Facultad de Psicología

Departamento de Psicología Social

Programa de Doctorado: Investigación en Psicología

**NUEVOS ACERCAMIENTOS METODOLÓGICOS AL ESTUDIO DE LA
VIOLENCIA DE GÉNERO**

**NEW METHODOLOGICAL APPROACHES TO THE STUDY OF INTIMATE
PARTNER VIOLENCE AGAINST WOMEN**

TESIS DOCTORAL

Manuel Martín-Fernández

Directores: Enrique Gracia Fuster y Marisol Lila Murillo

Valencia, 2019

Night, the world, it's mine, with nobody else out here
It's time, run wild and royally cavalier
To burn, ignite, I'd do it for so much less
When all is made clear there is nothing else
And here we stand, the sweet arresting duality
And I come to, it's resin all over me
Will I awake, will I get a ride with you
In this race of two

We're an endless stream of choices
We're the softest murmur of voices

Without names we're fantasizing
Dancing like flames, mesmerizing
My dark disquiet playing such eerie harmonies
Making waves and diving under
Lightning to the sound of thunder
My dark disquiet singing such haunting melodies

So white, so still, so bright, it's almost too painful now
I'm ready to fight, to run from the light
And here now comes the sweet corrupting reality
While now I'm free, will I once cease to be?
Will I awake, will I get a ride with you
In this race of two?

We are momentary masters
We're false kings and bastards

Without names we're fantasizing
Dancing like flames, mesmerizing
My dark disquiet playing such eerie harmonies
Making waves and diving under
Lightning to the sound of thunder
My dark disquiet singing such haunting melodies

We are marionettes by strings animated
Yet like lovers of strings liberated
We are marionettes by strings animated
Yet like lovers of strings liberated
Without names we're fantasizing

Brothers by blood separated
Marionettes animated
Lovers of strings liberated

—Poets of the Fall, "My Dark Disquiet"

Agradecimientos

Escribo estas líneas preso de una gran emoción, pues éstas constituyen el inicio del fin de una etapa. Y es que una tesis doctoral, aunque técnicamente no es más que el producto, el manuscrito que sostiene ahora entre sus manos, para el autor de estas líneas ha supuesto más. Mucho más. Y dado que esta es la única sección del manuscrito donde puedo permitirme que impere la lírica a la claridad expositiva, le advierto que estas líneas van a ocupar una extensión considerable. Por suerte, las reflexiones aquí vertidas no son más que una digresión, y si su interés está más centrado en el producto técnico, no tiene más que saltarse los agradecimientos y pasar directamente al prefacio.

En primer lugar, esta tesis ha supuesto la culminación de un proceso académico largo y no exento de vicisitudes, que comenzó con un joven entusiasta cuyo principal objetivo era comprender mejor las motivaciones y acciones de esos entes tan ajenos que lo rodeaban: los demás. Esta primera aproximación ingenua a la psicología fue mutando hasta convertirse en un amor ciego, hasta fanático, por la ciencia y por el método científico. Y es que el futuro de nuestra disciplina y de nuestros modelos teóricos sólo puede pasar por unos diseños experimentales claros, que permitan responder a las preguntas relevantes que nos formulemos los psicólogos, pasando por una recogida de datos sólida y un tratamiento estadístico de los mismos asimismo apropiado y riguroso.

Ha supuesto, en segundo lugar, toda una travesía vital de aprendizaje y crecimiento personal, y por qué no, también de crecimiento interpersonal. Un viaje que no hubiera podido recorrer en solitario y por el que estoy inmensamente agradecido a todas aquellas personas, mentores, mentoras, colegas, compañeros y compañeras, amigos, amigas, y familiares que me han ayudado y apoyado durante el camino.

La primera de estas personas es Enrique, el director de esta tesis, el capitán del navío de este proyecto sin quien, sencillamente, nada de todo esto hubiera sido posible. Enrique, gracias por acogerme con los brazos abiertos entre los miembros de la tripulación de tu equipo de investigación, a pesar de que en nuestro primer intercambio de correos te cambiase el apellido —algo que ahora bien sé que detestas—, y a pesar de llegar con dos horas de retraso a nuestra primera reunión en persona. Gracias por haber apostado por mí y haberme apoyado tanto en los momentos de alegría, como en los de amargura. Gracias por haberme iniciado en los principios del noble arte de la venta de cabras, por enseñarme a tocar de oído esa melodía endiablada que suponen los artículos en inglés, y por recibir todas mis locas ideas con la mente abierta. Gracias también por todos esos temidos y a la vez tan atinados azules, que tantos quebraderos de cabeza me han llegado a dar, y por todas las discusiones y reflexiones distendidas, cervezas y tintos de verano mediante, sobre nuestra propia práctica investigadora. Ha sido un auténtico placer trabajar contigo, codo con codo, durante esta etapa.

La segunda persona, y desde luego no por ello menos relevante, es Marisol, la codirectora de esta tesis. Porque todo navío precisa de una contramaestre hábil y audaz, capaz de coordinar a todos los miembros de la tripulación que es Contexto y el grupo Viostrategy para que la nave llegue a buen puerto. Sin tu guía, calidad humana y supervisión nada de esto hubiera sido posible. Gracias por acogerme entre los miembros del equipo y haberme hecho sentir uno más, como si llevase toda la vida en la Universidad de Valencia. Gracias por tu apoyo y tus ánimos constantes en lo académico y en lo personal. Gracias por enseñarme a domar a las fieras y hacer de las clases algo más que una mera exposición irreflexiva de contenidos teóricos. Mil gracias por tu apoyo y guía a la hora de construir las temidas introducciones, por mostrarme la relevancia y el impacto social que pueden tener a llegar las

labores que llevamos a cabo desde nuestros despachos —o casas—, no quedándose aisladas en el mundo académico. Ha sido un privilegio y un placer trabajar a tu lado durante la tesis.

No puedo olvidarme de mis compañeros predoctorales, algunas ya doctoras y otros a puntos de serlo. Me gustaría agradecer, más allá de la mera cortesía profesional, a Miriam, Vivi y Faraj todas sus contribuciones a la presente tesis doctoral. Porque sin ellos, los estudios que componen esta tesis no hubieran sido posibles.

Gracias asimismo a todas las personas de Contexto, Elena, Alba, Mavi, Raquel, Aritz, y Arabella, entre muchas otras, por vuestra labor encomiable durante todos estos años en las trincheras. Gracias a Ángel y a Ignacio por los cafés durante el FIPU, por compartir vuestras perspectivas, inquietudes y experiencias investigadoras, con este híbrido entre metodólogo y psicólogo social. Y gracias a Sergio, el buen doctor, por todas esas charlas y cafés tan refrescantes sobre *lo que verdaderamente importa*.

Moving to Sweden, I would really like to thank Juan Merlo for all his support, guidance and aid during my short period over the Nordic lands. I immediately feel welcome in your team, and I truly learn the value of multilevel models there. Thanks also for giving me the opportunity to learn from the best of the field, such as George Leckie and Harvey Goldstein. And I promise, I will never more misspell the MAIHDA. I would also like to thank Anna-Karin Ivert and Maria Wemrell for all their insights and contributions to the work we conducted there. Y gracias también a ti, Raquel, por guiarme y ayudarme a entender un poco mejor el loco sueco-mundo, y por haberme tratado como a un hermano durante mi incursión nórdica. I am looking forward to meet you all again.

Tampoco puedo olvidarme de aquellos que me guiaron durante mi etapa en la Universidad Autónoma de Madrid. Y sólo puedo empezar por

vosotros, Javier y Carmen. Os traslado desde aquí mi más sincero y enorme agradecimiento, pues sin vosotros hoy no habría podido llegar al punto en el que me encuentro en mi carrera académica. Gracias por iniciarme en los misterios de la estadística bayesiana y por enseñarme a descifrar los entresijos de los modelos factoriales y de los modelos de la teoría respuesta al ítem. Gracias por enseñarme el valor de los estudios de simulación y a ser crítico sobre los métodos empleados en mi propio trabajo. Gracias por vuestra guía y apoyo incondicional. Fue todo un honor haber trabajado con vosotros durante los inicios de mi carrera investigadora. Espero que nuestros caminos vuelvan a encontrarse.

Gracias también a Vicente, Paco, y Julio —a quien todos recordarnos con un inmenso cariño—, por darme la oportunidad de trabajar y aprender con vosotros en la cátedra de Modelos y Aplicaciones Psicométricas. Y gracias a mis compañeros del Laboratorio 17, Mori, Ale, Loli, y Miguel, por todos los momentos compartidos.

Tampoco puedo olvidarme de ti, Manuela. Quizá fue serendipia, pero fuiste tú quien me picó con el gusanillo de la investigación, y quien me mostró que los científicos son ante todo personas, con todo lo que ello implica. Fuiste tú quien guiaste mis primeros pasos investigadores y quien me enseñó a amar y a valorar el método científico, con sus cambios paradigmáticos y su evolución constante. Y por ello te estaré siempre agradecido.

Gracias también a todos mis colegas, compañeros y amigos de la Autónoma. Gracias a Blanca y David —oh, malvado, malvado, David— por nuestros contubernios en la mesa de las conspiraciones, allá dónde ésta estuviera. Gracias por compartir vuestras cuitas, sueños y aspiraciones. Habéis sido—y probablemente siempre lo seréis— todo un ejemplo y una inspiración para mí.

Gracias Lena, por nuestra correspondencia epistolar y literaria con la que tanto he disfrutado en los momentos de más amarga soledad, y con la que tanto he reflexionado sobre lo académico, sí, pero también sobre lo humano.

Gracias Ichi, madre adoptiva de Lu durante los tiempos más oscuros de la debacle, por compartir penas y glorias con este humilde metodólogo durante toda nuestra etapa universitaria. Gracias por mantener siempre la mano tendida y por mostrarte crítica con todo, ya sea lo propio o lo ajeno. Siempre consigues sacarle punta a nuestras conversaciones, no importa la distancia ni el tiempo. Por todo ello, mil gracias, Ichi.

Gracias Gina por nuestros intercambios de ideas peregrinas, por tus críticas siempre constructivas, no importa lo loca que sea la idea o la aproximación metodológica de turno, y por estar siempre disponible. Parece que fue ayer cuando empezamos el máster.

Pero no sólo de lo académico se nutre la etapa vital de la tesis, y por ello no sólo al mundo académico quiero dirigir mis agradecimientos. Gracias, Codigonautas, por estar siempre disponibles y dispuestos a emprender nuevas aventuras en otros mundos, ya sean más oscuros, más oníricos, o más fantasiosos. Gracias por las incontables horas evitando el advenimiento de los Primigenios más allá del espacio y del tiempo, maniobrando intrépidamente entre las filas del Sabbat, o intrigando con los antiguos de la Camarilla. Gracias a todos, de verdad, por mantener abierta siempre esa vía de escape. Gracias Dani, por tu ironía, talante y buen humor, siempre presente e igualmente contagioso. Gracias Pher, por tus malabares de agenda, y por tener siempre un hueco para un amigo cuando de verdad hace falta. Gracias, David Kendo, por tu apoyo y amistad incondicional durante todos estos años. Gracias, David Sertor, por poner un poco de orden a este mundo sin reglas; leer siempre estará *broken*. Y gracias a ti también, Sergio, codigonauta en el exilio, por las tardes y noches intrigando con Drahomanov.

estos años de tesis cuando por fin me he dado cuenta del valor incalculable que es tener una familia de semejante envergadura.

Gracias, Antonia. Sencillamente por todo. No hubiera sobrevivido a todos estos años sin tu constante apoyo ni tus cuidados, siempre presentes sin importar la distancia. Gracias por escuchar pacientemente todas mis cuitas y frustraciones, y por tener siempre palabras de ánimo y aliento en los momentos difíciles. Gracias por aguantarme, incluso en mis momentos más cáusticos, y por estar siempre dispuesta a perdonarme. Que sepas que yo también he puesto velas en favor de tu buena fortuna. Gracias Pablo por estar a su lado, y aunque a veces me queje, gracias por todas las fotos.

Gracias Ana, por todas esas horas de charleta, por tu apoyo, amor y comprensión durante todos los tragos amargos que hemos pasado juntos. Y por supuesto, gracias también por todos los momentos felices. Gracias por estar siempre a mi lado y entenderme cómo sólo tú puedes hacerlo.

Gracias a mis tíos, Isabel, Agustín, Miguel y Marisol por mantenernos a todos juntos durante todos estos años y por habernos dado tantísimo a todos los sobrinos. Gracias a todos mis primos, Ángel, Celia, Manuel y Hugo, por tantos y tantos momentos memorables. Gracias también a mi abuela Teresa, la matriarca de los Fernández Miranda, pues sin ella ninguno de sus nietos habríamos llegado a dónde estamos.

Gracias a mis abuelos, Justina y Jesús, por compartir conmigo la historia de nuestros orígenes, y por querernos y respaldarnos tantísimo a mi hermana y a mí desde siempre. Y perdonadme por no haberos llamado tanto como debería, prometo hacerlo más a menudo ahora que he terminado con la tesis —al fin y al cabo, ya no tengo excusa. Gracias a mi tío Jesús y a Maribel, por ese brillo radiante que siempre traen a casa después de sus viajes y aventuras, capaz de levantarnos el ánimo a todos. Gracias a mi hermana Irene, por tener siempre una sonrisa y un abrazo listos y dispuestos para su hermano.

Y gracias a Pedro, mi padre. Que aunque hoy no pueda leer estas líneas, estoy seguro de que se mostraría en el fondo orgulloso —aunque procurando, eso sí, que no me enterase; no se me fuera a subir demasiado a la cabeza. Gracias por todo.

Prometí que el penúltimo párrafo de los agradecimientos os lo dedicaría a todos vosotros. Así que vamos a ello: gracias a GG, príncipe de D, por ser la banda sonora que ha rodeado a esta tesis durante los últimos años, por inspirarme y animarme a buscar respuestas y motivaciones más allá de la psicología, y, en última instancia, por hacer que siempre trate de superarme. Gracias a MB y AS, pues fueron las primeras piedras angulares y gracias a ellas hemos llegado y evolucionado hasta donde estamos. Gracias también a los compañeros de cuadrilla, H, G, A, RF, SS, y TD, por acompañarnos durante los primeros momentos de este viaje. Gracias a C, JKL, GGi, MR, IP, JV, GR, AT, IS, CO, JW, y GF por los momentos de todo tipo que hemos compartido. Gracias a EJB, S, VH, y a SMV y LB por las líneas que sus relaciones nos han dejado explorar. Gracias a la señora B, y a M, A, J, C, Ch, W, y K por sus constantes servicios, pasados, presentes y futuros. Gracias muy especialmente a LA, por los retos de enorme complejidad que plantea y porque siempre hace que las cosas se vuelvan interesantes, sin importar lo mal que ya parezcan. Y muchísimas gracias a EC, porque sin ella nada de esto tendría sentido.

Y, por último, sólo faltas tú. Y en menudo brete me has colocado, pues soy bien consciente de lo poco aficionada que eres a las dedicatorias, agradecimientos, y a las palabras melosas en general. Así que prometo ser breve: gracias Carmen por ayudarme a organizar esta fiesta. Gracias por tu ayuda y apoyo en mis horas mayor necesidad. Gracias, de verdad. Por todo.

Prefacio

La presente tesis doctoral se basa en un compendio de publicaciones cuyo hilo conductor es la violencia de género, abordada desde un punto de vista principalmente cuantitativo. La violencia de género se define en nuestro contexto socio-cultural y jurídico como cualquier acto violento o agresión ejercido por un hombre contra una mujer que tenga o pueda tener como consecuencia un daño físico, sexual, o psicológico sobre la víctima (Ley Orgánica 1/2004, 28 de diciembre). Los estudios anglosajones emplean el término violencia contra la mujer en la pareja íntima (i.e., *intimate partner violence against women*) para añadir un componente de especial relevancia en esta tesis: la violencia está contextualizada dentro de una relación de pareja.

En esta tesis se empleará el término violencia de género para hacer referencia a los actos violentos ejercidos por un hombre contra su pareja o expareja. Quedan por tanto excluidos de esta definición operativa los actos de violencia contra las mujeres fuera del contexto de la relación de pareja, así como los actos de violencia que se den entre parejas no heteronormativas. Asimismo, tampoco se incluyen en el foco de esta tesis aquellos actos violentos o agresiones ejercidas por una mujer contra un hombre.

Las publicaciones incluidas en esta tesis doctoral abordan dos cuestiones relacionadas, si bien diferenciadas. Los tres primeros estudios se centran en las tasas de violencia de género en la Unión Europea y tratan de dar respuesta a un problema que las grandes macro-encuestas rara vez abordan: la comparabilidad entre países de los resultados obtenidos. Los tres estudios posteriores se centran en las actitudes públicas hacia la violencia de género,

con el objetivo principal de desarrollar nuevas medidas, fiables y válidas, para evaluar estas actitudes.

Se incluyen en esta tesis los textos completos de todos los artículos ya publicados, así como las versiones más recientes de dos estudios enviados a revistas para su publicación, que en el momento de maquetación de esta tesis se encuentran en proceso de revisión. Las publicaciones que componen la presente tesis doctoral son las siguientes:

Martín-Fernández, M., Gracia, E., & Lila, M. (2019). Ensuring the comparability of cross-national survey data on intimate partner violence against women. *BMJ Open. Manuscrito en segunda revisión.*

Martín-Fernández, M., Gracia, E., & Lila, M. (2019). Psychological intimate partner violence against women in the European Union: A cross-national invariance study. *BMC Public Health. Manuscrito en segunda revisión.*

Gracia, E., Martín-Fernández, M., Lila, M., Merlo, J., & Ivert, A. K. (2019). Prevalence of intimate partner violence against women in Sweden and Spain: A psychometric study of the 'Nordic paradox'. *PloS One*, *14*, e0217015.

Martín-Fernández, M., Gracia, E., Marco, M., Vargas, V., Santirso, F. A., & Lila, M. (2018). Measuring acceptability of intimate partner violence against women: Development and validation of the A-IPVAW scale. *European Journal of Psychology Applied to Legal Context*, *10*, 26-34.

Martín-Fernández, M., Gracia, E., & Lila, M. (2018). Assessing victim-blaming attitudes in cases of intimate partner violence against women: Development and validation of the VB-IPVAW scale. *Psychosocial Intervention*, *27*, 133-143.

Gracia, E., Martín-Fernández, M., Marco, M., Santirso, F. A., Vargas, V., & Lila, M. (2018). The Willingness to Intervene in Cases of Intimate Partner Violence Against Women (WI-IPVAW) scale: Development and validation of the long and short versions. *Frontiers in Psychology*, *9*, 1146.

XVIII

Índice

XXVI

Listado de Abreviaturas

A-IPVAW: Escala de actitudes de aceptabilidad de la violencia de género

APA: Asociación americana de psicología

ANOVA: Análisis de varianza

AREA: Asociación americana de investigación educativa

CFA: Análisis factorial confirmatorio

CFI: Índice de ajuste comparativo

CTS: *Conflict tactics scale*

DIF: Funcionamiento diferencial del ítem/de los ítems

EFA: Análisis factorial exploratorio

EIGE: Instituto europeo por la igualdad de género

FRA: Agencia de los derechos fundamentales de la Unión Europea

MG-CFA: Análisis factorial confirmatorio multi-grupo

NMCE: Consejo nacional de medición en educación

RMSEA: Raíz del error cuadrático medio por aproximación

TLI: Índice de Tucker-Lewis

TRI: Teoría de respuesta al ítem

UE: Unión Europea

UNDP: Programa de desarrollo de las Naciones Unidas

VB-IPVAW: Escala de actitudes de culpabilización a las víctimas de violencia de género

VG: Violencia de género

WEF: Foro económico mundial

WI-IPVAW: Escala de actitudes de disposición a intervenir en casos de violencia de género

WHO: Organización Mundial de la Salud

Abstract

Intimate partner violence against women (IPVAW) is a social and public health problem of global proportions. It is the most common form of violence suffered by women and has important consequences not only for the physical and psychological health of victims, but also for the well-being of their children. To achieve a better understanding of this problem, the use of robust, appropriate and up-to-date instruments and statistical models is crucial. In this regard, the present dissertation address two measurement problems related to IPVAW: the comparability of the measures assessing different forms of this violence used in a large survey across the European Union (EU), and the development of appropriate measures (i.e., reliable and valid), to evaluate three aspects of the public attitudes towards IPVAW, which are one of the key variables in the study of this type of violence. To address these issues, six studies have been designed and carried out.

To tackle the first problem, the first two studies tested the measurement invariance of the measures used in a survey about IPVAW conducted by the EU Agency for Fundamental Rights across the 28 member states, in order to examine how the levels of this type of violence are distributed across the EU. In the third study, valid comparisons were made between Sweden and Spain—two countries exemplifying the "Nordic paradox"— through a latent means analysis, with the aim of ruling out that this paradox is a product of a measurement bias. Taken together, the results of these three studies underlined the importance of using appropriate statistical methods to account for the cross-national comparability of the measures included in large socio-demographic surveys. The measurement invariance

approach allowed us to conduct a more appropriate and refined analysis of how physical and sexual violence levels are distributed across the European Union.

The last three studies of this dissertation addressed the lack of appropriate instruments to assess public attitudes towards IPVAW. Three new scales were developed and validated: the A-IPVAW scale, which evaluates attitudes of acceptability of IPVAW, the VB-IPVAW scale, which assess victim blaming attitudes in cases of IPVAW, and the WI-IPVAW scale, which measures the willingness to intervene in cases of IPVAW. The results of these studies showed that the three measures are reliable instruments with enough evidences of validity to be used for research and intervention purposes, since they are especially informative to detect persons with high levels of these types of attitudes and discriminate well among them. These scales are a step forward in the study of attitudes towards IPVAW, allowing researchers to extend knowledge about their conceptualization, measurement, prevalence, and the social factors that may influence these attitudes in order to improve prevention and intervention strategies.

Resumen

La violencia de género es un problema social y de salud pública de proporciones globales. Es la forma de violencia más frecuente que sufren las mujeres y tiene importantes consecuencias para el bienestar físico y psicológico no sólo de las víctimas, sino también para sus hijos. Para una adecuada comprensión de este fenómeno, resulta de vital importancia abordar su estudio aplicando herramientas y metodologías robustas, apropiadas y actualizadas. Y es precisamente en este contexto en el que se enmarca la presente tesis doctoral, donde se tratan dos problemas de medición relacionados con la violencia de género: la comparabilidad de las medidas de diferentes formas de esta violencia en una gran encuesta poblacional de la Unión Europea, y el desarrollo de medidas adecuadas —esto es, con suficiente precisión y evidencias de validez—, para evaluar tres aspectos diferenciados de las actitudes públicas hacia la violencia de género, una de las variables clave en el estudio de este fenómeno. Con el propósito de dar respuesta a estos dos problemas, se han diseñado y llevado a cabo seis estudios.

Para tratar el primer problema, en los dos primeros estudios se evaluó la invarianza factorial de las medidas incluidas en la encuesta sobre violencia de género de la agencia de los derechos fundamentales de la Unión Europea entre los 28 países miembros, para después examinar cómo se distribuyen los niveles de violencia física, sexual y psicológica entre ellos. El tercer estudio, por su parte, se centró en la realización de comparaciones válidas en los niveles de violencia de género física y sexual entre Suecia y España, dos países que ejemplifican la paradoja nórdica, con el objetivo específico de descartar que este fenómeno se deba a un sesgo de medida. Tomados en su conjunto, los

resultados de los tres primeros estudios de esta tesis subrayaron la importancia de emplear métodos estadísticos capaces de dar cuenta de la comparabilidad entre países de las medidas incluidas en las grandes encuestas poblacionales. Los análisis de invarianza factorial mostraron que las medidas de violencia física, sexual y psicológica de la encuesta de la agencia de los derechos fundamentales de la Unión Europea son invariantes entre los diferentes países miembros. Esto permitió a su vez examinar de manera más apropiada y sofisticada cómo se distribuyen los niveles de estas formas de violencia en la Unión Europea.

En los tres últimos estudios de la presente tesis doctoral se abordó directamente la falta de herramientas adecuadas para evaluar las actitudes públicas hacia la violencia de género, presentando el desarrollo y la validación de tres nuevas escalas para medir tres tipos de estas actitudes: la escala A-IPVAW, para evaluar las actitudes de aceptabilidad de la violencia de género, la escala VB-IPVAW, que da cuenta de las actitudes de culpabilización a las víctimas de violencia de género, y la escala WI-IPVAW, que mide la disposición a intervenir en casos de violencia de género. Los resultados de los tres últimos estudios de la presente tesis doctoral pusieron de manifiesto que estas tres medidas complementarias de actitudes hacia la violencia de género cuentan con una adecuada consistencia interna y con suficientes evidencias de validez para ser empleadas en contextos de investigación o de intervención, pues son especialmente informativas para discriminar con precisión entre aquellas personas con altos niveles de estos tipos de actitudes. Estas escalas suponen un avance en el estudio de las actitudes públicas hacia este tipo de violencia, pues posibilitan ampliar el conocimiento sobre las mismas ahondando en su conceptualización, evaluación, prevalencia y en los factores sociales que pueden influir en estas actitudes para mejorar las estrategias de prevención e intervención de la violencia de género.

CAPITULO 1

Introducción General

Introducción General

La violencia de género (VG) es un problema social y de salud pública de proporciones globales (Ali & Naylor, 2013; García-Moreno, Jansen, Ellsberg, Heise, & Watts, 2006; World Health Organization [WHO], 2013). Es la forma de violencia más frecuente que sufren las mujeres y tiene importantes consecuencias para el bienestar físico y psicológico no sólo de las víctimas, sino también para sus hijos (Campbell, 2002; Craparo, Gori, Petruccelli, Cannella, & Simonelli, 2014; Devries et al., 2013; Ellsberg, Jansen, Heise, Watts, & Garcia-Moreno, 2008; Guedes, Bott, Garcia-Moreno, & Colombini, 2016). Su prevalencia a nivel mundial es del 30%, mientras que en los países desarrollados es del 23,2% (WHO, 2013).

Se trata de un fenómeno complejo y multicausal, con profundas raíces en diferentes niveles que van desde factores individuales e interpersonales, hasta factores macro-sociales (Heise, 1998, 2011; Heise & Kotsadam, 2015; Little & Kaufman Kantor, 2014). Desde esta perspectiva, la VG no se explicaría por un único factor, sino más bien por la interacción de diferentes factores a distintos niveles (Heise, 2011; WHO, 2002).

Para una adecuada comprensión de la VG, resulta de vital importancia abordar su estudio aplicando herramientas y metodologías robustas, adecuadas y actualizadas. Y es precisamente en este contexto en el que se enmarca la presente tesis doctoral, abordando dos problemas de medición bien diferenciados, si bien relacionados.

El primero de ellos tiene que ver con la comparabilidad de los datos disponibles sobre VG de las grandes encuestas poblacionales en la Unión Europea (UE), ya que rara vez se examina en detalle la invarianza factorial de

las medidas empleadas en este tipo de estudios para obtener las tasas de prevalencia de VG. La ausencia de este tipo de análisis limita severamente la validez de las comparaciones realizadas entre distintos países o grupos poblacionales, pues los resultados obtenidos pueden estar enmascarando sesgos de medida o verse afectados por diferentes factores culturales, como las creencias o las expectativas sobre la VG (Davidov, Meuleman, Cieciuch, Schmidt, & Billie, 2014; Lubke, Dolan, Kelderman, & Mellenbergh, 2003; Milfont & Fischer, 2010; Putnick & Bornstein, 2016).

El segundo problema abordado en esta tesis está centrado en la medición de una de las variables clave para entender este fenómeno: las actitudes públicas hacia la VG (Copp, Giordano, Longmore, & Manning, 2016; Flood & Pease, 2009; Powell & Webster, 2018). Y es que a pesar de la relevancia que estas actitudes pueden llegar a tener para comprender y explicar mejor la VG, las medidas actitudinales que se han desarrollado hasta el momento no son lo suficientemente precisas ni cuentan con suficientes evidencias de validez para evaluarlas de forma adecuada (Gracia & Lila, 2015). Resulta por lo tanto necesario desarrollar nuevas herramientas o depurar las existentes para medir apropiadamente estas actitudes.

La violencia de género en la Unión Europea

La VG es un fenómeno presente en todas las sociedades y que incluye diferentes manifestaciones que van desde la violencia psicológica (e.g., abuso emocional, conducta controladora), hasta la violencia física y sexual (Cooker et al., 2002; WHO, 2012).

En Europa, una encuesta llevada a cabo por la agencia de derechos fundamentales de la UE (*European Union Fundamental Rights Agency* [FRA]) entre los 28 países miembros encontró que, en promedio, el 22% de las mujeres europeas habían sufrido VG física y sexual en algún momento de sus
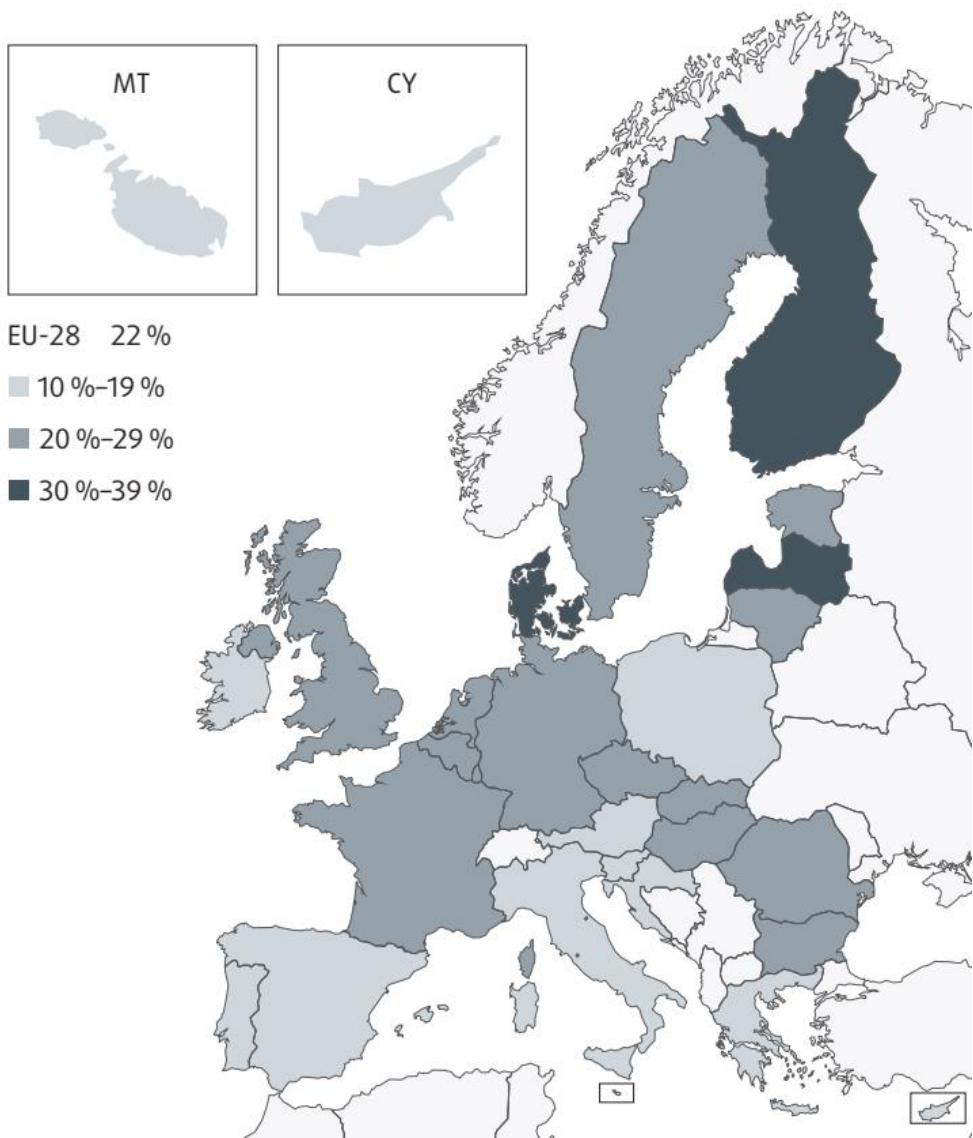
Figura 1. Prevalencia de VG física y sexual en la UE por país.

Fuente: *Violence against women: An EU-wide survey: Main results* (FRA, 2014a).

vidas, y que el 43% de ellas habían experimentado asimismo algún episodio de violencia psicológica (FRA, 2014a). Estas cifras de prevalencia, sin embargo, varían enormemente entre países. En el caso de la violencia física y sexual, la prevalencia varía entre un 13% en España hasta un 32% en Dinamarca y

Letonia (ver Figura 1). Con respecto a la violencia psicológica, las tasas de prevalencia varían entre un 31% en Irlanda y un 60% en Letonia.

No obstante, a pesar del alcance y relevancia de estos datos, hasta la fecha ningún estudio ha examinado la comparabilidad de las escalas empleadas en la encuesta del FRA para obtener estas prevalencias. Esto limita seriamente la validez de estos resultados, pues no es posible descartar que las diferencias en VG entre países se deban a diferentes factores dependientes de la cultura, como las expectativas y creencias culturales sobre la VG. En esta tesis doctoral se aborda precisamente este problema estudiando la invarianza factorial de las medidas de violencia física y sexual incluidas en la encuesta del FRA, por un lado, y las medidas de violencia psicológica por otro. Se plantea a su vez un estudio psicométrico en profundidad de la llamada *paradoja nórdica*, término empleado por Gracia y Merlo (2016) para reflejar las elevadas tasas de prevalencia de VG en los países nórdicos (e.g., Dinamarca, Finlandia, Suecia), a pesar de ser estos los países con mayores niveles de igualdad de género en el mundo (European Institute for Gender Equality [EIGE], 2017).

Comparabilidad de las medidas de violencia física y sexual

La VG puede ser un tema culturalmente sensible, y cómo es percibida, conceptualizada e interpretada puede variar entre los distintos países de la UE. Es importante, por lo tanto, evaluar la invarianza factorial de las medidas empleadas por las grandes encuestas poblacionales para poder realizar comparaciones entre países que sean válidas (Milfont & Fischer, 2010; Van de Schoot, Lugtig, & Hox, 2012; Vandenberg & Lance, 2000). La invarianza factorial es un prerrequisito fundamental en la investigación intercultural, ya que permite realizar comparaciones significativas entre países descartando la posibilidad de un sesgo debido a la cultura en las respuestas de los encuestados (Davidov et al., 2014). Cuando la invarianza factorial no se

cumple, ello implica que los encuestados de distintos países responden de manera diferencial a las preguntas, por lo que sus respuestas no podrían compararse directamente (Jang et al., 2017; Putnick & Bornstein, 2016). Si ese es el caso, entonces no puede asumirse que la VG se interprete de la misma manera en todos los países miembros de la UE, ya que la misma puntuación en un país puede reflejar un constructo diferente o dar como resultado diferentes niveles de VG en otro. Calcular pues tasas de prevalencia para comprar diferentes países sin primero asegurarse de que se cumple la invarianza factorial puede llevar a conclusiones imprecisas y distorsionadas, ya que la validez de tales comparaciones puede ser cuestionable (Lubke et al., 2003).

