



REVISTA IBEROAMERICANA DE PSICOLOGÍA Y SALUD

Revista Oficial de la Federación Iberoamericana de Asociaciones de Psicología (FIAP)
[Official Journal of the Latin-American Federation of Psychological Associations]

www.elsevier.es/rips



Daño psicológico en casos de víctimas de violencia de género: estudio comparativo de las evaluaciones forenses



Ramón Arce^{a,*}, Francisca Fariña^b y Manuel Vilariño^c

^a Departamento de Psicología Organizacional, Jurídica-Forense y Metodología, Universidad de Santiago de Compostela, Santiago de Compostela, La Coruña, España

^b Departamento de Análisis e Intervención Psicosocioeducativa, Universidad de Vigo, Vigo, Pontevedra, España

^c Unidad de Psicología Forense, Universidad de Santiago de Compostela, Santiago de Compostela, La Coruña, España

Recibido el 19 de noviembre de 2014; aceptado el 13 de abril de 2015

Disponible en Internet el 14 de mayo de 2015

PALABRAS CLAVE

Simulación;
Violencia de género;
Trastorno por estrés
postraumático;
Daño psicológico;
Evaluación forense

Resumen Se diseñó un estudio con el objetivo de comparar la eficacia de la evaluación forense basada en una medida psicométrica con la basada en el análisis de contenido de la entrevista del daño psicológico en casos de violencia de género. Para este cometido, la evaluación forense no solo ha de proporcionar una evaluación del daño psicológico, sino también establecer una relación causa-efecto entre hechos enjuiciados y daño, así como un diagnóstico diferencial de simulación. Para ello se solicitó a 101 mujeres, mentalmente sanas y sin antecedentes de violencia de género, que simularan daño psicológico derivado de la vivencia de violencia de género en el MMPI-2 y en una entrevista narrativa en recuerdo libre, la entrevista clínico-forense de Arce y Fariña (2001). Los resultados mostraron que las mujeres disponían de una capacidad muy elevada de simulación en el modelo psicométrico, el MMPI-2, del 79,6%. En cuanto al diagnóstico diferencial de simulación, si bien las escalas de validez clasificaron consistentemente de forma correcta como simuladoras al 80,2%, el 19,8% no fueron detectadas como tales. Como en la evaluación forense el error tipo II no es admisible (clasificación de un daño simulado como real), la evaluación psicométrica no es prueba suficiente de daño. El análisis de contenido de la entrevista clínico-forense puso de manifiesto que el 3% de las participantes lograron simular la huella psicológica, detectándose en el 97% narrativas no normativas, esto es, falsas. Si bien el análisis de contenido de la entrevista clínico-forense dificulta y detecta más la simulación que la evaluación psicométrica, tampoco es prueba suficiente por sí misma, ya que no controla totalmente el error tipo II. Se discuten las implicaciones para la práctica forense de estos resultados.

© 2014 Sociedad Universitaria de Investigación en Psicología y Salud. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

* Autor para correspondencia.

Correo electrónico: ramon.arce@usc.es (R. Arce).

KEYWORDS

Malingering;
Intimate partner
violence;
Post-traumatic stress
disorder (PTSD);
Psychological injury;
Forensic assessment

Psychological injury in intimate partner violence cases: A contrastive analysis of forensic measures

Abstract With the aim of comparing in the forensic assessment of the psychosocial injury in intimate partner violence cases, the efficacy of the psychometric measure with the content analysis of the clinical interview, a study was designed. For this, forensic assessment not only must provide a diagnostic of the psychological injury, but also a differential diagnosis of malingering. 101 mentally healthy and with no-history of intimate partner violence (IPV) females answered, under malingering instructions of psychological injury consequence of IPV, to the MMPI-2, and were submitted to a free narrative interview, the forensic-clinical interview (Arce & Fariña, 2001). The results showed that participants had a very high capacity, 79.6%, to malingering in the psychometric measure, the MMPI-2. As for the differential diagnosis of malingering, the validity scales and configurations classified correctly as malingerers to the 80.2% of the protocols, failing in the remaining 19.8%. As in forensic assessment, the Type II Error (classification of malingered injury as real) is not admissible; the psychometric measure is not sufficient evidence of psychological injury. The content analysis of the forensic-clinical interview revealed that 3% of the participants were capable to malingering the psychological injury, while in 97% were detected non-normative narratives, that is, were classified as false. Although the content analysis of the interview makes malingering significantly more difficult and detects more malingering than the psychometric measure, fails to comply with full control of the Type II Error. Thus, the content analysis of the clinical-forensic interview is not in itself sufficient forensic evidence. The implications of the results for forensic practice are discussed.

© 2014 Sociedad Universitaria de Investigación en Psicología y Salud. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introducción

La etiqueta legal de víctima de un delito no solo supone que una persona haya sido objeto de una acción constitutiva de tal, sino que es necesario que como consecuencia haya sufrido algún daño (victimización), que se demuestre este y que se establezca una relación inequívoca entre el daño y la acción delictiva a enjuiciar. La victimización, acorde a la Declaración de los Principios Básicos de Justicia para Víctimas de Crímenes y Abuso de Poder (United Nations, 1985), se concreta en el sufrimiento de daño físico, psicológico (lesión mental o sufrimiento emocional), material o un menoscabo en sus derechos (vulneración de derechos fundamentales). La carga de la prueba (esto es, la demostración del daño y el establecimiento del nexo causal entre los hechos a enjuiciar y el daño) corresponde a la acusación, habiendo de superar el estándar de demostración «más allá de toda duda razonable». A este respecto, la jurisprudencia (p.ej., Sentencia del Tribunal Supremo de 29 de octubre de 1981; Sentencia 213/2002 de 14 de febrero de 2002) ha concretado que la condena comporta la plena seguridad, no la alta probabilidad, ya que la condena de un inocente representa una quiebra absoluta de los principios básicos de libertad, seguridad y justicia. Deductivamente, la plena seguridad se traslada a la prueba de acusación. En consecuencia, en la prueba de daño, el error tipo II (falsos positivos, es decir, clasificar como daño real un daño improcedente o simulado, ya que la prueba de daño se relaciona directamente con la culpabilidad de un inocente) ha de ser cero. En términos de evaluación del daño psicológico, esto supone que la técnica forense no solo ha de medir este y vincularlo con los hechos, sino también controlar totalmente la

simulación. A esta obligación judicial se une la sospecha clínica de simulación en las evaluaciones forenses (American Psychiatric Association, 2002, 2014) y el diagnóstico diferencial de simulación siempre en la evaluación del daño psicológico cuando haya implicaciones forenses (American Psychiatric Association, 2002). En suma, tanto judicial como clínicamente se requiere una evaluación de la simulación (diagnóstico diferencial) previa a la evaluación del daño psicológico.

