

ESTUDIO DE LA EFICACIA DE LAS CATEGORÍAS DE REALIDAD DEL TESTIMONIO DEL SISTEMA DE EVALUACIÓN GLOBAL (SEG) EN CASOS DE VIOLENCIA DE GÉNERO

Manuel Vilariño*, Mercedes Novo** y Dolores Seijo**

* Universidade Lusófona do Porto (Portugal).

** Departamento de Psicología Social, Básica y Metodología, Universidad de Santiago de Compostela (España).

(Recibido 8 de Junio de 2010; revisado 30 de Septiembre de 2010; aceptado 4 de Octubre de 2010)

Abstract

With the aim of contrasting the efficiency of the reality criteria in the assessment of the credibility of the testimony in violence against woman cases, a study was design. In this line, the efficacy of the Global Evaluation System (GES) adapted to violence against woman cases was compared with the standard procedure used, the SVA/CBCA. For this, the statements of 25 real victims of gender violence, confirmed by a judicial sentence, and 25 from mock victims, were collected. The results showed that the reality criteria discriminated between real and mock victim statements. In relation to the validity of the statements, results revealed that the procedures based on two statements (i.e, GES) were more productive in the detection of fabricated statements than those based in one statement (i.e., SVA). Finally, data supported that the reality criteria of the GES specific for gender violence cases were more efficient in differentiate real from fabricated statements, than the reality criteria of the CBCA.

Keywords: Credibility, Reality criteria, Testimony, Statement, Violence against women, Maltreatment.

Resumen

Se diseñó un estudio para contrastar la eficacia de los criterios de realidad en la evaluación de la credibilidad del testimonio en casos de violencia de género. En concreto, se contrastó la efectividad del Sistema de Evaluación Global (SEG) para casos de violencia de género con el procedimiento estándar basado en el SVA/CBCA. Para ello, se tomaron las declaraciones de 25 mujeres víctimas de violencia de género, confirmadas por sentencia judicial, y 25 de mujeres no víctimas, que inventaban una declaración. Los resultados mostraron que los criterios de realidad discriminaban entre declaraciones reales e inventadas. Por lo que se refiere a la validez de las declaraciones, los resultados pusieron de manifiesto que los procedimientos basados en dos declaraciones (i.e., SEG) eran más productivos en la detección de declaraciones inventadas que los basados en una declaración (i.e., SVA). Finalmente, los resultados apoyaron que los criterios de realidad para casos de violencia de género del SEG eran más eficaces en la discriminación entre declaraciones reales e inventadas, que los criterios de realidad del CBCA.

Palabras clave: Credibilidad, Criterios de realidad, Testimonio, Declaración, Violencia de género, Maltrato.

Introducción

La asunción, por parte de la sociedad en su conjunto, del maltrato a la mujer como una lacra, ha motivado la puesta en funcionamiento de un amplio abanico de medidas dirigidas a atajarla desde múltiples ámbitos. Un claro ejemplo de esta sensibilización ha sido el reconocimiento legal de la violencia de género como un tipo delictivo diferente de la violencia doméstica, que alcanza su máxima expresión en la Ley 1/2004, de 28 de diciembre, de Medidas de Protección Integral contra la Violencia de Género (para una revisión amplia de la evolución de ambos conceptos en el Derecho español véase Martín, 2009). Esta Ley recoge en el art. 1.3 que la violencia de género es “todo acto de violencia física y psicológica, incluidas las agresiones a la libertad sexual, las amenazas, las coacciones o la privación arbitraria de la libertad”, y el art. 1.1 establece que este tipo de violencia se ejerce “sobre las mujeres, por parte de quienes sean sus cónyuges o de quienes estén o hayan estado ligados a ellas por relaciones similares de afectividad, aún sin convivencia, como manifestación de la discriminación, la situación de desigualdad y las relaciones de poder de los hombres sobre las mujeres”.

No obstante, a pesar de estas medidas y de la conciencia social existente, el número de muertes por violencia de género continúa siendo elevado. En el año 2009, 55 mujeres murieron a manos de su pareja o ex-pareja (Centro Reina Sofía para el Estudio de la Violencia, 2009) y en el 2009 se han registrado un total de 55 muertes (Ministerio de Igualdad, 2009). Asimismo, el número de denuncias se mantiene estable desde el año 2007, rondando la cifra de 35.000 por trimestre.

Por otro lado, en relación con las decisiones judiciales, las estadísticas señalan que no más del 70% de los casos enjuiciados culminan en sentencia condenatoria (Europa Press, 2010; Ministerio de Igualdad, 2008), cuando en otros tipos delictivos el promedio se sitúa en el 90% (Novo, 2000). Aunque no puede establecerse como única causa de esta reducida probabilidad las dificultades dimanantes de la carga de la prueba, éstas parecen ser las más importantes. Así, el carácter íntimo y privado de este tipo de violencia obliga, en no pocas ocasiones, al juzgador a dictar sentencia en base al testimonio de las partes; fenómeno que evidencia la insuficiencia probatoria como uno de los principales problemas del proceso penal en casos de violencia de género (Arce, Seijo, y Novo, 2010). Esta contingencia, pone de manifiesto el importante papel que el psicólogo forense podría ejercer en esta casuística, de cara a facilitar la labor del

juzgador en la asignación de la credibilidad (Novo, y Seijo, 2010). De hecho, en un estudio de archivo se ha hallado que este tipo de corroboración periférica se practicaba en cerca del 20% de los casos (Arce, Vilariño, y Alonso, 2008). El peritaje psicológico, además de estudiar la potencial presencia de un daño psíquico originado por los hechos denunciados (Vilariño, Fariña, y Arce, 2009), puede analizar la realidad de los mismos mediante técnicas de análisis de contenido de la declaración.

La investigación psicológica sobre el testimonio tiene sus orígenes en las primeras décadas del siglo XX en Europa Central, expandiéndose rápidamente (Spörer, 2008) y tomando diferentes enfoques, que la literatura, en función del valor y de la relevancia judicial, ha concretado en correlatos del comunicador, indicios no verbales, indicios fisiológicos y análisis de la declaración (p.e., Spörer 1997; Vrij, 2008). En esta línea, la amplia difusión que los medios de comunicación han dado al polígrafo, ha conllevado el conocimiento masivo del mismo. Esta técnica presenta un funcionamiento que consiste en el registro de los cambios producidos a nivel psicofisiológico en diversas variables, ante la formulación de una serie de cuestiones relativas a los hechos investigados (i.e., tasa cardíaca, conductancia electrodérmica, presión sanguínea). Sin embargo, no existe un patrón único de respuesta fisiológica asociado al engaño (Lykken, 1981). Además, presenta importantes limitaciones como puede ser la realización de artimañas de tipo físico o mental por parte del sujeto, en el momento de responder a las cuestiones planteadas que provocan alteraciones en los registros (Sierra, Jiménez, y Bunce, 2006). En cualquier caso, en nuestro ordenamiento jurídico sería una prueba indiciaria, debido a que no está recogida como medio de prueba en la LECrim. De modo semejante, forman parte del conocimiento popular una serie de indicios no verbales de engaño (e.g., no mirar a los ojos al interlocutor durante el relato), que también han sido objeto de estudio científico, con el objetivo de dilucidar su alcance en la detección de la realidad de una declaración (v.gr., Eaton, McKeen, y Campbell, 2001; Vrij, 2008; Vrij, Akehurst, Soukara, y Bull, 2004; Vrij, Mann, y Fisher, 2006). Los resultados hallados dentro de este enfoque no son excesivamente concluyentes e, incluso, resultan contradictorios. En esta línea, Spörer y Schwandt (2007) llevaron a cabo un meta-análisis en el que hallaron que sólo el asentimiento, los movimientos de pies y piernas, y de manos presentan, de manera fiable, una relación negativa con la mentira. Esto es, los hallazgos contradicen la creencia general de que las conductas no verbales se incrementan durante la mentira, detectándose, por el contrario, un descenso en las mismas. Por otro lado, Arce, Fariña, y Freire (2002) en una comparación de los

métodos de análisis de contenido con los procedimientos de análisis no verbales y extralingüísticos, observaron un valor más alto para los primeros. De facto, algunos de estos métodos de análisis de contenido de la declaración gozan de un amplio desarrollo teórico y empírico, así como un importante uso en la práctica forense.