El primer estudio de la presente tesis doctoral aborda este vacío en la literatura sobre VG, realizando un análisis de invarianza factorial en profundidad para asegurar la comparabilidad de las medidas de VG física y sexual empleadas en la encuesta del FRA, evaluando así si las mujeres encuestadas de cada país conceptualizan e interpretan las preguntas sobre este tipo de violencia de la misma manera. Una vez establecida la invarianza factorial, se compararán mediante un análisis de medias latentes las diferencias entre países en sus niveles de VG física y sexual.

Comparabilidad de las medidas de violencia psicológica

La violencia psicológica es la forma de VG más frecuente, que puede darse tanto en solitario como en conjunción con otras formas de este tipo de violencia (Hamby & Sugarman, 1999; Karakurt & Silver, 2013; Liles et al., 2012; Murphy & O'Leary, 1989). Existe, no obstante, una fuerte asociación entre la VG psicológica y la física, ya que la violencia psicológica suele preceder a la física, por lo que se la considera uno de sus principales factores de riesgo (Salis, Salwen, & O'Leary, 2014; Schumacher & Leonard, 2005).

La prevalencia de la VG psicológica en las grandes encuestas sociodemográficas varía considerablemente dependiendo de cómo se defina y se mida este tipo de violencia, con algunos estudios estimándola en torno a un 10-20%, mientras que otros sitúan la prevalencia media de la VG psicológica alrededor de un 80-90% (Coker et al., 2002; Kaukinen 2004; Romans, Forte, Cohen, Du Mont, & Hyman, 2007; Stets & Straus, 1990; Straight, Harper, & Arias, 2003; Tjaden & Thoennes, 2000). Asimismo, la frecuencia y la severidad de la VG psicológica puede variar bastante de un país a otro (Carney & Barnes, 2012; García-Moreno et al., 2006), lo que implica que la manera en que este tipo de violencia es percibida e interpretada puede ser diferente entre distintos países y culturas.

La mayoría de las encuestas poblacionales abordando la VG psicológica siguen la tradición de la *Conflict Tactis Scale* (CTS), definiendo este tipo de violencia a través de conductas concretas (Garcia-Moreno et al., 2006; Straus, 1979; Straus, Hamby, Boney-McCoy & Sugarman, 1996). Normalmente se consideran dos aspectos en la evaluación de la VG psicológica: el abuso emocional y la conducta controladora (Beeble, Bybee, & Sullivan, 2007; Follingstad, Coyne, & Gambone, 2005; Hamberger, Larsen, & Lehrner, 2017; Murphy & O'Leary, 1989; Salis, Salwen, O'Leary, 2014). El abuso emocional incluye conductas cuyo propósito es generar daño emocional o amenazas de daño físico en la víctima, como el menosprecio, la humillación, la amenaza o la intimidación. Por su parte, la conducta controladora incluye actos orientados a monitorizar, aislar o limitar las conductas de la víctima, como prohibirle abandonar la casa, restringir su contacto con otras personas, o insistir continuamente en conocer su paradero (Carney & Barnes, 2012; Garcia-Moreno et al., 2006; FRA, 2014a; Smith et al., 2018).

La sección de la encuesta del FRA sobre VG psicológica sigue esta tradición del CTS y emplea el mismo conjunto de preguntas sobre abuso

emocional y conducta controladora en los 28 países de la UE. Sin embargo, al igual que sucede con las medidas sobre VG física y sexual de esta encuesta, todavía ningún estudio ha examinado con propiedad la invarianza factorial de las preguntas de VG psicológica, lo que limita asimismo la validez de las comparaciones entre países realizadas con esta medida (Davidov et al., 2014; Lubke, et al., 2003; Putnick & Bornstein, 2016).

El segundo estudio de esta tesis se centra precisamente en este problema, evaluando la invarianza factorial de la medida de VG psicológica incluida en el FRA y poniendo así a prueba si los participantes de cada país conceptualizan e interpretan las preguntas de esta encuesta del mismo modo. Al igual que con las medidas de VG física y sexual, se compararán los niveles de VG psicológica entre países mediante un análisis de medias latentes una vez la comparabilidad de esta medida haya sido asegurada.

La paradoja nórdica: ¿un artefacto metodológico?

Gracia y Merlo (2016) proponen el término paradoja nórdica para reflejar las altas tasas de prevalencia de VG en los países nórdicos (e.g., Dinamarca, Finlandia, Suecia) a pesar de contar estos con los mayores niveles de igualdad de género en toda la UE (European EIGE, 2017; United Nations Development Program [UNDP], 2017; World Economic Forum [WEF], 2017). La encuesta realizada por el FRA sobre VG apoya esta idea, pues Dinamarca, Finlandia y Suecia se cuentan entre los países con mayor prevalencia de violencia física y sexual (i.e., un 32%, 30%, y 28%, respectivamente) (FRA, 2014a).

Las razones que podrían explicar este sorprendente resultado todavía son desconocidas, y apenas se han realizado estudios que aborden específicamente esta paradoja (Lundgren, Heimer, Westerstrand, Kalliokoski, 2001). Si bien Gracia y Merlo (2016) plantean una serie de razones teóricas y

metodológicas que podrían explicar la paradoja nórdica, todas ellas pasan por una cuestión previa de vital importancia: descartar la posibilidad de que estas diferencias en VG con el resto de países europeos se deban a un sesgo de medida. En otras palabras, que la paradoja nórdica no se trata de un artefacto metodológico. Para ello es necesario establecer la invarianza factorial entre los diferentes países de la UE.

En el tercer estudio de la presente tesis doctoral se tomarán Suecia y España como dos países que ejemplifican esta paradoja. Suecia ocupa la tercera posición en el índice global de inequidad (WEF, 2017), la quinta en el índice global de la brecha de género (UNDP, 2017), y la primera en el índice de igualdad de género (EIGE, 2017). España, por su parte, ocupa la 13ª (índice global de inequidad) la 24ª (índice global de la brecha de género,) y la 11ª (índice de igualdad de género). Sin embargo, a pesar de estas diferencias, España se encuentra entre los países de la UE con menor prevalencia de VG (FRA, 2014a). Partiendo de los datos de la encuesta del FRA, se pondrá a prueba la invarianza factorial de las medidas de violencia física y sexual entre ambos países. Asimismo, se llevará a cabo un análisis diferencial de los ítems (DIF, por sus siglas en inglés: *differential ítem functioning*) para examinar individualmente en cada ítem la presencia de sesgos de medida entre ambos países. Finalmente, una vez asegurada la equivalencia métrica de las medidas empleadas, se compararán los niveles de VG física y sexual entre España y Suecia para poner a prueba si la paradoja nórdica se trata de un artefacto metodológico.

Actitudes públicas hacia la violencia de género

Las actitudes públicas hacia la VG son factor clave para entender el contexto social en el que se produce este tipo de violencia, contribuyendo a fomentar o disuadir su perpetración en la sociedad (Carlson & Worden, 2005;

10

Copp, Giordano, Longmore, & Manning, 2016; Flood & Pease, 2009; Gracia, 2014; Powell & Webster, 2018). La investigación realizada hasta el momento subraya la importancia de prestar atención a las actitudes hacia la VG, pues están estrechamente relacionadas con las tasas de incidencia, las respuestas de los profesionales de los sectores públicos, y las respuestas de las propias víctimas (Browning, 2002; Gracia, Garcia, & Lila, 2011, 2014; Gracia, Herrero, Lila, & Fuente, 2009; Rizo & Macy, 2011; West & Wandrei, 2002). Abordar estas actitudes es, por tanto, uno de los principales objetivos para las estrategias de prevención e intervención, tanto con víctimas como con agresores de VG (García-Moreno et al., 2015; Fernández-González, Calvete, & Orue, 2017; Jewkes, Flood, & Lang, 2015; Lila & Gracia, 2015; Lila, Gracia, & Catalá-Miñana, 2018).

Las actitudes públicas hacia la VG pueden manifestarse de diferentes maneras, bien desde su componente cognitivo, afectivo o conductual, o mediante una interacción de varios o todos ellos (Eagly & Chaiken, 1998, 2005, 2007; Maio, Olson, Bernard, & Luke, 2003). Las actitudes públicas hacia la VG pueden influir, por lo tanto, en cómo este tipo de violencia es percibida y justificada (e.g., aceptabilidad de la VG), en las atribuciones que se realizan (e.g., culpabilización de las víctimas), y en las intenciones de actuar en casos de VG (e.g., disposición a intervenir) (Gracia & Lila, 2015).

Dada la relevancia que las actitudes públicas hacia la VG pueden llegar a tener, se hace necesario contar con herramientas de evaluación fiables y válidas, que permitan llevar a cabo investigaciones que mejoren el conocimiento existente sobre las mismas y que puedan ser a su vez empleadas en contextos de intervención. Si bien existen algunas escalas y test diseñados para medir diferentes componentes de estas actitudes, ninguna de estas medidas cuenta con las suficientes garantías psicométricas para su utilización en tales circunstancias. La presente tesis trata de dar respuesta a este vacío en

la literatura presentando tres nuevas herramientas para evaluar diferentes aspectos de las actitudes públicas hacia la VG.

Actitudes de aceptabilidad de la violencia de género

Las actitudes de aceptabilidad de la VG están estrechamente relacionadas con la perpetración de este tipo de violencia y son, a su vez, uno de sus principales factores de riesgo (Abramsky et al., 2011; Archer & Graham-Kevan, 2003; Copp et al., 2016; Gracia, Rodríguez, Martín-Fernández, & Lila 2015; Sugarman & Frankel, 1996; WHO, 2002). Altos niveles de aceptabilidad de la VG pueden llevar a la percepción de que este tipo de conductas son normativas, incrementando así el riesgo de que los hombres incurran en la VG y su justificación posterior por las víctimas y sus círculos sociales más próximos (Waltermaurer, 2012).

Resulta por lo tanto de vital importancia contar con medidas fiables y válidas de la aceptabilidad de la VG, capaces de proporcionar nuevos conocimientos sobre las condiciones sociales que contribuyen a la VG (Gracia & Lila, 2015; Muehlenhard & Kimes, 1999). Existen algunos estudios que han empleado escalas actitudinales para medir la aceptabilidad de la VG en población joven (Coop et al., 2016; Fincham et al., 2008), o en poblaciones rurales (Schwab-Reese & Renner, 2017). Entre la población general, varias encuestas demográficas de salud han incluido escalas breves e ítems individuales evaluando la aceptabilidad de algunas conductas de VG (Wang, 2016; WHO 2013; Yount, Halim, Hynes, & Hillman, 2011). Sin embargo, la mayoría de estas encuestas han sido diseñadas para su empleo en países subsaharianos, y han mostrado resultados diferentes cuando se realizan cambios menores en los enunciados de los ítems que utilizan (Tsai et al., 2017). En esta misma línea, Gracia y Lila (2015) encontraron en su revisión de las actitudes públicas hacia la VG en la UE, que la información disponible en las

encuestas europeas no es sólo escasa —y principalmente basada en ítems únicos—, sino que tampoco está basada en instrumentos psicométricamente sólidos (i.e., fiables y válidos) ni lo suficientemente apropiados. Se hace por lo tanto necesario contar con medidas de aceptabilidad de la VG adecuadas para este tipo de encuestas.

El cuarto estudio de la presente tesis aborda precisamente este problema, donde se desarrolla y valida una nueva escala de veinte ítems para medir las actitudes de aceptabilidad de la VG en población general. Para ello, se examinaron en profundidad el contenido de los ítems mediante un panel de expertos, así como su estructura latente a través del empleo de modelos factoriales y de la teoría de respuesta al ítem (TRI). Para evaluar la validez de esta nueva medida, se relacionaron las actitudes de aceptabilidad de la VG con otras variables actitudinales, como la severidad percibida de la VG o las actitudes sexistas, así como con variables sociodemográficas, como el sexo o la edad. Estudios anteriores han encontrado que altos niveles de aceptabilidad están negativa y estrechamente relacionados con la severidad percibida de este tipo de violencia (Gracia & Herrero, 2006a; Martín-Fernández, Gracia & Lila, 2018; Taylor & Sorenson, 2005). También se han encontrado correlatos entre las actitudes sexistas y la aceptabilidad de la VG (Flood & Pease, 2009; Herrero, Rodríguez, & Torres, 2017; Lila, Gracia, & García, 2013). Con respecto al sexo y la edad, investigaciones previas han puesto de manifiesto que los hombres y las personas de mayor edad tienden a aceptar y justificar más la VG que las mujeres o las personas más jóvenes, respectivamente (Bryant & Spencer, 2003; Carlson & Worden, 2005; Fincham, Cui, Braithwaite, & Pasley, 2008; Gracia et al., 2015; Gracia & Tomás, 2014; Taylor & Sorenson, 2005). Se utilizó además una pequeña muestra de agresores con penas por VG para examinar las diferencias existentes entre este grupo y el resto de participantes en el estudio.

Actitudes de culpabilización a las víctimas a violencia de género

Las actitudes de culpabilización a las víctimas se encuentran también entre los factores de riesgo de la VG, pues reflejan la tolerancia pública que puede haber hacia este tipo de violencia y se utilizan a menudo para explicarla o, incluso, justificarla (Gracia, 2014; Gracia & Tomás, 2014; WHO, 2002). Estas actitudes de culpabilización no sólo afectan a las respuestas de los agentes públicos y a la disposición personal a intervenir en casos conocidos de VG, sino que pueden afectar además a las respuestas de los agresores y de las propias víctimas. Las actitudes de culpabilización manifestadas por las personas más próximas a las víctimas de VG pueden promover y facilitar las conductas de los agresores, a la vez que hacen más arduo para las víctimas admitir una situación de violencia, dificultando así la búsqueda y la obtención de ayuda de fuentes formales o informales (Ansara & Hindin, 2010; Garrido-Macías, Valor-Segura, & Expósito, 2017; Liang, Goodman, Tummala-Narra, & Weintraub, 2005; Valor-Segura, Exposito, & Moya, 2011; Voith, 2017; West & Wandrei, 2002).

Disponer de medidas psicométricamente adecuadas para evaluar estas actitudes resulta pues de una importancia crucial para la investigación y la intervención en VG (Gracia & Lila, 2015; Powell & Webster, 2018; Muehlenhard & Kimes, 1999). Si bien existen investigaciones previas que ya han abordado el estudio de las actitudes de culpabilización a las víctimas de VG, la mayoría de las medidas que emplean estos estudios cuentan con algunas desventajas o presentan serias limitaciones. Algunos de estos instrumentos sólo consideran VG aquellos actos de violencia física que sucedan dentro de una pareja casada, ignorando completamente otros tipos de violencia relevantes como la psicológica o la emocional (Jackson et al., 1994; Petretic-Jackson et al., 1994). Otras medidas han sido validadas sólo con estudiantes universitarios, lo que dificulta su generalización a poblaciones más amplias (Fox & Cox, 2011; Scott & Strauss, 2007). Otros de los instrumentos

14

disponibles se basan enteramente en el planteamiento de viñetas o escenarios de un caso único (Koepke et al, 2014; Vidal-Fernández & Megías, 2014; Yamawaki et al., 2009), o bien se han utilizado en estudios exploratorios y precisan de más investigación para establecer sus propiedades psicométricas en muestras más representativas y de mayor tamaño (Fox & Cox, 2011; Yun & Vonk, 2011). Existen asimismo otros instrumentos que incluyen sub-escalas para la evaluación de las actitudes de culpabilización a las víctimas de VG, pero han sido diseñados para ser utilizados con agresores de VG y no en población general (Henning & Holdfold, 2006; Henning, Jones, & Holdfold, 2006; Lila, Oliver, Catalá-Miñana, Galiana, & Gracia, 2014). Por último, los datos de las grandes encuestas poblacionales, aunque incluyen algún ítem de actitudes de culpabilización a las víctimas de VG, rara vez suelen basarse en instrumentos de evaluación con suficiente fiabilidad o suficientes evidencias de validez (Gracia & Lila, 2015).

Se hace por lo tanto evidente la necesidad de una herramienta psicométricamente adecuada que evalué las actitudes de culpabilización a las víctimas de VG, tanto para propósitos de investigación como de intervención, y que sea a su vez apropiada para ser utilizada en grandes encuestas poblacionales.

El quinto estudio de esta tesis doctoral hace frente a este vacío en la literatura y plantea el desarrollo y la validación de una nueva escala para evaluar este tipo de actitudes. Partiendo del conjunto de ítems identificados por Gracia & Lila (2015) en su revisión sobre las actitudes hacia la VG en la UE, se plantea un estudio de validación cruzada de una nueva medida de actitudes de culpabilización a las víctimas de VG — con dos versiones: la completa de doce ítems, y una breve de cinco ítems—, analizando además su estructura interna y su fiabilidad. Para examinar la validez de esta medida, se correlacionaron las puntuaciones de las dos versiones de esta nueva escala con

otros constructos relacionados, como son la aceptabilidad de la VG, la severidad percibida de la VG, o las actitudes sexistas, además de con otras variables sociodemográficas como el sexo o la edad (Capezza & Arriaga, 2008; Keopke, et al., 2014; Scott & Straus, 2007; Taylor & Sorenson, 2005; Yun & Vonk, 2011). Investigaciones recientes apuntan a que aquellas personas con mayores niveles de actitudes de culpabilización tienden a su vez a percibir la VG como menos severa, más aceptable, y muestran también mayores niveles de sexismo hostil y benevolente (Gracia et al., 2018; Juarros-Basterretxea, Overall, Herrero, & Rodríguez-Díaz, 2019; Martín-Fernández, Gracia, & Lila, 2018; Vargas, Lila, Catalá-Miñana, & Gracia, 2017). Asimismo, los hombres y las personas de mayor edad son más propensas a mostrar mayores niveles de este tipo de actitudes (Bryant & Spencer, 2003; Gracia & Tomás, 2014). Adicionalmente, se empleó una segunda muestra de agresores condenados por VG y que participan en un programa de intervención para comparar sus niveles de actitudes de culpabilización con los niveles del resto de participantes.

Disposición a intervenir en casos de violencia de género

A pesar de que la VG sigue siendo un delito que se denuncia poco ante las autoridades, es al mismo tiempo una circunstancia ampliamente conocida por los círculos sociales más próximos a las víctimas (Gracia, 2004, Taylor & Sorenson, 2005; Taylor et al., 2016). Por ejemplo, en una encuesta realizada entre los 28 países miembros de la UE se encontró que cerca del 23% de los participantes conocía a una mujer que hubiera sido víctima de VG entre sus amigos o familiares, el 17% en su comunidad de vecinos o en su barrio, y el 9% en su centro de trabajo o estudios (European Comission, 2016). Son precisamente las personas que conocen o presencian un caso de VG las que están en mejor posición para intervenir y tratar de ayudar a las víctimas — bien avisando a la policía, ofreciéndose para ayudar, llamando la atención al

agresor, etc.—, pero lo cierto es que también pueden decidir no hacer nada e ignorar completamente la situación (Banyard y Moynihan, 2011; Taylor et al., 2016).

La disposición a intervenir en casos de VG puede funcionar tanto como un factor de riesgo como un factor de protección de la VG (Gracia, 2014; Koepsell et al., 2006; Voith, 2017). Actitudes no favorables hacia la intervención entre los círculos sociales que rodean a las víctimas pueden facilitar o reforzar la conducta de los agresores, además de hacer más difícil que las víctimas reconozcan este problema y busquen ayuda o recursos que les permitan escapar de la VG. En cambio, las actitudes favorables hacia la intervención (e.g., llamar a la policía o intervenir directamente) entre aquellas personas cercanas a las víctimas de VG pueden tener un efecto protector en favor de las víctimas, ya sea inhibiendo o reduciendo el riesgo de VG, o bien incrementando los costes sociales y legales para los agresores (McDonnell et al., 2011). Las actitudes favorables hacia la intervención son asimismo especialmente relevantes, ya que las víctimas de este tipo de violencia son proclives a buscar ayuda primero entre fuentes informales, como familiares y amigos, antes de acudir a las fuentes formales como la policía (Liang et al., 2005; Ansara y Hindin, 2010; McCart, Smith, y Sawyer, 2010; Wee et al., 2016).

Incrementar la probabilidad de que la gente esté dispuesta a intervenir para ayudar a las víctimas de VG debería ser, por lo tanto, uno de los principales objetivos de cualquier estrategia de prevención que busque transformar la percepción pública de este problema en un sentido de mayor responsabilidad e implicación personal, contribuyendo así al control social informal de la VG (Gracia, Herrero, Lila, & Fuente, 2009).

Por todo ello, se hace evidente la necesidad de potenciar y avanzar en el conocimiento sobre la disposición a intervenir en casos de VG y en los aspectos más relevantes que la rodean, como son la elevada prevalencia de las

actitudes no favorables a la intervención, las diferentes preferencias a la hora de intervenir, sus correlatos y determinantes, o la evaluación de la efectividad de las intervenciones que abordan estas actitudes. Y para realizar este tipo de investigación resulta de vital importancia contar con herramientas fiables y válidas que midan la disposición a intervenir en casos de VG.

Si bien se han desarrollado algunas herramientas para examinar la disposición a ayudar en casos de violencia, la mayoría de la investigación realizada hasta el momento ha estado centrada en la conducta de los espectadores de violencia entre parejas jóvenes y adolescentes, espectadores de acoso sexual, o espectadores de situaciones de violación (Banyard & Moynihan, 2011; Banyard, Moynihan, Cares, & Warner, 2014; Branch, Richard, & Dretsch, 2013; McMahon et al 2014; Stein, 2007). Otros estudios que han evaluado la disposición de intervenir presentan serias limitaciones de generalizabilidad, pues emplean sobre todo muestras reducidas de estudiantes universitarios, o instrumentos con baja fiabilidad (Baldry & Pagliaro, 2014; Baldry et al., 2015; Cinquegrana et al., 2018). Asimismo, los datos de las grandes encuestas poblacionales sobre actitudes hacia la intervención en casos de VG o bien se basan en ítems únicos, o en herramientas sin una adecuada fiabilidad ni suficientes evidencias de validez (Gracia & Lila, 2015). Hace falta todavía una herramienta psicométricamente adecuada para evaluar la disposición a intervenir en casos de VG, apropiada tanto para su empleo en muestras comunitarias como en encuestas poblacionales.

En el sexto y último estudio de esta tesis se presenta el desarrollo y la validación de la escala WI-IPVAW (por sus siglas en inglés: *Willingness to Intervene in Cases of Intimate Partner Violence*) para evaluar este tipo de actitudes públicas hacia la VG. En este estudio, además de la versión completa de la escala se proponen dos versiones breves, de nueve y seis ítems, para ser utilizadas en contextos de investigación con diferentes necesidades (e.g., la

versión completa para contextos de investigación e intervención, y las versiones breves para encuestas poblacionales o estudios que evalúen un gran número de variables). Mediante el empleo de diferentes modelos estadísticos se abordan diferentes sesgos de medida, como la deseabilidad social o la invarianza factorial por sexos. Para poner a prueba su validez, en este estudio se evaluaron también las relaciones entre la disposición a intervenir en casos de VG con otras variables actitudinales, como la aceptabilidad de la VG, las actitudes de culpabilización a las víctimas de VG, la severidad percibida de la VG, y el sexismo hostil (Flood & Pease, 2009; Gracia & Herrero, 2006b; Gracia, García & Lila, 2008; Herrero, Rodríguez, & Torres, 2017; Lila, Gracia, & García, 2013; Martín-Fernández et al., 2018; Taylor & Sorenson, 2005). Se exploró a su vez el efecto que el sexo, la edad, o el nivel educativo pueden tener en la disposición a intervenir en casos de VG (Carlson & Worden, 2005; Fincham et al., 2008; Flood & Pease, 2009; Gracia et al., 2009, 2015).

En resumen, la presente tesis doctoral trata de dar respuesta a dos problemas de medición relacionados con la VG mediante el empleo de modelos y procedimientos estadísticos actuales. El primero de ellos tiene que ver con las medidas empleadas para obtener las prevalencias de diferentes tipos de VG en las grandes encuestas poblacionales europeas y su comparabilidad entre diferentes países. Asegurar la comparabilidad de estas medidas no es una cuestión menor, especialmente cuando el objetivo de dichas encuestas es precisamente realizar comparaciones válidas entre los países miembros de la UE. El segundo problema se trata de la falta de herramientas de evaluación para medir una de las variables clave en el estudio de este tipo de violencia: las actitudes públicas hacia la VG. En esta tesis se desarrollan tres nuevas medidas para la evaluación de tres tipos distintos de estas actitudes, y se proporcionan

además versiones breves de las mismas que podrían ser potencialmente incluidas en las grandes encuestas poblacionales.

CAPITULO 2

Objetivos

Objetivos

El objetivo general de la presente tesis doctoral es doble: realizar, por un lado, comparaciones válidas y apropiadas de los niveles de VG física, sexual y psicológica, entre los distintos países miembros de la UE; y desarrollar, por el otro lado, nuevas medidas para evaluar diferentes tipos de actitudes públicas hacia la VG que resulten psicométricamente adecuadas, esto es: que sean fiables y que cuenten con suficientes evidencias de validez.

Objetivo 1

El primer objetivo general trata de dar respuesta a un problema de medición de las grandes encuestas poblacionales que rara vez se aborda en los estudios de VG: la comparabilidad de las medidas utilizadas para evaluar los diferentes tipos de VG. Para ello se han diseñado y llevado a cabo tres estudios. Los dos primeros evalúan la invarianza factorial entre los 28 países miembros de la UE de las medidas incluidas en la encuesta del FRA, para después examinar cómo se distribuyen los niveles de VG física, sexual y psicológica entre ellos. El tercer estudio, por su parte, se centra en la realización de comparaciones válidas en los niveles de VG física y sexual entre Suecia y España, dos países que ejemplifican la paradoja nórdica, con el objetivo específico de descartar que este fenómeno se deba a un sesgo de medida.

En concreto, los estudios realizados para acometer este primer objetivo son los siguientes:

- Estudio 1: Martín-Fernández, M., Gracia, E., & Lila, M. (2019). Ensuring the comparability of cross-national survey data on intimate partner violence against women. *BMJ Open. Manuscrito en segunda revisión.*

- Estudio 2: Martín-Fernández, M., Gracia, E., & Lila, M. (2019). Psychological intimate partner violence against women in the European Union: A cross-national invariance study. *BMC Public Health. Manuscrito en segunda revisión.*
- Estudio 3: Gracia, E., Martín-Fernández, M., Lila, M., Merlo, J., & Ivert, A. K. (2019). Prevalence of intimate partner violence against women in Sweden and Spain: A psychometric study of the 'Nordic paradox'. *PloS One*, *14*, e0217015.

Objetivo 2

El segundo objetivo general de esta tesis está orientado a proporcionar nuevas medidas para evaluar adecuadamente las actitudes de aceptabilidad de la VG, de culpabilización a las víctimas de VG, y de disposición a intervenir en casos de VG. Con este fin, se han planteado los siguientes estudios:

- Estudio 4: Martín-Fernández, M., Gracia, E., Marco, M., Vargas, V., Santirso, F. A., & Lila, M. (2018). Measuring acceptability of intimate partner violence against women: Development and validation of the A-IPVAW scale. *European Journal of Psychology Applied to Legal Context*, *10*, 26-34.
- Estudio 5: Martín-Fernández, M., Gracia, E., & Lila, M. (2018). Assessing victim-blaming attitudes in cases of intimate partner violence against women: Development and validation of the VB-IPVAW scale. *Psychosocial Intervention*, *27*, 133-143.
- Estudio 6: Gracia, E., Martín-Fernández, M., Marco, M., Santirso, F. A., Vargas, V., & Lila, M. (2018). The Willingness to Intervene in Cases of Intimate Partner Violence Against Women (WI-IPVAW) scale: Development and validation of the long and short versions. *Frontiers in Psychology*, *9*, 1146.

CAPITULO 3

Metodología

Metodología

En este capítulo se tratará de dar una visión de conjunto de la metodología empleada para acometer los dos objetivos generales de la presente tesis doctoral, así como los objetivos específicos de cada estudio. Dado que gran parte de las estrategias analíticas de los seis estudios comparten elementos comunes, se presentan a continuación las dos metodologías principales utilizadas en cada uno de ellos: métodos para evaluar la comparabilidad de las medidas sobre VG de la encuesta del FRA y métodos para la construcción y validación de las escalas de actitudes de aceptabilidad de la VG (A-IPVAW, por sus siglas en inglés: *acceptability of intimate partner violence against women*), de culpabilización a las víctimas de VG (VB-IPVAW, por sus siglas en inglés: *victim blaming in cases of intimate partner violence against women*), y de disposición a intervenir en casos de VG (WI-IPVAW, por sus siglas en inglés: *willingnes to intervene in cases of intimate partner violence against women*).

Métodos para evaluar la comparabilidad de las medidas de violencia de género en la Unión Europea

Para poner a prueba que los ítems incluidos en las medidas de VG física, sexual, y psicológica, miden efectivamente el mismo constructo en los 28 países de la UE, se llevó a cabo un análisis de invarianza factorial. La invarianza factorial es un prerrequisito crucial en la investigación intercultural o entre países, pues permite establecer si las herramientas utilizadas para evaluar el constructo de interés, en este caso los diferentes tipos de VG, mantienen el mismo modelo factorial con los mismos parámetros estructurales (e.g., saturaciones factoriales, umbrales) en todos los grupos, en este caso los

diferentes países de la UE (Milfont & Fisher, 2010; Putnick & Bornstein, 2016). O en otras palabras, si los participantes de cada país conceptualizan e interpretan los ítems de la misma manera (Lubke, Dolan, Kelderman, & Mellenbergh, 2003). Si la invarianza factorial no se cumple, entonces no sería posible comparar los resultados obtenidos con las medidas del FRA, ya que las respuestas de los participantes de un país podrían estar evaluando un constructo diferente o dar como resultado puntuaciones completamente diferentes a las de los participantes de otro país (Davidov, Meuleman, Cieciuch, Schmidt, & Billie, 2014).

Antes de realizar este análisis, se llevó a cabo un estudio descriptivo de los ítems de la encuesta del FRA, obteniendo para ello las medias y desviaciones típicas de cada ítem, así como sus estadísticos de asimetría y kurtosis. Estos ítems miden la frecuencia de ocurrencia de diferentes conductas violentas (e.g., "¿Te ha abofeteado?", "¿Ha intentado forzarte para mantener relaciones sexuales sujetándote o haciéndote daño?", "¿Te ha menospreciado o humillado delante de otras personas?") con una escala Likert de cuatro puntos (1 = "Nunca", 2 = "Una vez", 3 = "De dos a cinco veces", 4 = "Más de seis veces"). Para realizar los análisis de los dos primeros estudios, estos ítems fueron dicotomizados ("0 = Nunca, 1 = Una o más veces"), mientras que en el tercer estudio se mantuvieron las categorías de respuesta originales. En esta etapa preliminar de análisis se obtuvieron, asimismo, los estadísticos de consistencia interna α de Cronbach y ω de McDonald de cada medida de VG además de las correlaciones corregidas entre cada ítem y el resto del test. Se considera normalmente que la consistencia interna de una escala es adecuada cuando los valores de α y ω son superiores a .70.

Invarianza factorial

Para poner a prueba la invarianza factorial de las medidas de VG física, sexual, y psicológica, y examinar cómo se distribuyen los niveles de estos tipos de violencia en los 28 países de la UE, fue necesario determinar en primer lugar la estructura latente de cada una de estas medidas. Para ello se realizó un análisis factorial confirmatorio (CFA, por sus siglas en inglés: *confirmatory factor analysis*) en cada país por separado. Dada la naturaleza categórica de los ítems, se utilizó como método de estimación mínimos cuadrados ponderados con medias y varianzas ajustadas (WLSMV, por sus siglas en inglés: *weighted least squares mean- and variance- adjusted*), pues este método resulta especialmente apropiado cuando se analizan ítems categóricos y con una fuerte asimetría (Li, 2016; Rhemtulla, Brosseau-Liard, & Savalei, 2012). La bondad de ajuste de los modelos resultantes se evaluó con una combinación de índices de ajuste: CFI, TLI, y RMSEA. Valores del CFI y TLI ≥ .95 son considerados como indicativo de un buen ajuste (Hu & Bentler, 1999), mientras que valores del RMSEA ≤ .08 y .06 se consideran, respectivamente, señal de un ajuste mediocre y excelente (MacCallum, Browne, & Sugawara, 1996).

Una vez establecida la estructura latente de cada medida de VG, se efectuó una serie de CFA multi-grupo (MG-CFA) entre los 28 países para poner a prueba diferentes niveles de invarianza factorial: invarianza configural, métrica y escalar. Estos niveles son modelos anidados, y cada uno de ellos añade nuevas restricciones a los ítems para asegurar su comparabilidad entre los distintos países (Milfont & Fischer, 2010; Vandenberg & Lance, 2000). El nivel de invarianza configural pone a prueba si la misma estructura factorial puede aplicarse a todos los países, sin imponer ninguna restricción en los parámetros del modelo. El nivel de invarianza métrica añade una restricción de igualdad entre las saturaciones factoriales, imponiendo que los valores de

estos parámetros sean iguales en todos los países. El nivel de invarianza escalar añade además una restricción de igualdad entre los umbrales de los ítems, fijando al mismo valor en todos los países tanto las saturaciones factoriales como los umbrales de los ítems. De esta manera se garantiza que los participantes de distintos países con un mismo patrón de respuestas en los ítems obtengan como resultado la misma puntuación en el factor. Si se cumple el nivel de invarianza escalar, entonces es posible comparar las medias de los factores de las medidas de VG entre los diferentes países de la UE.

Siguiendo las recomendaciones de Cheung y Rensvold (2002) y Chen (2007), para evaluar la bondad de ajuste de los MG-CFA, se calculó el cambio en los índices de ajuste CFI (ΔCFI) y RMSEA (ΔRMSEA). Sin embargo, estas recomendaciones fueron desarrolladas para datos de naturaleza continua y para modelos estimados mediante procedimientos de estimación basados en máxima verosimilitud, por lo que la interpretación de estos índices de ajuste debe de hacerse con extrema cautela cuando se emplean datos categóricos (Sass, Schmidt, & Marsh, 2014). Por esta razón se utilizaron los puntos de corte propuestos por Meade, Johnson, y Brady (2008), pues su aproximación es más conservadora a la hora de evaluar el cambio en los índices de ajuste: ΔCFI ≤ .002 y ΔRMSEA ≤ .007. El desempeño de estos puntos de corte con datos categóricos suele ser similar a los de los procedimientos basados en máxima verosimilitud cuando el tamaño muestral es elevado y las respuestas de los ítems no se distribuyen normalmente (Sass, Schmidt, & Marsh, 2014).