La prueba del daño psicológico tiene un carácter transversal, pues este tipo de daño emana de todo tipo de violencia (sexual, física, vulneración de derechos fundamentales y, por supuesto, psicológica), siendo inexcusable cuando se juzga un caso de violencia psicológica, pues sin daño no hay víctima y, por extensión, caso. Esta contingencia convierte la prueba del daño psicológico en prueba continua (y única en violencia psicológica) en los casos de denuncia de violencia de género. La literatura ha identificado como daño psicológico, en función de los efectos en la salud mental del sufrimiento de una acción delictiva, el trastorno por estrés postraumático (TEP) (Kessler, Sonnega, Hughes y Nelson, 1995). Sin embargo, esta operacionalización fue tachada de limitada por Arce y Fariña (2009, 2014) ya que, cuando el estresor es de naturaleza psicosocial, tal como cuando la violencia es psicológica, procede, como huella, el trastorno adaptativo (TA) (American Psychiatric Association, 2014). El daño psicológico es comórbido con otros trastornos, estimándose la tasa de comorbilidad en el 92% (Brown, Campbell, Lehman, Grisham y Mancill, 2001). En violencia de género, los trastornos comórbidos son depresión, inadaptación social, ansiedad y disfunciones sexuales (Bargai, Ben-Shakhar y Shalev, 2007; Kessler et al., 1995).

Ahora bien, estos síndromes sin TEP/TA, así como los sub-síndromes, no son prueba suficiente de daño psicológico (O'Donnell, Creamer, Bryant, Schnyder y Shalev, 2006).

Para el doble propósito de medida del daño psicológico y diagnóstico diferencial de simulación, la evaluación clínica tradicional no es válida, pues no informa sistemáticamente de esta porque no la sospecha (Rogers, 2008). De hecho, el instrumento clínico de medida de referencia, la entrevista clínica (p. ej., First, Spitzer, Gibbon, Spitzer y Williams, 1999), no incluye mecanismos estandarizados de control de la simulación. Aún es más, la facilita al someter a los sujetos a una prueba de reconocimiento de síntomas (Resnick, West y Payne, 2008; Fariña, Arce, Vilariño y Novo, 2014). Como consecuencia de esta inadecuación de la entrevista clínica tradicional para el contexto forense, Arce y Fariña (2001) han propuesto y validado una entrevista para el contexto forense: la «entrevista clínico-forense». Esta somete a los evaluados a una tarea de conocimiento; no contamina la prueba con preguntas del entrevistador; posibilita el establecimiento de un nexo causal entre los hechos a enjuiciar y el daño psicológico, y contiene una herramienta estandarizada para el diagnóstico diferencial de simulación, un sistema categorial metódico de análisis de las estrategias de simulación. Estas son: no cooperación con la evaluación, síntomas sutiles, síntomas improbables/absurdos, síntomas obvios, síntomas (cuasi)raros, combinación de síntomas y patrones espurios de psicopatología, severidad de síntomas, inconsistencia de síntomas, estereotipos erróneos y agrupación indiscriminada de síntomas. Aunque esta entrevista ha sido validada científicamente y se ha mostrado altamente eficaz en el diagnóstico del daño psicológico y el diferencial de simulación (Arce, Fariña, Carballal y Novo, 2006; Arce, Fariña, Carballal y Novo, 2009; Arce, Pampillón y Fariña, 2002; Vilariño, Arce y Fariña, 2013; Vilariño, Fariña y Arce, 2009), no es totalmente válida para el campo forense, pues permite una tasa de error tipo II entre el 3 y el 5%, según el estresor causante del daño. Asimismo, aunque cumple los criterios establecidos por la Corte Suprema de Estados Unidos como prueba científica (Daubert vs. Merrell Dow Pharmaceuticals, 1993), no cuenta con el respaldo de un modelo teórico. De la gran amalgama de modelos explicativos, los que mejor se ajustan a las demandas de un sujeto simulador en una tarea de entrevista son los modelos narrativos, que asumen que los sujetos crean narraciones que tienen por objeto describir un evento (Novo y Arce, 2003). Los 2 modelos narrativos propuestos, «Anchored Narratives» (Wagenaar, 1995; Wagenaar, van Koppen y Crombag, 1993) y «Story Model» (Hastie, Penrod y Pennington, 1983), presentan como característica diferencial el modo de ajuste de la narrativa. En concreto, el *Story Model* prescribe la construcción de un evento en una tarea «paso a paso» que, en nuestro caso, se concretaría en: evaluación de la necesidad de demostrar el daño psicológico; aplicación de la metamemoria a la simulación (el sujeto reconoce que va a ser sometido a un análisis para la detección de simulación y se anticipa a él con técnicas de metamemoria), y concreción del daño en un evento narrativo (finalmente, el sujeto construye una historia narrativa de su estado mental ajustado al daño psicológico). Por su parte, el modelo *Anchored Narratives* no solo prevé la construcción de eventos, sino que estos los clasifica en normativos y no-normativos. Las narrativas no-normativas son aquellas que contienen inexactitudes, tal

como las que generan la impresión social de credibilidad a pesar de que no existe realidad de las mismas; narrativas claramente falsas; narrativas carentes de anclaje; narrativas erróneas. Por su parte, las narrativas normativas informan de la realidad. El ajuste de la entrevista clínico-forense al modelo de *Anchored Narratives* quedaría como sigue:

- a) Narrativas reales: la narrativa incluye el daño psicológico directo (TEP/TA).
- b) Narrativas falseadas (anómalas):
 - 1) Narrativas carentes de anclaje:
 - No cooperación con la evaluación.
 - Narrativa sin el anclaje esperado.
 - 2) Narrativas claramente falsas: inconsistencia interna en la narrativa.
 - 3) Narrativas erróneas: narrativa inválida.
 - 4) Narrativas que generan la impresión social de realidad, pero sin realidad en las mismas:
 - Síntomas sutiles.
 - Síntomas improbables/absurdos.
 - Síntomas obvios.
 - Síntomas (cuasi) raros.
 - Combinación de síntomas/patrones psicopatológicos espurios.
 - Severidad de síntomas.
 - Agrupación indiscriminada de síntomas.