Dentro de esta perspectiva, centrada en el análisis de contenido de la declaración, destacan dos técnicas: el Reality Monitoring (RM; Johnson y Raye, 1981) en la versión ampliada por Spörer (1997); y el Statement Validity Assessment (SVA; Steller, 1989; Steller y Boychuck, 1992). El RM considera que las memorias de lo real y de lo imaginado varían en una serie de rasgos. De este modo, aquellas que poseen un origen externo, esto es, que están basadas en la observación de un suceso, presentan más cantidad de información sensorial, más detalles contextuales y menos referencias a procesos cognitivos que las de origen interno, es decir, imaginadas. Por su parte, el SVA, en la actualidad, constituye probablemente el instrumento más usado en el estudio de la credibilidad (Bensi, Gambetti, Nori, y Giusberti, 2009). Así, las evaluaciones mediante esta técnica son aceptadas como prueba en algunas cortes norteamericanas y en varios países del norte de Europa como Suecia, Alemania y Holanda (Vrij, 2005). Su aplicación se estructura en tres fases: a) obtención de la declaración mediante una entrevista semi-estructurada, b) estudio de la credibilidad del testimonio a través del Criteria Based Content Analysis (CBCA; Steller y Köhnken, 1994), y c) aplicación de un listado de validez.

El CBCA representa el elemento central del SVA. Inicialmente fue desarrollado con objeto de analizar la declaración de niños víctimas de abusos sexuales, aunque se ha comprobado su utilidad para la detección del engaño en testimonios de adultos (Köhnken, Schimossek, Aschermann, y Hofer, 1995; Landry y Brigham, 1992; Spörer, 1997) y en casos diferentes al abuso sexual (Steller, 1989; Tye, Amato, Honts, Devitt, y Peters, 1999; Yuille, 1988). El instrumento está compuesto por 19 criterios de credibilidad estructurados en torno a 5 grandes categorías. Estos criterios pueden analizarse como presentes o ausentes, o bien atendiendo a la fuerza o grado en que aparecen en el testimonio. Así, su presencia se interpretará en el sentido de veracidad de la declaración, aunque de su ausencia no se puede inferir la falsedad de la misma. En general, la investigación ha puesto de manifiesto que esta técnica cuenta con aval estadístico para discriminar entre testimonios reales e inventados (p.e., Arce, Fariña, y Vilariño, 2010; Parker y Brown, 2000; Tye et al., 1999; Vrij, 2005, 2008); aunque, no todos los resultados van en esta misma dirección (i.e., Porter y Yuille, 1996). Por otro

lado, se ha observado que factores tales como la familiaridad con el evento relatado (Blandon-Gitlin, Pezdek, Rogers, y Brodie, 2005; Pezdek et al., 2004), el entrenamiento en los criterios (Vrij, Akehurst, Soukara, y Bull, 2002), la edad (Vrij et al., 2004), y determinados patrones o trastornos de personalidad (Böhm, 2005; Böhm y Steller, 2008; Lee, Klaver, y Hart, 2008), pueden afectar las puntuaciones obtenidas mediante esta técnica. Además, no todos los criterios del CBCA parecen ser igual de útiles en la discriminación de ambos tipos de testimonios (p.e., Lee et al., 2008; Vrij, 2005). Estas discrepancias en los resultados podrían deberse, de acuerdo con Bensi et al. (2009), a los diferentes enfoques metodológicos empleados en los estudios experimentales.

Sin embargo, a pesar de la potencialidad de este instrumento para la práctica forense, dado que su diseño original estaba dirigido al análisis de la realidad de declaraciones de menores víctimas de abusos sexuales, su traslado y aplicación a casuísticas distintas ha de hacerse con suma cautela. Este hecho motivó que Arce y Fariña (2006a, 2009; Arce, 2010; Arce, Fariña, y Vivero, 2007) desarrollaran y validaran una adaptación del Sistema de Evaluación Global (SEG) (Arce y Fariña, 2005, 2006b, 2006c) para la evaluación forense de la credibilidad del testimonio en casos de violencia de género. Las categorías de validez, resultantes de la combinación de criterios del Statement Reality Analysis, SRA (Undeutsch, 1967, 1988); del Statement Validity Analysis, SVA (Steller, 1989; Steller y Boychuk, 1992; Steller, Raskin, Yuille, y Esplin, 1990) y de los criterios que Arce y Fariña (2006a, 2009; Arce, 2010; Arce, Fariña, y Vivero, 2007) concretaron a partir el estudio de la jurisprudencia española, fueron creadas y confeccionadas siguiendo las normas redactadas al respecto por Anguera (1990), dando lugar a un sistema categorial mutuamente excluyente, fiable y válido, en lo que Weick (1985) ha denominado como *sistemas de categorías metódicas*. Por su parte, adoptando esta misma metodología, crearon un sistema metódico de criterios de realidad fruto de la combinación de las categorías del Criteria Based Content Analysis, CBCA (Steller y Köhnken, 1994) a las que añadieron categorías complementarias de Arntzen (1970), Szewczyk (1973), Trankell (1972), del SRA (Undeutsch, 1967, 1988), del listado de criterios ampliados del Reality Monitoring, RM, por Spörer (1997), y de los criterios de realidad de Arce y Fariña específicos de casos de violencia de género.

Ante este estado de la literatura, hemos llevado a cabo un estudio con el objeto de comparar, en casos de violencia de género, la eficacia del modelo de evaluación de la realidad de la declaración más contrastado, el CBCA, pero que no es específico de casos

de violencia de género, ni ha sido verificado en este contexto, con el SEG adaptado para casos de violencia de género.

Método

Participantes

Las declaraciones sobre los hechos objeto de la denuncia fueron tomadas de 50 mujeres, 25 de ellas provenientes de casos reales acorde a una resolución judicial firme, con una edad media de 40,46 años, que fluctuaba entre los 19 y los 64 años ($M = 40,46$; $DE = 10,46$). Las otras 25, con historial de convivencia en pareja y que sostenían no haber sido objeto de violencia de género, tenían una edad promedio de 32 años, con un rango que oscilaba entre los 22 y los 73 años ($M = 32,24$; $DE = 13,71$).

Diseño

La metodología de investigación empleada fue del tipo experimental con casos de archivo y de un contexto experimental. Concretamente, se planificó un diseño para el contraste de la memoria de casos reales de violencia de género, tomada de casos de archivo, con la memoria de declaraciones inventadas, tomada de víctimas irreales.