Después de establecer la invarianza factorial de las diferentes medidas de VG de la encuesta del FRA, se utilizaron las medidas del modelo invariante para poner a prueba la validez basada en sus relaciones con otras variables en todos los países (Asociación Americana de Investigación en Educación [*American Educational Research Association* [AERA], *American Psychological Association* [APA], & *National Council on Measurement in Education* [NCME],

1999). Para ello se obtuvieron las correlaciones entre los factores de VG física, sexual y psicológica. Se compararon además las diferencias en los niveles de cada tipo de violencia entre los participantes de la encuesta por salud percibida, experiencias de abuso infantil y nivel de ingresos, todo ello mediante un ANOVA de un factor. El tamaño del efecto de estas comparaciones se evaluó con el estadístico η^2 parcial, indicando valores superiores a .01, .06 y .14 tamaños del efecto pequeños, moderados o grandes, respectivamente (Miles & Shevlin, 2001).

Finalmente se compararon las medias factoriales de VG física, sexual y psicológica entre los 28 países a través de un análisis de medias latentes. Este procedimiento tiene en cuenta los diferentes pesos de cada ítem (i.e., las saturaciones factoriales) para medir el constructo y todas las restricciones de los análisis de invarianza factorial, obteniendo así una comparación más robusta y sofisticada de los niveles de los distintos tipos de VG entre todos los países de la UE. La magnitud de estas comparaciones entre países fue evaluada mediante la d de Cohen. Valores de este estadístico superiores a .20, .50 y .80 son indicativo, respectivamente, de un tamaño del efecto pequeño, moderado o grande (Cohen, 1988). Además de la d de Cohen, se calculó la U_3 de Cohen, un estadístico que refleja el porcentaje de casos de un país que presentan mayores niveles de VG que el promedio de otro (Ruscio, 2008; Hanel & Mehler, 2019).

Funcionamiento diferencial de los ítems

Adicionalmente, en el tercer estudio de esta tesis doctoral se efectuó un análisis de funcionamiento diferencial de los ítems (DIF) para datos categóricos para las medidas de VG física y sexual entre España y Suecia. Al igual que la invarianza factorial, el análisis DIF evalúa si hay un efecto del grupo (i.e., país) en las puntuaciones factoriales de ambos tipos de VG. Sin embargo, hay una sutil diferencia entre la invarianza factorial y el análisis DIF:

mientras que el primero pone el énfasis en la igualdad de los parámetros estructurales entre países (i.e., saturaciones factoriales, umbrales de los ítems), el segundo lo pone en la equivalencia de las puntuaciones factoriales de cada medida.

El análisis DIF se realizó mediante el método de regresión logística (Choi, Gibbons, & Crane, 2011; French & Miller, 1996). Un ítem presenta DIF cuando la probabilidad de elección de una de sus categorías no es la misma para los participantes de los distintos grupos (i.e., países) con la misma puntuación en el factor, indicando por lo tanto que los participantes de cada grupo están respondiendo a los ítems de manera diferencial.

Métodos para el desarrollo y validación de nuevas escalas

Para desarrollar y validar las diferentes medidas de actitudes públicas hacia la VG presentadas en esta tesis, se siguieron los estándares planteados por la APA, AREA, & NCME (1999) para la construcción de test. En estos estándares se identifican cinco fuentes de evidencia de validez que puede cumplir una herramienta de evaluación psicológica: evidencias de validez basadas en el contenido del test, en los procesos de respuesta, en la estructura interna, en las relaciones con otras variables, y en las consecuencias de la evaluación. Los análisis acometidos en el cuarto, quinto y sexto estudios tratan de proporcionar evidencias en al menos tres de estas fuentes de validez.

Evidencias de validez basadas en el contenido del test

Las evidencias de validez basadas en el contenido del test engloban todos aquellos análisis orientados a comprobar la operativización del constructo de interés, así como la relevancia y pertinencia del contenido de los ítems que lo componen (Messick, 1989). Con este fin, para desarrollar los ítems de las escalas A-IPVAW, VB-IPVAW, y WI-IPVAW, se partió del conjunto de

ítems identificados por Gracia y Lila (2015) en su revisión sobre las actitudes públicas en la UE. En concreto, el conjunto de ítems inicial fue de 102 para la escala de aceptabilidad, de 60 para la escala de culpabilización a las víctimas, y de 96 para la escala de disposición a intervenir. Estos ítems fueron traducidos al inglés de su lenguaje original por los expertos europeos en el campo de la VG, y traducidos al castellano por el equipo de investigación.

A continuación, se solicitó a un panel de seis expertos en VG que evaluasen la relevancia y la claridad de cada conjunto de ítems (Beck & Gable, 2001; Delgado-Rico, et al., 2012; Lynn, 1986; Polit & Beck, 2006). Para ello, los expertos utilizaron una escala Likert de 5 puntos para indicar su relevancia (e.g., "¿Es este ítem relevante para medir actitudes de aceptabilidad/culpabilización a las víctimas/disposición a intervenir en casos de VG?"; 1 = "Muy en desacuerdo", 5 = "Muy de acuerdo"). Finalmente se seleccionaron aquellos ítems que fueron valorados de media por los expertos por encima de 4 (i.e., la categoría "de acuerdo"). En concreto, se seleccionaron 20 ítems para la escala A-IPVAW, 14 para la escala de VB-IPVAW —si bien posteriormente se eliminaron dos de ellos, dejando la escala completa en 12 ítems—, y 28 para la escala WI-IPVAW. El formato de respuesta de los ítems seleccionados fue una escala Likert de 3 puntos en el caso de la escala A-IPVAW (i.e., 1= "Nada aceptable", 3 = "Aceptable"), de 4 puntos para la escala VB-IPVAW (i.e., 1 = "Muy en desacuerdo", 4 = "Muy de acuerdo"), y de 6 puntos para la escala WI-IPVAW (i.e., 1 = "Nada probable"; 2 = "Extremadamente probable").

Evidencias de validez basadas en la estructura interna

Los análisis dirigidos a evaluar la estructura latente y la consistencia interna de cada escala se consideran una fuente de validez basada en la estructura interna del test, pues gracias a ellos es posible determinar la

precisión de las medidas obtenidas con cada escala, el número de factores que las componen, y si los constructos que evalúan tienen o no respaldo empírico y sentido teórico (APA, AREA, NMCE, 1999).

Análisis descriptivo de los ítems y consistencia interna

Para determinar la distribución de las respuestas de los ítems, se obtuvieron primero sus medias, desviaciones típicas, el mínimo, el máximo, y los estadísticos de asimetría y kurotsis. Este paso previo es importante ya que al ser los ítems variables de respuesta ordenada, distribuciones muy asimétricas o con una pronunciada kurtosis pueden condicionar la elección del método factorial para analizar la estructura latente de cada escala (Schmitt, 2011).

Con respecto a la consistencia interna, se calcularon los α de Cronbach para cada una de las escalas, así como las correlaciones corregidas entre cada ítem y el resto del test. Estos análisis permiten determinar el grado en que cada ítem contribuye a la consistencia interna de su escala, y si es necesario eliminar alguno de ellos.

Estructura latente

En los tres últimos estudios de esta tesis se emplea un diseño de validación cruzada para determinar la estructura latente de cada una de las escalas de actitudes públicas hacia la VG. Es decir, se dividió la muestra en dos mitades para llevar a cabo un análisis factorial exploratorio (EFA, por sus siglas en inglés: *exploratory factor analysis*) en la primera mitad, y replicarlo en la segunda mitad mediante un CFA.

Antes de realizar el EFA, se puso a prueba la idoneidad de las matrices de datos de cada escala mediante el test de esfericidad de Bartlett y el estadístico Kaiser-Meyer-Olkin (KMO). A continuación, se efectuó un análisis

paralelo basado en el análisis factorial de rangos mínimos, utilizando la matriz de correlaciones policóricas (Timmerman & Lorenzo-Seva, 2011). Este método ha demostrado ser uno de los mejores procedimientos para determinar el número de factores que es necesario extraer en un EFA con datos categóricos (Garrido, Abad, & Ponsoda, 2013; 2016). El análisis paralelo utiliza simulación por Monte Carlo para generar matrices de datos aleatorios similares a la matriz de datos empíricos. Se utiliza entonces el análisis factorial de rangos mínimos para obtener el porcentaje de varianza explicado por diferentes modelos factoriales (i.e., modelo de un factor, de dos factores, de tres factores, etc…) en las matrices de datos aleatorios. Se obtuvieron asimismo la media y el percentil 95 de estos porcentajes de varianza explicada. Cuando el porcentaje de varianza explicada por un modelo factorial determinado en la matriz de datos empíricos está por debajo del porcentaje esperado para ese mismo modelo en las matrices simuladas, se asume entonces que ese modelo está añadiendo más factores de los necesarios. De este modo es posible establecer el número mínimo de factores necesarios para explicar más varianza de la que sería esperable por azar en la matriz de datos empíricos de cada escala.

A continuación, se realizó un EFA utilizando WLSMV como método de estimación, ya que es capaz de estimar de manera más consistente los parámetros del modelo factorial cuando se emplean datos categóricos (Li, 2016; Rhemtulla, Brosseau-Liard, & Savalei, 2012). La bondad de ajuste del modelo se evaluó mediante la misma combinación de índices de ajuste y sus puntos de corte que en los análisis de invarianza factorial de los estudios anteriores, esto es: CFI y TLI ≥ .95, y RMSEA ≤ .06 (Hu & Bentler, 1999; MacCallum, Browne, & Sugawara, 1996).

Se replicaron acto seguido los resultados obtenidos en el EFA mediante un CFA para cada escala. El CFA se llevó a cabo de nuevo utilizando el WLSMV

como método de estimación, así como la misma combinación de índices de ajuste (CFI y TLI ≥ .95; RMSEA ≤ .06).

Teoría de Respuesta al Ítem

En el cuarto y quinto estudios de esta tesis, además de los análisis factoriales, se estimó un modelo TRI para las escalas A-IPVAW y VB-IPVAW. Los modelos TRI tienen dos ventajas importantes sobre los modelos factoriales al emplearse datos categóricos: incluye estimadores de las puntuaciones factoriales más robustos, y no asume que la precisión (i.e., fiabilidad) de cada medida es constante a lo largo de los diferentes niveles del rasgo latente —en este caso, los diferentes tipos de actitudes públicas hacia la VG. La combinación de estos dos aspectos permite a los investigadores identificar qué niveles del rasgo latente son mejor evaluados por sus instrumentos mediante la función de información (Embrestone & Reise, 2000; Chernyshenko, Stark, Chan, Drasgow, & Williams, 2001).

Dado el formato de respuesta ordinal de los ítems de las escalas A-IPVAW y VB-IPVAW se empleó el modelo de respuesta graduada (Samejima, 1969). Este modelo fue estimado mediante el algoritmo Metropolis-Hastings Robbins-Monroe (Cai, 2010). La bondad de ajuste del modelo resultante fue evaluada mediante el estadístico M_2 para variables ordinales de Maydeu-Olivares y Joe (2006), ya que fue específicamente desarrollado para evaluar el ajuste general de los modelos de la TRI (Maydeu-Olivares & Garcia-Forero, 2010).

Evidencias de validez basadas en la relación con otras variables

Las evidencias de validez basadas en la relación con otras variables incluyen principalmente dos fuentes de validez clásicas: validez referida al criterio y validez concurrente (APA, AREA, NMCE, 1999; Messick, 1989). La

validez referida al criterio engloba todo análisis dirigido a evaluar si el constructo de interés se relaciona de manera esperada con otros constructos teóricamente relevantes. Por su parte, la validez concurrente, incluye aquellos análisis cuyo objetivo es o bien comparar grupos de participantes que teóricamente debieran mostrar puntuaciones significativamente diferenciadas en el constructo de interés, o bien relacionar diferentes medidas alternativas del mismo constructo.

Para evaluar la validez referida al criterio de las tres escalas de actitudes públicas hacia la VG desarrolladas en esta tesis doctoral, se correlacionaron las puntuaciones TRI de los participantes —o las puntuaciones factoriales en el caso de la escala de disposición a intervenir— con sus puntuaciones en severidad percibida de la VG, en sexismo hostil y en sexismo benevolente. Se obtuvieron asimismo las correlaciones entre las tres escalas. Se compararon además por sexo y por edad los niveles de los participantes de aceptabilidad de la VG, culpabilización a las víctimas de VG, y disposición a intervenir en casos de VG. Adicionalmente, en el cuarto y quinto estudio de esta tesis se evalúa la validez concurrente de las escalas A-IPVAW y VB-IPVAW, comparando las puntuaciones TRI de los participantes de la muestra general con las puntuaciones de una pequeña muestra de 50 agresores condenados por VG.

Métodos para la reducción de escalas

Finalmente, en el quinto y sexto estudios de esta tesis doctoral se presenta una versión breve de cinco ítems de la escala VB-IPVAW y dos versiones breves de nueve y seis ítems de la escala WI-IPVAW. En ambos estudios se siguieron las pautas recomendadas por Goetz et al. (2013). Para ello, se seleccionaron los ítems más relevantes atendiendo a su consistencia interna —y a su función de información en el caso de la escala VB-IPVAW—,

los modelos factoriales estimados, y a las evaluaciones de los expertos sobre el contenido de los ítems. Se examinaron a continuación las propiedades psicométricas de los ítems seleccionados para estas versiones breves, obteniendo así sus estadísticos de consistencia interna y sus correlaciones con otras medidas de actitudes públicas hacia la VG (e.g., aceptabilidad de la VG, severidad percibida de la VG), y con medidas de sexismo hostil y benevolente.

CHAPTER 4

Studies

Study 1

Ensuring the comparability of cross-national survey data on intimate partner violence against women: A measurement invariance study in the European Union

Ensuring the comparability of cross-national survey data on intimate partner violence against women: A measurement invariance study in the European Union[*]

Manuel Martín-Fernández, Enrique Gracia, & Marisol Lila
Department of Social Psychology, University of Valencia, Valencia, Spain

Abstract

Objectives To ensure the cross-national comparability of the set of questions addressing physical and sexual intimate partner violence against women (IPVAW) included in the European Union (EU) Agency for Fundamental Rights (FRA) survey. Once the measurement invariance of these measures is established, we aim to make appropriate and valid comparisons of the levels of physical and sexual IPVAW across the EU countries.

Design Cross-sectional, population-based study.

Participants Data were drawn from the survey conducted by the FRA on violence against women, including the responses of 42,002 adult women from the 28 countries of the EU.

Main outcome measures The set of questions addressing life-time prevalence of physical and sexual IPVAW used in the FRA survey. The psychometric properties (i.e., reliability and validity) of these measures were examined, as well as their latent structure and their measurement invariance across the 28 EU countries.

Results The physical and sexual IPVAW measures presented adequate internal consistency and validity evidence based on their relations to other variables in all countries. A latent two-factor structure was supported and scalar invariance was established across countries. Our results showed that the average levels of physical and sexual IPVAW were higher in countries like Denmark, Finland, Sweden and United Kingdom compared to the rest of the EU countries. In many of the other countries the levels of these types of violence overlapped, especially in the case of sexual IPVAW.

Conclusions The findings of this study underline the importance of using appropriate statistical methods to make valid cross-national comparisons in large socio-demographic surveys. The measurement invariance approach allowed us to conduct a more robust and refined analysis of how levels of physical and sexual IPVAW were distributed across all EU countries.

Strengths and limitations of this study

- This paper is the first to examine the cross-national comparability of survey data addressing physical and sexual intimate partner violence against women (IPVAW) in all European Union (EU) countries

- Appropriate analyses were applied to examine the psychometric properties, latent structure, and the measurement invariance of the measures included in a large socio-demographic survey on IPVAW

- A latent means analysis was conducted to compare the levels of physical and sexual IPVAW across all EU countries, an approach that takes into account the latent structure of the IPVAW measures and the magnitude of the contribution of each item to the measured construct

- The cross-sectional design of the survey did not allow for testing measurement invariance across different periods of time

Intimate partner violence against women (IPVAW) has been globally recognized as a major public health problem of epidemic proportions.[1] IPVAW is the form of violence most commonly suffered by women,[2-4] and it has severe physical and psychological consequences not only for the well-being of victims and their children, but for society as a whole.[5-9]

According to the survey conducted by the European Union (EU) Agency for Fundamental Rights (FRA), the estimated prevalence of physical and sexual IPVAW in the EU is 22%, varying across countries from 13% to 32%.[10] One of the main strengths of this survey is that women from the 28 EU countries answered the same set of questions assessing different types of IPVAW. However, the measurement equivalence of these questions across the EU countries has not yet been tested. This hampers the generalizability of these cross-national comparisons, as it not possible to ascertain whether the differences in IPVAW prevalence across all EU member states reflect actual differences between countries, or whether they are the result of cultural expectations or beliefs about intimate partner violence that may affect the interpretation of the FRA survey questions.

IPVAW can be a culturally sensitive issue, and how it is perceived, conceptualized and interpreted may vary across countries. Thus, to make valid cross-national comparisons based on survey data addressing IPVAW, it is important to test their measurement invariance.[11-13] Measurement invariance is a crucial prerequisite in cross-national research, since it allows meaningful comparisons across countries ruling out the possibility of cultural bias in the respondents' answers.[14] When measurement invariance is not supported, it means that respondents from different countries interpret and respond differently to the questions, and thus their scores cannot be directly compared.[15] [16] If that is the case, it cannot be assumed that IPVAW is interpreted in the same way in all EU member states, since the same score in

one country may reflect a different construct or yield different levels of IPVAW in another. Therefore, computing prevalence rates to compare samples from different countries without first ensuring measurement invariance could lead to unreliable and distorted conclusions, as the validity of such comparisons may be questionable.[17]

The main objective of this study is to ensure the cross-national comparability of the set of questions addressing physical and sexual IPVAW included in the FRA survey, by assessing whether respondents of each country conceptualize and interpret these questions in the same manner. Once the measurement invariance of these measures is established, we aim to make appropriate and valid comparisons of the physical and sexual IPVAW levels across the EU countries.

Method

Participants

The data were drawn from the FRA survey on violence against women.[10] This survey includes the responses of 42,002 adult women from the 28 EU countries who were currently or had previously been in an intimate relationship. The responses were collected in person through structured interviews and self-reports, following a two-stage clustered stratified sampling design with equal probability of selection of households within clusters. Additional information on sample collection and procedures can be found in the FRA survey technical report.[18] A special license for secondary data analysis was requested and granted from FRA (Reference No. 102577).

The sample used in this study was selected using the answers from respondents who did not omit any of the survey questions addressing physical and sexual intimate partner violence. The final sample is composed of 39,403

Table 1. Sociodemographic characteristics of the sample in each country (N = 39403)

		Age (%)				Self-perceived Health		Exp. of child abuse (%)	Income (%)			
	N	18-24	25-34	35-59	60-74	M	SD		Q1	Q2	Q3	Q4
Austria	1403	7.1	20.0	48.5	24.4	1.77	0.89	30.4	38.5	31.5	21.0	9.0
Belgium	1456	3.5	11.3	52.4	32.8	1.94	0.92	26.7	5.0	28.0	23.1	43.9
Bulgaria	1424	12.5	20.9	44.0	22.7	1.79	0.86	28.0	34.6	23.0	26.4	16.0
Croatia	1438	10.6	16.6	51.2	21.6	2.03	0.87	32.3	11.1	25.0	37.3	26.6
Cyprus	1390	5.2	14.8	53.4	26.6	2.12	0.90	12.4	33.7	28.7	22.9	14.7
Czech Rep.	1566	5.2	13.5	44.5	36.8	2.35	1.01	32.6	13.3	37.7	34.4	14.6
Denmark	1453	7.8	17.2	49.9	25.1	2.13	1.09	42.7	26.1	22.4	28.4	23.1
Estonia	1360	18.4	25.9	45.6	10.1	1.65	0.81	48.3	7.9	27.8	45.6	18.7
Finland	1457	10.4	16.6	48.1	24.9	2.07	0.90	52.4	31.3	29.6	22.9	16.2
France	1410	5.7	14.6	57.9	21.8	2.15	0.78	44.4	57.2	32.6	7.3	3.0
Germany	1478	12.0	15.3	48.5	24.1	1.85	0.91	43.0	26.6	33.9	24.9	14.6
Greece	1425	6.9	14.1	46.2	32.9	2.52	0.92	23.3	27.1	22.5	27.4	23.1
Hungary	1451	6.9	15.3	45.2	32.6	2.07	0.81	24.4	21.8	23.3	25.8	29.1
Ireland	1428	6.5	15.8	49.3	28.4	1.92	0.95	28.4	37.9	8.7	9.3	44.2
Italy	1483	7.5	15.9	48.5	28.1	2.63	0.85	34.1	31.8	24.2	21.4	22.5
Latvia	1440	6.0	13.6	55.5	24.8	2.03	0.85	31.9	49.8	32.2	10.8	7.2
Lithuania	1308	6.2	14.1	54.7	25.0	2.13	0.83	17.7	24.2	25.3	29.5	20.9
Luxembourg	889	8.1	11.5	51.1	29.2	2.57	0.85	43.3	37.6	19.7	19.9	22.8
Malta	1365	8.8	16.7	50.8	23.6	1.69	0.89	24.2	21.4	42.3	31.9	4.4
Netherlands	1451	6.2	14.3	57.4	22.0	2.19	0.93	32.9	31.4	25.2	23.0	20.3
Poland	1405	5.2	14.3	50.8	29.7	2.05	0.67	17.8	31.2	39.8	21.2	7.7
Portugal	1424	7.7	15.0	46.4	30.9	2.44	0.98	28.5	20.1	29.2	24.0	26.8
Romania	1465	4.4	12.8	58.4	24.4	2.15	0.77	22.6	19.2	28.8	26.4	25.7
Slovakia	1287	10.3	23.1	46.7	20.0	2.25	1.00	33.9	10.7	24.6	30.3	34.4
Slovenia	1372	5.0	12.5	47.5	34.9	2.61	0.92	13.2	21.0	41.1	22.3	15.6
Spain	1439	9.2	16.6	48.5	25.6	2.42	1.03	29.4	28.7	21.4	20.3	29.6
Sweden	1474	7.5	18.8	52.1	21.6	2.15	0.98	43.8	20.0	23.3	26.1	30.6
UK	1462	6.6	13.7	51.6	28.0	2.13	0.91	37.8	14.7	31.9	33.7	19.6

Note: The percentages and descriptive statistics are unweighted. M = Mean; SD = Standard Deviation. Income: Q1 = under lowest quartile, Q2 = between lowest quartile and median, Q3 = between median and highest quartile, Q4 = above highest quartile

women, aged from 18 to 74 years old, from the 28 EU countries. Socio-demographic information of the sample by country is shown in Table 1.

Patient Involvement. No patient involved.

Measures

Physical violence. Intimate partner physical violence is addressed in the FRA survey with 10 items describing episodes of physical violence perpetrated by either the current or any previous partner (e.g., "Has your current/previous partner ever slapped you?"). Respondents were asked to indicate how often they had experienced that episode on a 4-point Likert-type scale (1: "Never", 2: "Once", 3: "2-5 times", 4: "6 or more times"). The responses to these items were dichotomized (0: "Never", 1: "Once or more times"), as in some countries the frequencies of the upper categories were very low (i.e., less than 2% when the responses to the "2-5 times" and "6 or more times" categories are collapsed in Austria, Cyprus, Greece, Italy, Latvia, Malta, Slovakia, and Slovenia).

Sexual Violence. The FRA survey included four items addressing sexual violence committed by the current or any previous partner (e.g., "Has your current/previous partner made you take part in any form of sexual activity when you did not want to or you were unable to refuse?"). The response format of the items was a 4-point Likert-type scale indicating frequency (1: "Never", 2: "Once", 3: "2-5 times", 4: "6 or more times"). The responses to these items were also dichotomized (0: "Never", 1: "Once or more times").

Validity evidence based on relations to other variables

To test validity evidence of the physical and sexual IPVAW measures based on relations to other variables,[19] the following variables were used:

Self-perceived health. The respondents were asked to respond about their health in general at the beginning of the interview using a 5-point Likert-type graded scale (from 1 = "Very Bad" to 5 = "Very Good").

48

Experiences of child abuse. The FRA survey included a set of 11 questions addressing experiences of childhood physical and sexual abuse before the age of 15 (e.g., "did an adult who was 18 years or over hit you very hard so that it hurt?", "did an adult who was 18 years or over expose their genitals to you?"). If any of these questions were answered affirmatively, we considered that the respondent has experienced child abuse during their childhood.

Income. Reported income was measured in each country as household income quartiles (i.e., "below lowest quartile", "between lowest quartile and median", "between median and highest quartile", "above highest quartile").

Statistical Analyses

The main purpose of this study was to ensure the measurement equivalence of the physical and sexual violence measures of the FRA survey across the 28 EU member states in order to make appropriate and valid comparisons between these countries. To this end, the psychometric properties (i.e., reliability and validity) and the latent structure of these measures was assessed, and then measurement invariance was tested across the 28 EU countries.

First, the mean, standard deviation, skew, and kurtosis statistics were obtained. The internal consistency of the physical and sexual violence measures in each country was evaluated by computing the Cronbach's α and McDonald's ω, and using the correlation of each item with the rest of its set of questions (i.e., item-test corrected correlation).

Secondly, a confirmatory factor analysis (CFA) was conducted to determine the latent structure of the two IPVAW measures in each country. Two models were compared: a one-factor model, in which all the items loaded onto a single intimate partner violence factor, and a two-factor model, setting one factor for the physical violence items and another one for the sexual

violence items. Given the categorical nature of the items, weighted least squares with adjusted means and variances (WLSMV) was used as estimation method.[20] Model fit was evaluated using a combination of fit indices: CFI, TLI, and RMSEA. CFI and TLI values ≥ .95, and RMSEA values ≤ .06 are indicative of good fit.[21][22]

After establishing the factor model, a series of multi-group confirmatory factor analyses (MG-CFA) were carried out, testing three different levels of measurement invariance: configural, metric, and scalar.[12][23] Configural invariance applies the same factor model to all groups (i.e., countries), with no equality constraints for any parameters, ensuring that the same conceptualization of the construct is supported across countries. Metric invariance specifies that the values of the factor loadings are equal across countries, implying that each item is contributing equally and having a similar degree of importance on the factor in all the countries. Scalar invariance holds both factor loadings and item thresholds to be invariant, ensuring that respondents from different countries with the same response pattern on the items will yield the same factor score. If scalar invariance is supported, then the scores on the physical and sexual IPVAW measures can be compared across countries.

To assess which of these invariance levels was best supported by the data, the change in the CFI (ΔCFI) and RMSEA (ΔRMSEA) fit indices was computed, following the general guidelines of Cheung and Rensvold.[24] These guidelines, however, were developed for continuous data, and as Saas, Schmidt and Marsh noted,[25] the interpretation of these fit indices should be taken with caution in the case of categorical data. For this reason, instead of Chen's usual cut-off values for the changes in CFI and RMSEA (i.e., ΔCFI ≤ .010 and ΔRMSEA ≤ .015),[26] we used the cut-offs values proposed by Meade, Johnson, and Brady: ΔCFI ≤ .002 and ΔRMSEA ≥ -.007.[27] These cut-offs are more restrictive and tend

to perform as well as maximum-likelihood based procedures when the sample size is large and the items are not normally distributed.[25]

Having established an invariant factor model, validity evidence based on relations to other variables was tested using the invariant factor scores of the IPVAW items. A one-way ANOVA was computed for both physical and sexual violence measures of the FRA survey, testing differences by self-perceived health, experiences of child abuse, and income. The size effect of the variables was assessed with the partial eta-squared statistic, using values above .01, .06, and .14, as indicative of small, medium, and large size effects, respectively.[28]

Finally, after determining that both physical and sexual IPVAW measures were psychometrically sound and share an invariant latent structure in all EU countries, the factor means of physical and sexual IPVAW were compared across countries conducting a latent means analysis. This procedure is more appropriate and statistically sophisticated than simply computing the average prevalence for each country, as it does not assume that all the items have the same relevance to assess the construct and, moreover, it takes into account all the constraints for the invariant measurement model.[16 17 23] The magnitude of these cross-national comparisons was evaluated using Cohen's d statistic, indicating d values above .20, .50, and .80, small, medium and large size effects, respectively.[29] This statistic can be used to obtain the Cohen U_3 statistic, which indicates the percentage of cases of one country that is higher than the average of another.[30-32] The Cohen's d and U_3 statistics of the comparisons between each pair of countries can be found in the supplementary material.

Descriptive analyses, internal consistency, and validity analyses were conducted using the statistical package R and the *psych* library.[33 34] CFA and MG-CFA analyses were computed using *Mplus 7*.[35]

Results

Descriptive Analysis and Internal Consistency

Descriptive statistics of the physical and sexual violence measures are presented in Table 2. All the items showed mean values close to zero, and high skew and kurtosis values, especially in the sexual violence measure, indicating that most of the respondents reported having no experience of the episodes described by the items. The corrected item-test correlations were in general high, pointing to a strong relationship between the items and the rest of the questions. Regarding the internal consistency, both measures showed an adequate internal consistency in the complete dataset, with alpha and omega values above .70. Separating the sample by countries yields similar results in both the physical violence and sexual violence measures (ranging, respectively, from α=. 84 in Sweden to α=. 90 in Ireland and Lithuania, and from α = .69 in Slovenia to α = .91 in the Netherlands).

The items 7 and 9 of the physical violence measure presented an extremely low variance (i.e., "being burned", and "being cut, stabbed or shot"), below 0.01. Given this lack of variability, we decided to remove these items from this measure. As a result, the Cronbach's alpha of the physical violence measure increased to .89 (ranging from α = .84 in Slovenia to α = .92 in Romania and Lithuania), and McDonald's omega increased to .92.

Confirmatory Factor Analysis

Two CFA solutions were tested separately in each country: a one-factor model, assuming that all the items load onto a single IPVAW factor, and a two-factor model, setting one separate factor for the physical and sexual violence items. The CFI and TLI indices were excellent for both the one- and two-factor models, with values around .98 and .99 in all 28 EU countries, respectively. However, the residuals for the one-factor model were above the RMSEA ≤ .06 cut-off criterion for a well-fitting model in Belgium, Demark, Finland, France,

Table 2. Psychometric properties of the Physical Violence and Sexual Violence measures.

	M	SD	Skew	Kurtosis	$r_{item\text{-}test}$
Physical Violence (Cronbach's α= .87, McDonald's ω = .90)					
Threatened to hurt you physically	0.14	0.35	2.05	2.18	0.70
Pushed you or shoved you	0.16	0.36	1.88	1.53	0.73
Slapped you	0.13	0.33	2.23	2.99	0.73
Thrown a hard object at you	0.06	0.24	3.66	11.43	0.65
Grabbed you or pulled your hair	0.08	0.28	3.03	7.18	0.74
Beat you with a fist or a hard object, or kicked you	0.07	0.25	3.50	10.26	0.72
Burned you	0.01	0.09	11.55	131.34	0.27
Tried to suffocate you or strangle you	0.03	0.18	5.25	25.60	0.53
Cut or stabbed you, or shot at you	0.01	0.09	11.37	127.23	0.29
Beat your head against something	0.04	0.18	5.03	23.32	0.60
Sexual Violence (Cronbach's α= .85, McDonald's ω = .88)					
Forced you into sexual intercourse by holding you down or hurting you in some way	0.04	0.19	4.93	22.34	0.73
Attempted to force you into sexual intercourse by holding you down or hurting you in some way	0.04	0.19	4.94	22.44	0.73
Mad you take part in any form of sexual activity when you did not want to or you were unable to refuse	0.04	0.19	4.94	22.42	0.66
Consented to sexual activity because you were afraid of what your current partner might do if you refused	0.05	0.21	4.28	16.35	0.68

Note: M = Mean; SD = Standard Deviation; $r_{item\text{-}test}$ = corrected item-test correlation. Skew and kurtosis standard error were below .01.

Germany, Latvia, the Netherlands, Portugal, Sweden, and United Kingdom (the highest was Denmark, with .077, and the closest to the cut-off was Portugal,

Table 3. Measurement invariance fit indices.

Model	χ^2	_df_	CFI	TLI	RMSEA [95% CI]
Configural	2461.08	1484	0.999	0.998	0.022 [0.020 — 0.023]
Metric	3819.28	1808	0.997	0.997	0.028 [0.027 — 0.029]
Scalar	5530.65	2078	0.995	0.996	0.034 [0.033 — 0.035]

Note: χ2 = adjusted chi-squared test for model fit; _df_ = degrees of freedom;
CFI = Comparative Fit Index; TLI = Tucker-Lewis Index;
RMSEA = Root Mean Squared Error of Approximation.

with .061), whereas in the two-factor model the RMSEA were below this cut-off (the highest was .031 in Denmark and the lowest, .011 in Poland). For this reason, we kept the two-factor model as the latent structure of the physical and sexual IPVAW measures.

Measurement Invariance

The analysis of measurement invariance supported the configural, metric, and scalar invariance levels for the physical and sexual violence measures across all EU countries (Table 3). Attending to the fit indices, when the loadings were fixed to have the same value across all countries they did not differ substantially from the configural model (ΔCFI = .002, ΔRMSEA = -.006), supporting the metric invariance level. In the same direction, constraining the item thresholds as well as the item loadings did not substantially reduce the fit of the model (ΔCFI = .002, ΔRMSEA = -.006), indicating that the scalar invariance level could hold.

The resulting item parameters are displayed in Figure 1. All the items presented high standardized loadings, indicating a strong relationship of the items to the factor. The correlation between factors varies across countries, as the factor covariance matrix was freed in each country. The value for this correlation was around .75 in most of the countries, ranging from .56 in Denmark to .91 in Croatia.
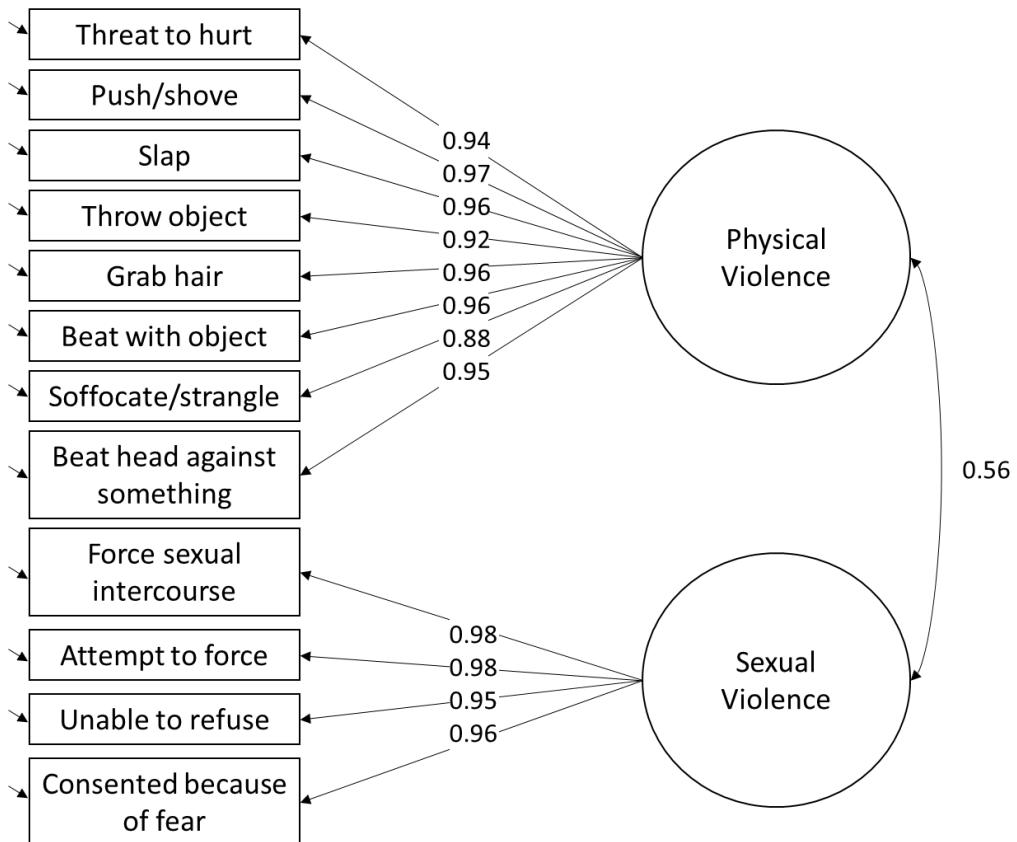
Figure 1. *Multi-group confirmatory factor analysis: scalar invariance model.*

Note: The standardized factor loadings belong to the reference country (i.e., Denmark). As in the standardized solution the factor variances are fixed to 1, there are slight differences in the decimals of the loadings in each country.