Establecer la evaluación forense por medio de la entrevista no es prueba suficiente, pues no controla totalmente el error tipo II, requiriendo una aproximación multimétodo en la que se combina entrevista e instrumentación psicométrica, el MMPI-2 (Graham, 2011; Greene, 2011). Este instrumento evalúa el daño psicológico directo (TEP/TA) e indirecto, así como la simulación (escalas de control de la validez). En todo caso, la evaluación psicométrica no es prueba suficiente para la evaluación forense porque: a) no proporciona diagnósticos sino impresiones diagnósticas; b) no clasifica correctamente a todos los simuladores, error tipo II (Rogers, Sewell, Martin y Vitacco, 2003); c) el diagnóstico de simulación es compatible con la formulación de otras hipótesis alternativas (Graham, 2011; Greene, 2011); d) no permite establecer un nexo causal entre los hechos y daño, y e) constituye una tarea conducente a la simulación (Arce et al., 2002, 2006, 2009).

En este contexto nos planteamos un estudio con los objetivos de conocer la capacidad de simulación del daño psicológico consecuencia de la violencia en la entrevista clínico-forense y el MMPI-2; contrastar el ajuste de los modelos de narrativas a la tarea de la entrevista clínico-forense, y comparar la eficacia predictiva del modelo psicométrico con el narrativo de la entrevista.

Método

Participantes

La muestra la compusieron un total de 101 participantes, todas ellas mujeres mayores de 18 años, con convivencia actual o pasada en pareja (> 1 año), legas en psicología, mentalmente sanas y que manifestaron no haber sufrido

violencia de género (la muestra inicial era de 108 mujeres, eliminándose 7 por sospecha de haber sido objeto de actos de violencia psicológica). El 3% estaban viudas, el 41,58% estaban casadas y el 55,45% solteras (relación de pareja), con edades entre los 19 y los 61 años ($M = 32,4$; $DE = 12,5$). Por nivel de estudios, el 14,85% habían cursado estudios primarios; el 57,43%, bachillerato, y el 27,72%, universitarios.

Diseño y procedimiento

La metodología de investigación empleada fue del tipo cuasi-experimental en un contexto de simulación de alta fidelidad. Para alcanzar la alta fidelidad se ofreció una recompensa de tipo económico (150 euros) a las 4 mejores simulaciones. En concreto, se planificó un diseño para conocer la capacidad de simulación y la validez de los protocolos en el instrumento psicométrico de referencia para la evaluación de daño psicológico en la práctica forense (Resnick et al., 2008), el MMPI-2. Para ello se las instruyó en la simulación, «instrucciones de simulación», de daño psicológico consecuencia de la victimización de violencia de género. Estas consistían en solicitarles que se pusiesen en el lugar de una mujer que simulaba haber sufrido malos tratos en el ámbito doméstico y a la que se le van a valorar las secuelas psicológicas de los mismos. La denuncia falsa de malos tratos se justificaba con base en la obtención de algún beneficio asociado (p. ej., guarda y custodia de los hijos, resentimiento, venganza, compensaciones económicas). A su vez, se planificó un contraste de la ejecución de los sujetos en 2 formatos de tarea distintos: tarea de reconocimiento, evaluación psicométrica, y tarea de conocimiento, entrevista clínico-forense (Arce y Fariña, 2001).

En un primer momento, la tarea de las participantes consistió en responder al cuestionario MMPI-2 (Hathaway y McKinley, 1999) bajo las «instrucciones estándar» del propio instrumento. Fueron reclutadas para una evaluación clínica estándar. Los pases se produjeron individualmente. Al terminar de cumplimentar el MMPI-2, se les dijo que estábamos realizando un estudio sobre el daño psicológico en violencia de género y se les pidió la participación voluntaria en él. Una vez aceptada la participación (todas aceptaron), fueron informadas de que volverían a ser evaluadas transcurrida una semana. Para esta nueva medida se las instruyó en instrucciones de simulación. Se las alentaba para que en este intervalo de tiempo se formaran o entrenaran en lo que estimaran conveniente, al tiempo que se les informaba de la recompensa a las mejores simulaciones (150 euros). Una semana más tarde, las participantes fueron evaluadas por medio de la entrevista clínico-forense (Arce y Fariña, 2001), que fue grabada en vídeo para los subsecuentes análisis. Dos fueron los entrevistadores que recabaron todas las entrevistas, repartiéndose las mismas a partes alícuotas. Finalmente, en torno a una semana después, las participantes fueron evaluadas de nuevo a través del MMPI-2 bajo las instrucciones de simulación.

Instrumentos de medida

Como instrumento de medida en la tarea de reconocimiento utilizamos el MMPI-2 (Hathaway y McKinley, 1999), tomando como medidas clínicas las escalas clínicas básicas y las

2 escalas específicas de medida del TEP/TA: las escalas Pk y Ps. Dado que el objetivo de este estudio es obtener conclusiones para la práctica forense, y los forenses han de tomar el manual original del instrumento como fuente (obligación legal de objetividad y persistencia en los criterios), para el control de la validez de los protocolos (diagnóstico diferencial de simulación) recurrimos a las escalas de validez referidas en el mismo: escala de Interrogantes, escala K, escala F, escala Fb y escala L. Como modelos y criterios para la clasificación de la simulación de estas escalas se tomaron los del manual español del MMPI-2, ya que los forenses han de valerse de estos modelos y criterios. Estos son: las escalas de Interrogantes ($PD \geq 30$), F y Fb ($T > 80$) y K ($T < 50$) son indicativas de simulación, y las escalas L ($T \geq 80$) y K ($T > 70$), de disimulación.