Procedimiento

Las evaluaciones y declaraciones verdaderas de casos reales fueron tomadas de casos judiciales de violencia de género, en el que se recabó el testimonio de la denunciante, pero no formaba parte la prueba forense demandada por el juzgado y, por tanto, no formó parte de las pruebas de acusación. Se consideraron evaluaciones y declaraciones verdaderas todas aquellas en las que convergían dos criterios: la conformidad del acusado con la condena solicitada por la Fiscalía, esto es, la aceptación de los hechos imputados, y la existencia en el procedimiento de pruebas claras de culpabilidad (e.g., pruebas documentales, testimoniales, quebrantamientos de la orden de alejamiento) que llevaron a la condena del acusado. Por su parte, las evaluaciones o declaraciones inventadas fueron recabadas entre mujeres mayores de edad, que convivían o habían convivido en pareja. Para ello, se las había instruido en lo que denominamos “instrucciones de invención”. Éstas consistían en pedirles que se pusiesen

en el lugar de una mujer que inventaba haber sufrido malos tratos y que iba a presentar una denuncia por éstos. La denuncia falsa de malos tratos se justificaba en la obtención de algún beneficio asociado (v.gr., guarda y custodia de los hijos, venganza, compensaciones económicas). Se les dio una semana para que planificaran la denuncia. Transcurrida ésta, se inició el proceso de evaluación consistente en la obtención de la declaración. Se obtuvieron dos declaraciones sobre los hechos objeto de la denuncia espaciadas una semana entre ambas. El procedimiento seguido para la obtención de las declaraciones fue el de la entrevista cognitiva mejorada (Fisher y Geiselman, 1992).

Obtención de los protocolos

Las declaraciones sobre los hechos que motivaban la denuncia fueron obtenidas por medio de la entrevista cognitiva en la versión mejorada (Fisher, Geiselman, y Amador, 1989; Fisher y Geiselman, 1992), es decir, adaptada al contexto forense. Ésta comprende cuatro técnicas generales de recuperación de memoria: reinstauración de contextos, recuerdo libre, cambio de perspectiva y recuerdo en orden inverso. La aplicación de la entrevista cognitiva no está limitada a la reproducción de un suceso de episodio único, ya que permite extender su utilización para recordar hechos que ocurren frecuentemente de una manera similar (Mantwill, Köhnken, y Ascherman, 1995). No obstante, en este trabajo, la entrevista cognitiva mejorada ha sido aplicada con los pasos recuerdo libre y reinstauración de contextos (es decir, no tomamos el cambio de perspectiva y recuerdo en orden inverso), ya que ofrece los mismos resultados que la entrevista cognitiva completa (Davis, McMahon, y Greenwood, 2005), al tiempo que controla la potencial contaminación proveniente de los interrogatorios. Por esto mismo, tampoco se tomaron notas, ni los entrevistadores llevaban anotaciones de los pasos a dar, guiándose por un principio de escucha activa. Las entrevistas se grabaron en vídeo para los posteriores análisis.

Análisis de los protocolos

Las declaraciones sobre los hechos objeto de la denuncia, grabadas en vídeo, fueron sometidas a un análisis de contenido en el que establecer la validez y realidad de las mismas, por medio del procedimiento creado para casos de violencia de género por Arce y Fariña (2006a, 2009; Arce, 2010; Arce, Fariña, y Vivero, 2007).

Las categorías de validez, resultantes de la combinación de criterios del Statement Reality Analysis, SRA (Undeutsch, 1967, 1988); del Statement Validity

Analysis, SVA (Steller, 1989; Steller y Boychuk, 1992; Steller, Raskin, Yuille, y Esplin, 1990) y de los criterios propios de violencia de género de Arce y Fariña (2006a, 2009; Arce, 2010; Arce, Fariña, y Vivero, 2007), fueron creadas y confeccionadas siguiendo las normas redactadas al respecto por Anguera (1990), dieron lugar a la confección de un sistema categorial mutuamente excluyente, fiable y válido en lo que Weick (1985) ha denominado sistemas de categorías metódicas. Las categorías así como su definición se muestran a continuación:

- Prueba (in)suficiente (¿supera la capacidad de memoria del testigo? ¿contiene toda la información necesaria de los hechos?).
- Prueba (in)válida:
 - (In)consistencia interna (¿Tiene contradicciones internas en el relato?).
 - (In)consistencia externa (¿Es consistente con otras pruebas robustas o incontrovertibles?).
 - Persistencia en las declaraciones (¿Son estables las declaraciones en hechos y contextos?).
 - (In)consistente con la anterior (¿Hay consistencia inter-declaraciones, aparecen/desaparecen detalles o hechos de la agresión centrales para la víctima?).
 - (In)consistencia con las leyes científicas y de la naturaleza (¿Contiene el relato hechos incompatibles con las leyes científicas o de la naturaleza?).

La unidad de análisis en todas las categorías fue el protocolo, registrándose las categorías como prueba suficiente o insuficiente, y, para las categorías de validez, en consistente o inconsistente. A su vez, se creó una variable resumen de la validez de la prueba, de modo que se codificaba como prueba inválida si alguna de las categorías de validez invalidaba la prueba, y como válida, si ninguna de las categorías de validez la invalidaba.

Con la misma metodología para la creación de un sistema categorial metódico, Arce y Fariña (2006a, 2009; Arce, 2010; Arce, Fariña y Vivero, 2007), habían formulado y contrastado la validez de un modelo general de realidad de las declaraciones adaptado a casos de violencia de género, partiendo de las categorías del Criteria Based Content Analysis, CBCA (Steller y Köhnken, 1994) a las que añadieron categorías complementarias de Arntzen y Szewczyk, del SRA (Undeutsch, 1967, 1988), del listado de criterios ampliados al Reality Monitoring, RM, por Spörer (1997), y de los criterios de realidad de Arce y Fariña aplicables a casos de violencia de género. Las

categorías de este sistema, Sistema de Evaluación Global-General (SEG-G), se muestran a continuación:

CARACTERÍSTICAS GENERALES. Los criterios aquí englobados se refieren a la declaración tomada en su totalidad.

- 1.- Estructura lógica.
- 2.- Elaboración inestructurada.
- 3.- Cantidad de detalles.
- 4.- Concreción (SRA).

CONTENIDOS ESPECÍFICOS. Se evalúan partes específicas del testimonio referidas a la presencia o fuerza de ciertos tipos de descripciones.

- 5.- Engranaje contextual.
- 6.- Descripción de interacciones.
- 7.- Reproducción de conversaciones.
- 8.- Complicaciones inesperadas durante el incidente.
- 9.- Información perceptual (RM).
- 10.- Operaciones cognitivas (RM, codificación inversa)

PECULIARIDADES DEL CONTENIDO. Se incluyen aquí aquellas características de una declaración que aumentan su concreción o viveza.

- 11.- Detalles inusuales.
- 12.- Detalles superfluos.
- 13.- Incomprensión de detalles relatados con precisión.
- 14.- Originalidad de las expresiones.
- 15.- Asociaciones externas relacionadas.
- 16.- Relatos del estado mental subjetivo.
- 17.- Atribución del estado mental del autor del delito.

CONTENIDOS REFERENTES A LA MOTIVACIÓN. Estos criterios desvelan la motivación del testigo para hacer la declaración.

