falsas alarmas igual o menor que el margen de error asumible científicamente (≤.05). Si, de acuerdo con las observaciones llevadas a cabo por Steller (1989) en casos reales confirmados, tomamos como punto de corte para validar una declaración como verdadera la presencia de 7 o más criterios, la probabilidad de errores al dictaminar como verdaderos testimonios falsos sería de .155.

Si, de acuerdo con las observaciones llevadas a cabo por Steller (1989) en casos reales confirmados, tomamos como punto de corte para validar una declaración como verdadera la presencia de 7 o más criterios, la probabilidad de errores al dictaminar como verdaderos testimonios falsos sería de .155.

Tabla 4. Contraste de la eficacia de los criterios del *CBCA* inter-contextos

Variable	p_{faa}	p_{faas}	X^2	p	Phi
Estructura lógica	.94	.963	.004	.948	.051
Elaboración inestructurada	.216	.057	4.36	.037	-.233
Cantidad de detalles	08	.06	.006	.938	-.046
Engranaje contextual	.451	.537	.471	.493	.086
Descripción de interacciones	.725	.685	.057	.812	-.044
Complicaciones inesperadas	.12	.019	2.796	.095	-.202
Asociaciones externas relacionadas	.196	.093	1.527	.217	-.148
Estado mental subjetivo	.647	.593	.139	.709	-.056
Estado mental autor delito	.451	.222	5.19	.023	-.243
Reproducción de conversaciones	.471	.389	.420	.517	-.083
Detalles inusuales	0	.019	.000	1	.095
Superfluos	.725	.796	.387	.534	.083
Correcciones espontáneas	.216	.259	.087	.768	.051
Admisión de falta de memoria	.137	.407	8.272	.004	.302
Dudas sobre el testimonio	0	.019	.000	1.0	.095
Autodesaprobación	.059	.13	.815	.367	.121

Arce, R., Seijo, D., y Rodríguez, M. J. (2005). Falsas alarmas de los criterios de realidad en delitos de agresiones sexuales e intimidación en muestras de adultos. En J. Sobral, G. Serrano, y J. Regueiro (Comps.), *Psicología jurídica, de la violencia y de género* (pp. 31-43). Madrid: Biblioteca Nueva. **ISBN: 84-9742-452-2.**

Perdón al autor del delito	.059	.037	.004	.098	-.051

Nota: p_{faa}= Probabilidad de falsa alarma en amenzas; p_{faas}= Probabilidad de falsa alarma en agresión sexual; G.L.(1).

Además, como puede verse en la Tabla 4, los criterios no son consistentes inter-contextos. De hecho, los criterios “elaboración inestructurada”, “atribución del estado mental del autor del delito” y “admisión de falta de memoria” presentan una probabilidad de falsas alarmas significativamente distinta en casos de amenazas y de agresiones sexuales a las que hay que sumar los dos criterios no productivos, en estos casos con estas muestras: “incomprensión de detalles relatados con precisión" y "detalles característicos de la ofensa”. En suma, los criterios no gozan de estabilidad inter-contextos lo que implica una debilidad de los mismos.

Discusión

Previamente a entrar en la discusión de los resultados presentados es preciso revisar las limitaciones de este estudio que pueden condicionar las mismas. Primera, se trata de un estudio de simulación en el que es más difícil discriminar la verdad y la mentira que en la realidad (Vrij, 2000). Segunda, las predicciones de los modelos no se plantean para adultos ni para el delito de amenazas aunque, para que un sistema de análisis categorial sea fiable, ha de ser consistentes inter-muestras e inter-contextos (Wicker, 1975). Tercera, las demandas que se realizan a los evaluadores, de acuerdo con los dos modelos tratados, responden no a criterios aislados sino a una impresión de conjunto que se transforma en una respuesta probabilística. Ahora bien, las Salas de Justicia nos demandan una respuesta categórica. Cuarta, la variable de medida no es la misma que en la mayoría de los estudios. En nuestro caso es la presencia o ausencia del criterio o atributo en tanto en muchos es el grado de presencia. Quinta, los resultados son difícilmente comparables a otros estudios porque el procedimiento de validación seguido es totalmente inverso y los criterios del *Reality Monitoring* no son equiparables inter-estudios. Con estas limitaciones en mente, se puede concluir que:

Arce, R., Seijo, D., y Rodríguez, M. J. (2005). Falsas alarmas de los criterios de realidad en delitos de agresiones sexuales e intimidación en muestras de adultos. En J. Sobral, G. Serrano, y J. Regueiro (Comps.), *Psicología jurídica, de la violencia y de género* (pp. 31-43). Madrid: Biblioteca Nueva. **ISBN: 84-9742-452-2.**

a. El *Reality Monitoring* no cuenta con un único sistema de análisis categorial; los atributos han sido definidos de distintas formas; las predicciones del modelo no son homogéneas inter-investigadores; no cuenta con criterios fijados para la evaluación ni se ha ajustado la medida de los mismos a las demandas forense en variables categóricas. Sobre esta última, los estudios se han llevado a cabo con variables de medida discretas o continuas de los criterios y basados en análisis de diferencias de medias, en ecuaciones de regresión o funciones discriminantes. Por ende, la respuesta que se obtiene en torno a la verdad o mentira no es judicialmente operativa y, en términos científicos, se enmarca dentro de la lógica difusa o borrosa. Por su parte aquellos estudios orientados a la validez predictiva no especifican los criterios seguidos por los evaluadores de lo que se infiere que no están sistematizados y que, por lo tanto, carecen de una capacidad de generalización al contexto forense. El presente estudio, además, nos advierte del peligro inherente de tomar los atributos sin haber fijado explícita y claramente los puntos de corte. De hecho, los atributos están presentes en la mayoría de los casos lo que, en una primera aproximación, nos llevaría a concluir que las declaraciones eran mayoritariamente verdaderas cuando no lo son. Incluso la probabilidad observada de error es superior al simple azar.
b. El análisis de la validez predictiva del *CBCA*, tomando la probabilidad acumulada de falsas alarmas y como punto de contraste para confirmar una declaración verdadera en la presencia de 7 criterios (Steller, 1989), nos informa que el porcentaje de falsas alarmas (clasificar como declaración verdadera una fabricada) sería del 15.5%.
c. Los criterios del CBCA no son igualmente productivos en diferentes contextos lo que debe llevar ineludiblemente bien a una especificidad contextual de los evaluadores bien a un aprendizaje de los efectos de contexto. Esta debilidad de predicción no se observa en los atributos del *Reality Monitoring* de lo que se desprende, en principio y a la espera de contrastes en otros contextos, que la destreza de un evaluador se puede generalizar. Ahora bien, la lógica de los modelos contamina estos resultados pues mientras en *Reality Monitoring* se centra en el análisis de la actividad

cognitiva general, el *CBCA* lo hace en la específica, en la que interacciona con los hechos. Por eso, las predicciones del CBCA interaccionan con el caso concreto (p.e., en una amenaza ha de haber casi indefectiblemente conversaciones en tanto en un robo no tiene porqué) e incluso con la muestra (v. gr., los adultos producen más actividad cognitiva que los menores).

d. Esta falta de operatividad de los atributos y criterios pudiera estar motivada porque no se ha tomado como covariable la complejidad del evento y las capacidades del testigo a fin de controlar sus efectos. Incluso sería conveniente estudiar sus trazos de memoria. Aún así, a no ser que se formularan unos criterios cuantitativos muy concretos, estos abrirían aún más la puerta a la discrecionalidad del evaluador.

e. Los sistemas de análisis de las declaraciones mediante estos atributos y criterios no conforman auténticos sistemas de categorías metódicos (p.e., las categorías no están bien definidas, las definiciones no son coincidentes, no son mutuamente excluyentes, se confunden criterios de fiabilidad y validez, no se han formulado puntos de corte). En resumen, los sistemas no son metódicos sino semi-objetivos y es preciso realizar un esfuerzo para encauzarlos en línea con una mayor objetividad (Bakeman y Gottman, 1989; Hebert y Attridge, 1975; Wicker, 1975) al tiempo que ajustar mejor las categorías para mejorar la capacidad predictiva del sistema.

Esta inconsistencia interna de los sistemas de evaluación de la credibilidad puesta a prueba limitan su valor para concluir tanto judicialmente (v. gr., Wigmore, 1937; Sentencia del TS de 8 de febrero de 1995, RJ 808; Sentencia del TS de 18 de noviembre, RJ 7987) como en la prueba psicológico-judicial (p.e., Hastie et al., 1983). En suma, no conforman una prueba inequívoca con lo que es necesario complementarlos con otras pruebas judiciales o adoptar una aproximación multimétodo que convierta en más robustos los resultados. Además, en la práctica forense el sistema descansa, en una gran medida, en las habilidades del evaluador lo que también le resta valor de prueba porque se no puede contrastar la efectividad de dicho evaluador. En resumen, el sistema por sí solo no es prueba judicial suficiente. Una solución podría ser completar el estudio con la huella psicológica, el análisis de la simulación o de las dos

partes. En esta línea, nosotros (Arce y Fariña, 2005) han confeccionado un protocolo de medida global, el Sistema de Evaluación Global.

Referencias

Alonso-Quecuty, M. L. (1990). Recuerdo de la realidad percibida vs. imaginada: Buscando la mentira. *Boletín de Psicología*, 29, 73-86.

Alonso-Quecuty, M. L. (1993). Información post-evento y reality monitoring. Cuando el testigo "no puede" ser honesto. En M. Diges, y M. L. Alonso-Quecuty. *Psicología forense experimental*. Valencia: Promolibro.