En la tarea de conocimiento, las participantes fueron sometidas a una entrevista clínico-forense (Arce y Fariña, 2001), esto es, a una entrevista de orden clínico en formato de discurso libre que se ha mostrado fiable y productiva en el contexto forense (Arce et al., 2006, 2009; Vilariño et al., 2009, 2013). El procedimiento consistió en pedir a las participantes que relataran todo aquello que había cambiado en su vida (p. ej., síntomas, conductas, pensamientos) en relación a antes de haber sufrido malos tratos, o lo que había mejorado tras dejar de sufrirlos. Si no informaban motu proprio sobre las relaciones interpersonales, el contexto de trabajo o académico, relaciones familiares y, en su caso, de pareja, el entrevistador procedía a reinstaurar esos contextos específicos de evaluación (eje V del DSM-IV-TR; American Psychiatric Association, 2002).

Análisis de los protocolos

Las entrevistas, grabadas en vídeo, fueron sometidas a un análisis de contenido sistemático. El primer objetivo de este análisis de contenido era detectar criterios diagnósticos del daño psicológico tanto directo como indirecto. Para ello, tomamos como categorías de análisis los criterios diagnósticos del TEP, ansiedad generalizada, trastorno depresivo mayor, disfunciones sexuales e inadaptación social recogidos en el DSM-V (American Psychiatric Association, 2014). Así, confeccionamos un sistema categorial, fiable y válido.

Tras el análisis de contenido de la sintomatología clínica de cada protocolo, los codificadores evaluaron, en una segunda tarea, las narrativas. Como sistema categorial se tomaron las estrategias de simulación de la entrevista clínico forense, ajustado al modelo de las narrativas, siendo la unidad de análisis, el protocolo. Este sistema se había mostrado fiable, válido y productivo en la detección de la simulación (Arce et al., 2002, 2006, 2009; Vilariño et al., 2009, 2013). Las categorías y su ajuste al modelo, así como su definición, se muestran a continuación:

- a) Narrativa real: si, a tenor de los criterios detectados, se cumplían los criterios diagnósticos del TEP/TA (daño psicológico directo).
- b) Narrativas falsas (anómalas):
 - 1) Narrativas carentes de anclaje:
 - No cooperación con la evaluación. Se codificaba esta posibilidad cuando el sujeto no mostraba interés en

la evaluación o no respondía a la misma ([American Psychiatric Association, 2014](#)), de modo que no explicaba un evento narrativo.

- Narrativa sin el anclaje esperado: la narrativa no informaba del daño psicológico directo (TEP/TA).
- 2) Narrativas claramente falsas: inconsistencia entre síntomas observados y manifestados. Esta categoría tiene por objeto analizar la correspondencia entre los síntomas elicitados por el sujeto y las observaciones del codificador sobre si esos síntomas se corresponden con la actitud, presencia y comportamiento del sujeto.
 - 3) Narrativas erróneas: la narrativa se concreta en daño psicológico no esperado o en estereotipos erróneos (concepciones equivocadas de las características clínicas asociadas con los trastornos mentales).
 - 4) Narrativas que generan la impresión social de realidad, pero que no hay realidad en las mismas:
 - La narrativa informa de síntomas sutiles. Los síntomas sutiles no son síntomas reales, sino problemas cotidianos que se informan como síntomas asociados a una enfermedad mental (p.ej., ser desorganizado, falta de motivación, dificultades en la toma de decisiones ordinarias).
 - La narrativa se relaciona con síntomas improbables/absurdos: son síntomas improbables aquellos que tienen un carácter fantástico o ridículo (opiniones, actitudes o creencias extrañas) y que no gozan de referentes reales (esto es, se excluyen aquí los síntomas raros).
 - Síntomas obvios (muy accesibles a la simulación). Estos son síntomas que se relacionan fácilmente con el daño psicológico.
 - Síntomas (cuasi)raros (no frecuentes, propios de la locura). Esta contingencia se da cuando el sujeto dice padecer síntomas que raramente se dan incluso entre poblaciones psiquiátricas reales o muy infrecuentemente en poblaciones normativas.
 - Combinación de síntomas (síntomas que no se manifiestan juntos) o patrones psicopatológicos espurios. Este indicador de simulación se codifica cuando el participante informa de síntomas reales que difícilmente coocurren, sin consistencia interna entre ellos o configuraciones infrecuentes en poblaciones clínicas.
 - Severidad de síntomas. Esta categoría analiza el grado de severidad de los síntomas manifestados. Es frecuente que los simuladores atribuyan a la sintomatología que dicen padecer una severidad extrema.
 - Agrupación indiscriminada de síntomas. La narrativa incluye problemas psiquiátricos de forma indiscriminada (asume todo tipo de síntomas).

Entrenamiento de codificadores

En este estudio participaron 2 codificadores con conocimientos de evaluación psicopatológica y forense, entrenados en este y otros sistemas de codificación, y uno de ellos con experiencia previa en codificación de este tipo de material ([Arce et al., 2006, 2009](#)). El entrenamiento consistió, tras presentarles y ejemplificarles cada categoría de análisis, en la ejecución con material del estudio que se desechó, utilizando el índice de concordancia como instrumento de

cotejo de la ejecución, lo que permitía, constatada la inconsistencia, subsanar errores de codificación a través de la contrastación y homogeneización de criterios.

Las definiciones de las categorías de análisis del TEP/TA, trastorno depresivo mayor, ansiedad generalizada, disfunciones sexuales e inadaptación social se correspondían con las de los criterios diagnósticos del DSM-V. Por ello, los codificadores contaban en la codificación con dicho manual, así como con un material con ejemplos de cada categoría obtenido de otros estudios. Además, para la codificación de las narrativas se les proporcionó un manual de creación propia con la definición de las categorías de análisis y ejemplos de cada categoría.