- 18.- Correcciones espontáneas.
- 19.- Admisión de falta de memoria.
- 20.- Plantear dudas sobre el propio testimonio.
- 21.- Auto-desaprobación.
- 22.- Perdón al autor del delito.

ELEMENTOS ESPECÍFICOS DE LA AGRESIÓN. Elementos del testimonio que no se relacionan con la viveza general de la declaración, sino con el crimen.

23.- Detalles característicos de la agresión.

CRITERIOS ESPECÍFICOS PARA VIOLENCIA DE GÉNERO (Arce y Fariña).

Elementos del testimonio.

24.- Manifestación de padecimiento de síntomas sutiles clínicos.

25.- Justificación de la agresión (provocadora).

26.- Intento de ocultar hechos que agravan la agresión.

27.- Comportamiento protector del agresor (“que no vaya a la cárcel”, “que no le pase nada”).

28.- Exculpación del agresor (“no me quería hacer daño”, “me pagaba para que me relajara”, “autodenuncia para exculpar al agresor”).

29.- Propuesta de una medida educativa para el agresor (“que no beba”, “métnle miedo”, “hagan que se porte bien”) (codificación inversa).

30.- Echar de menos la agresión.

31.- “Vuelta atrás” en la acusación.

Partiendo de este modelo general, Arce y Fariña formularon uno basado en los datos sobre la base de que, si bien estas categorías son propias de realidad, hay variables intervinientes que podrían mediar su eficacia y productividad. Los resultados de investigaciones con simulaciones dieron como resultado que las variables estructura lógica, elaboración inestructurada, cantidad de detalles, concreción, complicaciones inesperadas, originalidad de las expresiones, correcciones espontáneas, admisión de falta de memoria, planteamiento de dudas sobre el testimonio, autodesaprobación y sintomatología sutil, diferenciaban significativamente entre declaraciones reales e inventadas. Por su parte, los restantes criterios específicos de violencia de género, aunque eran poco productivos, en la mayoría de las variables sólo eran característicos de casos reales de violencia de género. Las restantes variables no discriminaban significativamente entre declaraciones reales e inventadas. Al respecto, Arce y Fariña hallaron que el modelo categorial ajustado, Sistema de Evaluación Global-Ajustado (SEG-A), era más efectivo en la discriminación entre declaraciones reales e inventadas que el original.

La unidad de análisis en todas las categorías fue el protocolo, registrándose las categorías como presentes o ausentes. Descartamos otras escalas de medida como el grado o intensidad de los criterios, no sólo porque la evaluación categórica se ajusta perfectamente al contexto forense, sino también porque las otras alternativas de medida abundan más en la carencia de objetividad del procedimiento, al no poderse establecer

puntos de corte sistemáticos, ya que la productividad de la memoria está mediatizada por numerosos factores externos difícilmente controlables como la complejidad del evento, o la capacidad intelectual o narrativa de la persona evaluada. Sometido el SEG-G a un contraste de la consistencia interna obtuvimos un coeficiente *alpha de Cronbach* ($n = 50$) de ,789.

Entrenamiento de los codificadores

En este estudio participaron dos codificadores, ciegos en relación a las hipótesis de trabajo, con experiencia forense en la codificación de este tipo de material (Vilariño, 2010, Vivero, 2006). Los codificadores fueron exhaustivamente entrenados en éste y otros sistemas de codificación. El entrenamiento consistió, tras presentarles y ejemplificarles cada categoría de análisis, en la ejecución con material que no iban a codificar posteriormente, utilizando el índice de concordancia como instrumento de cotejo de la ejecución, lo que permitía, constatada inconsistencia, subsanar errores de codificación a través de la homogeneización de criterios. Se confeccionó un manual con las definiciones, ejemplos de cada categoría de análisis y con la resolución de las fuentes previamente detectadas de dudas y errores en la codificación.

Análisis de la fiabilidad de las entrevistas y codificaciones

Varios fueron los entrevistadores del estudio por lo que era preciso conocer si el factor entrevistador pudiera mediar la productividad de la entrevista. Para ello dividimos los protocolos de víctimas reales e irreales en dos grupos al azar. Si los protocolos son iguales, no debería haber diferencias en las categorías de realidad y validez detectadas en las entrevistas de hechos. Los resultados del análisis de contenido de las entrevistas de hechos mostraron unos protocolos similares en los registros tanto en las categorías de realidad, $F(1;48) = 0,01$; $MC = 0,08$; ns , $\eta^2 = ,000$; $1-\beta = ,051$, como en las de validez, $F(1;48) = 0,31$; $MC = 0,08$; ns , $\eta^2 = ,006$; $1-\beta = ,085$. En suma, las entrevistas no están contaminadas por el factor entrevistador. A su vez, estos entrevistadores también se habían mostrado consistentes y productivos en otros estudios (p.e., Arce, Fariña, Carballal y Novo, 2006).

Los dos codificadores dividieron el material entre ellos. La distribución del material a los codificadores fue aleatoria y combinando entrevistas de casos reales con entrevistas de casos falsos.

Para calcular la fiabilidad intra-jueces, los codificadores repitieron, una semana después del fin de la codificación original, en torno al 20% de las entrevistas que ellos mismos habían evaluado. Asimismo y para obtener la fiabilidad inter-jueces, cada codificador examinó alrededor del 20% de los protocolos que inicialmente había codificado el otro codificador.

Se consideran evaluaciones concordantes aquellas que superan el punto de corte de ,80 (Tversky, 1977), que es más restrictivo que los valores kappa. En todas las variables y condiciones de evaluación (e.g., intra-evaluador, inter-evaluadores) el IC fue mayor de ,80 por lo que podemos sostener que las evaluaciones son consistentes inter-, intra-evaluadores y en el tiempo. Además, los codificadores también fueron consistentes en otros contextos de evaluación (i.e., Vilariño, 2010; Vivero, 2006). En consecuencia, los datos obtenidos del análisis de contenido de las entrevistas cognitivas son fiables (Wicker, 1975).

Resultados

Estudio de la validez de las declaraciones

Tabla 1. Prueba de χ^2 para el estudio de la combinación de los criterios de validez combinados del SVA, SRA y los criterios de validez de Arce y Fariña.

Variable	χ^2	p	ϕ	%_{dr}	%_{di}
Prueba insuficiente	3,56	,059	,333	0	20
Falta de consistencia interna	13,27	,001	,562	0	52
Persistencia de la declaración	27,5	,001	,783	0	76
Inconsistencia inter-declaraciones	18,67	,001	,655	0	60
Inconsistencia leyes naturaleza y científicas	----	----	----	0	0

Nota. $gl(1)$. %_{dr} = % en las declaraciones reales; %_{di} = % en las declaraciones inventadas. La categoría “carencia de consistencia externa” no se pudo codificar en las declaraciones inventadas. En todo caso, ha de tenerse presente que no se detectó ninguna contingencia entre las declaraciones verdaderas.