Validity evidence based on relations to other variables

The factor scores of the invariant model were used for the validity analyses in all countries. Significant and substantive differences were found in the physical IPVAW scores when self-perceived health ($F(4) = 94.0$, $p < .001$, $\eta^2 = .01$), experiences of child abuse ($F(1) = 984.7$, $p < .001$, $\eta^2 = .03$), and income ($F(1) = 127.3$, $p < .001$, $\eta^2 = .01$) were taken into account, showing that

women with higher scores on the physical IPVAW factor had lower self-perceived health levels, lower income and experiences of child abuse.

In the case of the sexual IPVAW factor scores, substantive differences were also found in self-perceived health ($F(4) = 82.6$, $p < .001$, $\eta^2 = .01$), experiences of child abuse ($F(1) = 311.4$, $p < .001$, $\eta^2 = .03$), and income ($F(1) = 113.4$, $p < .001$, $\eta^2 = .01$). These differences were in the same direction as in the physical IPVAW, as the women with higher scores on the sexual IPVAW factor presented lower self-perceived health levels, lower income, and experiences of child abuse.

Latent Means Analysis

After determining that the physical and sexual IPVAW measures are psychometrically sound, and after ensuring their measurement equivalence across all EU countries, the means of the physical and sexual violence factors can now be compared by conducting a MG-CFA. Denmark was used as the reference group, as it was one of the countries presenting the highest levels of physical and sexual IPVAW. The factor means for Denmark were fixed to zero and their variances fixed to one. In the rest of the countries these parameters were freely estimated. In this way, the estimated mean parameter of the standardized solution represents the difference in factor means between each country and Denmark. The resulting model presented a very good fit to the data (CFI = .997, TLI = .997, RMSEA = .027 [.026 - .029]). The standardized factor means for each country are displayed in Figures 2 and 3.

Although Denmark was the country with the highest levels of physical IPVAW, the differences between this country and Latvia ($z = -0.01$, $p = .850$, $d = 0.01$), Finland ($z = -0.04$, $p = .587$, $d = 0.04$), United Kingdom ($z = -0.12$, $p = .059$, $d = 0.13$), and Sweden ($z = -0.12$, $p = .115$, $d = 0.12$) were negligible, with Cohen's d values below .20. Large substantive differences were found between Denmark and Ireland ($z = -0.70$, $p < .001$, $d = 0.85$), Slovenia ($z = -0.75$, $p < .001$,
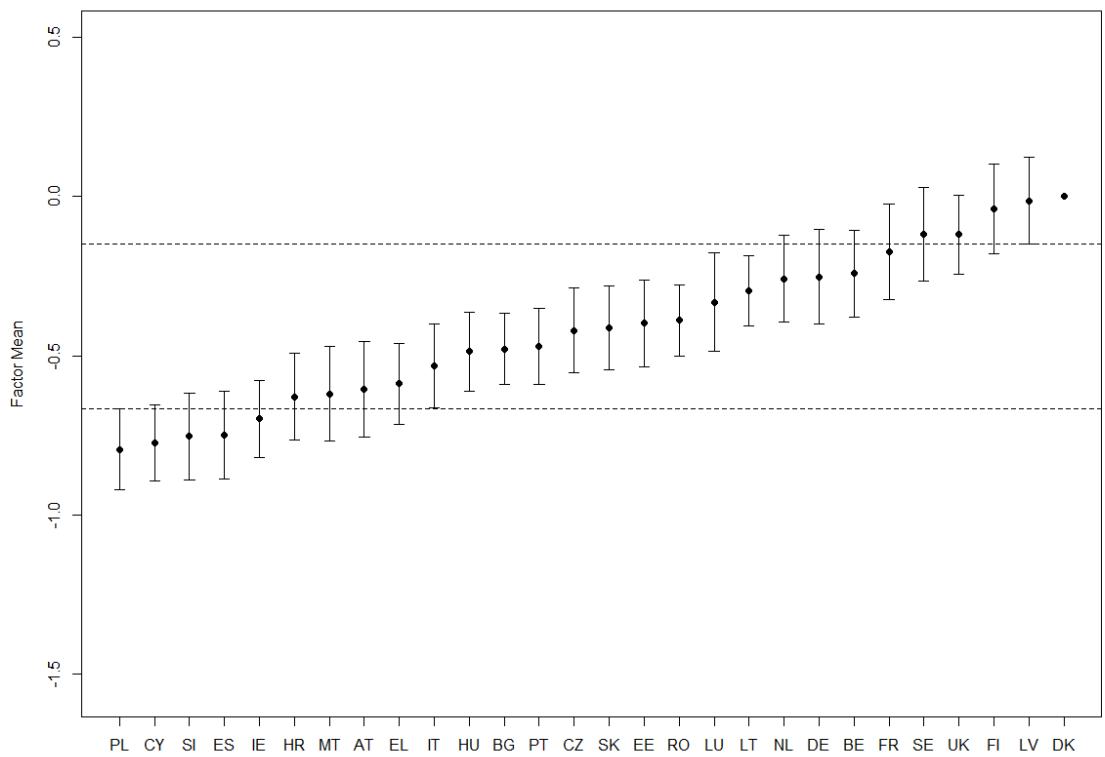
Figure 2. *Physical IPVAW latent means across the EU.*

Note: IPVAW: intimate partner violence against women. AT: Austria, BE: Belgium, BG: Bulgaria, CY: Cyprus, CZ: Czech Republic, DK: Denmark, EE: Estonia, FI: Finland, FR: France, DE: Germany, EL: Greece, ES: Spain, HR: Croatia, HU: Hungary, IE: Ireland, IT: Italy, LT: Lithuania, LV: Latvia, LU: Luxembourg, MT: Malta, NL: Netherlands, PL: Poland, PT: Portugal, RO: Romania, SE: Sweden, SI: Slovenia, SK: Slovakia, UK: United Kingdom. Denmark is the reference country.

d = 0.84), Spain (z = -0.75, p < .001, d = 0.86), Cyprus (z = -0.77, p < .001, d = 0.94), and Poland (z = -0.79, p < .001, d = 0.96).

It should be noted that the confidence intervals (CI) of the physical IPVAW factor means overlapped between many of the EU countries, indicating no significant differences between them. However, the countries could be grouped according to the confidence intervals. Thus, countries whose factor mean falls within the CI of Latvia showed on average higher levels of physical IPVAW (Denmark, Finland, United Kingdom, and Sweden). On the other hand,

countries with factor means falling within the CI of Poland—the country with the lowest factor mean—presented lower levels of physical IPVAW (Ireland, Slovenia, Spain, and Cyprus). The rest of the countries fell somewhere in between, with some countries like Lithuania or the Netherlands closer to the countries with higher levels of physical IPVAW, and others such as Greece or Italy closer to the countries with lower levels.

Regarding the sexual IPVAW measure, the countries with higher latent means in this factor were, along with Denmark, Finland ($z = 0.00$, $p = .986$, $d = 0.00$), Sweden ($z = -0.10$, $p = .537$, $d = 0.10$), Luxembourg ($z = -0.31$, $p = .082$, $d = 0.34$), Bulgaria ($z = -0.32$, $p = .072$, $d = 0.34$), and Slovenia ($z = -0.34$, $p = .130$, $d = 0.32$). Large differences in the sexual IPVAW factor were found between Denmark and Lithuania ($z = -0.87$, $p < .001$, $d = 1.06$), Croatia ($z = -0.88$, $p < .001$, $d = 1.04$), Austria ($z = -0.90$, $p < .001$, $d = 1.13$), Ireland ($z = -0.97$, $p < .001$, $d = 1.29$), Romania ($z = -0.98$, $p < .001$, $d = 1.24$), Poland ($z = -1.04$, $p < .001$, $d = 1.31$), Spain ($z = -1.07$, $p < .001$, $d = 1.32$), Greece ($z = -1.08$, $p < .001$, $d = 1.39$), Portugal ($z = -1.12$, $p < .001$, $d = 1.43$), and Cyprus ($z = -1.97$, $p < .001$, $d = 1.71$).

The CI of the sexual IPVAW factor means were, however, much wider than in the case of physical IPVAW, as this measure is comprised of only four items. As a result, most of the countries' factor means CI greatly overlapped. In this case the CI of Finland was used to group the countries that tend to present higher levels of sexual IPVAW (Denmark, Sweden, Belgium, Luxembourg, United Kingdom, and Slovenia), whereas the CI of Cyprus was used to delimit those countries with lower levels in this factor (Portugal, Greece, Spain, and Poland). The rest of the countries fall into the intermediate area, with some countries such as France and Germany nearer to the countries with higher levels of sexual IPVAW, and others like Romania and Ireland closer to the countries with lower levels of this type of violence. Italy was in the middle of the distribution, and was the only country whose CI did not overlap with the
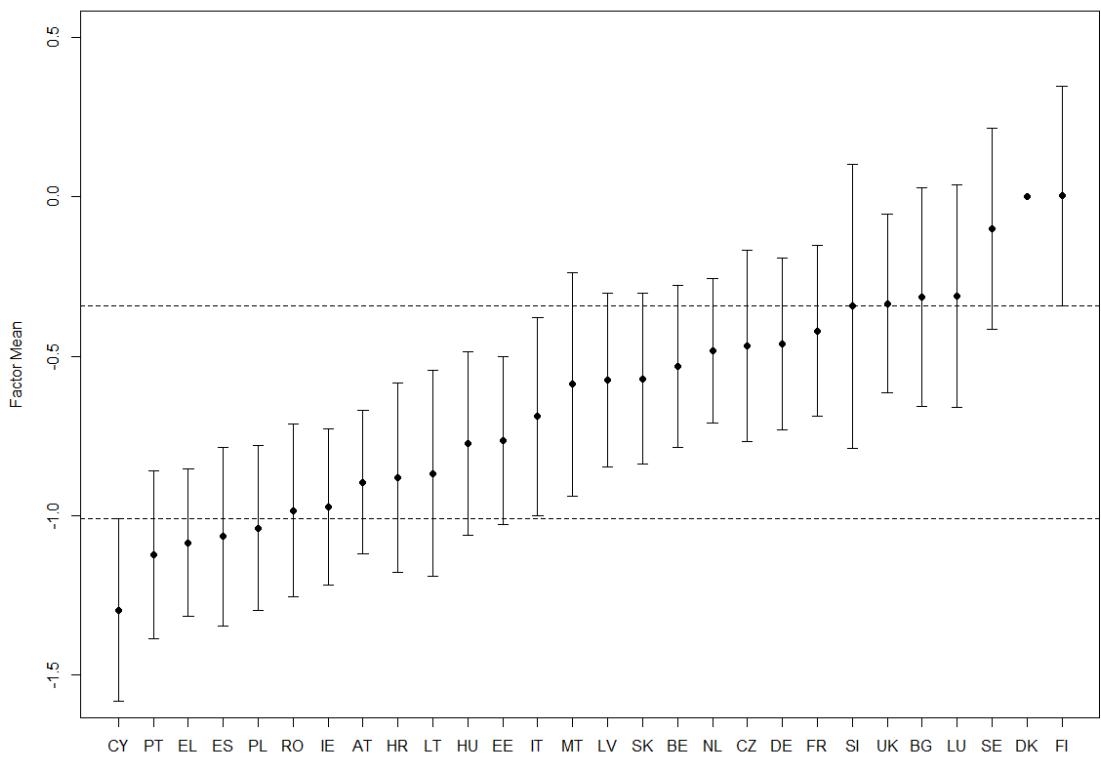
Figure 3. *Sexual IPVAW factor means across the EU.*

Note: IPVAW: intimate partner violence against women. AT: Austria, BE: Belgium, BG: Bulgaria, CY: Cyprus, CZ: Czech Republic, DK: Denmark, EE: Estonia, FI: Finland, FR: France, DE: Germany, EL: Greece, ES: Spain, HR: Croatia, HU: Hungary, IE: Ireland, IT: Italy, LT: Lithuania, LV: Latvia, LU: Luxembourg, MT: Malta, NL: Netherlands, PL: Poland, PT: Portugal, RO: Romania, SE: Sweden, SI: Slovenia, SK: Slovakia, UK: United Kingdom. Denmark is the reference country.

intervals for either Finland or Cyprus.

Comparisons of the physical and sexual IPVAW factor scores can also be made between each pair of countries. For instance, we compare Finland—one of the countries with highest levels of physical and sexual IPVAW—with the Netherlands and Spain—countries with moderate and low levels of both types of IPVAW, respectively. The differences between Finland and the

Netherlands were small in the case of physical IPVAW, showing that 59.9% of the Finnish sample had higher values in this factor than the average of the Dutch sample ($d = 0.24$, $U_3 = .595$). Medium differences were found in the case of sexual IPVAW, as 72.6% of the Finnish sample presented higher scores in the sexual IPVAW factor than the average of the Dutch sample ($d = 0.60$, $U_3 = .726$). When Finland is compared to Spain, large differences were found in both factors: 79.1% of the Finnish sample presented higher values in the physical IPVAW factor than the average of the Spanish sample ($d = 0.81$, $U_3 = .791$), and 90.7% in the case of sexual IPVAW ($d = 1.32$, $U_3 = .907$). The comparisons between each pair of countries together with their associated size effect can be found in the supplementary material.

Discussion

In this measurement invariance study, we conducted a set of analyses to ensure the cross-national comparability of the questions addressing physical and sexual IPVAW used in the FRA survey, and examined how physical and sexual IPVAW levels were distributed across all EU countries.

The first set of analyses aimed to study the psychometric properties (i.e., reliability and validity) of the questions addressing physical and sexual IPVAW across the EU. Our results showed that the measures of the FRA survey were measuring two different constructs: physical and sexual IPVAW. The two factors were related and showed an adequate internal consistency in all EU countries (with α and ω values above .70 in almost all countries). Regarding the validity evidence based on relations with other variables, we found, as expected, that women who reported having experiences of child abuse were more likely to show higher levels of physical and sexual IPVAW. This finding is consistent with previous research showing that women who were victimized in childhood have a higher risk of being victimized during adulthood.[36-41] In

addition, we found that the women presenting higher levels of physical and sexual IPVAW also reported lower levels of self-perceived health and income, a well-established finding in the IPVAW literature.[42-46]

The second set of analyses aimed to address the comparability of the physical and sexual IPVAW measures across the 28 EU countries, to eliminate the possibility of measurement bias. To do so, a series of MG-CFA analyses were conducted to test and establish measurement invariance. Although previous studies have acknowledged the difficulties in establishing the metric or scalar invariance levels when several groups are used,[14 47 48] we were able to demonstrate scalar invariance (i.e., equal factor loadings and thresholds) across the 28 EU countries. This may be due to the type of questions included in the FRA survey, as they are mostly behavioral oriented (e.g., being stabbed, cut, slapped, or being forced into sexual intercourse).

Once an invariant model had been established for the physical and sexual IPVAW measures, we were able to make proper comparisons between countries. Instead of computing the prevalence using the set of questions from the FRA survey, an approach that ignores the latent structure of the IPVAW measures and the magnitude of the contribution of each item to the construct, we conducted a latent means analysis, comparing the factor means of each country.[12 15 16] This is one of the main strengths of this study, and our findings suggest that there were almost no differences between many of the EU countries, since their confidence intervals overlapped considerably, implying that the levels of sexual and physical IPVAW were fairly similar between them. There were, however, substantial differences between the countries with higher levels of physical and sexual IPVAW and those with lower levels. In particular, countries such as Denmark, Finland, Sweden, and United Kingdom showed on average higher levels than countries like Cyprus, Poland, and Spain. These findings are in line with what is known as the Nordic paradox, as

Denmark, Finland, and Sweden—countries with the highest levels of gender equality in the EU—were among the countries with highest levels of physical and sexual IPVAW.[49] [50]

This study has some limitations. First, we used WLSMV as the estimation method for the measurement invariance analysis due to the asymmetrical distribution of the responses to the set of questions included in the FRA survey. According to Sass, Schmitt, and Marsh,[25] the use of fit indices with this method to test measurement invariance could lead to higher rates of Type I errors, assuming an invariant model when actually the instrument is non-invariant. To address this issue, we decided to use the cut-offs proposed by Meade et al. for the CFI and RMSEA,[27] as they are more restrictive than those of Chen.[26] Second, in this case, chi-squared based tests to compare the different levels of measurement invariance could not be used in conjunction with the fit indices, since this statistic is known to be sensitive to large sample sizes. Third, the cross-sectional design of the survey did not allow for testing measurement invariance across different periods of time. Fourth, the wide range of the confidence intervals of the physical and sexual IPVAW factor means suggest that the measures included in the FRA survey could be improved in order to obtain more accurate and reliable estimations of the IPVAW levels, especially in the case of sexual IPVAW.

Taken together, the results of this study underlined the importance of using appropriate statistical methods to account for the cross-national comparability of the measures included in large socio-demographic surveys. The measurement invariance approach allowed us to conduct a more appropriate and refined analysis of how physical and sexual IPVAW levels are distributed across the EU. This study is an important step towards a rigorous assessment of cross-national differences in physical and sexual IPVAW, and

further research is needed to evaluate the role that the country has, if any, in accounting for the differences across EU countries.

Declarations

Supplementary Material

The supplementary materials of the current version of the manuscript can be found at: http://tiny.cc/SupMat1

References

1. World Health Organization. Global and regional estimates of violence against women: Prevalence and health effects of intimate partner violence and non-partner sexual violence. Geneva: World Health Organization; 2013.
2. Devries KM , Mak JY , García-Moreno C , *et al.* The global prevalence of intimate partner violence against women. *Science* 2013;**340**:1527-8.
3. Garcia-Moreno C , Jansen HA , Ellsberg M , *et al.* Prevalence of intimate partner violence: findings from the WHO multi-country study on women's health and domestic violence. *Lancet* 2006;**368**:1260-9.
4. Stöckl, H., Devries, K., Rotstein, A., *et al.* The global prevalence of intimate partner homicide: a systematic review. *Lancet* 2013;**382**:859-65.
5. Campbell JC. Health consequences of intimate partner violence. *Lancet* 2002;**359:**1331-6.
6. Craparo G , Gori A , Petruccelli I , *et al*. Intimate partner violence: relationships between alexithymia, depression, attachment styles, and coping strategies of battered women. *J Sex Med* 2014;**11**:1484–94.
7. Ellsberg M , Jansen HA, Heise L, *et al.* Intimate partner violence and women's physical and mental health in the WHO multi-country study on women's health and domestic violence: an observational study. *Lancet* 2008;**371**:1165-72.

8. Guedes A , Bott S , Garcia-Moreno C , *et al.* Bridging the gaps: a global review of intersections of violence against women and violence against children. *Global Health Action* 2016;**9**:e31516.

9. Vilariño M , Amado BG , Vázquez MJ , Arce R. Psychological harm in women victims of intimate partner violence: Epidemiology and quantification of injury in mental health markers. *Psychosoc Interv* 2018;**27**:145–52.

10. European Union Agency for Fundamental Rights. Violence against women: An EU-wide survey. Main results. Luxembourg: Publications Office of the European Union; 2014.

11. Gregorich SE. Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Med Care* 2006;**44**:s78-94.

12. Milfont TL , Fisher R. Testing measurement invariance across groups: Applications in cross-cultural research. *Int J Psychol Res* 2010;**3**;111-30.

13. Vandenberg RJ , Lance CE. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organ Res Methods* 2000;**2**:4-69.

14. Davidov E , Meuleman B , Cieciuch J , *et al*. Measurement equivalence in cross-national research. *Annu Rev Sociol* 2014;**40**:55–75.

15. Jang S , Kim ES , Cao C , *et al*. Measurement invariance of the Satisfaction with Life Scale across 26 countries. *J Cross Cult Psychol* 2017;**48**:560–76.

16. Putnick DL , Bornstein MH. Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Dev Rev* 2016;**41**:71–90.

17. Van de Schoot R , Lugtig P , Hox J. A checklist for testing measurement invariance. *Eur J Dev Psychol* 2012;**9**:486-92.

18. European Union Agency for Fundamental Rights. Violence against women: an EU-wide survey: Survey methodology, sample and fieldwork. Luxembourg: Publications Office of the European Union; 2014.

19. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Standards for educational and psychological testing. Washington, DC: American Educational Research Association; 1999.

20. Beauducel A , Herzberg PY. On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Struct Equ Modeling* 2006;**13**:186-203.

21. Hu LT , Bentler PM. Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Modeling* 1999;**6**:1–55.

22. MacCallum RC , Browne MW , Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. *Psychol Methods* 1996;**1**:130–49.

23. Sass DA. Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *J Psychoeduc Assess* 2011;**29**:347–63.

24. Cheung GW , Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct Equ Modeling* 2002;**9**:233–55.

25. Sass DA , Schmitt TA , Marsh HW. Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. *Struct Equ Modeling* 2014;**21**:167–80.

26. Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct Equ Modeling* 2007;**14**:464–504.

27. Meade AW , Johnson EC , Braddy PW. Power and sensitivity of alternative fit indices in tests of measurement invariance. *J Appl Psychol* 2008;**93**:568-92.

28. Miles J , Shevlin M. Applying Regression and Correlation: A Guide for Students and Researchers. London: Sage; 2001.

29. Cohen J. Statistical power analysis for the behavioral sciences. Hilsdale: Lawrence Earlbaum Associates; 1988.

30. Grissom RJ , Kim JJ. Effect sizes for research: Univariate and multivariate applications. 2nd ed. New York: Routledge; 2012.

31. Ruscio J. A probability-based measure of effect size: Robustness to base rates and other factors. *Psychol Methods* 2008;**13**:19-30.

32. Hanel PH , Mehler DM. Beyond reporting statistical significance: Identifying informative effect sizes to improve scientific communication. *Public Underst Sci* 2019;**28**:468-85.

33. R Core Team. R: A language and environment for statistical computing. R foundation for Statistical Computing. 2019. https://www.R-project.org/

34. Revelle W. psych: Procedures for personality and psychological research. Northwestern University, Evanston, Illinois, USA. 2018. http://CRAN.R-project.org/package=psychVersion=1.8.12.

35. Muthén LK , Muthén BO. Mplus user's guide. Los Angeles, CA: Author; 2010.

36. Appel AE , Holden GW. The co-occurrence of spouse and physical child abuse: A review and appraisal. *J Fam Psychol* 1998;**12**:578-99.

37. Capaldi DM , Knoble NB , Shortt JW , Kim HK. A systematic review of risk factors for intimate partner violence. *Partner Abuse* 2012;**3**:231-80.

38. Clasen CC , Palesh OG , Aggarwal R. Sexual revictimization: A review of the empirical literature. *Trauma Violence Abus* 2005;**6**:103-29.

39. Margolin G , Gordis EB , Medina AM, Oliver PH. The co-occurrence of husband-to-wife aggression, family-of-origin aggression, and child abuse potential in a community sample. *J Interpers Violence* 2003;**18**:413-40.

40. Schumacher JA, Feldbau-Kohn S, Slep AMS , Heyman RE. Risk factors for male-to-female partner physical abuse. *Aggress Violent Beh* 2001;**6**:281-352.

41. Till-Tentschert U. The relation Between Violence Experienced in Childhood and women's exposure to violence in later life: Evidence from Europe. *J Interpers Violence* 2017;**32**:1874-94.

42. Coker AL , Davis KE , Arias I , *et al.* Physical and mental health effects of intimate partner violence for men and women. *Am J Prev Med* 2002;**23**:260-8.

43. Lövestad S , Krantz G. Men's and women's exposure and perpetration of partner violence: an epidemiological study from Sweden. *BMC Public* 2012;**12**:945.

44. Nybergh L , Taft C , Krantz G. Psychometric properties of the WHO violence against women instrument in a female population-based sample in Sweden: a cross-sectional survey. *BMJ Open* 2013;**3**;e002053.

45. Reichel D. Determinants of intimate partner violence in Europe: The role of socioeconomic status, inequality, and partner behavior. *J Interpers Violence* 2017;**32**:1853-73.

46. Sokoloff NJ, Dupont I. Domestic violence at the intersections of race, class, and gender: Challenges and contributions to understanding violence against marginalized women in diverse communities. *Violence Against Wom* 2005;**11**:38-64.

47. Davidov E , Schmidt P , Schwartz S. Bringing values back in: Testing the adequacy of the European Social Survey to measure values in 20 countries. *Public Opin Q* 2008;**3**:420–45.

48. Muthén B , Asparouhov T. Recent methods for the study of measurement invariance with many groups: alignment and random effects. *Sociol Method Res* 2018;**47**:637-64.

49. Gracia E , Merlo J. Intimate partner violence against women and the Nordic paradox. *Soc Sci Med* 2016;**157**:27–30.

50. Gracia E , Martín-Fernández M , Lila M , *et al.* Prevalence of intimate partner violence against women in Sweden and Spain: A psychometric study of the 'Nordic paradox'. *PLoS One* 2019;**14**:e0217015.

Study 2

Psychological intimate partner violence against women in the European Union:

A cross-national invariance study

Psychological intimate partner violence against women in the European Union: A cross-national invariance study[*]

Manuel Martín-Fernández, Enrique Gracia, & Marisol Lila
Department of Social Psychology, University of Valencia, Valencia, Spain

Abstract

Background. Intimate partner violence against women (IPVAW) is a worldwide public health problem. One of the most frequent forms of this type of violence in western societies is psychological IPVAW. According to the European Union (EU) Fundamental Rights Association (FRA) the prevalence of psychological IPVAW in the EU is 43%. However, the measurement invariance of the measure addressing psychological IPVAW in this survey has not yet been assessed. **Methods.** The aim of this study is to ensure the cross-national comparability of this measure, by evaluating its measurement invariance across the 28 EU countries in a sample of 37,724 women, and to examine how the levels of this type of violence are distributed across the EU. **Results.** Our results showed that the psychological IPVAW measure presented adequate psychometric properties (reliability and validity) in all countries. A latent structure of one factor was supported and scalar invariance was established in all countries. The average levels of psychological IPVAW were higher in countries like Finland, Latvia, Lithuania, and Sweden compared to the rest of the EU countries. In many of the other countries the levels of this type of

---

violence was overlapped. **Conclusion.** Our findings underlined the importance of using appropriate statistical methods to make valid cross-national comparisons in large population surveys.

Background

Intimate partner violence against women (IPVAW) is a worldwide social and public health problem[1,2,3], with serious consequences not only for the victims' physical and psychological well-being, but also for their children, and the wider community[4,5,6,7].

One of the most frequent forms of this type of violence in western societies is psychological IPVAW, which can occur either in isolation, or in conjunction with other forms of intimate partner violence[8,9,10,11]. There is, however, a strong link between psychological and physical IPVAW, since psychological violence often precedes physical IPVAW, and it is considered one of its main risk factors[12,13].

The average prevalence of psychological IPVAW in large population surveys varies largely depending on how this type of violence is defined and measured, with some studies estimating its prevalence at around 10-20% while others found prevalence rates of around 80-90%[14,15,16,17,18,19,20]. The frequency and severity of psychological IPVAW can also differ widely from one country to another [2,14], implying that the way in which this type of violence is perceived and interpreted can vary across countries and cultures.

Most surveys addressing psychological IPVAW have followed the tradition of the Conflict Tactics Scale (CTS) in defining it through specific behaviors[2,21,22]. Two aspects are usually considered in the assessment of this type of violence: emotional abuse and controlling behavior[11,12,23,24,25]. Emotional abuse involves behaviors intended to generate emotional harm or threat of harm, such as belittling, humiliating, threatening or intimidating the victim, whereas controlling behavior entails monitoring partner's behaviors or isolating them by limiting actions, such as forbidding them to leave the house, restricting contact with other people, or continually insisting on knowing the victim's whereabouts[2,14,26,27].

In the European Union (EU), a survey conducted by the EU Agency for Fundamental Rights (FRA) across the 28 member countries found that the average prevalence of psychological IPVAW was 43%, ranging from 31% in Ireland to 60% in Latvia[26]. The main advantage of this survey is that it also followed the CTS tradition, using the same set of questions addressing emotional abuse and controlling behavior in the 28 EU member states. Nevertheless, as is the case in most surveys, the measurement invariance of these questions has not yet been assessed, which calls into question the validity of these cross-national comparisons. It is therefore not possible to ensure whether the differences in psychological IPVAW prevalence across the EU countries reflects actual differences between countries, or whether they are the result of different cultural beliefs or expectations about intimate partner violence that may distort the interpretation of the FRA survey questions.

Therefore, before making any comparison across countries it is necessary to address the measurement invariance of the set of questions used in different countries[28,29,30]. Measurement invariance is an important prerequisite in cross-national research as it allows meaningful comparisons to be made across countries by ruling out the possibility of cultural bias in the respondents' answers[31,32,33]. When measurement invariance is not supported, it cannot be assumed that respondents from different countries interpret and answer the questions in the same way , and hence their scores cannot be directly compared[34]. Thus, obtaining prevalence rates to compare samples from different countries without first assessing measurement invariance could lead to inaccurate and biased conclusions, since the validity of such comparisons may become compromised.

The principal aim of this study is to ensure the cross-national comparability of the set of questions addressing psychological IPVAW used in

the FRA survey, by evaluating whether respondents of each country conceptualize and interpret these questions in the same way. For validity purposes, we also examined the relationships of psychological IPVAW to other forms of partner violence, such as physical and sexual IPVAW, and to other related sociodemographic and background variables, such as self-perceived health, household income, and experiences of child abuse[4,11,13,35,36,37,38,39]. Once the measurement invariance of this measure and his validity were established, we aim to make valid and appropriate comparisons of the psychological IPVAW levels across all EU countries.

Method

Participants

The sample used in the present study consisted of the responses of 37,724 women to the survey conducted by the European Union Agency for Fundamental Rights on violence against women[26]. Respondents to this survey were ever-partnered women, aged from 18 to 74 years old, from the 28 EU countries. The responses were collected following a two-stage clustered stratified sampling design with equal probability of selection of households within clusters; structured interviews were conducted in person[40]. The average response rate to the survey was 42.1%, ranging from 18.5% in Luxembourg to 84.0% in Hungrary[40]. Quality control checks were made for 10% of the interviewed women[40]. A license for secondary data analysis was granted by the FRA for all the analyses (Reference No. 102577).

The sample used in this study comprised the responses from respondents who answered all of the items addressing psychological partner violence. Socio-demographical information of the sample by country can be found in Table 1.

Table 1. Sociodemographic characteristics of the sample in each country (N = 37724)

	Age (%)				Income (%)				Self-perceived health		Experiences of child abuse (%)
	18-24	25-34	35-59	60+	Q1	Q2	Q3	Q4	M	SD	
Austria	12.4	21.3	45.1	21.1	34.1	23.0	26.6	16.3	1.77	0.86	26.7
Belgium	8.0	17.0	53.6	21.4	10.7	24.5	37.7	27.1	2.04	0.87	26.0
Bulgaria	4.9	13.5	44.2	37.3	12.5	39.1	35.0	13.5	2.34	1.00	27.7
Croatia	5.4	15.1	52.2	27.3	27.0	22.7	25.8	24.5	2.23	1.11	31.8
Cyprus	18.4	27.5	45.1	9.1	6.2	27.4	49.5	16.9	1.60	0.80	12.2
Czech Republic	9.8	16.6	49.0	24.6	29.8	29.2	22.8	18.2	2.06	0.89	32.3
Denmark	12.2	15.3	49.0	23.5	25.5	35.1	24.9	14.5	1.82	0.90	41.9
Estonia	5.9	12.7	47.0	34.4	28.0	23.6	26.1	22.3	2.56	0.92	48.4
Finland	5.9	15.5	45.2	33.5	19.3	22.4	27.6	30.7	2.07	0.79	52.2
France	5.4	14.7	57.0	22.9	52.7	32.5	9.7	5.1	1.99	0.84	44.9
Germany	5.7	13.3	55.9	25.2	23.5	26.1	29.2	21.3	2.14	0.84	42.8
Greece	7.4	16.4	50.9	25.2	21.9	40.7	33.1	4.3	1.69	0.88	23.7
Hungary	8.3	15.3	46.7	29.6	20.1	29.4	23.7	26.8	2.49	0.99	23.8
Ireland	6.2	20.1	49.7	24.1	39.7	30.3	20.9	9.2	1.69	0.83	27.7
Italy	4.4	14.6	60.8	20.1	62.5	30.0	5.4	2.0	2.23	0.76	33.6
Latvia	7.4	15.7	46.9	30.0	23.6	26.0	25.0	25.5	2.68	0.83	31.5
Lithuania	5.6	11.8	54.3	28.3	44.2	16.7	16.5	22.6	2.57	0.85	18.4
Luxembourg	3.8	14.0	63.7	18.5	15.5	32.6	31.4	20.5	1.94	0.89	44.0
Malta	3.9	13.6	52.3	30.1	30.8	41.3	20.6	7.2	2.06	0.66	23.4
Netherlands	3.2	12.1	59.5	25.2	18.9	28.7	26.9	25.6	2.15	0.77	33.0
Poland	9.9	23.7	47.5	18.8	10.4	24.7	30.7	34.2	2.23	0.98	17.6
Portugal	4.3	11.3	47.8	36.6	22.4	40.6	20.6	16.3	2.66	0.92	28.4
Romania	9.2	18.7	47.9	24.3	29.2	20.7	19.8	30.3	2.38	1.03	22.7
Slovakia	6.8	18.6	53.9	20.8	18.3	23.7	28.1	29.9	2.10	0.96	32.9
Slovenia	5.5	13.4	52.3	28.9	12.9	30.9	35.2	21.1	2.13	0.90	13.2
Spain	3.9	15.3	54.7	26.1	35.4	29.4	22.3	13.0	2.13	0.89	29.2
Sweden	2.7	10.8	54.3	32.2	3.5	27.2	23.6	45.7	1.93	0.92	43.2
United Kingdom	6.0	15.9	50.0	28.1	38.3	8.9	8.8	44.0	1.92	0.95	38.0

Note: The percentages and descriptive statistics are unweighted. M = Mean; SD = Standard Deviation. Income: Q1 = under lowest quartile, Q2 = between lowest quartile and median, Q3 = between median and highest quartile, Q4 = above highest quartile.