Estudio de la fiabilidad de las entrevistas y codificaciones

Los entrevistadores, entrenados y con experiencia en la entrevista clínico-forense, obtuvieron unos protocolos similares, tanto en términos de la sintomatología observada ($F(1, 99) = 1,24$; ns, como en las narrativas informadas, $F(1, 99) = 1,67$; ns. En suma, las entrevistas no están contaminadas por el factor entrevistador.

Los protocolos clínicos procedentes de las entrevistas se dividieron en 2 mitades equivalentes, de modo que cada codificador examinó en torno al 50% de las entrevistas, que se distribuyeron de manera aleatoria entre los codificadores. De este modo, uno de ellos analizó 51 protocolos y el otro los restantes 50. Para calcular la consistencia intrajueces, los codificadores repitieron, una semana después del fin de la codificación original, 10 (en torno al 20%) de las entrevistas que ellos mismos habían evaluado. Asimismo, y para obtener la medida de la consistencia interjueces, cada codificador examinó otros 10 protocolos que inicialmente había registrado el otro codificador. La consistencia se computó a través de la correlación intraclase para las variables discretas y la kappa de Cohen para las categóricas. Como quiera que el recuento de contingencias en sí no garantiza la consistencia, se verificó en las variables discretas la exacta correspondencia entre los registros. Constatamos, tanto para los síntomas clínicos como para las narrativas, unas $r_s > 0,70$ y $k_s > 0,60$, por lo que podemos sostener que las evaluaciones son consistentes interevaluador e intraevaluador, en el tiempo e intercontextos. En suma, los datos son fiables ([Wicker, 1975](#)).

Resultados

Análisis previos

Previamente al análisis de datos del estudio diseñado, procedimos con un cribado de los protocolos y participantes a fin de eliminar aquellos casos que pudieran contaminar los resultados. Tres son las fuentes de contaminación. Primera, que las participantes hubieran sido objeto de algún tipo de violencia de género. Para ello, y al margen de que todas habían indicado que no habían sufrido violencia de género, los entrevistadores les preguntaron por diversos tipos de violencia psicológica, ya que esta pudiera no ser informada como tal. Segunda, que los perfiles de respuestas fueran altamente inconsistentes en la evaluación, por una

aquiescencia extrema (TRIN Pd > 18); por ser respondidos al azar (VRIN Pd > 18; |F-Fb| > 19); o por *outliers* (L Pd > 10, y K Pd > 26) (Graham, 2011; Greene, 2011). Además, todos los protocolos en condiciones de simulación resultaron consistentes (TRIN y VRIN < 13). Tercera, por falta de implicación o comprensión de la tarea, que se verificó por la eficacia en la tarea, de modo que se desecharían aquellos casos en que no crearan una narrativa de su estado mental con, al menos, un síntoma de la medida directa del daño psicológico, o no simularan un trastorno relacionado con el daño psicológico en el MMPI-2. Se desestimaron 7 casos por potencial victimización psicológica de violencia de género.

Estudio de la capacidad de simulación en el MMPI-2

El siguiente objetivo era conocer si las participantes, en general, eran capaces de simular el daño psicológico propio de los malos tratos, esto es, un TEP/TA y los trastornos comórbidos (medidas indirectas) en el MMPI-2. Así, tomando el punto de corte de 70 en puntuaciones típicas, referente a partir del cual los resultados dan una impresión diagnóstica de psicopatología (desviadas +2 DE de la normalidad), se extrae qué daño psicológico asocian a victimización de violencia de género. Los resultados (tabla 1) muestran que reconocen como daño psicológico propio de la violencia de género (superan significativamente o igualan el criterio de decisión) un cuadro comórbido compuesto por hipocondriasis, depresión, histeria, desviación psicopática, paranoia, psicastenia, esquizofrenia, introversión social y TEP (escalas Pk y Ps). Complementariamente, discriminan como impropios de los malos tratos: hipomanía (trastorno bipolar, hipomaniaco).

Estos resultados ponen de manifiesto que potenciales víctimas de violencia de género disponen de las destrezas suficientes para simular el daño psicológico de los malos tratos (p. ej., TEP/TA, depresión, psicastenia, introversión social) y para discriminarlo efectivamente de otros trastornos no relacionados (p. ej., hipomanía).

Por su parte, el efecto de la manipulación de las respuestas (d = contraste de la media y desviación estándar empírica, con la media y desviación estándar de la

distribución de las puntuaciones T; $M = 50$; $DE = 10$) está próximo a las 3 desviaciones típicas en las medidas directas del daño psicológico (escalas Pk y Ps) y cercano o superando las 2 en las indirectas (p. ej., introversión social, depresión, psicastenia). En términos prácticos (BESD; Rosenthal y Rubin, 1982), la capacidad de simulación de daño psicológico directo derivado de violencia de género sería con total consistencia entre Pk y Ps, del 79,6%, y con una tasa máxima del 82,4% en Pk.

Estudio de la simulación en las escalas de control de validez del MMPI-2

Dado que la capacidad de simulación en la medida psicométrica es muy elevada, el diagnóstico diferencial de simulación, que descansa en las escalas de validez, es crítico. A este respecto, las escalas de validez (tabla 2) discriminan entre respuestas honestas y simuladas, aunque la escala No Respuestas lo hace en dirección contraria al modelo, es decir, la media de no respuestas es significativamente superior en la condición de respuestas honestas. Las escalas F y Fb, en las que puntuaciones elevadas se relacionan con simulación, y K, en la que puntuaciones bajas son indicativas de simulación, no solo diferencian en consonancia con el modelo entre protocolos honestos y simulados, sino que los tamaños del efecto son más que elevados (más de 3 desviaciones estándar para las escalas F y Fb, y 1,5 desviaciones para K). Por su parte, la escala L disminuye significativamente en su valor en las respuestas simuladas (valores alto en L indican disimulación, que es contrario a la simulación), lo que le confiere un potencial valor de criterio de no simulación. En el estudio de casos, las 3 medidas de simulación evidencian una probabilidad de superioridad muy elevada (0,85 y 0,99) (McGraw y Wong, 1992; la probabilidad de que la puntuación en la escala sea superior en condiciones de respuesta simulada frente a la honesta) y una sensibilidad elevada (> 70% clasificación correcta de la simulación), pero también con una elevada tasa de falsos positivos, inadmisibles en la evaluación forense. Como estrategia adicional de análisis de las escalas de validez, en el manual del MMPI-2