El estudio de la validez de las declaraciones (ver Tabla 1), esto es, de la prueba, advierte que todas las declaraciones reales constituyen prueba válida y suficiente, y de

la existencia de diferencias significativas entre las declaraciones verdaderas y las inventadas en los criterios: Prueba insuficiente (o sea, la falta de prueba), consistencia interna en la declaración, persistencia en la declaración, y consistencia entre las sucesivas declaraciones. Sucintamente, las declaraciones inventadas resultan en más pruebas insuficientes (20%), aunque todas las declaraciones comprendían actos de violencia física y psicológica; en relatos inconsistentes internamente (52%); y, en sucesivas declaraciones, en relatos no persistentes en hechos y contextos (76%) e inconsistentes inter-declaraciones (60%). Aditivamente, la prueba es inválida (la invalidez es el resultado del efecto aditivo de los criterios falta de consistencia interna, falta de persistencia en la declaración, falta de consistencia inter-declaraciones e inconsistencia con las leyes científicas y de la naturaleza) en el 88% de las declaraciones inventadas, que discrimina significativamente, $\chi^2(1) = 35,8; p < ,001; \phi = ,886$, entre éstas y las verdaderas (0). En resumen, mientras las declaraciones reales son sistemáticamente prueba válida y suficiente, el 20% de las declaraciones inventadas son prueba insuficiente y el 88% prueba inválida (no se observó que la interacción de ambas variables mejorase la detección del engaño). Por su parte, el modelo de estimación de la validez basado en una única declaración (las categorías de análisis aplicables son la consistencia interna de la declaración y la consistencia con las leyes científicas y de la naturaleza), invalidaría el 52% de las declaraciones inventadas y ninguna de las verdaderas, resultando significativamente más efectivo un protocolo basado en dos declaraciones (modelo SEG), frente a una (modelo SVA), $\chi^2(1) = 6,1; p < ,05; \phi = ,393$.

Estudio de la realidad de las declaraciones

El contraste del modelo general de Arce y Fariña, el SEG-G (2006a, 2009; Arce, 2010; Arce, Fariña, y Vivero, 2007) para casos de violencia de género, consistente en una combinación de criterios de realidad de Arntzen y Szewczyk, del CBCA, RM, SRA y criterios específicos de violencia de género de los propios Arce y Fariña, ha puesto de manifiesto que en las declaraciones verdaderas hay más criterios de realidad ($n = 390$) que en las inventadas ($n = 247$), $\chi^2(1) = 31,1; p < ,001$. No obstante, con este modelo para dejar en cero la identificación de una declaración falsa como verdadera, falsos positivos, (en el contexto judicial no se puede asumir error alguno en la clasificación como culpable de un inocente; es decir, es preferible absolver a un culpable a condenar a inocente) se precisa que la declaración contenga, al menos, 14 criterios de realidad,

con el coste de un 20% de falsos negativos (clasificaciones erróneas de declaraciones verdaderas como carentes de realidad) (ver Tabla 2).

Tabla 2. Criterios de realidad del SEG-G en las declaraciones reales e inventadas.

Nº de criterios/condición	Real	Inventada
6		2(8%)
7		1(4%)
8		3(12%)
9	1(4%)	4(16%)
10		6(24%)
11	1(4%)	4(16%)
12		1(4%)
13	3(12%)	4(16%)
14	2(8%)	
15	7(28%)	
16	3(12%)	
17	3(12%)	
18	1(4%)	
19	1(4%)	
20	1(4%)	
21	2(8%)	

Los resultados del contraste de la eficacia discriminativa entre declaraciones reales e inventadas de los criterios de realidad del SEG-G (ver Tabla 3), mostraron que las categorías estructura lógica, elaboración inestructurada, cantidad de detalles, concreción, complicaciones inesperadas, originalidad de las expresiones, correcciones espontáneas, admisión de falta de memoria, planteamiento de dudas sobre el testimonio, autodesaprobación y sintomatología sutil, diferenciaban entre ambas. En concreto, la presencia de estos criterios se relaciona con realidad de la declaración. Estos resultados dan soporte al modelo ajustado propuesto por Arce y Fariña, el SEG-A.

Tabla 3. Prueba de chi cuadrado para el estudio de la combinación de los criterios del CBCA, RM, SRA y los criterios de violencia de género de Arce y Fariña.

Variable	χ^2	p	ϕ	%_{dr}	%_{di}
CARACTERÍSTICAS GENERALES					
Estructura lógica	13,27	,001	-,562	100	52
Elaboración inestructurada	8,88	,003	-,468	96	56
Cantidad de detalles	3,72	,014	-,403	100	72
Concreción	28,93	,001	-,801	88	8
CONTENIDOS ESPECÍFICOS					
Engranaje contextual	----	----	----	100	100
Descripción de interacciones	----	----	----	100	100
Reproducción de conversaciones	----	----	----	100	100
Complicaciones inesperadas	7,29	,007	-,436	32	0
Información perceptual	2,03	,154	-,242	68	44
Operaciones cognitivas	0	1	,143	0	4
PECULIARIDADES DEL CONTENIDO					
Detalles inusuales	0	1	,000	36	36
Detalles superfluos	1,71	,192	-,246	96	80
Incomprensión detalles relatados	----	----	----	0	0
Originalidad de las expresiones	5,26	,022	-,365	76	40
Asociaciones externas relacionadas	3,06	,080	-,288	76	48
Relatos estado mental subjetivo	----	----	----	96	96
Atribución estado mental autor delito	0,52	,470	,204	92	100
CONTENIDOS REFERENTES A LA MOTIVACIÓN					
Correcciones espontáneas	16,09	,001	-,608	72	12
Admisión de falta de memoria	5,15	,023	-,361	64	28
Plantear dudas sobre testimonio	9,52	,002	-,480	52	8
Auto-desaprobación	4,08	,043	-,327	56	24
Perdón al autor del delito	0	1	,000	12	12
ELEMENTOS ESPECÍFICOS DE LA AGRESIÓN					
<u>Detalles característicos del delito</u>	----	----	----	0	0

Tabla 3 (Continuación). Prueba de chi cuadrado para el estudio de la combinación de los criterios del CBCA, RM, SRA y los criterios de violencia de género de Arce y Fariña.

Variable	χ^2	p	ϕ	%_{dr}	%_{di}
CRITERIOS ESPECÍFICOS DE VIOLENCIA DE GÉNERO					
Síntomas sutiles	7,29	,007	-,436	32	0
Justificación	1,13	,289	-,200	28	12
Ocultar hechos	0,19	,663	-,123	16	8
Protector	0,52	,470	-,204	8	0
Exculpación	---	---	---	0	0
Medida educativa	1,42	,234	,253	12	0
<u>Echar de menos la agresión</u>	----	----	----	4	0

Nota. $gl(1)$. %_{dr} = % de declaraciones en la condición de memoria de realidad; %_{di} = % de declaraciones en la condición de memoria de imaginación. La categoría “Vuelta atrás en la denuncia” no se puede registrar en las declaraciones inventadas. En todo caso, no se detectó ninguna contingencia entre las declaraciones verdaderas.

Este modelo ajustado da lugar a que las declaraciones verdaderas ($n = 183$) contengan más criterios de realidad que las inventadas ($n = 75$), $\chi^2(1) = 45,21$; $p < ,001$. En el estudio de casos (ver Tabla 4) observamos que las declaraciones inventadas pueden llegar a contener hasta 5 de estos criterios, con lo que para no cometer falsos positivos, se ha de asumir una tasa de falsos negativos del 8%, esto es, dos declaraciones verdaderas no serían consideradas como tales al no contar con suficientes criterios de realidad. En cualquier caso, se observa una reducción de clasificaciones incorrectas en relación con el modelo general.