Measures

 Psychological violence. The FRA survey includes two sets of items addressing psychological IPVAW in the first part of the interview. The first set contains eight items assessing controlling behavior (e.g., "Insisting on knowing where she is in a way that goes beyond general concern"), and economic violence (e.g., "Preventing you from making decisions about family finances or from shopping independently"). The second set is comprised of five items evaluating emotional abuse (e.g, "Belittling or humiliating you in front of other people"). The response format for all the items was a 4-point Likert-type scale indicating the frequency of such behaviors (1: "Never", 2: "Sometimes", 3: "Often", 4: "All the times"). Given that in most of the countries the frequencies of the upper two categories were extremely low (less than 2%, even when merged together), we decided to dichotomize the responses in order to set the same metric in all the items (0: "Never", 1: "Sometimes or more often").

Validity evidence based on relations to other variables

 We used the following variables to test the validity of the psychological IPVAW measure[41]:

 Physical violence. Physical IPVAW is assessed in the FRA survey with a set of eight items describing episodes of physical violence perpetrated by either the current or any previous partner (e.g., "Has your current/previous partner ever slapped you?"). The response format of the items was dichotomized (0: "Never", 1: "Once or more times"). The factor scores of this measure were used for the validity analyses.

 Sexual violence. Sexual IPVAW is evaluated in the FRA survey with a set of four items addressing sexual violence committed by the current or any previous partner (e.g., "Has your current/previous partner made you take part in any form of sexual activity when you did not want to or you were unable to refuse?"). The responses to these items were also dichotomized (0: "Never", 1:

"Once or more times"). The factor scores of this measure were also used for the validity analyses.

Self-perceived health. The FRA survey used a single item inquiring about the general health of the respondents at the beginning of the interview, using a 5-point Likert type graded response (from 1 = "Very Bad" to 5 = "Very Good").

Experiences of child abuse. This variable was assessed in the FRA survey using a set of 11 questions asking about experiences of childhood physical and sexual abuse before the age 15 (e.g., "Did an adult who was 18 years or over hit you very hard so that it hurt?", "has an adult who was 18 years or over expose their genitals to you?"). If any of these questions were answered affirmatively, we considered that the participant had experienced abuse during their childhood.

Income. The FRA survey includes a single item of reported income in each country (i.e., "under lowest quartile", "between lowest quartile and median", "between median and highest quartile", "above highest quartile"). To answer this item, respondents informed the interviewer about their monthly income or chose between four income bands. These bands varied depending on the country to make the quartiles comparable across all EU member states[40] (e.g., in Austria: "up to €1,600", "€1,601-€2,300", "€2,300-€3,000", "over €3,000").

Statistical Analyses

The main objective of this study is to assess the measurement invariance of the items addressing psychological IPVAW included in the FRA Survey, and to examine how the levels of this type of violence are distributed across the EU. To do so, we carried out the following analyses.

We first conducted a descriptive analysis of the set of items addressing psychological IPVAW, obtaining the mean, standard deviation, skew, and

kurtosis statistics, as well as the correlation of each item with the rest of the scale (i.e., item-test corrected correlation). We then analyzed the latent structure of this measure carrying out a confirmatory factor analysis (CFA) for each country separately. We compared two models: a one-factor model, where all the items loaded onto a single factor, and a two-factor correlated models, in which all the items of the first set—controlling behavior and economic violence—loaded on one factor, and the items of the second set—emotional abuse—loaded onto a second factor. Although the first set of questions has items of controlling behavior and economic violence, we decided to maintain the structure used in the FRA survey. To this end, both types of items were included in the same factor, since there were not enough indicators to estimate a separate factor of economic violence—the FRA survey only included two items of economic violence, and a minimum of three indicators are usually required[42,43]—. Given the categorical nature of the data, we used weighted least squares means and variances adjusted (WLSMV) as the estimation method. The fit of the models was evaluated by a combination of fit indices: the comparative fit index (CFI), the Tucker-Lewis index (TLI), and the root mean squared error of approximation (RMSEA). CFI and TLI values above .95 are indicative of good fit[44], whereas RMSEA values below .08 and .06 were considered as mediocre and excellent fit, respectively[45]. The internal consistency of the resulting factors was assessed using Revelle's omega total[46], as it does not assume tau-equivalence for the items[47]. Omega values above .70 are indicative of good internal consistency.

Once the latent structure of the scale was established, we carried out a series of multi-group confirmatory factor analyses (MG-CFA) to test the measurement invariance of the psychological IPVAW measure across the 28 EU countries. To this end, a series of nested models were evaluated: configural, metric and scalar invariance models[28]. In the configural invariance model, the

same factor structure is applied for all countries, assuming no equality constraints for any parameters. The metric invariance model holds the item loadings to the same value across countries. The scalar invariance model constraints both item loadings and thresholds to be invariant across countries, ensuring that respondents from different countries with the same pattern of responses will obtain the same factor score. If the scalar invariance model is supported, then the factor means on the psychological IPVAW measure can be compared across countries.

To assess the fit of the invariance models, we computed the change in the CFI (ΔCFI) and RMSEA (ΔRMSEA), using the general guidelines of Cheung and Rensvold[48] and Chen[49]. These guidelines, however, were developed for continuous data, and thus the interpretation of such indices should be made with caution when dealing with categorical data[50]. For this reason, we used the cut-off values proposed by Meade, Johnson, and Brady[51], as it is currently the most conservative approach for assessing the change in the fit indices: ΔCFI ≤ .002 and ΔRMSEA ≤ .007. The performance of these cut-offs with categorical data tend to be similar to maximum likelihood based procedures when the sample size is large and the items are not normally distributed[50].

After establishing an invariant factorial model, we used the invariant factor scores of the psychological IPVAW measure to test validity evidence based on relationships to other variables in all countries. Pearson correlations were obtained between the psychological IPVAW measure and the measures of physical and sexual IPVAW. In addition, we computed a one-way ANOVA, testing differences in psychological IPVAW by self-perceived health, experiences of child abuse, and income. The size effect of the variables was assessed with the partial eta-squared statistic, using values above .01, .06, and .14, as indicative of small, medium, and large size effects, respectively[52].

Finally, we compared the factor means of psychological IPVAW across countries through a latent means analysis. This procedure takes into account the different weights of the items (i.e., item loadings) to measure the construct and all the constraints of the invariance analyses, leading to a more appropriate and sophisticated comparison of psychological IPVAW levels across all EU countries. We evaluated the magnitude of these cross-national comparisons using Cohen's d statistic. Values of this statistic above .20, .50, and .80 indicate small, medium and large size effects, respectively[53]. In addition to Cohen's d, we also computed the Cohen's U_3 statistic, which reflects the percentage of cases of one country that is higher than the average of another[54,55]. Cohen's U_3 between each pair of countries can be found on the supplementary material.

All analyses were conducted with the library *psych* of the statistical package R [46,56], with the exception of the CFA and the MG-CFA, which were computed using *Mplus 8.3*[57].

Results

Descriptive Analysis

All the items presented means close to zero, as well as high skew and kurtosis statistics, showing that most of the respondents reported never having experienced the episodes described by the items (Table 2). The item-test corrected correlations were high in general, indicating a strong relationship between the items.

Confirmatory Factor Analysis and Internal Consistency

Two models where tested in each country: a one-factor model assuming that all the items are grouped onto a single factor, and a two-factor model, distinguishing between the controlling behavior and the emotional abuse items.

Table 2. Descriptive statistics of the psychological violence items.

	M	SD	Skew	Kurtosis	$r_{item-test}$
Controlling Behavior and Economic Violence					
Try to keep you from seeing your friends?	0.18	0.38	1.68	0.81	0.73
Try to restrict your contact with your family of birth or relatives?	0.11	0.31	2.48	4.16	0.65
Insist on knowing where you are in a way that goes beyond general concern?	0.23	0.42	1.32	-0.27	0.72
Get angry if you speak with another man/woman?	0.22	0.42	1.33	-0.24	0.70
Become suspicious that you are unfaithful?	0.21	0.41	1.39	-0.05	0.68
Prevent you from making decisions about family finances and from shopping independently?	0.11	0.31	2.49	4.20	0.58
Forbid you to work outside the home?	0.05	0.21	4.26	16.16	0.47
Forbid you to leave the house, takes away your car keys or locks you up?	0.05	0.21	4.29	16.41	0.52
Emotional Abuse					
Belittled or humiliated you in front of other people?	0.19	0.39	1.62	0.61	0.66
Belittled or humiliated you in private?	0.26	0.44	1.10	-0.79	0.69
Done things to scare or intimidate you on purpose, for example by yelling and smashing things?	0.19	0.39	1.55	0.41	0.69
Made you watch or look at pornographic material against your wishes?	0.02	0.15	6.41	39.15	0.32
Threatened to hurt or kill someone you care about?	0.04	0.19	4.9	21.98	0.40

Note: M = Mean; SD = Standard Deviation; $r_{item-test}$ = corrected item-test correlation. Skew and kurtosis standard error were below .01.

The one-factor model yielded a very good fit in all countries, showing excellent CFI and TLI values in the 28 EU countries (Table 3). The RMSEA, however, indicated only excellent fit in Bulgaria, Estonia, Finland, Ireland, Luxembourg, Malta, Romania, and Spain, presenting a mediocre model fit in the rest of the countries, with the exception of Croatia, where the RMSEA for the one-factor model was poor. Regarding the two-factor model, all the fit indices were excellent, showing almost a perfect fit. Nevertheless, the

correlations between the two factors of this second model were very strong, yielding values around .85 in most of the countries (ranging from .73 in Croatia to .93 in Ireland), which in turn may be indicating that both factors are measuring the same construct. Given the suitability of both factorial solutions, we decided to keep both factor structures for the measurement invariance analysis in order to then choose the one better supported by the data.

The internal consistency was very good in both models. In particular, the omega total for the one-factor solution was $\omega = .90$ in the complete sample (ranging from .88 in Denmark to .93 in Bulgaria, Croatia, and Ireland). For the two-factor solution the omega total was $\omega = .91$ (ranging from .88 in Denmark to .95 in Croatia).

Measurement Invariance

The analysis of measurement invariance supported the configural, metric, and scalar invariance models for the psychological IPVAW measure across all EU countries for the one-factor solution (Table 3). When the factor loadings were constrained to have the same value in all the countries, the metric model fit did not differ substantially from the configural model (ΔCFI = .000, ΔRMSEA = .006). Similarly, constraining the item thresholds as well as the item loadings to be equal across countries did not substantially reduce the fit of the model (ΔCFI = .001, ΔRMSEA = .002), supporting the scalar invariance model. When the factor loadings were constrained to have the same value in all the countries, the metric model fit did not differ substantially from the configural model (ΔCFI = .000, ΔRMSEA = .006). In the same direction, constraining the item thresholds as well as the item loadings to be equal across countries did not reduce substantially the fit of the model (ΔCFI = .001, ΔRMSEA = .002), supporting the scalar invariance model.

Regarding the two-factor solution, however, the data only supported the configural invariance model. Constraining the loadings to be equal in all

Table 3. Confirmatory factor analysis fit indices by country

	One-factor model			Two-factor model		
	CFI	TLI	RMSEA	CFI	TLI	RMSEA
Austria	0.98	0.98	0.066	0.99	0.99	0.039
Belgium	0.99	0.98	0.070	0.99	0.99	0.049
Bulgaria	0.99	0.99	0.052	0.99	0.99	0.035
Croatia	0.99	0.98	0.101	0.99	0.99	0.056
Cyprus	0.98	0.98	0.071	0.99	0.99	0.040
Czech Republic	0.98	0.97	0.073	0.98	0.98	0.054
Denmark	0.97	0.97	0.071	0.98	0.98	0.048
Estonia	0.99	0.99	0.055	0.99	0.99	0.032
Finland	0.99	0.98	0.059	0.99	0.99	0.039
France	0.99	0.98	0.061	0.99	0.99	0.038
Germany	0.98	0.98	0.066	0.99	0.98	0.045
Greece	0.99	0.98	0.060	0.99	0.99	0.036
Hungary	0.99	0.99	0.080	0.99	0.99	0.043
Ireland	0.99	0.99	0.059	0.99	0.99	0.041
Italy	0.98	0.98	0.069	0.99	0.98	0.045
Latvia	0.98	0.98	0.070	0.99	0.99	0.036
Lithuania	0.98	0.97	0.074	0.99	0.99	0.039
Luxembourg	0.99	0.99	0.055	0.99	0.99	0.037
Malta	0.99	0.98	0.056	0.99	0.99	0.037
Netherlands	0.98	0.97	0.076	0.98	0.98	0.053
Poland	0.99	0.99	0.066	0.99	0.99	0.040
Portugal	0.99	0.98	0.069	0.99	0.99	0.038
Romania	0.99	0.99	0.058	0.99	0.99	0.033
Slovakia	0.99	0.99	0.069	0.99	0.99	0.045
Slovenia	0.99	0.98	0.067	0.99	0.99	0.037
Spain	0.99	0.99	0.056	0.99	0.99	0.030
Sweden	0.97	0.97	0.075	0.98	0.98	0.051
United Kingdom	0.99	0.99	0.063	0.99	0.99	0.038

Note: CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Squared Error of Approximation.

the countries reduced the fit of the metric invariance model (ΔCFI = .004, ΔRMSEA = .006), above the ΔCFI .002 cut-off suggested by Meade et al.[51]. We

Table 4. Measurement invariance fit indices.

Model	χ^2	*df*	CFI	TLI	RMSEA [95% CI]
One-factor model					
Configural	11855.74	1820	0.987	0.985	0.064 [0.063 – 0.065]
Metric	12131.20	2171	0.987	0.987	0.058 [0.057 – 0.059]
Scalar	13947.57	2360	0.985	0.986	0.060 [0.059 – 0.061]
Two-factor model					
Configural	6046.60	1792	0.995	0.993	0.042 [0.041 – 0.043]
Metric	8884.20	2143	0.991	0.991	0.048 [0.047 – 0.049]
Scalar	10739.80	2440	0.989	0.990	0.050 [0.049 – 0.051]

Note: χ2 = adjusted chi-squared test for model fit; *df* = degrees of freedom; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Squared Error of Approximation.

thus decided to keep the one-factor solution as the latent structure of the psychological IPVAW measure for the rest of the analyses.

The standardized item loadings of the one-factor model were high in general, with values above .90 for most of the items and low standard errors (Figure 1). There were, however, small differences between the items, pointing out that not all the items contribute equally to the factor, and that some items were more relevant than others to assess the construct. Regarding the item thresholds, they were around 1 for most of the items, ranging from 0,60 for the item "Get angry if you speak with another man/woman?" to 2.11 for the item "Made you watch or look at pornographic material against your wishes?", covering a wide area of the latent trait continuum. The items with the lowest factor loadings ("Made you watch or look at pornographic material against your wishes?", and "Threatened to hurt or kill someone you care about?"), were also the items with the highest thresholds (i.e., 2.11 and 2.08, respectively), indicating that they were addressing more severe forms of violence but with somewhat less precision.
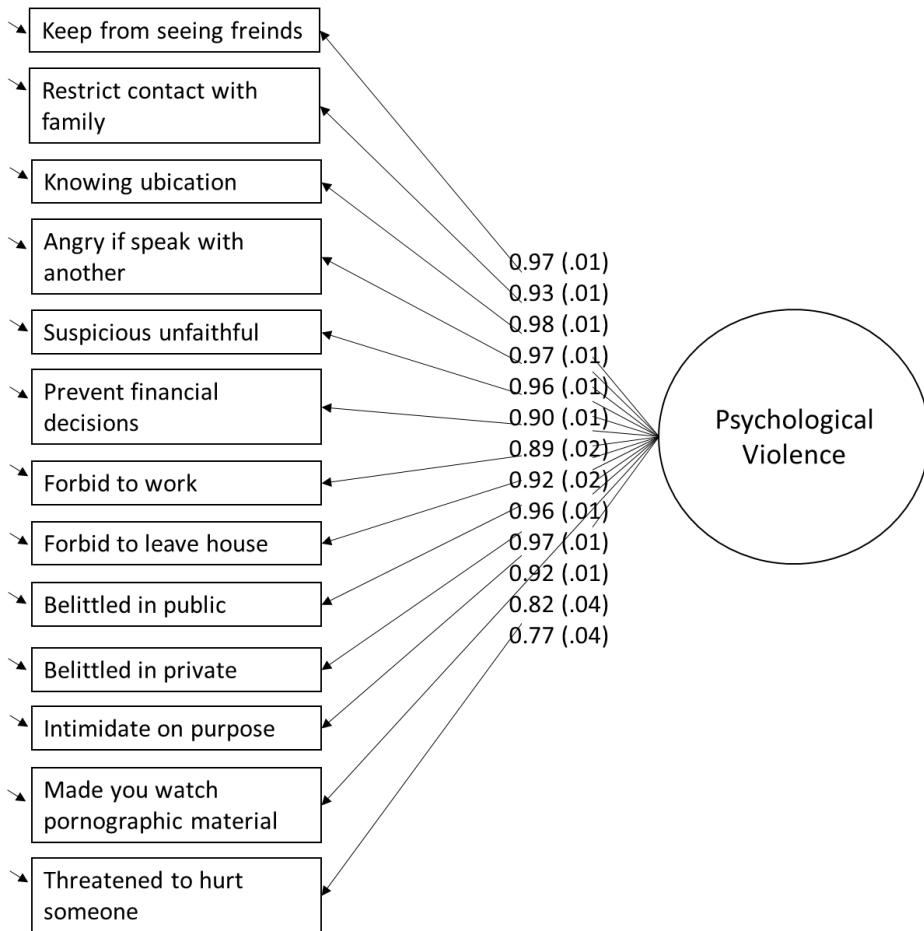
Figure 1. *Multi-group confirmatory factor analysis: scalar invariance model.*
Note: The standardized factor loadings belong to the reference country (i.e., Ireland). As in the standardized solution the factor variances are fixed to 1, there are slight differences in the decimals of the loadings in each country.

Validity evidence based on relations to other variables

We used the factor scores of the invariant one-factor model for the validity analyses in all countries. We found a positive and strong relationship between the psychological IPVAW factor scores and the physical and sexual IPVAW factor scores ($r = .85$ and $r = .75$, respectively), indicating that those

women with higher levels of psychological IPVAW also tend to show higher levels of physical and sexual IPVAW.

Significant differences with a small size effect were found in the psychological IPVAW scores when experiences of child abuse were taken into account ($F(1) = 1548$, $p < .001$, $\eta^2 = .040$), as respondents with a background of child abuse showed higher psychological IPVAW levels. Significant differences were found in self-perceived health ($F(4) = 47.4$, $p < .001$, $\eta^2 = .001$) and income ($F(1) = 138.9$, $p < .001$, $\eta^2 = .004$), although the size effect of both variables was negligible.

Latent Means Analysis

After determining that the psychological IPVAW measure is psychometrically sound and establishing an invariant model across all EU countries, the means of the psychological violence factor can now be properly compared across countries by carrying out a MG-CFA. We used Ireland as the reference group, as it was the country with the lowest rates of psychological IPVAW[26]. We fixed the latent mean for this country to zero and its variance to one, whereas in the rest of the countries these parameters were freely estimated. The model converged normally and showed a good fit (CFI = .986, TLI = .986, RMSEA = .061 [.060 - .062]). The standardized latent mean of each country represents the difference in standard deviations from Ireland. The estimated latent means for each country and their confidence intervals are displayed in Figure 2.

Croatia showed a lower latent mean than Ireland, although these differences were not significant ($z = -0.05$, $p = .554$, $d = 0.06$). Similarly, no significant differences were found between Ireland and Slovenia ($z = 0.05$, $p = .771$, $d = 0.05$), Spain ($z = 0.11$, $p = .405$, $d = 0.11$), Greece ($z = 0.16$, $p = .339$, $d = 0.14$), and Romania ($z = 0.18$, $p = .144$, $d = 0.17$). We found significant
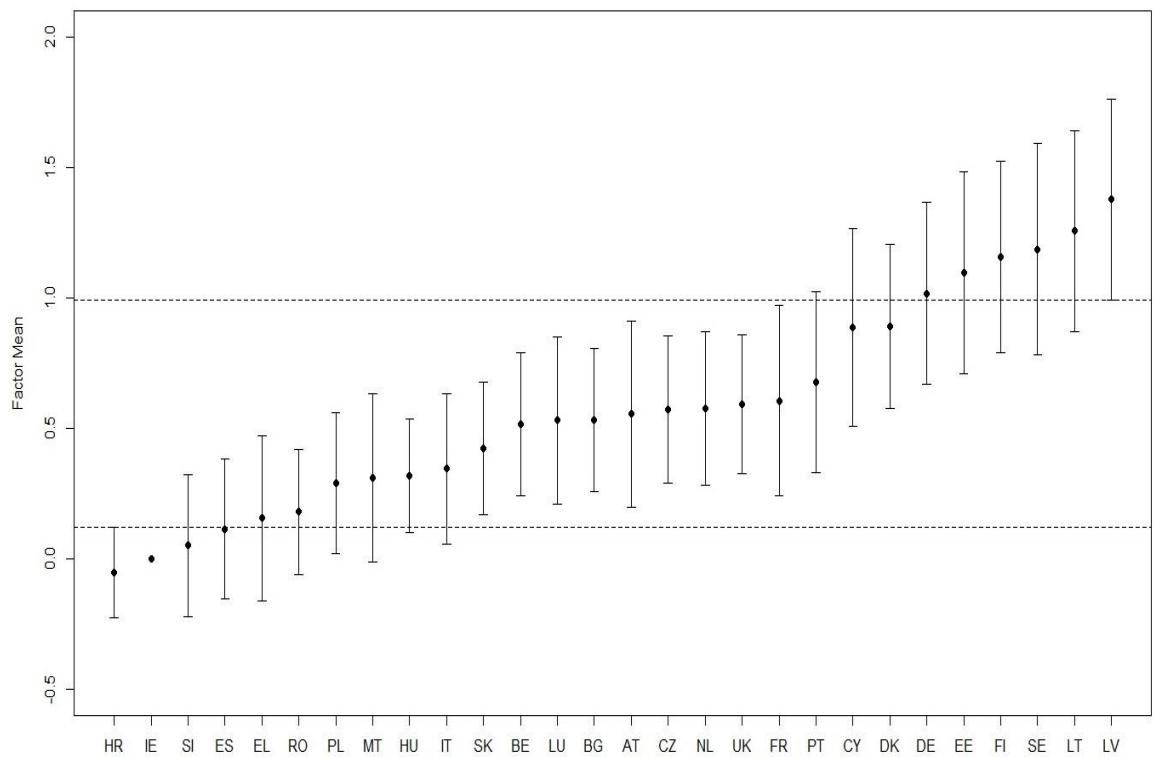
Figure 2. *Psychological IPVAW latent means across the EU.*

Note: IPVAW: Intimate partner violence against women. AT: Austria, BE: Belgium, BG: Bulgaria, CY: Cyprus, CZ: Czech Republic, DK: Denmark, EE: Estonia, FI: Finland, FR: France, DE: Germany, EL: Greece, ES: Spain, HR: Croatia, HU: Hungary, IE: Ireland, IT: Italy, LT: Lithuania, LV: Latvia, LU: Luxembourg, MT: Malta, NL: Netherlands, PL: Poland, PT: Portugal, RO: Romania, SE: Sweden, SI: Slovenia, SK: Slovakia, UK: United Kingdom. Ireland is the reference country.

differences with a large size effect between Ireland and Latvia ($z = 1.38$, $p <$ .001, $d = 0.93$), Lithuania ($z = 1.26$, $p < .001$, $d = 0.88$), Finland ($z = 1.16$, $p <$ .001, $d = 0.81$), and Sweden ($z = 1.19$, $p < .001$, $d = 0.80$).

The confidence intervals (CI) of the psychological IPVAW factor means were overlapped between many of the EU countries, showing that there were no significant differences between them. However, we grouped the countries using the CI of Croatia and Latvia, the countries with the lowest and highest levels of psychological IPVAW. Therefore, countries whose factor mean fell

within the CI of Croatia showed on average lower levels of psychological IPVAW (Ireland, Slovenia, and Spain). On the other hand, countries with factor means that fell within the CI of Latvia presented higher levels of this type of violence (Lithuania, Finland, Sweden, and Estonia). The rest of the countries fell in a middle area, with countries like Greece or Romania closer to those countries with the lowest levels of psychological IPVAW, and others like Germany or Denmark closer to those countries with the highest levels of this type of violence.

The psychological IPVAW factor scores can also be compared between each pair of countries. As an illustrative example, we compare the factor scores of Sweden, Austria, and Spain—countries with high, moderate and low levels of psychological IPVAW— using Cohen U_3 statistic. We found small differences between Sweden and Austria (d = .41, U_3 = .659), showing that the 65.9% of the Swedish sample has higher values in the psychological IPVAW factor than the average of the Austrian sample. We found moderate differences between Sweden and Spain (d = .74, U_3 = .770), with 77.0% of the Swedish sample presenting higher scores in this factor than the average of the Spanish sample. Smaller differences were found between Austria and Spain (d = .36, U_3 = .641), as the 64.1% of the Austrian sample had higher factor scores in the psychological IPVAW factor than the mean of the Spanish sample. A table detailing the comparisons between each pair of countries and as its associated size effect is provided in the supplementary material section.

Discussion

Although psychological IPVAW is one of the most extended forms of partner violence, little attention has been paid to the cross-national comparability of the data used in large population and health surveys. In this study we tackled this issue through a set of analyses to test the measurement

invariance of the set of questions addressing this type of violence in the FRA survey, and then examining how psychological IPVAW levels were distributed across all EU countries.

The first set of analyses aimed to assess the psychometric properties (i.e., latent structure, reliability, and validity) of the set of questions assessing psychological IPVAW in each of the EU countries. The initial CFA suggested that the items of this measure could be mapping either a one- or a two-factor latent structure. Currently there is not a clear consensus on how psychological IPVAW should be conceptualized and measured in cross-national research, and although a two-factor structure seems to be theoretically preferred, the number of studies addressing the latent structure of this type of IPVAW is rather scarce[2,58]. Even though previous studies have established a two-factor structure for psychological IPVAW, distinguishing between emotional abuse and other forms of controlling behavior with similar items, the correlation between emotional abuse and controlling behavior is usually strong, which in turn may indicate than a one-factor structure could be sufficient to account for the variability of the construct[12,23,58]. This idea was supported by the measurement invariance analyses, where we found that metric and scalar invariance only held under the one-factor structure, underlining that the items addressing emotional abuse and controlling behavior of the FRA survey could be grouped into a single factor across the EU (i.e., psychological IPVAW). The resulting factor showed high internal consistency, with ω values above .85 in all countries.

Regarding the validity evidence based on relations with other variables, our results pointed out that, as expected, psychological IPVAW was strongly related to physical and sexual IPVAW. The co-occurrence of psychological IPVAW with other forms of violence is a well-known phenomenon in the literature, and it is often considered as an antecedent of

physical IPVAW[11,13]. In addition, we found that the women who reported experiences of child abuse also presented higher levels of psychological IPVAW, which is in line with previous research indicating that women who have been victimized in childhood have higher risk of being victimized as adults[35,36,59].

The second set of analyses aimed to test the measurement invariance of the psychological IPVAW measure used in the FRA survey across all EU countries. Establishing measurement invariance is a necessary prerequisite in cross-national research before conducting any comparison between countries[31]. To this end, we carried out a series of MG-CFA models to examine whether the psychological IPVAW measure is comparable across the 28 EU countries, testing configural, metric, and scalar invariance. We were able to hold these three invariance models (i.e., same factors with equal loadings and thresholds) under the one-factor latent structure across countries, despite the difficulties frequently found into achieving this with many groups[60,61]. This may be due to the behavioral nature of the items used in the FRA survey, since they address the frequency of concrete episodes which may be difficult to misinterpret (e.g., "insisting on knowing where you are in a way that goes beyond general concern", "doing things to scare or intimidate her on purpose").

We were able to make appropriate and valid comparisons between countries once an invariant model was established for the psychological IPVAW measure. This is one of the main strengths of this study, since it allowed us to conduct a latent means analysis to compare the factor means of each country, a procedure that unlike computing the raw prevalence, takes into account the latent structure of the construct and the weight of each item to assess it[28,29]. Our findings showed that there were almost no differences between most of the countries, as the confidence intervals of the factor means

overlapped, indicating that the levels of psychological IPVAW were quite similar between them. We found, however, substantial differences between the countries with higher and lower levels of psychological IPVAW . For example, countries like Latvia, Lithuania, Finland, and Sweden presented average levels of psychological IPVAW significantly higher than countries such as Croatia, Ireland, Slovenia and Spain. Given the close relationship between different forms of IPVAW, this finding may further support the idea of the Nordic paradox[62,63], as despite being the EU member states with the highest levels of gender equality, the Nordic countries —Denmark, Sweden, and Finland—are also among the EU countries with the highest levels of psychological IPVAW.

This study has some limitations. The first one is the cross-sectional nature of the survey design, as it did not allow us to evaluate measurement invariance across different periods of time, neither to monitor whether the levels of psychological IPVAW change or remain constant over time. A second limitation is the tradeoff associated with the dichotomization of the items of the psychological IPVAW measure. This transformation was necessary to the set same metric for all the items, but we lose a small percentage of information (i.e., less than a 2%) about the most continued forms of this type of violence. The third limitation of this study refers to the assessment of the latent structure of the psychological IPVAW measure used in the FRA survey. Although this measure includes three forms of psychological violence (i.e., controlling behavior, economic violence, and emotional abuse), it was not possible to test a three-factor structure differentiating between these three aspects due to the reduced number of items addressing economic violence included in the FRA survey, since estimating a third factor with only two indicators loading on it could yield an unstable and unreliable solution[42,43]. The fourth limitation concern the estimation method used to establish the

measurement invariance of the set of questions included in the FRA survey, that is, WLSMV for categorical indicators. We decided to rely on the CFI and RMSEA fit indices to assess the goodness of fit of the invariance models rather than using chi-squared based tests, which are known to be sensible to large sample sizes. However, utilizing fit indices to study the measurement invariance with categorical data could lead to higher rates of Type I errors, assuming an invariant model when actually the instrument is non-invariant[50]. To tackle this issue, instead of the usual cut-offs for the TLI and RMSEA proposed by Chen[49], we decided to use the cut-off values proposed by Meade et al.[51], a much more restrictive approach. A fifth limitation of this study is the wide range of the confidence intervals of the psychological IPVAW factor means, which suggest that this measure could be improved in order to yield more accurate estimations of the IPVAW levels. Self-selection bias is another limitation, since only a 42% of the participants agreed initially to be interviewed. This is also reflected in the wide variability in the response rates across the EU countries, with some countries presenting response rates lower than 30% (e.g., Luxembourg, Netherlands, or Sweden), whereas others showed response rates above 60% (e.g., Cyprus, Hungary, or Latvia)[40]. Finally, the results of this study are limited to IPVAW, and the question of whether measurement invariance holds across countries when the perpetrators of the psychological violence against women are non-partners remains unexplored.

Conclusion

In this study we have tested the comparability of the psychological IPVAW measure used in the FRA survey by conducting a measurement invariance analysis. Although psychological IPVAW is one the most frequent form of intimate partner violence, and one of the main risk factors of physical IPVAW, to the best of our knowledge, no previous study has addressed the

measurement invariance of the measures used in any large survey to assess this construct. Our findings showed that the psychological IPVAW measure used in the FRA survey is invariant across all EU countries, allowing us to conduct a more refined analysis of how the levels of this variable are distributed across the EU. This is an important step towards a rigorous assessment of cross-national differences in psychological IPVAW, and further research is needed to evaluate the role that the country plays in accounting for the differences across EU countries.

List of Abbreviations

CFA: Confirmatory factor analysis.

CFI: Comparative fit index.

CI: confidence interval.

CTS: Conflict Tactics Scale.

EU: European Union.

FRA: European Union Fundamental Rights Agency.

IPVAW: Intimate partner violence.

MG-CFA: Multi-group confirmatory factor analysis.

RMSEA: root mean squared error of approximation.

TLI: Tucker-Lewis index.

WLSMV: weighted least squares means and variances adjusted.

Declarations

Ethics approval and consent to participate. The European Union Agency for Fundamental Rights supervised and approved the secondary data analyses conducted in this study, as well as the study design, granting a special license for this purpose (reference number 102577).

Supplementary Material

The supplementary materials of the current version of the manuscript can be found at: http://tiny.cc/SupMat2

References

1. Ali PA, Naylor PB. Intimate partner violence: A narrative review of the feminist, social and ecological explanations for its causation. Aggress Violent Beh. 2013;18;611-19.

2. Garcia-Moreno C, Jansen HA, Ellsberg M, Heise L, Watts CH. Prevalence of intimate partner violence: findings from the WHO multi-country study on women's health and domestic violence. Lancet. 2006;368;1260-9.

3. World Health Organization. Global and regional estimates of violence against women: prevalence and health effects of intimate partner violence and non-partner sexual violence. Geneva: World Health Organization; 2013.

4. Campbell JC. Health consequences of intimate partner violence. Lancet. 2002;359;1331-6.

5. Craparo G, Gori A, Petruccelli I, Cannella V, Simonelli C. Intimate partner violence: relationships between alexithymia, depression, attachment styles, and coping strategies of battered women. J Sex Med. 2004;11;1484-94.

6. Guedes A, Bott S, Garcia-Moreno C, Colombini M. Bridging the gaps: a global review of intersections of violence against women and violence against children. Global Health Action. 2016;9; doi:10.3402/gha.v9.31516.

7. Vilariño M, Amado BG, Vázquez MJ, Arce R. Psychological harm in women victims of intimate partner violence: Epidemiology and quantification of injury in mental health markers. Psychosoc Interv. 2018;27;145-52.

8. Hamby SL, Sugarman DB. Acts of psychological aggression against a partner and their relation to physical assault and gender. J Marriage Fam. 1999;61;959-70.

9. Karakurt G, Silver KE. Emotional abuse in intimate relationships: The role of gender and age. Violence Victims. 2013;27;804-21.

10. Liles S, Usita P, Irvin VL, Hofstetter CR, Beeston T, Hovell MF. Prevalence and correlates of intimate partner violence among young, middle, and older women of Korean descent in California. J Fam Violence. 2012;27;801-11.

11. Murphy CM, O'Leary KD. Psychological aggression predicts physical aggression in early marriage. J Consult Clin Psych. 1989;57;579-82.