Tabla 1 t para una muestra con el valor de prueba de T = 70. Respuestas simuladas

Escalas clínicas	t	p	M	Sx	d
Hipocondría	8,14	0,000	81,08	13,68	2,59
Depresión	5,73	0,000	77,21	12,65	2,39
Histeria (somatización)	3,55	0,001	73,99	11,31	2,24
Desviación psicopática	3,02	0,003	73,00	9,99	2,30
Paranoia	9,14	0,000	85,09	16,59	2,56
Psicastenia (ansiedad)	9,14	0,000	74,75	11,27	2,32
Esquizofrenia	11,09	0,000	88,80	17,04	2,77
Hipomanía	-6,76	0,000	63,61	9,49	1,40
Introversión social	1,31	0,193	71,56	12,01	1,95
Escala Pk	9,62	0,000	80,39	10,85	2,91
Escala Ps	7,96	0,000	78,63	10,90	2,74

Nota: $gl(100)$.

El tamaño del efecto de la manipulación de la respuesta se calculó con la media y desviación estándar de las puntuaciones T ($M = 50$, $DE = 10$).

Tabla 2 Efecto de las instrucciones en las escalas de control del MMPI-2. Pruebas de los efectos intrasujetos

Escalas de control de validez	F	p	M _{ie}	M _{is}	d	PS	Se
Escala ? (No Respuestas)	7,49	0,007	1,45	0,67	0,55	0,65	0,0
Escala L (Mentira)	5,2	0,025	53,03	50,22	0,45	0,62	0,0
Escala F (Incoherencia)	294,48	0,000	50,26	94,18	3,44	0,99	72,3
Escala K (Factor Corrector)	56,75	0,000	49,42	41,04	1,51	0,85	89,1
Fb Posterior	376,12	0,000	50,14	96,8	3,88	0,99	78,2

Nota: $g(1, 100)$.

M_{ie}: media de la condición de instrucciones estándar; M_{is}: media de la condición de instrucciones de simulación; PS: probabilidad de superioridad; Se: sensibilidad.

(Hathaway y McKinley, 1999, p. 42) se sugiere examinar las configuraciones de las anteriores escalas, mencionando explícitamente el índice de Gough (F-K). Sin embargo, no se aporta criterio de decisión alguno, remitiendo a fuentes externas. A este respecto, la referencia más potente es el metaanálisis de Rogers et al. (2003), en el que se advierte de una alta variabilidad en los criterios, tomando como normativo una PD > +18. Si bien este índice discrimina significativamente, $F(1, 100) = 234,68$, $p < 0,001$, y con un tamaño del efecto más que grande, $d = 3,06$, entre respuestas simuladas y honestas, no es eficiente en la detección de la simulación ($M = 19,34$), $t(100) = 1,04$; ns, $Se = 56,4$. En otras palabras, el índice F-K no informa sistemáticamente que las participantes en condiciones de respuesta simulada hayan seguido la estrategia general de dar una mala imagen. Una estrategia adicional derivada de las configuraciones de validez sería la combinación de estas escalas. A partir de esta, hallamos que el 6,9% de los protocolos de respuestas simuladas no son detectados por ninguna escala, el 12,9% por una única escala, el 13,95 por 2 y el 66,3% por 3. Sucintamente, como cualquier medida requiere de una estimación de la consistencia (también criterio legal de persistencia; Novo y Seijo, 2010), para concluir que el protocolo es simulado al menos 2 escalas deberían clasificarlo como tal. De este modo, el 80,2% de las respuestas simuladas serían clasificadas consistentemente como tales, en tanto el 19,8% no.

Estudio de capacidad y detección de la simulación en los eventos narrativos de la entrevista clínico-forense

El estudio de casos (tabla 3) advierte que el 3% de las participantes fueron capaces de crear una narrativa real, esto es, de simular efectivamente la sintomatología en la medida directa (y complementariamente, medidas indirectas) vinculada con las secuelas psicológicas de la victimización de violencia de género. En suma, la accesibilidad de la simulación en una narrativa (tarea de conocimiento) es igual al margen de error estadísticamente aceptable (0,05), $Z(101) = 0,95$; ns, $h = 0,171$. Complementariamente, si bien todas las participantes en condiciones de simulación crearon una narrativa de su estado mental, en el 97% crearon una narrativa con un anclaje erróneo (no se halló el daño psicológico esperado). Del total de narrativas, el 4% eran claramente falsas por *inconsistencia de síntomas* (se observa una inconsistencia entre lo manifestado en la narrativa y

lo observado). Por su parte, en el 19,8% se registró daño en la salud mental, pero que era errónea al circunscribirse a daño no esperado. Finalmente, se registraron narrativas que producen la impresión social de realidad (pero que no constituyen daño psicológico real) concretadas en las categorías «síntomas obvios» (47,5%) (referido a síntomas que se relacionan fácilmente con el daño psicológico, sin que se confirme el diagnóstico clínico de TEP/TA) y «severidad de síntomas» (21,7%) (atribución sistemática de una severidad extrema a los síntomas padecidos). En resumen, estos datos prestan apoyo al modelo *Anchored Narratives* en el campo forense, dado que las participantes crearon una narrativa de su estado mental (productividad), y por la eficacia discriminativa de las narrativas entre realidad y simulación (eficacia). Además, se infiere que siguieron los pasos prescritos en el *Story Model* en la creación de la narrativa: evaluación de la necesidad de demostrar el daño psicológico (todas las participantes intentaron simular daño); aplicación de la metamemoria a la simulación (el uso de estrategias de simulación advierte de indicios deliberados de ocultarla), y concreción del daño en un evento narrativo (todas las participantes construyeron un evento narrativo del daño sufrido en la salud mental con síntomas propios del daño psicológico).