Los criterios de realidad del CBCA, cuyos resultados del valor clasificatorio pueden verse en la Tabla 5, también aparecen en mayor medida en las declaraciones verdaderas ($n = 320$) que en las inventadas ($n = 231$), $\chi^2(1) = 14,38$; $p < ,001$. No obstante, para evitar la presencia de falsos positivos se precisan 13 criterios de realidad, con lo que el 44% de las declaraciones verdaderas no serían identificadas como tales (falsos negativos). De este modo, se observa un importante incremento en la tasa de falsos negativos en relación con los dos modelos anteriores.

Tabla 4. Criterios de realidad del SEG-A en las declaraciones reales e inventadas.

Nº de criterios/condición	Real	Inventada
0		1(4%)
1		2(8%)
2		7(28%)
3		6(24%)
4		6(24%)
5	2(8%)	3(12%)
6	8(28%)	
7	4(12%)	
8	5(20%)	
9	4(16%)	
10	1(4%)	
11	1(12%)	

Tabla 5. Criterios de realidad del CBCA en las declaraciones reales e inventadas.

Nº de criterios/declaración	Real	Inventada
5		1(4%)
6		2(8%)
7		2(8%)
8		3(12%)
9	1(4%)	4(16%)
10	2(8%)	5(20%)
11	4(16%)	6(24%)
12	4(16%)	2(8%)
13	6(24%)	
14	3(12%)	
15	1(4%)	
16	4(16%)	

Comparativamente, el modelo de realidad con las categorías del SEG-A, $\chi^2(1) = 6,65$; $p < ,01$; $\phi = ,410$ es estadísticamente más efectivo que el basado en el CBCA, e

igual que el SEG-G, $\chi^2(1) = 2,30$; *ns*; $\phi = ,257$. Aún así, sería preferible tomar el SEG-G con una tasa del 20% de falsos negativos que el CBCA, cuya tasa ascendería al 44%.

Estudio combinado de la validez y fiabilidad de la declaración

El estudio de la validez basado en el Sistema de Evaluación (SEG) advierte que el 88% de las declaraciones son prueba insuficiente o inválida, que lo convierte en un detector robusto de lo inventado, $\chi^2(1) = 28,88$; $p < ,001$. En todo caso, el modelo de validez aisladamente permitiría que el 12% de declaraciones inventadas fueran clasificadas como reales, en tanto los modelos derivados de los criterios de realidad del SEG-G y SEG-A tolerarían un margen de error del 20 y 8%, respectivamente. La combinación, de acuerdo con el SEG, de ambos estudios, validez y fiabilidad (esto es, el estudio de los criterios de realidad en la declaración), verificada la capacidad de clasificación correcta de los criterios de realidad del SEG-G y del SEG-A, entre declaraciones que son prueba válida y suficiente, mostró que, en el modelo general, las declaraciones inventadas pueden contener hasta 11 criterios (1), con lo que la tasa de falsos negativos se reduce a un 8%. Por su parte, en el modelo ajustado las declaraciones inventadas que eran prueba válida y suficiente, no incluían más de 3 criterios de realidad, mientras que las declaraciones verdaderas contienen, al menos, 5 criterios. En consecuencia, la combinación de la validez y suficiencia de la prueba con el modelo ajustado de la fiabilidad posibilita la discriminación total entre declaraciones verdaderas e inventadas.

Por su parte, el modelo del SVA, basado en el estudio de la validez con una única declaración y en los criterios de realidad del CBCA, dejaría, entre las declaraciones válidas, en la región de declaraciones verdaderas no clasificadas como tales (falsos negativos) al 44%, habiendo de tomarse como punto de corte 13 criterios de realidad, para evitar la comisión de falsos positivos.

Discusión

Previamente a la discusión de las conclusiones que se pueden derivar de los resultados hallados, se advierte de las potenciales limitaciones que han de ser consideradas para determinar el verdadero alcance. En primer lugar, el tamaño de la

muestra es reducido, lo que podría condicionar los resultados obtenidos. En esta línea, un incremento del tamaño muestral, posiblemente generase un aumento de la variabilidad de los casos vividos, lo que podría acarrear cambios en la productividad de las categorías. Segundo, aunque se tomaron medidas encaminadas a garantizar el carácter real de las víctimas reales de violencia de género, no es posible garantizar plenamente esta condición. De modo similar, carecemos de la certeza absoluta de que las víctimas irreales nunca hubieran sido víctimas directas o indirectas de violencia de género. En tercer lugar, se ha equiparado la prestación del testimonio en condiciones reales y de laboratorio en recreaciones de elevada fidelidad. Sin embargo, resulta imposible replicar totalmente el proceso legal y de evaluación, por el que pasarían las participantes irreales en caso de presentar en realidad una denuncia. Por último, dado que la casuística tomada es exclusiva de violencia de género, se ha de ser excesivamente cauteloso al generalizar estos resultados a otros tipos delictivos. Ateniéndonos a las limitaciones reseñadas, se pueden formular una serie de conclusiones dirigidas a dar respuesta a los objetivos establecidos en la parte introductoria del estudio.

Se ha registrado un evento narrativo que incluía violencia física y psicológica en todos los testimonios de las víctimas irreales de violencia de género. Este hecho evidencia que las víctimas irreales están capacitadas para prestar un testimonio que avale con hechos la denuncia. No obstante, una de cada cinco declaraciones inventadas no eran prueba suficiente, bien por no contener un evento narrativo de los hechos, bien por no superar la capacidad de retentiva del testigo. Asimismo, el estudio de la validez de las declaraciones acorde a los criterios del SEG, se ha mostrado sumamente efectivo a la hora de detectar estos testimonios como inventados. Concretamente, detectó el 88% de las declaraciones inventadas, frente a sólo el 52% detectado por el sistema tradicional, el SVA, basado en una declaración, evidenciándose una clara superioridad del estudio de la validez basado en dos declaraciones frente a una.

Por otro lado, tanto los modelos de realidad general y ajustado del SEG específicos para violencia de género, como el modelo del CBCA, cuentan con un aval estadístico a la hora de discriminar entre declaraciones reales e inventadas, esto es, los criterios que conforman los modelos aparecen significativamente en mayor medida en las declaraciones reales que en las inventadas. Estos resultados convergen con la literatura existente para el caso del CBCA (v.gr., Arce et al., 2010; Gödert, Gamer, Rill, y Vossel, 2005; Vrij, 2005). Sin embargo, para la práctica forense ha de establecerse un

punto de corte que discrimine entre ambos tipos de testimonios y que, además, evite la aparición de falsos positivos (criterio de decisión estricto). En el presente estudio, siguiendo el procedimiento descrito por Arce y Fariña, lo hemos establecido en el número máximo de criterios de realidad que puede contener una declaración falsa: 14 en el SEG-G, 5 en el SEG-A y 13 en el CBCA. Aunque con estos puntos de corte se consigue clasificar de manera correcta todos los testimonios falsos, su asunción supone obtener unas tasas de falsos negativos del 44% con el modelo del CBCA, del 20% con el SEG-G, y del 8% con el SEG-A. No obstante, mediante la combinación del estudio de la validez y de la realidad basados en el SEG, podemos reducir estas tasas a un 8% con el modelo general e, incluso, obtener una tasa nula con el ajustado. Por su parte, la combinación del CBCA con los criterios de validez del SVA no supone una reducción en el número de falsos negativos.