12. Salis KL, Salwen J, O'Leary KD. The predictive utility of psychological aggression for intimate partner violence. Partner Abuse. 2014;5;83-97.

13. Schumacher JA, Leonard KE. Husbands' and wives' marital adjustment, verbal aggression, and physical aggression as longitudinal predictors of physical aggression in early marriage. J Consult Clin Psych. 2005;73;28-37.

14. Carney MM, Barner JR. Prevalence of partner abuse: Rates of emotional abuse and control. Partner Abuse. 2012;3;286-335.

15. Coker AL, Davis E, Arias I, Desai S, Sanderson M, Brand HM, et al. Physical and mental health effects of intimate partner violence for men and women. Am J Prev Med. 2012;23;260-8.

16. Kaukinen C. Status compatibility, physical violence, and emotional abuse in intimate relationships. J Marriage Fam. 2004;66;452-71.

17. Romans S, Forte T, Cohen MM, Du Mont J, Hyman I. Who is most at risk for intimate partner violence? A Canadian population-based study. J Interpers Violence. 2007;22;1495-1514.

18. Stets JE, Straus MA. The marriage license as a hitting license: A comparison of assaults in dating, cohabiting, and married couples. J Fam Violence. 1989;4;161-80.

19. Straight ES, Harper FW, Arias, I. The impact of partner psychological abuse on health behaviors and health status in college women. J Interpers Violence. 2003;18;1035-54.

20. Tjaden P, Thoennes, N. Prevalence and consequences of male-to-female and female-to-male intimate partner violence as measured by the National Violence Against Women Survey. Violence Against Wom. 2006:6;142-61.

21. Straus MA. Measuring intrafamily conflict and violence: The Conflict Tactics Scales. J Marriage Fam. 1979;41;75-88.

22. Straus MA., Hamby SL, Boney-McCoy S, Sugarman DB. The revised conflict tactics scales (CTS2) development and preliminary psychometric data. J Fam Issues. 1996;17;283-316.

23. Beeble ML., Bybee D, Sullivan CM. Abusive men's use of children to control their partners and ex-partners. Eur Psychol. 2007;12;54-61.

24. Follingstad DR, Coyne S, Gambone L. A representative measure of psychological aggression and its severity. Violence Victims. 2005;20;25-38.

25. Hamberger LK, Larsen SE. Lehrner A. Coercive control in intimate partner violence. Aggress Violent Beh. 2017;37;1-11.

26. European Union Agency for Fundamental Rights. Violence against women: An EU-wide survey. Main results. Luxembourg: Publications Office of the European Union; 2014.

27. Smith, SG, Zhang, X, Basile, KC, Merrick, MT, Wang, J, Kresnow, M, et al. The National Intimate Partner and Sexual Violence Survey (NISVS): 2015 Data Brief – Updated Release. Atlanta, GA: National Center for

Injury Prevention and Control, Centers for Disease Control and Prevention; 2018.

28. Milfont TL, Fisher R. Testing measurement invariance across groups: Applications in cross-cultural research. Int J Psychol Res. 2010;3;111-30.

29. Putnick DL, Bornstein MH. Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. Dev Rev. 2016;41;71-90.

30. Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. Organ Res Methods. 2000;3; 4-70.

31. Davidov E, Meuleman B, Cieciuch J, Schmidt P, Billiet J. Measurement equivalence in cross-national research. Annual Rev Sociol. 2014;40;55-75.

32. Jang S, Kim ES, Cao C, Allen TD, Cooper CL, Lapierre LM, et al. Measurement invariance of the satisfaction with life scale across 26 countries. J Cross Cult Psychol. 2017;48;560-76.

33. Van de Schoot R., Lugtig P, Hox J. A checklist for testing measurement invariance. Eur J Dev Psychol. 2012;9;486-492.

34. Lubke GH, Dolan CV, Kelderman H, Mellenbergh GJ. On the relationship between sources of within-and between-group differences and measurement invariance in the common factor model. Intelligence. 2003;31;543-66.

35. Capaldi DM, Knoble NB, Shortt JW, Kim HK. A systematic review of risk factors for intimate partner violence. Partner Abuse. 2012;3;231-80.

36. Hamby S, Grych J. The web of violence: Exploring connections among different forms of interpersonal violence and abuse. London: Springer; 2013.

37. Renninson C, Planty M. Nonlethal intimate partner violence: Examining race, gender, and income patterns. Violence Victims. 2003;18;433-43.

38. Staggs SL, Riger S. Effects of intimate partner violence on low-income women's health and employment. Am J of Community Psychol. 2005;36;133-45.

39. Thompson MP, Arias I, Basile KC, Desai S. The association between childhood physical and sexual victimization and health problems in adulthood in a nationally representative sample of women. J Interpers Violence. 2002;17;1115-29.

40. European Union Agency for Fundamental Rights. Violence against women: an EU-wide survey: Survey methodology, sample and fieldwork. Luxembourg: Publications Office of the European Union; 2014.

41. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Standards for educational and psychological testing. Washington, DC: American Educational Research Association; 1999.

42. Hair J, Black W, Babin B, Anderson R. Multivariate Data Analysis. New Jersey: Person Educational, Inc; 2014.

43. Tomas JM, Hontangas PM, Oliver A. Linear confirmatory factor models to evaluate multitrait-multimethod matrices: The effects of number of indicators and correlation among methods. Multivar Behav Res. 2000;35;469-99.

44. Hu LT, Bentler PM. Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Modeling. 1999;6;1-55.

45. MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. Psychol Methods. 1996;1;130-149 .

46. Revelle W. Psych: procedures for personality and psychological research. Evanston: Northwestern University; 2018.

47. McNeish D. Thanks coefficient alpha, we'll take it from here. Psychol Methods. 2018;23;412-33.

48. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct Equ Modeling. 2002;9;233-55.

49. Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. Struct Equ Modeling. 2008;14;464-504.

50. Sass DA, Schmitt TA, Marsh HW. Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. Struct Equ Modeling. 2014;21;167-80.

51. Meade AW, Johnson EC, Braddy PW. Power and sensitivity of alternative fit indices in tests of measurement invariance. J Appl Psychol. 2008;93;568-92.

52. Miles J, Shevlin, M. Applying Regression and Correlation: A Guide for Students and Researchers. London: Sage; 2001.

53. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale: Lawrence Earlbaum Associates; 1988.

54. Ruscio J. A probability-based measure of effect size: Robustness to base rates and other factors. Psychol Methods. 2008;13;19-30.

55. Hanel PH, Mehler DM. Beyond reporting statistical significance: Identifying informative effect sizes to improve scientific communication. Public Underst Sci. 2019;28;468-85.

56. R Core Team. R: A language and environment for statistical computing. Austria: R foundation for Statistical Computing; 2019. Retrieved from: http://CRAN.R-project.org/package=psychVersion=1.8.12.

57. Muthén LK, & Muthén BO. Mplus user's guide. Los Angeles, CA: Muthén & Muthén; 2017.

58. Heise L, Pallitto C, García-Moreno C, Clark CJ. Measuring psychological abuse by intimate partners: Constructing a cross-cultural indicator for the Sustainable Development Goals. SSM-Population Health. 2019;100377. doi: https://doi.org/10.1016/j.ssmph.2019.100377

59. Till-Tentschert U. The Relation Between Violence Experienced in childhood and women's exposure to violence in later life: Evidence from Europe. J Interpers Violence. 2017;32;1874-94.

60. Davidov E, Schmidt, Schwartz SH. Bringing values back in: The adequacy of the European Social Survey to measure values in 20 countries. Public Opin Quart. 2008;72;420-45.

61. Muthén B, Asparouhov T. Recent methods for the study of measurement invariance with many groups: alignment and random effects. Sociol Methods Res. 2018;47;637-64.

62. Gracia E, Merlo J. Intimate partner violence against women and the Nordic paradox. Soc Sci Med. 2016;157;27-30.

63. Gracia E, Martín-Fernández M, Lila M, Merlo J, Ivert AK. Prevalence of intimate partner violence against women in Sweden and Spain: A psychometric study of the 'Nordic paradox'. PloS One. 2019;14; doi: 10.1371/journal.pone.0217015.

Study 3

Prevalence of intimate partner violence against women in Sweden and Spain: A psychometric study of the 'Nordic paradox'

Prevalence of intimate partner violence against women in Sweden and Spain: A psychometric study of the 'Nordic paradox'[*]

Enrique Gracia[1], Manuel Martín-Fernández[1], Marisol Lila[1], Juan Merlo[2] & Anna-Karin Ivert[3]

[1] Department of Social Psychology, University of Valencia, Valencia, Spain.
[2] Unit for Social Epidemiology, University of Lund, Malmö, Sweden.
[3] Department of Criminology, Malmö University, Malmö, Sweden.

Abstract

The high prevalence of intimate partner violence against women (IPVAW) in countries with high levels of gender equality has been defined as the "Nordic paradox". In this study we compared physical and sexual IPVAW prevalence data in two countries exemplifying the Nordic paradox: Sweden (N = 1483) and Spain (N = 1447). Data was drawn from the European Union Agency for Fundamental Rights Survey on violence against women. To ascertain whether differences between these two countries reflect true differences in IPVAW prevalence, and to rule out the possibility of measurement bias, we conducted a set of analyses to ensure measurement equivalence, a precondition for appropriate and valid cross-cultural comparisons. Results showed that in both countries items were measuring two separate constructs, physical and sexual IPVAW, and that these factors had high internal consistency and adequate

validity. Measurement equivalence analyses (i.e., differential item functioning, and multigroup confirmatory factor analysis) supported the comparability of data across countries. Latent means comparisons between the Spanish and the Swedish samples showed that scores on both the physical and sexual IPVAW factors were significantly higher in Sweden than in Spain. The effect sizes of these differences were large: 89.1% of the Swedish sample had higher values in the physical IPVAW factor than the Spanish average, and this percentage was 99.4% for the sexual IPVAW factor as compared to the Spanish average. In terms of probability of superiority, there was an 80.7% and 96.1% probability that a Swedish woman would score higher than a Spanish woman in the physical and the sexual IPVAW factors, respectively. Our results showed that the higher prevalence of physical and sexual IPVAW in Sweden than in Spain reflects actual differences and are not the result of measurement bias, supporting the idea of the Nordic paradox.

Intimate partner violence against women (IPVAW) remains a pervasive social and public health problem in western societies [1-8]. Increasing gender equality is at the core of the prevention efforts of this type of violence, as gender inequality is considered a main factor explaining IPVAW. Accordingly, rates of IPVAW are expected to drop as country-level gender equality increases [9-12, see 13, for a review]. However, in western countries, high country levels of gender equality are not always linked with low prevalence of IPVAW.

The high prevalence of IPVAW in countries with high levels of gender equality was defined by Gracia and Merlo as the "Nordic paradox" [14]. Nordic countries are, according to different international indicators (e.g., Global Inequality Index; Global Gender Gap Index; European Index of Gender Equality), the most gender equal countries in the world [15-17]. However, despite these high levels of gender equality, Nordic countries have high prevalence rates of IPVAW. The high prevalence of IPVAW in Nordic countries is illustrated by a European Union (EU) survey on violence against women conducted by the European Union Agency for Fundamental Rights (FRA) [18]. In this survey the average lifetime prevalence of physical and/or sexual violence by intimate partners in the 28 EU member states was 23%, with a range between 13% and 32%. However, Nordic countries in the EU were among the countries with higher lifetime prevalence of IPVAW, with rates of 32% (Denmark, the highest IPV prevalence in the EU), 30% (Finland), and 28% (Sweden). The high prevalence of IPVAW in Nordic countries is also supported by other studies and national surveys [19-25]. However, despite survey and research data pointing to a disproportionally high level of IPVAW in countries with the highest levels of gender equality like the Nordic ones, interestingly, this puzzling research question is rarely asked and, so far, remains unanswered.

The reasons explaining these high levels of IPVAW prevalence in Nordic countries, despite their high levels of gender equality, are not yet understood as almost no research has addressed specifically this paradox [22]. Gracia and Merlo [14], proposed a number of theoretical and methodological lines of inquiry towards understanding the Nordic paradox. However, as these authors noted [14], a first step to ascertain whether the Nordic paradox reflects true differences in IPVAW prevalence is to rule out the possibility that measurement bias is causing prevalence differences between Nordic and other countries. To eliminate this possibility, a key question is to ensure the comparability of IPVAW prevalence data across countries. In other words, comparisons of IPVAW data across countries should not be made without first ensuring measurement invariance.

IPVAW can be a culturally sensitive issue, and the way this type of violence is perceived or reported may vary across countries. Therefore, ensuring cross-cultural measurement invariance is critically important for appropriate and valid cross-cultural comparisons of self-reported IPVAW scores between respondents from different countries [26-32]. As Jang et al. noted [29], different perceptions of items or different interpretations of response scales can lead to measurement non-invariance (i.e., non-equivalence of measures). If this is the case, it cannot be assumed that the construct of interest, in our case IPVAW, is interpreted in the same way across countries because the same score in one country may have a different meaning or reflect different levels of IPVAW in another. Without ensuring measurement invariance, score comparisons across samples from different countries can be unreliable and inadequate, and the validity of comparing women's IPVAW experiences across countries becomes questionable [28,29,32,33].

Present Study

Sweden and Spain are two countries exemplifying the Nordic paradox. According to several international gender equality indices, Sweden is ranked third in the Global Inequality Index [17], fifth in the Global Gender Gap Index [16], and first in the EU in the European Index of Gender Equality [15]. According to the same sources, Spain is ranked 13th (Global Inequality Index) or 24th (Global Gender Gap Index) in the world, and 11th in the EU (European Index of Gender Equality). However, despite the higher gender equality in Sweden, Spain has a substantially lower prevalence of IPVAW.

The FRA survey provides a composite indicator of the prevalence of physical and/or sexual violence by any partners (current and/or previous) since the age of 15. According to this indicator, the lifetime prevalence of physical and/or sexual violence among women perpetrated by any partner is 28% in Sweden and 13% in Spain ([18], p. 28). That is, the lifetime prevalence of physical and/or sexual IPVAW is 15 percentage points higher in Sweden than in Spain, while, according to the European Index of Gender Equality, gender equality in Sweden is 14 points higher than in Spain (the updated Gender Equality Index data for the year when the survey was conducted was 64 in Spain and 78 in Sweden, and is currently 68 in Spain and 82 in Sweden) [15].

One of the advantages of the FRA survey is that respondents from the 28 EU member states answer the same set of questions addressing different types of IPVAW. Another advantage of this survey is that it includes questions regarding IPVAW that are acts-based or behavioral oriented (e.g., being stabbed, cut, slapped, or being forced into sexual intercourse). This type of questions addressing IPVAW have a clear advantage over simply asking women whether their partners or ex-partners have ever been violent towards them, which is a more subjective approach and can lead to underreporting

[4,34,35]. However, as the psychometric properties of the set of questions addressing physical and sexual IPVAW used in the FRA survey are unknown, and the measurement equivalence across countries (i.e., cross-cultural invariance) of these questions has never been tested, it is not possible to ascertain whether the differences between Sweden and Spain in lifetime prevalence of physical and sexual IPVAW reflect real differences or are the result of measurement bias.

A precondition to compare prevalence data on physical and sexual IPVAW across countries, in our case Sweden and Spain, is the availability of an equivalent measurement model. In this study, we aim to analyze whether the set of questions assessing physical and sexual IPVAW used in the FRA survey are reliable, valid and comparable measures of these types of violence across Sweden and Spain. If the measures of physical and sexual IPVAW are comparable and confirm higher levels of physical and sexual violence in Sweden than in Spain, these results would support the idea that the Nordic paradox (at least with respect to Sweden and Spain) reflects real prevalence differences.

Method

Participants

In this study we used the Spanish (N = 1447) and Swedish (N = 1483) samples from the survey conducted by the European Union Agency for Fundamental Rights on violence against women [18]. Respondents to this survey were ever-partnered women, aged 18 to 74. The sampling followed a two-stage clustered stratified design with the same probability of selection of households within clusters. The responses were collected in-person in both countries, although in Sweden the first contact was made telephonically. Further details on sample collection and procedures can be found in FRA [36].

Table 1. Socio-demographical variables (unweighted)

	Total (%)	Spain (%)	Sweden (%)
Age			
18-24	127 (4.3)	77 (5.3)	50 (3.4)
25-29	164 (5.6)	102 (7.0)	62 (4.2)
30-34	218 (7.4)	122 (8.4)	96 (6.5)
35-39	270 (9.2)	163 (11.2)	107 (7.2)
40-49	662 (22.6)	335 (23.2)	327 (22.0)
50-59	624 (21.3)	272 (18.8)	352 (23.7)
60+	865 (29.5)	376 (26.0)	489 (33.0)
Income			
Finding it very difficult on present income	172 (5.9)	132 (9.1)	40 (2.7)
Finding it difficult on present income	436 (14.9)	263 (18.2)	173 (11.7)
Coping on present income	1226 (41.8)	608 (42.0)	618 (41.7)
Living comfortably on present income	1075 (36.7)	428 (29.6)	647 (43.6)
Education			
Has not completed primary education	250 (8.5)	75 (5.2)	4 (0.3)
Completed primary education	719 (24.5)	221 (15.3)	62 (4.2)
Compulsory secondary education	422 (14.4)	461 (15.3)	84 (5.7)
Upper secondary education	620 (21.2)	282 (31.9)	331 (22.3)
Post-secondary education (but not university level)	549 (18.7)	135 (19.5)	280 (18.9)
Graduate studies	286 (9.8)	219 (15.1)	493 (33.2)
Post-graduate studies	250 (8.5)	53 (3.7)	195 (13.1)
Self-perceived Health			
Very Bad	28 (1.0)	16 (1.1)	12 (0.8)
Bad	173 (5.9)	94 (6.5)	79 (5.3)
Fair	555 (18.9)	295 (20.4)	260 (17.5)
Good	1254 (42.8)	684 (47.3)	570 (38.4)
Very Good	916 (31.3)	354 (24.5)	562 (37.9)
My partner or an ex-partner has been physically violent against me			
Yes	404 (15.9)	137 (9.7)	267 (23.6)
No	2143 (84.1)	1280 (90.3)	863 (76.4)

My partner or an ex-partner has been sexually violent against me			
Yes	127 (5.0)	59 (4.2)	68 (6.0)
No	2416 (95.0)	1357 (95.8)	1059 (94.0)

A special license for secondary data analysis was obtained from FRA (Reference No. 102577). Socio-demographical variables of both samples are described on Table 1.

Measures

Physical violence. The FRA survey included 10 items addressing physical IPV perpetrated by the current or any previous partner (e.g., "Your current/previous partner has slapped you?", "Your current/previous partner has grabbed you or pulled your hair?"). Participants have to answer in a 4-point Likert-type scale indicating how often have they experienced this type of violence (1: "Never", 2: "Once", 3: "2-5 times", 4: "6 or more times"). In this study respondents were considered victims of intimate partner physical violence when reported one of the episodes described by the items at least one time, whereas severe violence was considered in those cases where respondents have experienced the episodes more than one time.

Sexual violence. FRA survey addresses intimate partner sexual violence with 4 items describing episodes of sexual violence perpetrated by the current or any previous partner (e.g., "Your current/previous partner has forced you into sexual intercourse by holding you down or hurting you in some way?", "Your current/previous partner has made you take part in any form of sexual activity when you did not want to or you were unable to refuse?"). Respondents have to indicate how often have they experienced this type of violence using a 4-point Likert-type (1: "Never", 2: "Once", 3: "2-5 times", 4: "6 or more times"). Respondents were considered as victims of intimate partner sexual violence when reported one of the episodes described by the items at

114

least one time, whereas severe violence was considered in those cases where respondents have experienced the episodes more than one time.

Validity evidence based on relations to other variables

To test validity based on relations to other variables [37], we used two measures: (1) *Self-perceived health*. The FRA survey included an item in which respondents are asked how their health was in general, and they have to answer using a 5-point Likert-type ordinal scale (ranging from 1 = "Very Bad" to 5 = "Very Good"). "Do not know", "Not applicable" and "Refused" categories were treated as missing values. (2) *Self-reported physical and sexual IPVAW victimization*. At the end of the FRA survey, respondents are asked to complete, confidentially, two dichotomous items (Yes/No) about experienced life-time physical IPV ("My partner or an ex-partner has been physically violent against me"), and experienced life-time sexual IPV ("My partner or an ex-partner has been sexually violent against me").

Data Analyses

First, descriptive analyses of the set of items assessing physical and sexual violence included in the FRA survey were conducted. The mean, standard deviation, skewness, and kurtosis statistics were computed for each item. These statistics were obtained with the unadjusted responses of the participants, as the aim was to study the properties of the items.

A confirmatory factor analysis (CFA) was conducted to assess the latent structure (i.e., internal construct validity) of the set of questions used in the FRA survey to address physical and sexual violence. Two models were estimated and compared using robust weighted least squares (WLSMV), as this method tend to perform better with categorical data [38]. The first model was a one-factor model in which all items loaded onto a single violence factor, implying that all violent acts, regardless of their physical or sexual nature, pertained to the same construct. The second model was a two-factor model

where the items addressing physical violence loaded on one factor and the items assessing sexual violence on another factor, implying that each set of items were sampling different constructs. In this second model the factors are correlated, and thus these two constructs are assumed to be related. Model fit was tested with a combination of fit indices: comparative fit index (CFI), Tucker-Lewis index (TLI), and root mean square error of approximation (RMSEA). CFI and TLI values above .95 are indicative of good fit, whereas RMSEA values below .08 and .06 are considered indicative of mediocre and good fit, respectively [39,40]. Once the latent structure is determined, the internal consistency of the resulting factor or factors will be studied by computing Cronbach's α and McDonald's ω. McDonald's ω is more suitable when the items are not tau-equivalent (i.e., they do not have the same factor loadings) [41]. After establishing the latent structure of the items, validity based on relationships with other variables was tested conducting a set of mean comparisons and correlations with variables with expected links to IPVAW (i.e., self-perceived health, and self-reported physical and sexual IPVAW victimization).

Once these analyses have been carried out separately for both Sweden and Spain, to ensure the comparability of IPVAW scores across these countries two complementary analyses were conducted: A differential item functioning (DIF) analysis for categorical data, and a multi-group confirmatory factor analysis (MG-CFA) between countries to test measurement invariance [42-45]. Both procedures aim to assess whether there is a group effect (i.e., country) on IPVAW factor scores, but they focus on different issues. Whereas the DIF focuses on the equivalence of the latent scores, the MG-CFA focuses on the equivalence of the structural parameters of the model (e.g., loadings and intercepts). First, a DIF analysis was conducted using the logistic regression method [46,47]. An item presents DIF when the probability of endorsement of

an item category is not the same for respondents from different groups (i.e., countries) with equivalent scores in the factor, indicating that the respondents of each group are answering that item differentially. Second, a series of MG-CFA was conducted, testing configural, metric, and scalar measurement invariance levels across the Swedish and Spanish samples [27,48-51]. These levels of invariance are required for a meaningful comparison of IPVAW scores for Sweden and Spain. Configural invariance evaluate whether Swedish and Spanish women conceptualize the construct in the same way, testing if the same factorial model fits for both groups. Metric invariance constraint the factor loadings to be equal across groups, implying that Swedish and Spanish respondents interpret the items similarly. Scalar invariance test whether the same threshold parameters could be estimated for each group, indicating that the items yield the same factor score for Swedish and Spanish samples. Change in CFI (ΔCFI) and RMSEA (ΔRMSEA) was computed to test which of these invariance levels were better supported by the data. If the change in the CFI (ΔCFI) and in the RMSEA (ΔRMSEA) is below .010 or .015, respectively, then the most restrictive level of invariance is supported [26,52].

After assessing measurement invariance, the raw prevalences of the items were compared as a descriptive analysis of the differences between Sweden and Spain. Finally, a MG-CFA latent means analysis was also conducted, to analyze IPVAW differences across countries. Factor scores on latent variables provides a more refined approach to assess differences in IPVAW between two countries. They are continuous variables that take into account how relevant for the factor is each item, and can capture more variability. To assess the magnitude of the latent mean differences, Cohen's *d* effect size index was obtained using the resulting factor scores [49]. Cohen *d* can also be used to compute the Cohen U_3 statistic, which evaluate the percentage of cases of one group that is higher than the average of the other

group, and the probability of superiority, which indicates the probability that a person selected at random from one group will have a higher score than a person randomly selected from the other group [53-55].

Descriptive, DIF, and validity analyses were carried out with the statistical software package R [56], using the *psych* and *lordif* libraries [46,57]. The CFA and the measurement invariance analyses were conducted with the software package *Mplus* [58].

Results

Descriptive Analyses

The descriptive statistics of the items addressing physical violence can be found in Table 2. The means of the items were around 1, the lowest category (i.e., "Never"), with standard deviations around 0.4 and 0.5 for the Spanish and Swedish women, respectively. Both groups presented positive skew statistics and high kurtosis values, indicating that most of the responses were centered in the lower categories. The variance of the items 7 and 9 (i.e., "being burned", and "being cut, stab or shot") was extremely low, indicating that almost none of the respondents reported experiencing this type of violence. Given the lack of variability in the responses on these items in both countries (1% or less), they were removed for subsequent analyses.

Regarding the sexual violence items (Table 3), the means were also centered on the lower category (i.e., "Never"), with standard deviations around 0.40 and 0.50 for the Spanish and the Swedish respondents, and showed a positive skew and had high kurtosis values. As in the physical violence items, the respondents tended to select the lower categories in the sexual violence items.

Confirmatory factor analysis and internal consistency

A one-factor model and a two-factor model were then estimated to

Table 2. Physical violence items descriptive statistics

Spain

	M	SD	Skew	Kurtosis
Threatened to hurt you physically?	1.19	0.66	3.55(.02)	11.39(.02)
Pushed you or shoved you?	1.20	0.65	3.39(.02)	10.46(.02)
Slapped you?	1.13	0.55	4.28(.01)	17.61(.01)
Thrown a hard object at you?	1.07	0.40	6.25(.01)	39.70(.01)
Grabbed you or pulled your hair?	1.09	0.45	5.24(.01)	27.54(.01)
Beat you with a fist or a hard object, or kicked you?	1.08	0.44	5.55(.01)	30.97(.01)
Burned you?	1.01	0.11	18.91(.00)	356.01(.00)
Tried to suffocate you or strangle you?	1.04	0.26	8.26(.01)	75.82(.01)
Cut or stabbed you, or shot at you?	1.01	0.11	18.73(.00)	407.65(.00)
Beat your head against something?	1.04	0.29	8.15(.01)	70.02(.01)

Sweden

	M	SD	Skew	Kurtosis
Threatened to hurt you physically?	1.30	0.79	2.53(.02)	5.08(.02)
Pushed you or shoved you?	1.37	0.83	2.12(.02)	3.19(.02)
Slapped you?	1.21	0.65	3.2(.02)	9.36(.02)
Thrown a hard object at you?	1.13	0.52	4.24(.01)	17.79(.01)
Grabbed you or pulled your hair?	1.17	0.61	3.64(.02)	12.31(.02)
Beat you with a fist or a hard object, or kicked you?	1.16	0.60	3.84(.02)	13.80(.02)
Burned you?	1.01	0.10	22.90(.00)	616.22(.00)
Tried to suffocate you or strangle you?	1.05	0.29	7.23(.01)	56.92(.01)
Cut or stabbed you, or shot at you?	1.00	0.07	15.24(.00)	230.51(.00)
Beat your head against something?	1.08	0.40	5.47(.01)	31.10(.01)

Note: M = Mean, SD = Standard Deviation, In brackets: Skew and Kurtosis statistics standard error.

determine the latent structure of the items for each country separately, using WLSMV as the estimation method. Both models converged successfully. The one- and two-factor models fitted adequately in the Spanish sample (Table 4).

Table 3. Sexual violence items descriptive statistics

SPAIN

	M	SD	Skew	Kurtosis
has forced you into sexual intercourse by holding you down or hurting you in some way?	1.06	0.38	6.86(.01)	47.00(.01)
has attempted to force you into sexual intercourse by holding you down or hurting you in some way?	1.05	0.36	7.31(.01)	53.59(.01)
has made you take part in any form of sexual activity when you did not want to or you were unable to refuse?	1.05	0.35	7.57(.01)	57.29(.01)
Have you consented to sexual activity because you were afraid of what your current partner might do if you refused?	1.06	0.39	6.85(.01)	46.56(.01)

SWEDEN

	M	SD	Skew	Kurtosis
has forced you into sexual intercourse by holding you down or hurting you in some way?	1.10	0.48	4.91(.01)	23.65(.01)
has attempted to force you into sexual intercourse by holding you down or hurting you in some way?	1.09	0.45	5.15(.01)	26.49(.01)
has made you take part in any form of sexual activity when you did not want to or you were unable to refuse?	1.13	0.54	4.17(.01)	16.76(.01)
Have you consented to sexual activity because you were afraid of what your current partner might do if you refused?	1.11	0.51	4.88(.01)	22.95(.01)

Note: M = Mean, SD = Standard Deviation. In brackets: Skew and Kurtosis statistics standard error.

In the Swedish sample the one-factor model showed a good fit to the data, although the RMSEA was mediocre. Adding a second factor improved substantially the RMSEA in the Swedish group, being below the .06 cut-off for a well fitted model. For this reason, we decided to keep the two-factor solution in both samples, as both countries showed similar fit indices. The factor loadings of the items in the Spanish and the Swedish samples were high,

Table 4. CFA fit indices

Model		χ^2	_df_	CFI	TLI	RMSEA [95% CI]
One-factor	Spain	243.24	54	0.99	0.99	0.049 [0.043; 0.056]
	Sweden	630.74	54	0.97	0.97	0.086 [0.080; 0.092]
Two-factor	Spain	125.51	53	0.99	0.99	0.031 [0.024; 0.038]
	Sweden	159.71	53	0.99	0.99	0.037 [0.030; 0.037]

Note: CFI: Comparative fit index, TLI: Tucker-Lewis index, RMSEA = Root mean squared error of approximation.

showing values above .80 in both factors (Figure 1). This indicates that the items were strongly related to the measured construct. The correlations between the factors were also high, .84 and .72 for the Spanish and the Swedish groups, respectively.

Regarding the internal consistency, both factors showed a high internal consistency. In particular, the physical IPVAW factor showed a Cronbach's α = .91 in both countries, and a McDonald's ω = .92 in the Spanish sample and .91 in the Swedish sample. The sexual IPVAW factor had a Cronbach's α = .88 and .86, and a McDonald's ω = .90 and .86 in the Spanish and Swedish groups, respectively.

Validity evidence based on relations to other variables

The standardized factor scores from the two-factor model were used to conduct the validity analyses, as the items did not contribute equally to their factor (i.e., are not tau-equivalent).

The scores on the physical IPVAW factor were compared by self-perceived health categories in each country separately. In Spain, we found significant differences in this factor, $F(4) = 6.39$, $p < .001$, $\eta2 = 017$. Post-hoc analysis showed that the differences in self-perceived health were between the upper
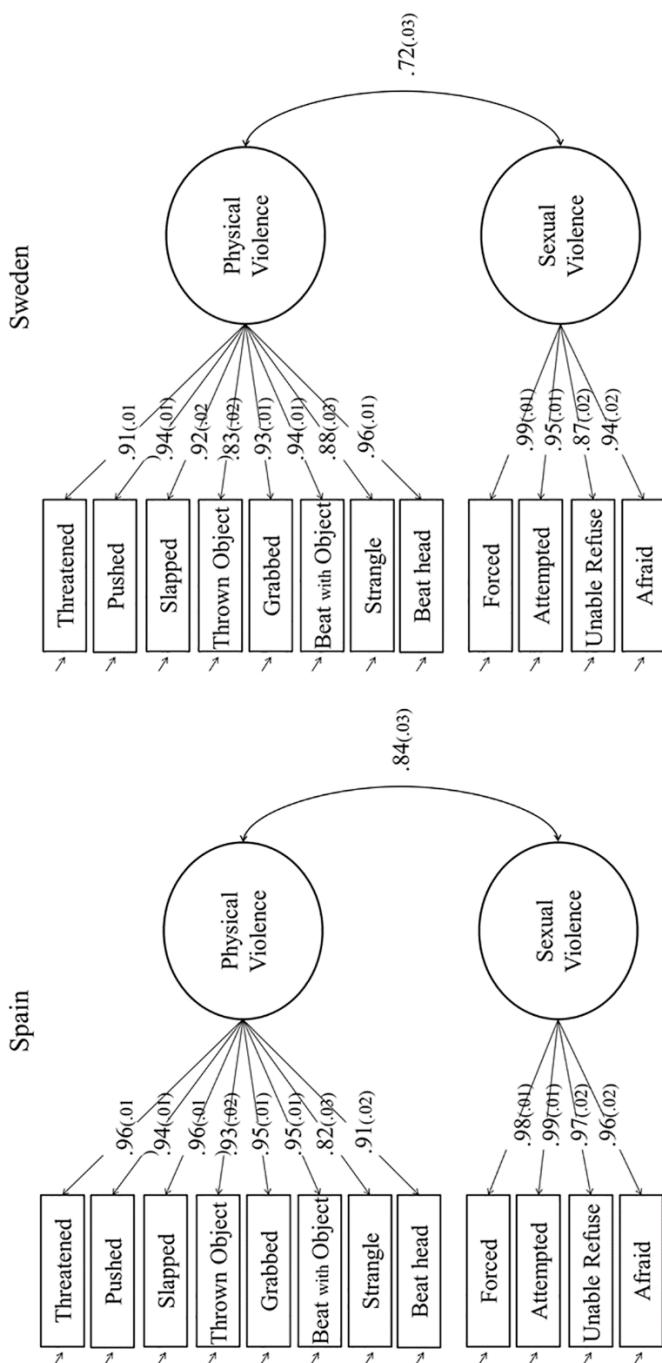
Figure 1. CFA two-factor model

two categories (i.e., "Very good" and "Good") and the lower two categories (i.e., "Bad" and "Very bad"), implying that respondents who indicated a positive self-perceived health showed lower scores on this factor. We also found significant differences in the Swedish sample by self-perceived health, $F(4) = 10.26$, $p < .001$, $\eta^2 = .027$. We found in the post-hoc analyses that respondents who chose the upper category in self-perceived health (i.e. "Very Good") showed lower scores in the physical IPVAW factor in comparison with the other response categories.

Regarding the sexual IPVAW factor, we found significant differences in the Spanish sample when the scores in this factor were compared by self-perceived health categories, $F(4) = 6.63$, $p < .001$, $\eta^2 = .018$. In particular, post-hoc analyses indicated that respondents who chose the lowest category of self-perceived health (i.e. "Very Bad") showed higher scores on the physical IPVAW factor than those respondents who chose the upper three categories (i.e., "Very Good", "Good", and "Fine"). Significant differences by self-perceived health were also found in the Swedish sample, $F(4) = 6.52$, $p < .001$, $\eta^2 = .017$. These differences were between the upper two and the lower two categories of this variable (i.e., "Very Good" and "Good" vs. "Very Bad" and "Bad").