Tabla 3 Tabla de contingencia de narrativas

Narrativa	P _o
<i>Reales</i>	0,03
<i>Falseadas (anómalas)</i>	0,970
<i>Narrativas carentes de anclaje</i>	
Carentes de anclaje (errónea)	0,970
Carente de anclaje (sin evento)	0
<i>Narrativas erróneas (daño no esperado)</i>	0,198
<i>Narrativas claramente falsas (inconsistencia observado-manifestado)</i>	0,040
<i>Impresión social realidad</i>	
Síntomas obvios	0,475
Severidad síntomas	0,217

Nota: $n = 101$.

p_o: proporción observada; las restantes categorías resultaron improductivas.

Contraste de la efectividad de la evaluación en tarea de reconocimiento (psicométrica) y conocimiento (entrevista)

La tasa de simulación es significativamente mayor, $\chi^2(1, n = 101) = 118,13, p < 0,001$, en la medida del daño psicológico en tarea de reconocimiento (evaluación psicométrica, modelo psicométrico) (0,796) que en tarea de conocimiento (entrevista, modelo de narrativas) (0,030). Igualmente, la magnitud del efecto, $\phi = 0,775$, es elevada. En consecuencia, una evaluación psicométrica facilita, frente a la entrevista narrativa, un 77,5% la tarea de simulación al sujeto no honesto. Adicionalmente, la técnica psicométrica clasifica correctamente de forma consistente (al menos 2 medidas) el 80,2% de las simulaciones, en tanto la técnica de entrevista (modelo de narrativas), el 97%, resultando este más productivo, $\chi^2(1) = 12,56, p < 0,001, \phi = 0,265$, manifestándose en una tasa de eficacia superior en la clasificación de la simulación del 26,5%.

Discusión

Los anteriores resultados están sujetos a limitaciones en cuanto a su alcance, que es preciso tener presente. Primera, las participantes cumplieron una tarea de simulación, con lo que la implicación en la tarea es presumiblemente menor que en casos reales (Konecni y Ebbesen, 1992) y los resultados han de generalizarse a contextos reales con ciertas reservas (Fariña, Arce y Real, 1994). Segunda, los resultados de este estudio no pueden generalizarse a otros instrumentos de medida. Tercera, las participantes se circunscriben a personas de la población normativa en condiciones de respuesta honesta y simulada, no incluyendo víctimas reales, por lo que se desconocen los efectos en los falsos negativos (p. ej., casos reales clasificados como simulados).

Con estas salvaguardas en mente, de los anteriores resultados se pueden derivar las siguientes conclusiones:

- La evaluación psicométrica del daño psicológico derivado de la vivencia de violencia de género facilita la simulación efectiva de daño psicológico hasta tasas en torno al 80%. Además, las respuestas en esta medida son consistentes (TRIN y VRIN < 13). Sucintamente, la capacidad de simulación del daño psicológico es plena en la medida psicométrica: fingimiento de forma consistente de la huella psicológica. A su vez, también facilita la discriminación de la secuela psicológica de otros daños en la salud mental no relacionados con esta, esto es, inoperativiza la estrategia identificada de simulación, la agrupación indiscriminada de síntomas (asunción de todo tipo de síntomas).
- En cuanto al control de los falsos positivos (diagnóstico diferencial de simulación; error tipo II), las escalas de validez son estadísticamente significativas y efectivas, pero cometen un falso positivo en torno a 1 de cada 5 evaluaciones de respuestas simuladas, cuando en la práctica forense no es admisible ningún falso positivo. En resumen, los indicadores del diagnóstico diferencial de simulación del MMPI-2 son fiables y válidos, pero insuficientes para la tarea forense.

- Los sujetos en condiciones de simulación de daño psicológico ajustan totalmente (100%) su proceder a los modelos de narrativas, creando un evento narrativo de los daños en la salud mental e intentando anticiparse a la detección de la simulación. A su vez, los resultados no prestan apoyo alguno al supuesto del modelo clínico de la simulación (American Psychiatric Association, 2014) que la relaciona con la falta de cooperación en la evaluación (todas cooperaron) y un apoyo marginal (3,9%) al postulado de falta de consistencia entre lo observado y lo manifestado (narrativas claramente falsas). Asimismo, el modelo de narrativas es muy potente en la clasificación de la simulación (97% de narrativas falseadas). Los resultados del análisis de contenido de las narrativas en la entrevista (p. ej., todas crearon narrativas; el 97% de las simuladas fueron detectadas como anómalas/carentes de anclaje; el 19,8% como erróneas; el 4% como claramente falsas; el 69,2% como generadoras de impresión social de realidad, pero irreales) prestan apoyo a la validez del modelo de narrativas como soporte teórico para la entrevista clínico-forense. Sin embargo, esta no es totalmente válida por sí misma para la tarea forense, pues deja paso a un 3% de falsos positivos, cuando dicha tasa en el campo forense ha de ser cero.
- Comparativamente, la técnica psicométrica facilita el acceso a la simulación, en tanto la de entrevista con análisis de las narrativas es superior a la psicométrica en la clasificación de la simulación. Aun así, la técnica de entrevista con análisis de las narrativas es insuficiente por sí misma en la correcta y total clasificación de la simulación. Por ello, la tarea forense requiere la adopción de una estrategia multimétodo (Graham, 2011; Greene, 2011) en la que se combinen ambas técnicas (evaluación psicométrica y entrevista) y de la concreción de un criterio de decisión que anule totalmente la posibilidad de falsos positivos.