En conclusión, aunque los tres modelos cumplen los objetivos de la tarea forense al posibilitar el establecimiento de un criterio de decisión estricto, esto es, que detecte todas las declaraciones inventadas, el modelo SVA/CBCA ofrece unas tasas de ejecución forense extremadamente pobres, ofreciendo un porcentaje muy elevado de falsos negativos. Por su parte, el modelo de análisis de credibilidad de la declaración del SEG, basado en un estudio combinado de la validez y la realidad del testimonio, responde más adecuadamente a las demandas de la práctica forense, especialmente en el caso del modelo de realidad ajustado, el SEG-A.

Referencias

- Anguera, M.T. (1990). Metodología observacional. En J. Arnau, M.T. Anguera, y J. Gómez (Eds.), *Metodología de la investigación en las ciencias del comportamiento* (pp. 125-236). Murcia: Servicio de Publicaciones de la Universidad de Murcia.
- Arce, R. (2010). El Sistema de Evaluación Global en casos de violencia de género: Huella psíquica y testimonio. *Informació Psicológica*, 99, 19-35.
- Arce, R., y Fariña, F. (2005). Peritación psicológica de la credibilidad del testimonio, la huella psíquica y la simulación: El Sistema de Evaluación Global (SEG). *Papeles del Psicólogo*, 26, 59-77.

- Arce, R., y Fariña, F. (2006a). *Construcción y validación de un sistema categorial metódico de realidad para casos de violencia de género*. Inédito, Unidad de Psicología Forense, Universidad de Santiago de Compostela.
- Arce, R., y Fariña, F. (2006b). Psicología del testimonio: Evaluación de la credibilidad y de la huella psíquica en el contexto penal. En Consejo General del Poder Judicial (Ed.), *Psicología del testimonio y prueba pericial* (pp.39-103). Madrid: Consejo General de Poder Judicial.
- Arce, R., y Fariña, F. (2006c). Psicología del testimonio y evaluación cognitiva de la veracidad de testimonios y declaraciones. En J.C. Sierra, E.M. Jiménez, y G. Buela (Coords.), *Psicología forense: Manual de técnicas y aplicaciones* (pp. 563-601). Madrid: Biblioteca Nueva.
- Arce, R., y Fariña, F. (2009). Evaluación psicológica forense de la credibilidad y daño psíquico en casos de violencia de género mediante el sistema de evaluación global. En F. Fariña, R. Arce, y G. Buela (Eds.), *Violencia de género. Tratado psicológico y legal* (pp. 147-168). Madrid: Biblioteca Nueva.
- Arce, R., Fariña, F., Carballal, A., y Novo, M. (2006). Evaluación del daño moral en accidentes de tráfico: Desarrollo y validación de un protocolo para la detección de la simulación. *Psicothema*, 18, 278-283.
- Arce, R., Fariña, F., y Freire, M.J. (2002). Contrastando la generalización de los métodos empíricos de detección del engaño. *Psicologia: Teoria, Investigação e Prática*, 7(2), 71-86.
- Arce, R., Fariña, F., y Vilarinho, M. (2010). Contraste de la efectividad del CBCA en la evaluación de la credibilidad en casos de violencia de género. *Intervención Psicosocial*, 19(2), 109-119.
- Arce, R., Fariña, F., y Vivero, A. (2007). Estudio exploratorio de la efectividad de las técnicas de análisis de contenido de las declaraciones en casos de violencia de género. En C. Guillén y R. Guil (Coords.), *Psicología social: Un encuentro de perspectivas* (Vol. I, pp. 590-604). Cádiz: Asociación de Profesionales de la Psicología Social.
- Arce, R., Seijo, A., y Novo, M. (2010). Testimony validity: A comparative study of legal and empirical criteria. *Psychology in Spain*, 14(1), 1-7.
- Arce, R., Vilarinho, M., y Alonso, M.A. (2008, Diciembre). *Estudio de sentencias de violencia de género con menores implicados: Carga de la prueba y decisiones*

sobre menores. I Congreso de Multidisciplinar sobre Interferencias Parentales tras la Ruptura de Pareja, Santiago de Compostela.

- Arntzen, F. (1970). *Psychologie der zeugenaussage* [La psicología del testigo]. Göttingen: Verlag für Psychologie.
- Bensi, L., Gambetti, E., Nori, R., y Giusberti, F. (2009). Discerning truth from deception: The sincere witness profile. *The European Journal of Psychology Applied to Legal Context, 1*, 101-121.
- Blandon-Gitlin, I., Pezdek, K., Rogers, R., y Brodie, L. (2005). The effects of event familiarity on criterion-based content analysis ratings: An experimental study. *Law and Human Behavior, 29*, 187-197.
- Böhm, C. (2005). El peritaje psicológico sobre la credibilidad de declaraciones testimoniales en Alemania. Nuevos desafíos. En R. Arce, F. Fariña, y M. Novo (Eds), *Psicología jurídica* (pp. 93-101). Santiago de Compostela: Consellería de Xustiza, Interior e Administración Local.
- Böhm, C., y Steller, M. (2008). Testimonio y trastorno límite de personalidad. En F.J. Rodríguez, C. Bringas, F. Fariña, R. Arce, y A. Bernardo (Eds.), *Psicología Jurídica. Entorno judicial y delincuencia* (pp. 135-147). Oviedo: Ediciones de la Universidad de Oviedo.
- Centro Reina Sofía para el Estudio de la Violencia. *Mujeres asesinadas por su pareja, España* (2009). Recuperado el 02 de Enero de 2010, de <http://www.centroreinasofia.es/paneldecontrol/est/pdf/EST014-3291.pdf>
- Davis, M.R., McMahon, M., y Greenwood, K.M. (2005). The efficacy of mnemonic components of the cognitive interview: Towards a shortened variant for time-critical investigations. *Applied Cognitive Psychology, 19*, 75-93.
- Eaton, W.O., McKeen, N.A., y Campbell, D.W. (2001). The waxing and waning of movement: Implications for psychological development. *Developmental Review, 21*, 205-223.
- Europa Press (2010, Febrero, 16). *Condenados casi 100.000 hombres por violencia de género en tres años y medio*. Recuperado el 16 de Febrero de 2010, de <http://www.europapress.es/nacional/noticia-condenados-casi-100000-hombres-violencia-genero-tres-anos-medio-20090421214004.html>
- Fisher, R.P., y Geiselman, R.E. (1992). *Memory-enhancing techniques for investigative interview*. Springfield, IL: Charles C. Thomas.