The scores on both physical and sexual violence IPVAW factors were also correlated with the self-reported physical and sexual IPVAW victimization items of the FRA survey for each country separately. Biserial correlations were used. We found that the physical violence factor scores were positively related to the single item of the FRA survey addressing physical violence in Sweden and Spain ($r = .33$, $p < .001$, and $r = .37$, $p < .001$, respectively). The scores on this factor were also related to the sexual violence item, especially in Spain ($r = .16$, $p < .001$, in Sweden, and $r = .35$, $p < .001$, in Spain). Regarding the sexual violence, we found a positive relationship between the factor scores and the single item of the FRA survey in both countries ($r = .38$, $p < .001$, in Sweden,

and $r = .26$, $p < .001$, in Spain). The correlations between the sexual violence factor scores and the single item addressing physical violence from the FRA survey were also positive in Sweden and Spain ($r = .26$, $p < .001$, and $r = .25$, $p < .001$, respectively).

Measurement equivalence analyses

First, a series of nested logistic regression models for categorical data was conducted to identify, at item-level, if there was a group effect due to pertaining to different countries for both factors, and whether that effect was constant—uniform DIF—or varies across the factor scores—non-uniform DIF— [46]. We detected non-uniform DIF only for item 4 (i.e., "Thrown a hard object at you") in the physical IPVAW factor, $\chi^2(1) = 12.11$, $p < .001$, $R^2_{Nagelkerke} = .008$. The effect size in this item was, however, below 0.02, and thus this could be considered negligible, as adding the DIFF effect to the model does not improve their fit substantially [46,53]. No DIF was detected for any items of the sexual IPVAW factor.

Second, measurement invariance across countries was explored using a MG-CFA for the two-factor model. The configural, metric, and scalar invariance levels were tested (see Table 5). The configural invariance level was supported by the data, entailing that the same factorial model can be applied in both countries. Constraining the factor loadings to have the same value in both groups did not substantially decrease the fit of the CFI and the RMSEA fit indices (ΔCFI = .000, ΔRMSEA = .001), indicating that the metric invariance level could be assumed. Finally, when the threshold parameters of the items were constrained to be equal across groups, the change of the CFI and the RMSEA indices were below the ΔCFI = .10 and ΔRMSEA = .15 cut-offs, supporting the scalar invariance level.

Table 5. Measurement Invariance fit indices

	χ^2	df	CFI	TLI	RMSEA [95% CI]
Invariance level					
Configural	281.80	106	0.996	0.996	0.034 [0.029; 0.038]
Metric	305.73	116	0.996	0.996	0.033 [0.029; 0.038]
Scalar	551.18	152	0.992	0.993	0.042 [0.039; 0.046]
Latent Means	352.31	150	0.996	0.996	0.030 [0.026; 0.034]

Note: CFI: Comparative fit index, TLI: Tucker-Lewis index, RMSEA = Root mean squared error of approximation.

Raw prevalences

As no DIF was detected, item-based comparisons across countries can be made. All items, both in the physical and sexual IPVAW factors, had a higher prevalence in Sweden than in Spain (Table 6). These differences held for both general prevalence (physical IPVAW: Sweden: 7.9%, Spain: 4.3%; sexual IPVAW: Sweden: 5.5%, Spain: 2.3%) and severe prevalence (physical IPVAW: Sweden: 5.1%, Spain: 2.8%; sexual IPVAW: Sweden: 4%, Spain: 1.8%). As for the raw prevalence considering all items, in Sweden the prevalence of physical and sexual IPVAW was also higher than in Spain.

Latent means analysis

Once the scalar invariance level was established, the differences between Spanish and Swedish women were assessed estimating a new MG-CFA. This model assumes that the structural parameters (i.e., slopes and thresholds) are equal, and thus the means of the factor scores can be compared assuming that respondents interpret the items similarly in both groups. For the Spanish sample the mean was fixed to zero in both physical and sexual IPVAW factors, whereas in the Swedish sample these parameters were freed. The Swedish sample showed a higher latent mean in the physical IPVAW factor than the Spanish sample ($z = 0.72$, $p < .001$). The effect size of these differences between Sweden and Spain was large, *d = 1.23, Cohen's U_3 = .891, probability of*

Table 6. Physical and Sexual Violence Item Prevalences

	Spain		Sweden	
	Prev (%)	Sev Prev (%)	Prev (%)	Sev Prev (%)
Physical Violence				
Threatened to hurt you physically?	8.43	6.35	15.30	10.85
Pushed you or shoved you?	9.95	6.50	19.84	13.16
Slapped you?	6.56	4.49	11.59	6.94
Thrown a hard object at you?	3.52	2.14	7.35	4.25
Grabbed you or pulled your hair?	4.77	3.04	8.62	6.20
Beat you with a fist or a hard object, or kicked you?	4.22	2.76	7.88	5.59
Tried to suffocate you or strangle you?	2.49	0.09	3.10	1.35
Beat your head against something?	2.21	1.24	4.78	2.62
Total	12.43	8.03	27.86	16.76
Sexual Violence				
has forced you into sexual intercourse by holding you down or hurting you in some way?	2.63	1.94	5.25	3.70
has attempted to force you into sexual intercourse by holding you down or hurting you in some way?	2.21	1.80	4.85	3.43
has made you take part in any form of sexual activity when you did not want to or you were unable to refuse?	2.00	1.73	7.14	4.98
have you consented to sexual activity because you were afraid of what your current partner might do if you refused?	2.49	2.00	4.98	4.04
Total	4.28	3.09	10.90	7.45

superiority = .807. This means that 89.1% of the Swedish sample presented higher values in the physical IPVAW factor than the average of the Spanish sample, and if one woman is randomly selected from each country, there is an

80.7% probability that the Swedish woman will score higher in this factor than a Spanish woman.

Regarding the sexual violence factor, we found that the latent mean was also higher in the Swedish group ($z = 1.99$, $p < .001$). In this case the effect size was extremely large, $d = 2.5$, *Cohen's U_3* = .994, *probability of superiority* = .961, which means that the 99.4% of the Swedish women presented higher values on the sexual IPVAW factor scores than the Spanish women. Also, if one woman is randomly selected from each country, there is a 96.1% probability that the Swedish woman will score higher in the sexual IPVAW factor than a Spanish woman.

Discussion

In this study we compared physical and sexual IPVAW prevalence data in two countries exemplifying the Nordic paradox [14]: Sweden and Spain. To ascertain whether differences between these two countries reflect true differences in IPVAW prevalence, and to rule out the possibility of measurement bias, we conducted a set of analyses to ensure measurement equivalence, as a precondition for appropriate and valid cross-cultural comparisons. Once an equivalent measurement model had been established, we compared physical and sexual IPVAW scores between the two countries. Our results showed that the higher levels of physical and sexual IPVAW in Sweden than Spain reflect actual differences in IPVAW prevalence and are not the result of measurement bias, supporting the idea of the Nordic paradox.

The first set of analyses conducted in this study aimed to examine whether the series of questions assessing physical and sexual IPVAW used in the FRA survey were reliable and valid measures of this type of violence in both Sweden and Spain. First, results from CFA examining the latent structure of the items used in the FRA survey supported a two-factor model in the two

countries. That is, these items were measuring two separate constructs: physical and sexual IPVAW. Once the latent structure of the physical and sexual violence items had been established, reliability analyses (computing Cronbach's α and McDonald's ω) were conducted, showing that these scales had high internal consistency in both countries (all values ranging from .86 to .92). In this first set of analyses, we also addressed the validity of physical and sexual IPVAW factors based on their relations to other variables in the two countries. In both Sweden and Spain, scores in the physical and sexual IPVAW factors were significantly associated, as expected, to self-perceived health. The physical and sexual IPVAW scores were also correlated with two single-item measures of self-reported (not act-based measures) physical and sexual IPVAW victimization.

Once the psychometric properties of these measures had been established for each country, the next set of analyses aimed to ensure the comparability of these measures across Sweden and Spain by conducting different measurement equivalence tests. In the present study, to test the comparability of the physical and sexual IPVAW scales between Sweden and Spain, two complementary analyses were conducted: a DIF analysis and a MG-CFA. The joint use of these two techniques is one of the main strengths of the current manuscript, as they provide complementary information. In particular, both analyses showed that the country had no effect on the physical and the sexual IPVAW scores. No DIF was detected, indicating that the probability of endorsing a category of response in each item was the same for Swedish and Spanish respondents and, therefore, factors scores were comparable (i.e., no recalibration of item parameters was needed). Regarding MG-CFA, configural, metric, and scalar measurement invariance levels were supported, indicating that respondents in Sweden and Spain used the same conceptual framework to respond to the items (i.e., configural invariance), that

the items were interpreted in a similar way, contributing equally to the scale scores (i.e., metric invariance), and that differences across countries in the observed items were the result of actual differences in the corresponding latent factors of physical and sexual IPVAW (i.e., scalar invariance). Results from these measurement invariance analyses ensured the comparability of physical and sexual IPVAW scores between Spanish and Swedish respondents.

When we examined the raw prevalence of the items, both in the physical and sexual IPVAW scales, all had a higher prevalence in Sweden than in Spain (both for general and severe IPVAW). Considering all items together, the general lifetime prevalence of IPVAW was higher in Sweden (physical: 27.86%, sexual: 10.9%) than in Spain (physical: 12.43%, sexual: 4.3%). The same pattern was also found for severe physical (16.76% Sweden vs. 8.03% Spain) and sexual (7.4% Sweden vs. 3.1% Spain) IPVAW. However, although comparisons based on the raw prevalence can be useful as a first descriptive step, they provide a limited description of the phenomenon, as this measure does not consider the differential contribution of each item to its corresponding factor (i.e., not all items have the same importance), and cannot capture as much variability as a continuous measure like the factor scores on latent variables.

Latent means comparisons between the Spanish and the Swedish samples showed that the standardized factor scores on both the physical and sexual IPVAW factors were higher in Sweden than in Spain, and that these differences were substantially higher for sexual IPVAW. The effects size of these differences was large for both types of IPVAW, and particularly remarkable in the case of sexual IPVAW. If we transform the effect size into percentages, 89.1% of the Swedish sample had higher values in the physical IPVAW factor than the Spanish average, whereas 99.4% of the Swedish women presented higher values in the sexual IPVAW than the Spanish latent mean in

that factor. When we analyze these effect sizes in terms of probability of superiority (i.e., the probability that a woman from one country will score higher than a woman from the other country, if both are randomly selected), there was an 80.7% probability that a Swedish woman would score higher than a Spanish woman in the physical IPVAW factor, and a 96.1% probability that the Swedish woman would score higher than the Spanish woman in the sexual IPVAW factor. These results clearly illustrate the importance of using appropriate measurement approaches for cross-country comparisons, as they provide a more accurate picture of country differences. Prevalence indicators based on raw prevalences provide a more restricted view of the phenomenon, and can distort or conceal important differences, such as those found in this study regarding sexual IPVAW differences between Sweden and Spain.

Summing up, our results showed that the prevalence of physical and sexual IPVAW is clearly higher in Sweden than in Spain, that these differences are more evident in the case of sexual violence, and that these differences are not the result of measurement bias. Taken together, these results support the idea of the Nordic paradox, that is, the puzzling fact that despite the high levels of gender equality achieved in countries like Sweden, the prevalence of physical and, in particular, sexual IPVAW remains disproportionately high. The higher rates of physical and sexual IPVAW in countries with high levels of country-level gender equality such as Sweden––regardless of whether we consider its prevalence on its own, or in comparison with another country with lower levels of gender equality such as Spain––remains unexplained, and clearly invites further research. The psychometric study conducted in this paper was not designed to explain the Nordic paradox, but to eliminate the possibility that this phenomenon was due to measurement bias. Once measurement bias has been ruled out, the research question posited by the Nordic paradox remains unanswered.

The reasons explaining the high levels of IPVAW prevalence in countries like Sweden, despite their high levels of gender equality, are not yet understood. Although research supports the link between country-level gender equality and violence against women, the nature and direction of this relationship appears to be complex [13]. For example, a systematic review analyzed the evidence supporting different hypotheses regarding the relationships between country-level gender equality and violence against women: increased gender equality decreased violence (amelioration hypotheses), increased gender equality increased violence (backlash hypotheses), and increased gender equality equals men and women in experiencing violence (convergence hypothesis [59]. This review concluded that none of these relationships could be assumed, and that this association is complex and should be further investigated. For example, to shed light on the Nordic paradox, future research should examine a number of potential lines of enquiry such as those proposed by Gracia and Merlo [14]. Future research should also extend the type of analysis conducted in this study to include other Nordic countries, as well as other countries with low levels of gender equality and also lower levels of IPVAW. This type of research should acknowledge the complex and multidetermined nature of IPVAW [60-62], with appropriate methodological approaches such as multilevel analyses of individual heterogeneity and discriminatory accuracy [63-65].

This study has clear implications regarding cross-country comparisons on key issues such as IPVAW. For adequate cross-cultural comparisons, international surveys should use reliable and valid measures, and most importantly, ensure measurement invariance. Establishing cross-cultural measurement invariance is a precondition for appropriate and valid comparisons across countries [26,28-32]. As Davidov noted [48], "absent invariance, observed differences in means or other statistics might reflect

differences in systematic biases of response across countries or different understanding of the concepts, rather than substantive differences" (p. 429). Lack of evidence of measurement invariance can cast doubts on how cross-country comparisons are interpreted. Using reliable, valid and comparable measures (i.e., using an equivalent measurement model) prevents uncertainty or ambiguous interpretations, and ensures that we reach the right conclusions when comparing countries on key issues such as IPVAW.

References

1. Campbell JC. Health consequences of intimate partner violence. Lancet. 2002;359:1331-6. https://doi.org/10.1016/S0140-6736(02)08336-8

2. Craparo G, Gori A, Petruccelli I, Cannella V, Simonelli C. Intimate partner violence: relationships between alexithymia, depression, attachment styles, and coping strategies of battered women. J Sex Med. 2014;11:1484-94. http://doi.org/10.1111/jsm.12505

3. Devries KM, Mak JY, Garcia-Moreno C, Petzold M, Child JC, Falder G, et al. The global prevalence of intimate partner violence against women. 2013;340:1527-28. https://doi.org/10.1126/science.1240 937

4. Ellsberg M, Heise L, Pena R, Agurto S, Winkvist A. Researching domestic violence against women: methodological and ethical considerations. Stud Fam Plann. 2001;32:1-16. https://doi.org/ 10.1111/j.1728-4465.2001. 00001.x

5. Stöckl H, Devries K, Rotstein A, Abrahams N, Campbell J, Watts C, et al. The global prevalence of intimate partner homicide: A systematic review. Lancet. 2013;382:859-65. https://doi.org/10.1016/S0140-6736(13)61030-2

6. Vilariño M, Amado BG, Vázquez MJ, Arce R. Psychological harm in women victims of intimate partner violence: Epidemiology and quantification of

injury in mental health markers. Psychosoc Interv. 2018;27:145-52. https://doi.org/10.5093/pi2018a23

7. Wong JY, Tiwari A, Fong DY, Bullock L. A cross-cultural understanding of depression among abused women. Violence Against Women. 2016;22:1371-96. https://doi.org/10.1177/ 1077801215624791

8. World Health Organization. Global and regional estimates of violence against women: Prevalence and health effects of intimate partner violence and non-partner sexual violence. Geneva: World Health Organization; 2013.

9. Archer J. Cross-cultural differences in physical aggression between partners: A social-role analysis. Pers Soc Psychol Rev. 2006;10:133-53. http://dx.doi.org/10.1207/s15327957pspr1002_3

10. World Health Organization. Promoting gender equality to prevent violence against women. Geneva: WHO; 2009.

11. World Health Organization/London School of Hygiene and Tropical Medicine. Preventing intimate partner and sexual violence against women: taking action and generating evidence. Geneva: World Health Organization; 2010.

12. Yodanis CL. Gender inequality, violence against women, and fear: A cross-national test of the feminist theory of violence against women. J Interpers Violence. 2004;19:655-75. https://doi.org/10.1177/ 0886260504263868

13. Latzman NE, D'Inverno AS, Niolon PH, Reidy DE. Gender inequality and gender-based violence: Extensions to adolescent dating violence. In: Wolfe DA, Temple JR, editors. Adolescent dating violence: Theory, research, and prevention. New York: Academic Press; 2019. p. 283-314.

14. Gracia E, Merlo J. Intimate partner violence against women and the Nordic paradox. Soc Sci Med. 2016;157:27-30. https://doi.org/ 10.1016/j.socscimed.2016.03.040

15. European Institute for Gender Equality. Gender Equality Index, 2017. Measuring gender equality in the European Union 2005-2015. Luxembourg: Publication Office of European Union; 2017.

16. United Nations Development Program. Human Development Reports. Gender Inequality Index. 2017. Available from http://hdr.undp.org/en/content/gender-inequality-index-gii

17. World Economic Forum. Global Gender Gap. 2017. Retrieved from https://www.weforum.org/reports/the-global-gender-gap-report-2017

18. European Union Agency for Fundamental Rights. Violence against women: An EU-wide survey. Main results. Luxembourg: Publications Office of the European Union; 2014.

19. Aebi MF, Akdeniz G, Barclay G, Campistol C, Caneppele S, Gruszczynska B, et al. European sourcebook of crime and criminal justice statistics. 5th Edition. Helsinki: Akateeminen Kirjakauppa; 2014.

20. Brå [National Council for Crime Prevention]. [Offences in close relationships: A national survey]. Stockholm: Brå; 2014.

21. Heiskanen M, Piispa M. Faith, hope and battering: A survey of men's violence against women. Helsinki: Statistics Finland; 1998.

22. Lundgren E, Heimer G, Westerstrand J, Kalliokoski AM. Captured queen. Men's violence against women in "equal" Sweden - a prevalence study. Stockholm: Fritzes Offentliga Publikationer; 2001.

23. Nerøien AI, Schei B. Partner violence and health: results from the first national study on violence against women in Norway. Scand J Public Health. 2008;36:161-8. https://doi.org/10.1177/1403494 807085188

24. Sanz-Barbero B, Corradi C, Otero-García L, Ayala A, Vives-Cases C. The effect of macrosocial policies on violence against women: A multilevel study in 28 European Countries. Int J Public Health [Preprint]. 2018. Available from: https://doi.org/10.1007/s00038-018-1143-1

25. Wijma B, Schei B, Swahnberg K, Hilden M, Offerdal K, Pikarinen U, et al. Emotional, physical, and sexual abuse in patients visiting gynaecology clinics: a Nordic cross-sectional study. Lancet. 2003;361:2107-13. https://doi.org/10.1016/S0140-6736(03)137 19-1

26. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct Equ Modeling. 2002;9:233-55. https://doi.org/10.1207/S15328007SEM0902_5

27. Davidov E, Schmidt P, Schwartz S. Bringing values back in: Testing the adequacy of the European Social Survey to measure values in 20 countries. Public Opin Q. 2008;3:420-45. https://doi.org/10.1093/ poq/nfn035

28. Davidov E, Meuleman B, Cieciuch J, Schmidt P, Billiet J. Measurement equivalence in cross-national research. Annu Rev Sociol. 2014;40:55-75. https://doi.org/10.1146/annurev-soc-071913-043137

29. Jang S, Kim ES, Cao C, Allen TD, Cooper CL, Lapierre LM, et al. Measurement invariance of the Satisfaction with Life Scale across 26 countries. J Cross Cult Psychol. 2017;48:560-76. https://doi.org/ 10.1177/0022022117697844

30. Putnick DL, Bornstein MH. Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. Dev Rev. 2016;41:71-90. https://doi.org/ 10.1016/j.dr.2016.06.004

31. Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. Organ Res Methods. 2000;3:4-70. https://doi.org/10.1177/109442810031002

32. Xu H, Tracey TJ. Use of Multi-Group Confirmatory Factor Analysis in examining measurement invariance in counseling psychology research.

The European Journal of Counselling Psychology. 2017;6:75-82. https://doi.org/10.5964/ejcop.v6i1.120

33. Nybergh L, Taft C, Krantz G. Psychometric properties of the WHO Violence Against Women instrument in a female population-based sample in Sweden: A cross-sectional survey. BMJ Open. 2013;3:e002053. http://dx.doi.org/10.1136/bmjopen-2012-00205 3

34. Ellsberg M, Heise L, Shrader E. Researching violence against women: a practical guide for researchers and advocates. Washington, DC: Center for Health and Gender Equity; 1999.

35. Heise L, García-Moreno C. Violence by intimate partner. In: Krug EG, Dahlberg LL, Mercy JA, Zwi AB, Lozano R, editors. World report on violence and health. Geneva: World Health Organization; 2002. p. 87-122.

36. European Union Agency for Fundamental Rights. Violence against women: An EU-wide survey. Survey methodology, sample and fieldwork. Luxembourg: Publications Office of the European Union; 2014.

37. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Standards for educational and psychological testing. Washington, DC: American Educational Research Association; 1999.

38. Asparouhov T, Muthén B. Weighted least squares estimation with missing data (Technical Report). 2010. Available from http://www.statmodel.com/download/GstrucMissingRevision.pdf

39. Hu LT, Bentler PM. Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Modeling. 1999;6:1-55. https://doi.org/10.1080/ 10705519909540118

40. MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. Psychol Methods. 1996:1;130-49.

41. Trizano-Hermosilla I, Alvarado JM. Best alternatives to Cronbach's alpha reliability in realistic conditions: Congeneric and asymmetrical measurements. Front Psychol. 2016;7:769. https://doi.org/10.3389/fpsyg.2016.00769

42. Sass DA. Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. J Psychoeduc Assess. 2011;29:347-63. http://dx.doi.org/10.1177/0734282911406661

43. Sass DA, Schmitt TA, Marsh HW. Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. Struct Equ Modeling. 2014;21:167-80. https://doi.org/10.1080/10705511.2014.882658

44. Sireci SG, Rios JA. Decisions that make a difference in detecting differential item functioning. Educ Res Eval. 2013;19:170-87. https://doi.org/10.1080/13803611.2013.767621

45. Zumbo BD. Does item-level DIF manifest itself in scale-level analyses? Implications for translating language tests. Language Testing. 2003;20:136-47. https://doi.org/10.1191/0265532203 lt248oa

46. Choi SW, Gibbons LE, Crane PK. Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. J Stat Softw. 2011,39:1-30.

47. French AW, Miller TR. Logistic regression and its use in detecting differential item functioning in polytomous items. J Educ Meas. 1996;33:315-32. https://doi.org/10.1111/j.1745-3984.1996.tb00 495.x

48. Davidov E. A cross-country and cross-time comparison of the human values measurements with the second round of the European Social Survey. Surv Res Methods. 2008;2:33-46. http://dx.doi.org/10.18148/srm/2008.v2i1.365

49. Hong S, Malik ML, Lee MK. Testing configural, metric, scalar, and latent mean invariance across genders in sociotropy and autonomy using a non-Western sample. Educ Psychol Meas. 2003;63:636-54. http://dx.doi.org/10.1177/0013164403251332

50. Milfont TL, Fischer R. Testing measurement invariance across groups: Applications in cross-cultural research. Int J Psychol Res. 2010;3:112-31. http://dx.doi.org/10.21500/20112084.857

51. Wu AD, Li Z, Zumbo BD. Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. Practical Assessment, Research & Evaluation. 2007;12:1-26. https://doi.org/10.22237/jmasm/1225512660

52. Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. Struct Equ Modeling. 2007;14:464-504. https://doi.org/10.1080/10705510701301834

53. Cohen J. Statistical power analysis for the behavioral sciences. Hilsdale: Lawrence Earlbaum Associates; 1988.

54. Grissom RJ, Kim JJ. Effect sizes for research: Univariate and multivariate applications. 2nd ed. New York: Routledge; 2012.

55. Ruscio J, Mullen T. Confidence intervals for the Probability of Superiority effect size measure and the Area Under a Receiver Operating Characteristic Curve. Multivariate Behav Res. 2012;47:201-23. https://doi.org/10.1080/00273171.2012.658329

56. R Core Team. R: A language and environment for statistical computing. R foundation for Statistical Computing. 2017. Available from https://www.R-project.org/

57. Revelle W. psych: Procedures for personality and psychological research. Northwestern University, Evanston, Illinois, USA. 2016. Available from: http://CRAN.R-project.org/package=psych Version = 1.5.4.

58. Muthén LK, Muthén BO. Mplus user's guide. Los Angeles, CA: Author; 2010.

59. Roberts SC. What can alcohol researchers learn from research about the relationship between macro-level gender equality and violence against women?. Alcohol and Alcoholism. 2011;46(2):95–104.

60. Heise L. What works to prevent partner violence? An evidence overview. London, STRIVE, London School of Hygiene and Tropical Medicine. 2011

61. Heise L, Fulu E. What works to prevent violence against women and girls? State of the field of violence against women and girls: what do we know and what are the knowledge gaps. Pretoria, South Africa, Medical Research Council. 2014

62. Heise LL, Kotsadam A. Cross-national and multilevel correlates of partner violence: An analysis of data from population-based surveys. Lancet Glob Heal. 2015; 3(6): e332-e340. https://doi.org/10.1016/S2214-109X(15)00013-3

63. Merlo, J. Multilevel analysis of individual heterogeneity and discriminatory accuracy (MAIHDA) within an intersectional framework. Soc Sci Med. 2018;203:74-80. https://doi.org/10.1016/ j.socscimed.2017.12.026

64. Merlo J, Wagner P, Ghith N, Leckie G. An original stepwise multilevel logistic regression analysis of discriminatory accuracy: The case of neighbourhoods and health. PLoS One. 2016;11(4):e0153778. https://doi.org/10.1371/journal.pone.0153778

65. Ivert AK, Merlo J, Gracia E. Country of residence, gender equality and victim blaming attitudes about partner violence: A multilevel analysis in EU. Eur. J. Public Health. 2018;28(3):559–564. https://doi.org/10.1093/eurpub/ckx138

Study 4

**Measuring Acceptability of Intimate
Partner Violence against Women:
Development and validation of the
A-IPVAW scale**

Measuring Acceptability of Intimate Partner Violence against Women: development and validation of the A-IPVAW scale[*]

Manuel Martín-Fernández, Enrique Gracia, Miriam Marco, Viviana Vargas, Faraj A. Santirso & Marisol Lila
Department of Social Psychology, University of Valencia, Spain

Abstract

Intimate partner violence against women (IPVAW) is a major social and public health problem of global proportions. Public attitudes toward IPVAW shape the social environment in which such violence takes place, and attitudes of acceptability of IPVAW are considered a risk factor to actual IPVAW. The aim of this study was to develop and validate a scale measuring acceptability of IPVAW (A-IPVAW). To this end, a sample of 1800 respondents was recruited via social media. A second sample of 50 IPVAW offenders was used for concurrent validity analyses. Following a cross-validation approach and using item response theory analyses, we found that the latent structure of the scale was one-dimensional and very informative for high and very high levels of acceptability of IPVAW. Regarding criterion-related validity, we found that (a) our measure was related to perceived severity of IPVAW and ambivalent sexism, (b) men showed higher levels of acceptability than women, and (c) IPVAW offenders reported higher levels of acceptability than men from the

general population. Taken together, our results provide evidence that the A-IPVAW is a reliable and valid instrument to assess acceptability of IPVAW.

Acknowledgments

Intimate partner violence against women (IPVAW) is a major social and public health problem of global proportions (Ali & Naylor, 2013; García-Moreno, Jansen, Ellsberg, Heise, & Watts, 2006; WHO 2013). IPVAW is one of the most frequent forms of violence suffered by women (Devries et al., 2013; Stockl et al., 2013), with important consequences for their physical, social and psychological well-being (Campbell, 2002; Craparo, Gori, Petruccelli, Cannella, & Simonelli, 2014; Ellsberg, Jansen, Heise, Watts, & Garcia-Moreno, 2008; Wong, Tiwari, Fong, & Bullock, 2016).

IPVAW is a complex phenomenon, deeply rooted in the socio-cultural context, and in this regard, it needs to be understood within the social and cultural norms that permeate it (Gracia 2014a). Public attitudes and responses regarding violence against women reflect these norms and are key to better understand its root causes and, therefore, developing more effective intervention measures (Gracia & Lila, 2015; Gracia & Merlo, 2016). Public attitudes toward IPVAW can shape the social environment in which such violence takes place, contributing to either foster or discourage IPVAW in societies (Browning, 2002; Campbell & Manganello, 2006; Frye, 2007; Gracia, 2014b). These attitudes have been related to incidence and reporting rates, public and professional responses, and the victims' own responses (Carlson & Worden, 2005; Frye, 2007; Gracia & Herrero, 2006a; Gracia, Garcia, & Lila, 2008, 2011, 2014).

Attitudes of acceptability of IPVAW have been linked to the perpetration of this type of violence (Copp, Giordano, Longmore, & Manning, 2016; Gracia, Rodríguez, & Lila, 2015; Sugarman & Frankel, 1996). High levels of acceptability of IPVAW can lead to the perception of this type of behavior as normative, increasing the risk of men perpetrating IPVAW and of this type of violence being justified by victims and their social circles (Waltermaurer, 2012). In this regard, attitudes of acceptability of IPVAW have been considered

as a risk factor of actual IPVAW (Archer & Graham-Kevan, 2003; Capaldi, Knoble, Shortt, & Kim, 2012; Flood & Please, 2009; Lila, Gracia, & Murgui, 2013; Stith, Smith, Penn, Ward, & Tritt, 2004; WHO, 2002). Attitudes toward IPVAW are hence one of the main targets for intervention and prevention strategies (García-Moreno et al., 2015; Gracia & Lila, 2015; Jewkes, Flood, & Lang, 2015; Lila, Gracia, & Herrero, 2012).

The availability of reliable and valid measures of the acceptability of IPVAW is important for research and intervention purposes, as they can provide knowledge on the social conditions that contribute to IPVAW (Gracia & Lila, 2015; Muehlenhard & Kimes, 1999). Some studies have used attitudinal scales to measure the acceptability of IPVAW among young people (Coop et al., 2016; Fincham, Cui, Braithwaite, & Pasley, 2008) and rural populations (Schwab-Reese, & Renner, 2017). Among the general population, several demographic health surveys have included some brief scales and individual items measuring acceptability of certain IPVAW behaviors (Yount, Halim, Hynes, & Hillman, 2011; Wang, 2016; WHO 2013). However, most of these demographic surveys were designed for sub-Saharan countries and have yielded different results when minor changes are made to the wording of the items (Tsai et al., 2017). In this line, in their review of attitudes toward IPVAW in the European Union, Gracia and Lila (2015) found that the available information from European surveys was not only scarce (mostly based on single items), but neither was it supported by reliable and valid instruments. There is still a need for a reliable, valid and concise measure of the acceptability of IPVAW suitable for this kind of surveys.

The Present Study

The aim of the current study is to address this gap in the literature and develop and validate a scale measuring acceptability of IPVAW. Drawing from the exhaustive pool of items identified in a review of European Union surveys

(Gracia & Lila, 2015), we develop a measure to assess acceptability of IPVAW, including items tapping the acceptability of physical, verbal, and emotional violence (Capezza & Arriaga, 2008). We sought to cross-validate this scale following up-to-date guidelines for factor analyses (Schmitt, 2011), and fitting an item response theory (IRT) model. Although IRT models were originally developed for aptitude evaluation, in the last decades they have been increasingly utilized for personality, behavioral and attitudinal measures. IRT models allow researchers to improve the development of their psychological instruments by assessing the quality and suitability of individual items. Given these advantages, in this study we sought to provide a unified and valid measure of the acceptability of IPVAW from an IRT framework.

For validity purposes, we will explore the relationships between our measure of acceptability and other related constructs, such as perceived severity of IPVAW and sexist attitudes, and gender and age differences. Perceived severity of IPVAW has been negatively related to acceptability of IPVAW (Gracia & Herrero, 2006b; Taylor & Sorenson, 2005). Sexist attitudes have been found to be closely related to attitudes toward IPVAW (Flood & Pease, 2009; Herrero, Rodríguez, & Torres, 2017; Lila, Gracia, & García, 2013). On the other hand, gender is one of the more consistent predictors of attitudes toward IPVAW, with research showing greater justification and acceptability of IPVAW among men; and regarding age, research suggests that attitudes supporting the use of IPVAW tend to be more prevalent among older people (Carlson & Worden, 2005; Fincham et al., 2008; Gracia et al., 2015; Gracia & Tomás, 2014). Finally, for concurrent validity we will use a sample of male offenders court-ordered to an intervention program for IPVAW batterers, as this population is expected to have higher rates of acceptability of IPVAW (Gracia et al., 2015; Ruiz-Hernández, García-Jiménez, Llor-Esteban, & Godoy-Fernández, 2015).

Method

Sample

Online sampling recruitment was used for the current study. Data were collected through social media and e-mail snowballing. These methods have proven to be effective and cost-efficient in previous studies (for systematic reviews see Thornton et al., 2016; Topolovec-Vranic & Natarajan, 2016). A total pool of 2698 responses was collected, most of which were from women (67.6% of the respondents). Although previous studies have found a similar proportion of participation in social media by gender (Thornton et al., 2016), we opted to use a representative sample balanced by gender. To do so, we selected a random sample that maintained a similar proportion of male and female participants.

The final sample consisted of 1800 Spanish-speaking respondents residing in Spain (52.8% females), aged from 18 to 82 years old (M_{age} = 34.55, SD_{age} = 14.54). Following a cross-validation approach this sample was divided in two subsamples, each of 900 participants, with similar ratios of sex, age, nationality, and educational level categories. Socio-demographic characteristics of the sample are shown in Table 1.

A second sample of 50 male offenders court-ordered to an intervention program for IPVAW was selected for validity purposes. These offenders had a suspended sentence conditioned to their attending an intervention program. This sample had a mean age of 39.84, ranging from 21 to 69 years old. Most of these participants had completed compulsory secondary education (86%).

Instruments

Acceptability of IPVAW scale (A-IPVAW). A pool of 102 items tapping acceptability of IPVAW was drawn from a recent review of European surveys on violence against women (Gracia & Lila, 2015). The items drawn from the pool were translated to English from their original language by European

Table 1. Socio-demographic variables of the general sample (N = 1800).

	N	%
Sex		
Women	950	52.8
Men	850	47.2
Age		
18-24	724	40.2
25-54	877	48.7
55+	199	11.1
Nationality		
Spanish	1665	92.7
Immigrant	135	7.3
Educational level		
Compulsory	276	15.3
Upper Secondary	502	27.9
University: Undergraduate	394	21.9
University: Postgraduate	628	34.9

experts on the field of IPVAW who provided the survey data for the review. For this study, a panel of six experts on IPVAW rated the relevance of each item on a 5-point Likert-type scale (i.e., "Is this item relevant to measure attitudes of acceptability of IPVAW?"; 1 = Strongly Disagree, 5 = Strongly Agree). Those items rated with a 4 (i.e., the agree category) or more by the six experts were selected to compose a twenty-item measure of acceptability of IPVAW. The items were translated to Spanish by the same authors of the review. Respondents had to rate how acceptable are a range of men's behaviors against their female partners on a 3-point Likert-type scale (0 = Not acceptable, 1 = Somewhat acceptable, 2 = Acceptable). A fourth category, Very Acceptable, was merged with the Acceptable category since it was selected by almost no respondents in both the general and offenders samples. In particular, the frequencies of the upper category in both samples were quite similar. The items reflected physical violence (e.g., it is acceptable for a man "to hit his

partner if she has been unfaithful"), coercion or verbal violence (e.g., it is acceptable for a man "to threaten his partner with hurting her or others if she leaves him"), and emotional violence such as controlling behaviors (e.g., it is acceptable for a man "to set limits on how his partner dresses"). The full A-IPVAW scale is provided in the Appendix 1.