Alonso-Quecuty, M. L. (1995). Psicología y testimonio. En M. Clemente (ed.), *Fundamentos de la psicología jurídica* (pp. 171-184). Madrid: Fundación Universidad Empresa.

Arce, R. y Fariña, F. (2005). El Sistema de Evaluación Global (SEG) de la crediibilidad del testimonio: Hacia una propuesta integradora. En R. Arce, F. Fariña, y M. Novo (eds.). *Psicología jurídica* (pp. 101-118). Santiago de Compostela: Consellería de Xustiza, Interior e Administración Local. Arce, R., Fariña, F. y Freire, Mª. J. (2002). Contrastando la generalización de los métodos empíricos de detección del engaño. *Psicologia: Teoria, Investigação e Prática*, 7(2), 71-86.

Arce, R., Novo, M. y Alfaro, E. (2000). La obtención de la declaración en menores y discapacitados. En A. Ovejero, Mª de la V. Moral y P. Vivas (eds.), *Aplicaciones en psicología social* (pp. 147-151). Madrid: Biblioteca Nueva.

Bakeman, R. y Gottman, J. M. (1989). *Observación de la interacción: Introducción al análisis secuencial*. Madrid: Morata.

Fariña, F., Arce, R. y Novo, M. (2002). Heurístico de anclaje en las decisiones judiciales. *Psicothema*, 14(1), 39-46.

Fisher, R. P. y Geiselman, R. E. (1992). *Memory-enhancing techniques for investigative interview*. Sprinfield: Charles C. Thomas.

Hebert, J. y Attridge, (1975). A guide for developers and users of observation systems and manuals. *American Educational Resources Journal*, 12(1), 1-20.

Johnson, M.K. y Raye, C.L. (1981). Reality monitoring. *Psychological Review*, 88, 67-85.

Jones, D. P. y McGraw, J. M. (1987). Reliable and fictitious accounts of sexual abuse in children. *Journal of Interpersonal Violence*, 2, 27-45.

Hastie, H., Penrod, S. D. y Pennington, N. (1983). *Inside the jury*. Cambridge, Mass.: Harvard University Press [trad. castellana: *La institución del jurados en los Estados Unidos. Sus intimidades.* Madrid, Cívitas, 1986].

Landry, K. L. y Brigham, J. C. (1992). The effect of training in criteria-based content analysis on the ability to of detect deception in adults. *Law and Human Behavior, 16*, 663-676.

Porter, S. y Yuille, J. C. (1996). The language of deceit: An investigation of the verbal clues in the interrogation context. *Law and Human Behavior, 20*, 443-458.

Sporer, S. L. y Küpper, B. (1995). Realitätsüberwachung und die Beurteilung des wahrheitsgehaltes von erzählungen: Eine experimentelle studie. *Zeitschrift für Sozialpsychologie*, 26, 173-193.

Steller, M. (1989). Recent developments in statement analysis. En J. Yuille (ed.), *Credibility assessment* (pp. 135-154). Dordrecht: Kluwer Academic Publishers.

Steller, M. y Köhnken, G. (1989/1994). Análisis de declaraciones basados en criterios. En D.C. Raskin (ed.), *Métodos psicológicos en la investigación y pruebas criminales* (pp. 217-245). Bilbao: Desclée de Brouwer (Orig. 1989).

Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327-352.

Undeutsch, U. (1967). Beurteilung der glaubhaftigkeit von zeugenaussagen. En U. Undetsch (ed.), *Handbuch der psychologie, Vol. II: Forensische psychologie* (pp. 26-181). Göttingen: Verlag für Psychologie.

Vrij, A. (2000). *Detecting lies and deceit*. Chichester: Wiley and sons.

Vrij, A. y Akehurst, L. (1998). Verbal communication and credibility: Statement validity. En A. Memon, A. Vrij y R. Bull (eds.), *Psychology and law. Truthfulness, accuracy and credibility* (pp. 3-31). Londres: McGraw-Hill.

Vrij, A., Edward, K., Roberts, K. P. y Bull, R. (2000). Detecting deceit via analysis of verbal and analysis of non verbal behavior. *Journal of Nonverbal Behavior*, 24, 239-263.

Arce, R., Seijo, D., y Rodríguez, M. J. (2005). Falsas alarmas de los criterios de realidad en delitos de agresiones sexuales e intimidación en muestras de adultos. En J. Sobral, G. Serrano, y J. Regueiro (Comps.), *Psicología jurídica, de la violencia y de género* (pp. 31-43). Madrid: Biblioteca Nueva. **ISBN: 84-9742-452-2.**

Wicker, A. W. (1975). An application of a multipletrait-multimethod logic to the reliability of observational records. *Personality and Social Psychology Bulletin*, 4, 575-579.

Wigmore, J. H. (1937). *The science of judicial proof*. Boston: Litltle Brown.