- Fisher, R.P., Geiselman, R.E., y Amador, M. (1989). Field test of the cognitive interview: Enhancing the recollection of actual victims and witness of crime. *Journal of Applied Psychology*, 74, 722-727.
- Gödert, H.M., Gamer, M., Rill, H.G., y Vossel, G. (2005). Statement validity assessment: Inter-rater reliability of criteria-based content analysis in the mock-crime paradigm. *Legal and Criminological Psychology*, 10, 225-245.
- Johnson, M.K., y Raye, C.L. (1981). Reality monitoring. *Psychological Review*, 88, 67-85.
- Köhnken, G., Schimossek, E., Aschermann, E., y Höfer, E. (1995). The cognitive interview and the assessment of the credibility of adults' statements. *Journal of Applied Cognitive Psychology*, 80, 671-684.
- 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.
- Lee, Z., Klaver, J.R., y Hart, S.D. (2008). Psychopathy and verbal indicators of deception in offenders. *Psychology, Crime and Law*, 14, 73-84.
- Lykken, D.T. (1981). *A tremor in the blood*. Nueva York, NY: McGraw-Hill.
- Mantwill, M., Köhnken, G., y Ascherman, E. (1995). Effects of the cognitive interview on the recall of familiar and unfamiliar events. *Journal of Applied Psychology*, 80, 68-78.
- Martin, J.J. (2009). Evolución histórica en nuestro derecho: De la violencia intrafamiliar a la violencia de género. En F. Fariña, R. Arce, y G. Buena (Eds.), *Violencia de género. Tratado psicológico y legal* (pp. 19-38). Madrid: Biblioteca Nueva.
- Memon, A., Cronin, O., Eaves, R., y Bull, R. (1993). The cognitive interview and the child witness. En G.M. Stephenson y N.K. Clark (Eds.), *Issues in criminology and legal psychology: Vol. 20. Children, evidence and procedure* (pp. 3-9). Leicester: British Psychological Society.
- Ministerio de Igualdad (2008). *Evaluación de la aplicación de la ley orgánica 1/2004, de 28 de diciembre, de medidas de protección integral contra la violencia de género*. Recuperado el 30 de Octubre de 2009, de http://www.migualdad.es/noticias/pdf/informe_ejecutivo-14_julio_2008def.pdf
- Ministerio de Igualdad (2009). II Informe Anual del Observatorio Estatal de Violencia sobre la Mujer 2009. Recuperado el 02 de Enero de 2010, de <http://www.migualdad.es/ss/Satellite?blobcol=urldata&blobheader=application%2>

[Fpdf&blobheadername1=Content-disposition&blobheadervalue1=inline&blobkey=id&blobtable=MungoBlobs&blobwhere=1244652571786&ssbinary=true](#)

- Novo, M. (2000). *Heurísticos y evidencia en la toma de decisiones en la sala de justicia*. Tesis Doctoral, Universidad de Santiago de Compostela.
- Novo, M., y Seijo, D. (2010). Judicial judgement-making and legal criteria of testimonial credibility. *The European Journal of Psychology Applied to Legal Context*, 2, 91-115.
- Parker, A.D., y Brown, J. (2000). Detection of deception: Statement Validity Analysis as a means of determining truthfulness or falsity of rape allegations. *Legal and Criminological Psychology*, 5, 237-259.
- Pezdek, K., Morrow, A., Blandon-Gitlin, I., Goodman, G.S., Quas, J.A., Saywitz, K. et al. (2004). Detecting deception in children: Event familiarity affects criterion-based content analysis ratings. *Journal of Applied Psychology*, 89, 119-126.
- 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.
- Sierra, J.C., Jiménez, E.M., y Bunce, D. (2006). Técnicas de evaluación en psicología jurídica y forense. En J.C. Sierra, E.M. Jiménez, y G. Buela (Eds.), *Psicología forense: Manual de técnicas y aplicaciones* (pp. 485-525). Madrid: Biblioteca Nueva.
- Spörer, S.L. (1997). The less travelled road to truth: Verbal cues in deception detection in accounts of fabricated and self-experienced events. *Applied Cognitive Psychology*, 11, 373-397.
- Spörer, S.L. (2008). Lessons from the origins of eyewitness testimony research in Europa. *Applied Cognitive Psychology*, 22, 737-757.
- Spörer, S.L., y Schwandt, B. (2007). Moderators of nonverbal indicators of deception. A meta-analytic synthesis. *Psychology, Public Policy, and Law*, 13, 421-446.
- Steller, M. (1989). Recent developments in statement analysis. En J.C. Yuille (Ed.), *Credibility assessment*. (pp. 135-154). Dordrecht: Kluwer Academic Publishers.
- Steller, M., y Boychuck, T. (1992). Children as witness in sexual abuse cases: Investigative interview and assessment techniques. En H. Dent y R. Flin (Eds.), *Children as witness*. Chichester: Wiley and Sons.

- Steller, M., y Köhnken, G. (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).
- Steller, M., Raskin, D.C., Yuille, J.C., y Esplin, P. (1990). *Child sexual abuse: Forensic interviews and assessment*. Nueva York, NY: Springer.
- Szewczyk, H. (1973). Kriterien der beurteilung kindlicher zeugenaussagen [Criterios de evaluación de los testimonios de los niños]. *Probleme und Ergebnisse der Psychologie*, 46, 46-66.
- Trankell, A. (1972). *Reliability of evidence. Methods for analyzing and assessing witness statements*. Stockholm: Beckman (Orig. 1963).
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327-352.
- Tye, M.J.C., Amato, S.L., Honts, C.R., Deveitt, M.K., y Peters, D. (1999). The willingness of children to lie and the assessment of credibility in an ecologically relevant laboratory setting. *Applied Developmental Science*, 3, 92-109.
- Undeutsch, U. (1967). Beurteilung der glaubhaftigkeit von zeugenaussagen [Evaluación de la credibilidad de los testigos]. En U. Undetsch (Ed.), *Handbuch der psychology: Vol. II. Forensische psychologie* (pp. 26-181). Göttingen: Verlag für Psychologie.
- Undeutsch, U. (1988). The development of statement reality analysis. En J. Yuille (Ed.), *Credibility assessment* (pp. 101-119). Dordrecht: Kluwer Academic Publishers.
- Vilariño, M. (2010). *¿Es posible discriminar declaraciones reales de imaginadas y huella psíquica real de simulada en casos de violencia de género?* Servizo de Publicacións e Intercambio Científico de la Universidade de Santiago de Compostela.
- Vilariño, M., Fariña, F., y Arce, R. (2009). Discriminating real victims from feigners of psychological injury in gender violence: Validating a protocol for forensic setting. *The European Journal of Psychology Applied to Legal Context*, 1, 221-243.
- Vivero, A. (2006). *Contraste de declaraciones reales e imaginadas en formato de recuerdo libre*. Trabajo Tutelado de tercer Ciclo, Universidad de Santiago de Compostela.
- Vrij, A. (2005). Criteria-based content analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy and Law*, 11, 3-41.
- Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. Chischester: John Wiley and Sons.

- Vrij, A., Akehurst, L., Soukara, S., y Bull, R. (2002). Will the truth come out? The effect of deception, age, status, coaching and social skills on CBCA scores. *Law and Human Behavior*, 26, 261-283.
- Vrij, A., Akehurst, L., Soukara, S., y Bull, R. (2004). Detecting deceit via analyses of verbal and nonverbal behavior in children and adults. *Human Communication Research*, 30, 8-41.
- Vrij, A., Mann, S., y Fisher, R.P. (2006). An empirical test of the Behaviour Analysis Interview. *Law and Human Behavior*, 30, 329-345.
- Weick, K.E. (1985). Systematic observational methods. En G. Lindzey y E. Aronson (Eds.), *The handbook of social psychology* (Vol. 1, pp. 567-634). Hillsdale, NJ: LEA.
- Wicker, A.W. (1975). An application of a multiple-trait-multimethod logic to the reliability of observational records. *Personality and Social Psychology Bulletin*, 4, 575-579.
- Yuille, J.C. (1988). The systematic assessment of children's testimony. *Canadian Psychology*, 29, 247-262.