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Referencias

- American Psychiatric Association. (2002). *DSM-IV-TR. Manual diagnóstico y estadístico de los trastornos mentales*. Barcelona, España: Masson.
- American Psychiatric Association. (2014). *Manual diagnóstico y estadístico de los trastornos mentales (DSM-V)*. Madrid, España: Panamericana.
- Arce, R. y Fariña, F. (2001). *Construcción y validación de un procedimiento basado en una tarea de conocimiento para la medida de la huella psíquica en víctimas de delitos: la entrevista forense [manuscrito inédito]*. Universidad de Santiago de Compostela.
- 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. In F. Fariña, R. Arce, & G. Buela-Casal (Eds.), *Violencia de género. Tratado psicológico y legal* (pp. 147-168). Madrid, España: Biblioteca Nueva.
- Arce, R. y Fariña, F. (2014). Evaluación psicológico-forense de la credibilidad y daño psíquico mediante el Sistema de Evaluación Global. In P. Rivas y G. L. Barrios (Eds.), *Violencia de género:*

- perspectiva multidisciplinar y práctica forense* (pp. 411–441). Pamplona, Navarra, España: Thomson Aranzadi.
- Arce, R., Pampillón, M. C. y Fariña, F. (2002). Desarrollo y evaluación de un procedimiento empírico para la detección de la simulación de enajenación mental en el contexto legal. *Anuario de Psicología*, 33, 385–408.
- 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., Carballal, A. y Novo, M. (2009). Creación y validación de un protocolo de evaluación forense de las secuelas psicológicas de la violencia de género. *Psicothema*, 21, 241–247.
- Bargai, N., Ben-Shakhar, G. y Shalev, A. Y. (2007). Posttraumatic stress disorder and depression in battered women: The mediating role of learned helplessness. *Journal of Family Violence*, 22, 267–275.
- Brown, T. A., Campbell, L. A., Lehman, C. L., Grisham, J. R. y Mancill, R. B. (2001). Current and lifetime comorbidity of the DSM-IV anxiety and mood disorders in a large clinical sample. *Journal of Abnormal Psychology*, 110, 585–599.
- Daubert vs. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993).
- Fariña, F., Arce, R. y Real, S. (1994). Ruedas de identificación: de la simulación y la realidad. *Psicothema*, 7, 395–402.
- Fariña, F., Arce, R., Vilariño, M. y Novo, M. (2014). Assessment of the standard forensic procedure for the evaluation of psychological injury in intimate-partner violence. *The Spanish Journal of Psychology*, 17(e32), 1–10.
- First, M. B., Spitzer, R. L., Gibbon, M., Spitzer, R. L. y Williams, J. B. W. (1999). *Guía del usuario para la entrevista clínica estructurada para los trastornos del eje I del DSM-IV: SCID-I*. Barcelona, España: Masson.
- Graham, J. R. (2011). *MMPI-2: Assessing personality and psychopathology* (5th ed.). New York, NY: Oxford University Press.
- Greene, R. L. (2011). *The MMPI-2/MMPI-2-RF: An interpretive manual* (3rd ed.). Boston, MA: Allyn & Bacon.
- Hastie, R., Penrod, S. y Pennington, N. (1983). *Inside the Jury*. Cambridge, Mass: Harvard University Press.
- Hathaway, S. R. y McKinley, J. C. (1999). *Inventario multifásico de personalidad de Minnesota-2. Manual*. Madrid, España: TEA.
- Kessler, R. C., Sonnega, A., Hughes, M. y Nelson, C. B. (1995). Posttraumatic stress disorder in the national comorbidity survey. *Archives of General Psychiatry*, 52, 1048–1060.
- Konecni, V. J. y Ebbesen, E. B. (1992). Methodological issues in research on legal decision-making, with special reference to experimental simulations. In F. Lösel, D. Bender, & T. Blieseener (Eds.), *Psychology and law. International perspectives* (pp. 413–423). Berlín, Alemania: Walter de Gruyter.
- McGraw, K. O. y Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin*, 111, 361–365.
- Novo, M. y Arce, R. (2003). *Jueces: formación de juicios y sentencias*. Granada, España: GEU.
- Novo, M. y Seijo, D. (2010). Judicial judgement-making and legal criteria of testimonial credibility. *European Journal of Psychology Applied to Legal Context*, 2, 91–115.
- O'Donnell, M. L., Creamer, M., Bryant, R. A., Schnyder, U. y Shalev, A. (2006). Posttraumatic disorders following injury: Assessment and other methodological considerations. In G. Young, A. W. Kane, & K. Nicholson (Eds.), *Psychological Knowledge in Courts. PTSD, Pain and TBI* (pp. 70–84). New York, NY: Springer.
- Resnick, P. J., West, S. y Payne, J. W. (2008). Malingering of post-traumatic disorders. In R. Rogers (Ed.), *Clinical Assessment of Malingering and Deception* (3rd ed., pp. 109–127). New York, NY: The Guilford Press.
- Rogers, R. (2008). Structured interview and dissimulation. In R. Rogers (Ed.), *Clinical Assessment of Malingering and Deception* (3rd ed., pp. 301–322). New York, NY: The Guilford Press.
- Rogers, R., Sewell, K. W., Martin, M. A. y Vitacco, M. J. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI-2 and malingering. *Assessment*, 10, 160–177.
- Rosenthal, R. y Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, 74, 166–169.
- United Nations (1985). Declaration of Basic Principles of Justice for Victims of Crime and Abuse of Power (General Assembly resolution 40/34 of 29 November 1985). Viena, Austria: United Nations. Recuperado de <http://www.un.org/documents/ga/res/40/a40r034.htm>
- Vilariño, M., Fariña, F. y Arce, R. (2009). Discriminating real victims from feigners of psychological injury in gender violence: Validating a protocol for forensic settings. *European Journal of Psychology Applied to Legal Context*, 1, 221–243.
- Vilariño, M., Arce, R. y Fariña, F. (2013). Forensic-clinical interview: Reliability and validity for the evaluation of psychological injury. *European Journal of Psychology Applied to Legal Context*, 5, 1–21.
- Wagenaar, W. A., van Koppen, P. y Crombag, H. F. M. (1993). *Anchored Narratives. The Psychology of Criminal Evidence*. Hemel Hempstead, UK: Harvester Wheatsheaf.
- Wagenaar, W. A. (1995). Anchored Narratives: A theory of judicial reasoning. In G. Davies, S. Lloyd-Bostock, M. McMurrain, & C. Wilson (Eds.), *Psychology, Law and Criminal Justice* (pp. 267–285). Berlín, Alemania: Walter de Gruyter.
- 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.