Perceived severity of IPVAW (PS-IPVAW; Gracia, García, & Lila, 2008, 2009, 2011). This instrument presents eight IPVAW situations (e.g., "During an argument, a man hits his partner and afterwards he asks for her forgiveness") to which respondents had to rate the severity of each situation (ranging from 1, "Not severe at all", to 10, "Extremely severe"). This instrument was validated in the general Spanish population, and with police officers and intimate partner violence offenders, showing acceptable psychometric properties, and it has been related to IPVAW victim blaming attitudes, personal responsibility, sexism, and empathy (Gracia, García, & Lila, 2009; Gracia & Tomás, 2014; Lila, Gracia, & García, 2013; Vargas, Lila, & Catalá-Miñana, 2015). The scale showed a good internal consistency in the recruited general sample (Cronbach's α = .89).

Ambivalent Sexism Inventory (ASI; Glick & Fiske, 1996). The Spanish adaptation of the ASI was used (Expósito, Moya, & Glick, 1998). The ASI is comprised of two subscales, hostile and benevolent sexism, each with 11 items. Hostile sexism is related to attitudes of prejudice and discrimination against women based on the assumption of women's inferiority and their differences from men (e.g., "Women are too easily offended"). Benevolent sexism, on the other hand, is based on men's need for and reliance on women (e.g., "No matter how accomplished he is, a man is not truly complete as a person unless he has the love of a woman"). This inventory has been validated and adapted to more than twenty countries (Glick et al, 2000; Glick, Sakall, Urgurlu, Ferreira, & Aguilar de Souza, 2002), and has proved to be strongly related to IPVAW

responsibility attribution and to attitudes toward intervention in IPVAW cases among police officers (Lila et al., 2013; Lila, Oliver, Catalá-Miñana, Galiana, & Gracia, 2014). Both subscales showed a good internal consistency in the sample of this study (Cronbach's α = .88 and .89 for hostile and benevolent sexism, respectively).

Procedure

An online survey was built presenting the A-IPVAW scale, the PS-IPVAW scale, and the ASI inventory. The survey was open for a four-week recruitment period and participation was anonymous. A message offering some information about the study and calling for participation was posted in several social media groups.

Data Analyses

To assess the psychometric properties of the A-IPVAW scale, the following analyses were carried out. First, item descriptive statistics and the overall internal consistency of the scale were examined. Second, a cross-validation approach was followed to evaluate the factorial structure of the A-IPVAW scale. The sample was divided into two subsamples, each one of 900 participants. An exploratory factor analysis (EFA) was conducted on the first subsample, and the second subsample was used to replicate the EFA results taking a confirmatory factor analysis (CFA) approach.

To decide the number of factors needed for extraction, a parallel analysis based on minimum rank factor analysis with polychoric correlations (Timmerman & Lorenzo-Seva, 2011) and the minimum average partial (MAP) criterion (Velicer, 1976) was performed on the first subsample. These methods have shown a good performance determining the number of latent dimensions with categorical data (Garrido, Abad, & Ponsoda, 2011, 2016). Bartlett's sphericity test and the Kaiser-Meyer-Olkin (KMO) statistic were computed to evaluate the suitability of the data for an EFA. Given the

categorical nature of the data, the estimation method used was weighted least squares with adjusted means and variances (WLSMV), since it is more robust for ordinal and categorical data (Asparouhov & Muthén, 2010). The polychoric correlation matrix was used to conduct the EFA (Muthén & Kaplan, 1985, 1992). Model fit was assessed in terms of relative comparative fit, with the CFI and TLI indices. Model residuals were also evaluated with the SRMS and the RMSEA statistics. CFI and TLI values ≥ .95 are indicative of good model fit (Hu & Bentler, 1999), whereas SRMR values ≤ .08, and RMSEA values ≤ .06 are considered good fitting models (MacCallum, Browne, & Sugawara, 1996). Estimated chi-square values and their *df* were also provided, although this statistic tends to be affected by large sample sizes.

A CFA was then conducted with the second subsample, replicating the model suggested by the EFA. Again, WLSMV was selected as estimation method. Model fit was assessed using the same combination of fit indices and their aforementioned cut-off values (CFI & TLI ≥ .95; RMSEA ≤ .06).

Once the scale dimensionality was assessed, an item response theory (IRT) model was fitted to the whole sample. IRT models have some major advantages over CFA: (a) IRT includes improved factor scores estimates, and (b) does not assume measurement precision to be constant, allowing researchers to identify which factor scores of the latent construct are better assessed by their psychological instruments (Chernyshenko, Stark, Chan, Drasgow, & Williams, 2001). Samejima's graded response model (1969) was selected given the ordinal nature of the data. Then the model was estimated using the *MHRM* algorithm (Cai, 2010), and model fit was evaluated with the same cut-off values for the same fit indices (CFI & TLI ≥ .95; RMSEA ≤ .06). To compute these indices, the Maydeu's M_2 statistic for ordinal variables was used instead of other approximations of the chi-square for ordinal variables, since it was specifically developed to assess the overall fit for IRT models (Maydeu-

Olivares & Garcia-Forero, 2010; Maydeu-Olivares & Joe, 2006). The test information function was provided and the IRT scores —the person parameters of the model— were used as an acceptability estimate for the subsequent validity analyses.

Finally, the A-IPVAW scale validity was evaluated for the whole sample. First, acceptability estimates (i.e., IRT factor scores) were correlated with the PS-IPVAW scores and with the hostile and benevolent sexism scores from the ASI. Then comparisons were made of the A-IPVAW scale scores between men and women and age groups from the general sample, and between men from the general sample and the sample of male batterers.

Descriptive statistics, classical internal consistency, and IRT analyses were conducted with the statistical package R (R Core Team, 2016). The *psych* (Revelle, 2016) and the *mirt* (Chalmers, 2012) libraries were used for this purpose. Bartlett's sphericity test, the KMO statistic, parallel analysis and the MAP criterion were computed with the *factor* package (Lorenzo-Seva & Ferrando, 2006), whereas the EFA and CFA analysis were conducted with *Mplus 7.1* (Muthén & Muthén, 2010).

Results

Descriptive statistics and internal consistency

The item descriptive analyses and item-total correlations are shown in Table 2. All items presented mean values close to 0, with standard deviations around 0.30. In addition, the skew and kurtosis indices showed that the item distributions were strongly displaced to the left, displaying a leptokurtic distribution. Taken together, the descriptive analyses reveal that respondents tended to choose the "Not acceptable" category for almost all items. Regarding

Table 2. A-IPVAW items descriptive statistics.

	M	SD	Min	Max	Skew	Kurtosis	$r_{item\text{-}total}$
aipvaw1	0.42	0.61	0.00	2.00	1.15(.01)	0.26(.01)	0.54
aipvaw2	0.41	0.61	0.00	2.00	1.17(.01)	0.32(.01)	0.53
aipvaw3	0.10	0.33	0.00	2.00	3.63(.01)	13.53(.01)	0.63
aipvaw4	0.14	0.39	0.00	2.00	2.80(.01)	7.55(.01)	0.59
aipvaw5	0.10	0.33	0.00	2.00	3.48(.01)	12.33(.01)	0.51
aipvaw6	0.10	0.34	0.00	2.00	3.52(.01)	12.63(.01)	0.62
aipvaw7	0.04	0.22	0.00	2.00	6.63(.01)	47.37(.01)	0.56
aipvaw8	0.09	0.32	0.00	2.00	4.01(.01)	16.64(.01)	0.59
aipvaw9	0.05	0.24	0.00	2.00	5.66(.01)	34.61(.01)	0.51
aipvaw10	0.02	0.18	0.00	2.00	8.21(.00)	73.24(.00)	0.56
aipvaw11	0.04	0.21	0.00	2.00	6.48(.00)	45.79(.00)	0.50
aipvaw12	0.03	0.18	0.00	2.00	7.47(.00)	61.29(.00)	0.56
aipvaw13	0.03	0.20	0.00	2.00	7.97(.00)	67.68(.00)	0.63
aipvaw14	0.02	0.18	0.00	2.00	8.34(.00)	75.38(.00)	0.55
aipvaw15	0.09	0.34	0.00	2.00	3.77(.01)	14.57(.01)	0.60
aipvaw16	0.15	0.41	0.00	2.00	2.74(.01)	7.12(.01)	0.53
aipvaw17	0.12	0.37	0.00	2.00	3.27(.01)	10.64(.01)	0.48
aipvaw18	0.09	0.33	0.00	2.00	3.86(.01)	15.28(.01)	0.54
aipvaw19	0.02	0.17	0.00	2.00	8.59(.00)	80.42(.00)	0.57
aipvaw20	0.26	0.49	0.00	2.00	1.71(.01)	2.04(.01)	0.55

Note: Note: *M* = Mean, *SD* = Standard Deviation, *Min* = Minimum, *Max* = Maximum, $r_{item\text{-}total}$ = item-total corrected correlation. In brackets: the standard error for the *skew* and *kurtosis* statistics. a-ipvaw: acceptability of intimate partner violence against women.

the item-total corrected correlations, all items presented values above .40, indicating that the items were strongly related to the measured construct. The internal consistency of the scale was good (Cronbach's α = .89).

Exploratory Factor Analysis

An EFA was conducted using the first subsample. Bartlett's sphericity test was significant (*p* < .001), indicating that the items were dependent, and the KMO

index was acceptable (KMO = .928). Therefore, the correlations of the items and the data matrix were suitable for carrying out a factor analysis. The parallel analysis based on minimum rank factor analysis showed that the first factor accounted for 44.6% of the explained variance, far above the average 19.6% expected for a data matrix of random responses, whereas a hypothetical second factor accounted only for 8.4%, below the average 8.5% expected for a randomly generated data matrix. In addition, the MAP criterion indicated that the differences between the one-factor and the two-factor solutions were negligible; a one-factor solution was therefore considered more suitable. As both tests yielded similar results, only one factor was extracted using WLSMV as the estimation method. The estimated model converged normally and showed a good fit to the data. Although the chi-square test was significant (χ^2 (*170*) = 485.74, p < .001), which a priori implies that the model did not fit well the data, the comparative fit indices of the model were above the cut-off values (CFI = .96, TLI = .96), and the model residuals were fair (SRMR = .080, RMSEA = .045). These fit indices indicated that the model fitted well the data.

Confirmatory Factor Analysis

A CFA was conducted with the second subsample. A one-factor model was posited and estimated with the WLSMV method. As displayed in Figure 1, all standardized loadings were greater than .70, with standard estimation errors around .03. Again, the chi-square test was significant (χ^2 (*170*) = 502.94, p < .001) and the comparative fit indices of the model were above the cut-off values (CFI = .95, TLI = .95), and with well-fitted residuals (RMSEA = .047), showing an adequate fit to the data. The CFA results replicated the EFA results in a different subsample, yielding the same solution. This one-factor solution was kept as the latent structure of the A-IPVAW scale.
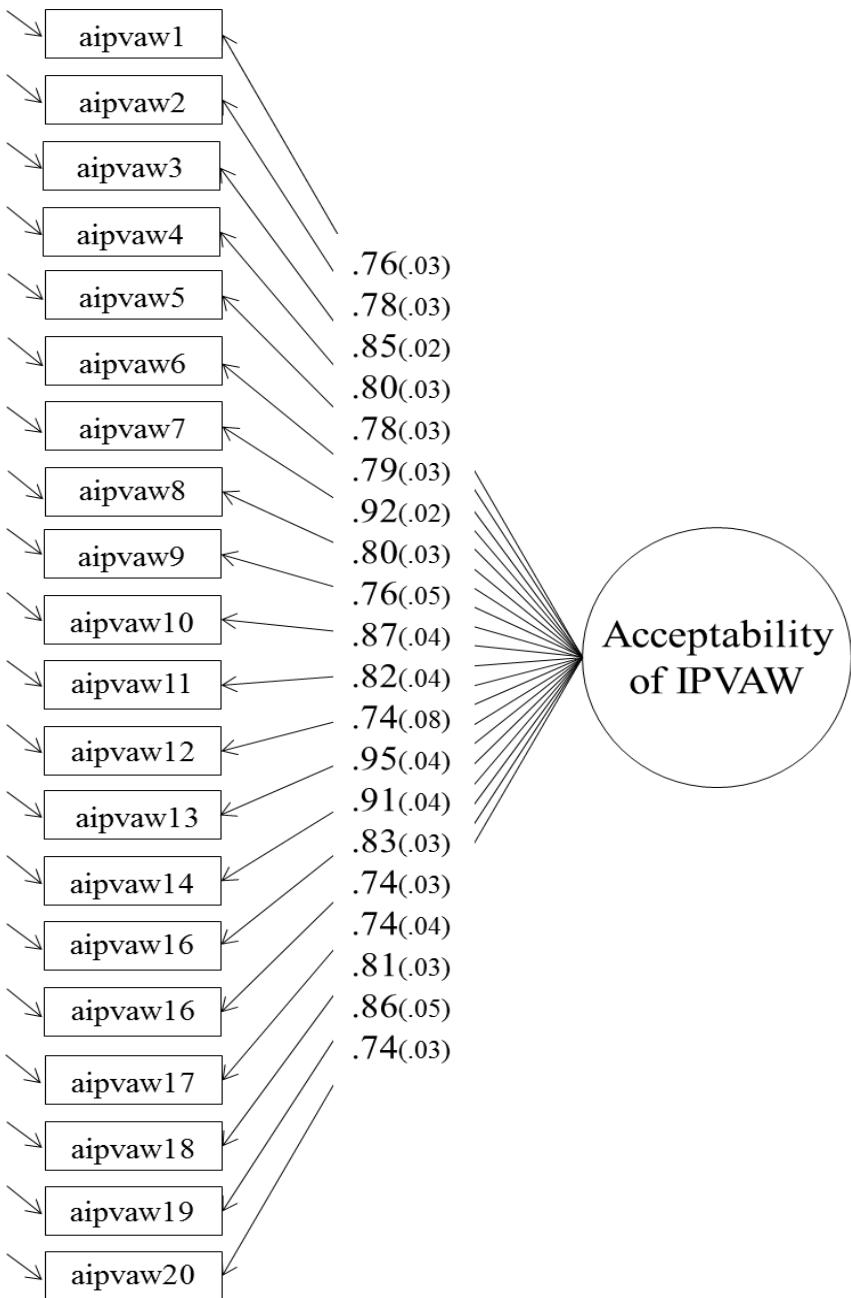
Figure 1. CFA One-factor Model.

Note. Standardized factor loadings (with their standard error in brackets) are depicted in the diagram for each item of the scale. aipvaw: Acceptability of intimate partner violence against women.

Table 3. A-IPVAW scale IRT item parameters.

	a	b_1	b_2
a-ipvaw1	2.29	0.52	1.95
a-ipvaw2	2.25	0.53	2.00
a-ipvaw3	2.82	1.68	2.86
a-ipvaw4	2.36	1.49	2.87
a-ipvaw5	1.96	1.87	3.33
a-ipvaw6	2.65	1.67	2.88
a-ipvaw7	2.98	2.25	3.08
a-ipvaw8	2.41	1.85	2.91
a-ipvaw9	2.25	2.32	3.45
a-ipvaw10	3.20	2.44	3.26
a-ipvaw11	2.44	2.40	3.48
a-ipvaw12	2.99	2.39	3.42
a-ipvaw13	3.66	2.57	3.34
a-ipvaw14	3.15	2.46	3.25
a-ipvaw15	2.59	1.75	2.84
a-ipvaw16	1.95	1.55	2.95
a-ipvaw17	1.71	1.87	3.23
a-ipvaw18	2.04	1.93	3.10
a-ipvaw19	3.60	2.44	3.25
a-ipvaw20	2.11	1.00	2.67

Note: a: discrimination parameter; b_k: threshold parameters. a-ipvaw: acceptability of intimate partner violence against women.

Item Response Theory

Once the dimensionality of the A-IPVAW scale was delimited, the items were calibrated under Samejima's graded response model (1969) for the full sample. Item parameter estimates are shown in Table 3. The threshold parameters (b_1 and b_2) indicate the point on the latent trait continuum (i.e., acceptability of IPVAW) where the probability of endorsement between two adjacent categories is .50 for any respondent with a person parameter θ (i.e., acceptability estimates) equal to the threshold parameter value. Therefore,

respondents with acceptability estimates lower than the b_1 parameter would be more likely to endorse the lowest category (i.e., "Not acceptable"), whereas those respondents with acceptability estimates higher than the b_1 parameter would tend to endorse one of the other two categories. Those respondents would more likely endorse the intermediate category (i.e., "Somewhat acceptable") if their acceptability estimate was lower than the b_2 parameter, and the upper category (i.e., "Acceptable") if their acceptability estimate was higher than the b_2 parameter. In general, the threshold parameters were high, indicating that the test was sampling high (above 1) and very high levels (above 2) of acceptability of IPVAW.

The discrimination parameters (a), in turn, provided information about the precision of each item. In particular, the greater this parameter result, the less likely it is that a given respondent will endorse a category above their acceptability estimate. The discrimination parameters of the A-IPVAW scale were very high, with values above 2 for almost all the items.

The information function of the test, as depicted in Figure 2, showed that the A-IPVAW scale was especially informative for respondents with high and very high acceptability estimates. The standard error of estimation (s.e.) informed about the accuracy of the scale for the different latent trait levels; the lower the s.e. is, the higher is the precision of the scale for a given latent trait level. In particular, s.e. values below 0.5 are equivalent to a a Cronbach's α of .75 or higher; and s.e. values below 0.3 are equivalent to a Cronbach's α of .91 or higher. This means that the test can estimate very accurately the attitudes toward acceptability of those respondents with moderate, high, and very high levels of acceptability of IPVAW, although it cannot discriminate well among respondents with low and very low levels of acceptability. Overall fit of the model was evaluated with the ordinal version of the M_2 statistic. This statistic works similarly to the χ^2 statistic in the factor analysis framework. Thus, the
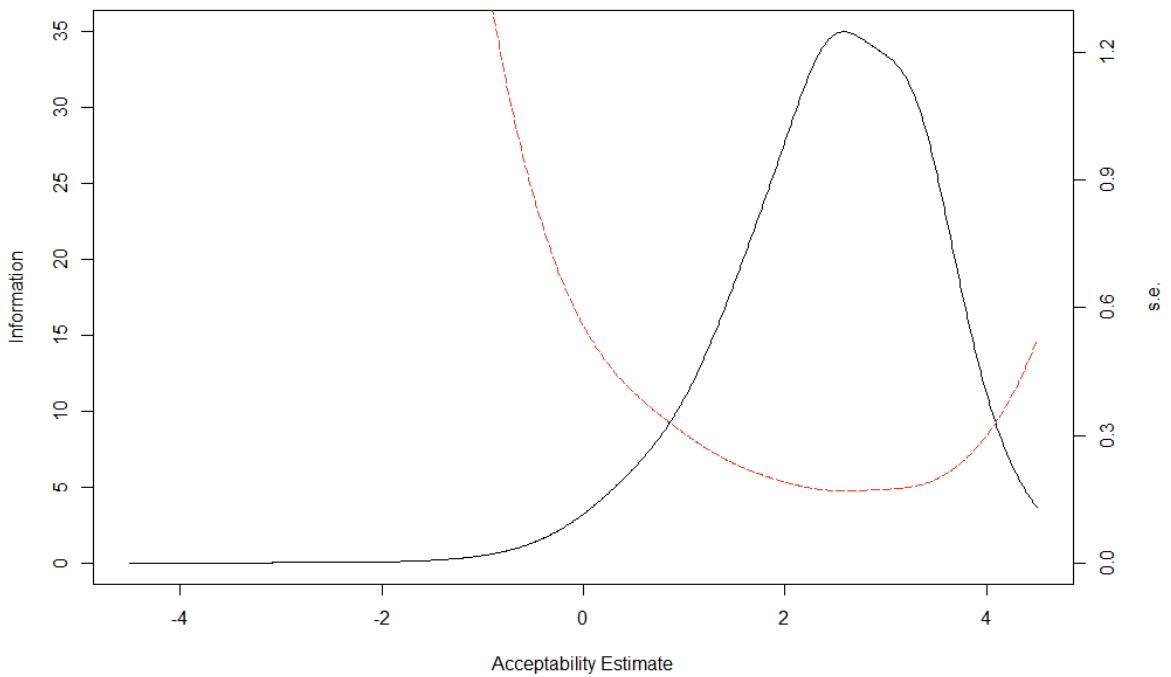
Figure 2. Test Information Function.

Acceptability Estimate: IRT scores of the scale. *Information*: accuracy of the measure over the latent trait continuum (i.e., acceptability). s.e.: Standard error of estimation (discontinuous line).

same fit indices (CFI, TLI, and RMSEA) could be computed for the IRT analyses using the M_2. Using this procedure, the model showed an adequate fit (M_2 (*150*) = 560.87, $p < .001$; CFI = .99, TLI = .98, and RMSEA = 0.036).

Validity analyses

The IRT scores (i.e., "acceptability estimates") from the A-IPVAW scale were used for the validity analyses. IRT scores are more appropriate than the raw sum of the items, as some items are more relevant to measure the latent construct than others. IRT scores were on the logistic metric, with an expected mean value of 0 and a standard deviation of 1 (see Chalmers, 2012).

Table 4. A-IPVAW scale correlation to other variables.

	Hostile Sexism	Benevolent Sexism	PS-IPVAW
A-IPVAW	.44**	.34**	-.47**
Hostile Sexism		.81**	-.39**
Benevolent Sexism			-.30**

Note: **: $p < .001$. A-IPVAW: Acceptability of intimate partner violence against women scale. PS-IPVAW: Perceived severity of intimate partner violence against women scale.

As shown in Table 4, the A-IPVAW scale was positively related with both subscales of the ASI, especially with the hostile sexism subscale, whereas the PS-IPVAW was negatively related with the test. Participants with higher acceptability estimates on the A-IPVAW scale tended to show more agreement with the ASI items and, moreover, tended to perceive the situations posited by the PS-IPVAW as less severe.

When the A-IPVAW scores were compared by gender, significant differences were found between men ($M = 0.27$, $SD = 0.95$) and women ($M = -0.17$, $SD = 0.77$), $t(1637.5) = 10.82$, $p < .001$, $d = .51$, with a moderate effect size. There were also significant differences between the men from the general sample and those from the offenders sample ($M = 0.98$, $SD = 1.35$), $t(51.88) = -3.67$, $p < .001$, $d = .61$, with a moderate effect size. Significant differences were found between age groups, $F(2) = 3.49$, $p = .03$, $\eta^2 = .004$; however the effect size was below the low cut-off value of .01 for the partial eta-squared, and therefore the effect of age on the A-IPVAW scores could be considered negligible (Miles & Shevlin, 2001).

Discussion

Attitudes of acceptability of IPVAW are an important risk factor of actual IPVAW (Abramsky et al., 2011; Archer & Graham-Kevan, 2003; Gracia et al., 2015; Copp et al., 2016; WHO, 2002). However, the measurement of these

attitudes in surveys in western countries tends to be based on single items and not supported by reliable and valid instruments (Gracia & Lila, 2015). The aim of this study was to develop and validate the A-IPVAW scale, a twenty-item instrument to assess the acceptability of IPVAW among the general population. Taken together, our results provide evidence that the A-IPVAW is a reliable and valid instrument to assess acceptability of IPVAW.

Regarding the latent structure of the scale, although items of different types of IPVAW (i.e., physical, verbal, and emotional) were included, results from the cross-validation analyses suggested that a single factor was sufficient to account for the variability of respondents' attitudes of acceptability of IPVAW, rather than a multidimensional model considering one factor for each different type of IPVAW addressed by the scale. Adding more dimensions to the model was not necessary, as it would not have improved the model fit to the data and would hinder the interpretation of the scale.

The use of IRT to study the psychometric properties of the scale constitutes a major strength of this paper, since little research has applied this analytical framework to the study and measurement of IPVAW. Some previous research has applied IRT to the study of intimate partner violence, either by fitting an IRT model to an existing scale (e.g., Beck, Menke, & Figueredo, 2013; Jose, Olino, & O'Leary, 2012; Reichenheim, Klein, & Moraes, 2007), or by testing differential item functioning across gender (e.g., Edelen, McCaffrey, Marshall, & Jaycox, 2009; Yount, VanderEnde, Zureick-Brown, Minh, Schuler, & Anh, 2014); however, none of these studies addressed the acceptability of IPVAW. Moreover, IRT offers improved factor scores that can be used to assess which latent trait levels (i.e., acceptability estimates) are measured more accurately. The test information function is a dynamic approach to study the reliability of a psychometric instrument. Unlike Cronbach's α, the IRT information function does not assume that the accuracy of the scale is constant across the entire

latent trait (i.e., acceptability estimates), and thus some latent trait levels are more accurately measured than others. In particular, the A-IPVAW scale is especially informative (i.e., accurate) for moderate, high and very high levels of acceptability of IPVAW. Our measure can detect individuals with high levels of acceptability and discriminate among them with high precision. On the other hand, the precision of the scale is lower for individuals with low and very low levels of acceptability.

Regarding the validity analyses, we found that respondents with higher scores on the A-IPVAW scale tend to evaluate less severely the IPVAW situations described by the PS-IPVAW items. This finding is congruent with previous research, since individuals who consider IPVAW as such only in cases of extreme or severe violence (e.g., physical violence) are more likely to perceive other kinds of violence (e.g., emotional or verbal violence) as more "tolerable" (Gracia & Herrero, 2006b; Muehlenhard & Kimes, 1999; Taylor & Sorenson, 2005). Our results also revealed that respondents with higher levels of acceptability presented higher scores of ambivalent sexism, in particular in the hostile sexism subscale. Sexism has also been previously related to attitudes justifying IPVAW (e.g. Glick, Sakalli-Ugurlu, Ferreira, & de Souza, 2002; Herrera, Expósito & Moya, 2012; Herrero et al., 2017; Valor-Segura, Expósito, & Moya, 2011).

Regarding gender differences, women showed lower scores in the A-IPVAW scale than men, which is also consistent with previous research (Carlson & Worden, 2005; Fincham et al., 2008; Flood & Pease, 2009; Gracia & Herrero, 2006a). Moreover, in line with the findings of Gracia et al. (2015), we found that convicted batterers are more prone to show higher levels of acceptability than men from the general population. Although this preliminary result should be taken with caution as the sample of batterers is small and somewhat limited, it highlights that the scale is indeed especially informative

for those individuals with higher levels of acceptability of IPVAW, and thus at higher risk of committing IPVAW (Archer & Graham-Kevan, 2003; Gracia et al., 2015). In this regard, the A-IPVAW scale can be used as a forensic tool that can detect cases of higher risk of IPVAW perpetration, as it can differentiate accurately between participants with low and high acceptability estimates (Andreu-Rodríguez, Peña-Fernández, & Loza, 2016). It can also be used as an evaluation instrument for intervention programs with IPVAW perpetrators, monitoring attitudinal changes during and after the intervention (Carbajosa, Catalá-Miñana, Lila, & Gracia, 2017; Ferrer-Perez, Ferreiro-Basurto, Navarro-Guzmán, & Bosch-Fiol, 2016; Lila, Gracia, & Catalá-Miñana, 2017; Lila, Oliver, Galiana, & Gracia, 2013).

This study is not without limitations. The scale was developed within the Spanish socio-cultural context, and thus further research is needed to generalize our results to other cultural settings. Another limitation is the sampling method. Although online recruitment has proven to be an effective and cost-effective sampling method (Thornton et al., 2016), it comes with some tradeoffs that limit the generalizability of the results. As noted by Topolovec-Vranic and Natarajan (2016), it is harder to verify the socio-demographic information provided by on-line participants than with more traditional sampling strategies. In addition, self-selection bias in the targeted sample can be an issue, since people who agree to participate in the study might be more motivated than the general population. However, the socio-demographic characteristics of the sample are similar to other internet-based demographic studies in Spain (Acebes Arribas, 2016), ensuring the representativeness of the sample at least across Spanish internet users. The effect of social desirability should be carefully examined in future studies, assessing the relationship between the A-IPVAW, a self-reported measure, and implicit measures of acceptability of IPVAW (Gracia, Rodríguez, & Lila, 2015).

Future research should also address the factorial invariance of the scale, ensuring that the gender differences encountered in the A-IPVAW scores are due to actual differences between latent means for men and women and not to different interpretations of the items. In the same way, factorial invariance between men and convicted batterers should be addressed with a larger sample of IPVAW perpetrators.

Despite these limitations, the development and validation of the A-IPVAW represents a step forward in the study of attitudes toward IPVAW. With the emergent importance of attitudes in the study of IPVAW in demographic surveys, the availability of psychometrically sound instruments becomes a key issue. The A-IPVAW aims to fill this need.

References

Abramsky, T., Watts, C. H., Garcia-Moreno, C., Devries, K., Kiss, L., Ellsberg, M., ... & Heise, L. (2011). What factors are associated with recent intimate partner violence? Findings from the WHO multi-country study on women's health and domestic violence. *BMC Public Health*, *11*, 109.

Acebes Arribas, B. (2016). *Estudio de medios de comunicación on line*. Madrid: IAB.

Ali, P. A., & Naylor, P. B. (2013). Intimate partner violence: A narrative review of the feminist, social and ecological explanations for its causation. *Aggression and Violent Behavior*, *18*, 611-619.

Andreu-Rodríguez, J. M., Peña-Fernández, M. E., & Loza, W. (2016). Predicting risk of violence through a self-appraisal questionnaire. *The European Journal of Psychology Applied to Legal Context*, *8*, 51-56.

Archer, J., & Graham-Kevan, N. (2003). Do beliefs about aggression predict physical aggression to partners? *Aggressive Behavior, 29,* 41-54.

Asparouhov, T., & Muthén, B. (2010). *Weighted least squares estimation with missing data*. Retrieved from: http://www.statmodel.com/download/GstrucMissingRevision.pdf

Beck, C. J., Menke, J. M., & Figueredo, A. J. (2013). Validation of a measure of intimate partner abuse (Relationship Behavior Rating Scale–Revised) using item response theory analysis. *Journal of Divorce & Remarriage*, *54*, 58-77.

Browning, C. R. (2002). The span of collective efficacy: Extending social disorganization theory to partner violence. *Journal of Marriage and Family*, *64*, 833-850.

Cai, L. (2010). High-dimensional exploratory item factor analysis by a Metropolis–Hastings Robbins–Monro algorithm. *Psychometrika*, *75*, 33-57.

Campbell, J. C. (2002). Health consequences of intimate partner violence. *The Lancet*, *359,* 1331-1336.

Campbell, J. C. & Manganello, J. (2006). Changing public attitudes as a prevention strategy to reduce intimate partner violence. *Journal of Aggression, Maltreatment & Trauma, 13*, 13-40.

Capaldi, D. M., Knoble, N. B., Shortt, J. W., & Kim, H. K. (2012). A systematic review of risk factors for intimate partner violence. *Partner Abuse*, *3*, 231-280.

Capezza, N. M., & Arriaga, X. B. (2008). Factors associated with acceptance of psychological aggression against women. *Violence Against Women*, *14*, 612-633.

Carbajosa, P., Catalá-Miñana, A., Lila, M., & Gracia, E. (2017). Differences in treatment adherence, program completion, and recidivism among batterer subtypes. *The European Journal of Psychology Applied to Legal Context*, 9, 93-101.

Carlson, B. E., & Worden, A. P. (2005). Attitudes and beliefs about domestic violence: Results of a public opinion survey I. Definitions of domestic violence, criminal domestic violence, and prevalence. *Journal of Interpersonal Violence*, *20*, 1197-1218.

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*, 1-29.

Chernyshenko, O. S., Stark, S., Chan, K. Y., Drasgow, F., & Williams, B. (2001). Fitting item response theory models to two personality inventories: Issues and insights. *Multivariate Behavioral Research*, *36*, 523-562.

Copp, J. E., Giordano, P. C., Longmore, M. A., & Manning, W. D. (2016). The development of attitudes toward intimate partner violence: An examination of key correlates among a sample of young adults. *Journal of Interpersonal Violence*. http://journals.sagepub.com/doi/pdf/10.1177/0886260516651311

Craparo, G., Gori, A., Petruccelli, I., Cannella, V., & Simonelli, C. (2014). Intimate partner violence: relationships between alexithymia, depression, attachment styles, and coping strategies of battered women. *The Journal of Sexual Medicine*, *11*, 1484-1494.

Devries, K. M., Mak, J. Y., Garcia-Moreno, C., Petzold, M., Child, J. C., Falder, G., ... & Pallitto, C. (2013). The global prevalence of intimate partner violence against women. *Science*, *340*, 1527-1528.

Edelen, M. O., McCaffrey, D. F., Marshall, G. N., & Jaycox, L. H. (2009). Measurement of teen dating violence attitudes an item response theory evaluation of differential item functioning according to gender. *Journal of Interpersonal Violence*, *24*, 1243-1263.

Ellsberg, M., Jansen, H. A., Heise, L., Watts, C. H., & Garcia-Moreno, C. (2008). Intimate partner violence and women's physical and mental health in

the WHO multi-country study on women's health and domestic violence: an observational study. *The Lancet*, *371*, 1165-1172.

Expósito, F., Moya, M. C., & Glick, P. (1998). Sexismo ambivalente: medición y correlatos [Ambivalent sexism: measurement and correlates]. *Revista de Psicología Social*, *13*, 159-169.

Ferrer-Perez, V. A., Ferreiro-Basurto, V., Navarro-Guzmán, C., & Bosch-Fiol, E. (2016). Programas de intervención con maltratadores en España: la perspectiva de los/as profesionales [Batterer intervention programs in Spain: The professionals' perspective]. *Psychosocial Intervention*, *25*, 159-168.

Fincham, F. D., Cui, M., Braithwaite, S., & Pasley, K. (2008). Attitudes toward intimate partner violence in dating relationships. *Psychological Assessment*, *20*, 260.

Flood, M., & Pease, B. (2009). Factors influencing attitudes to violence against women. *Trauma, Violence, & Abuse*, *10*, 125-142.

Frye, V. (2007). The informal social control of intimate partner violence against women: Exploring personal attitudes and perceived neighborhood social cohesion. *Journal of Community Psychology*, *35*, 1001-1018.

Garcia-Moreno, C., Jansen, H. A., Ellsberg, M., Heise, L., & Watts, C. H. (2006). Prevalence of intimate partner violence: findings from the WHO multi-country study on women's health and domestic violence. *The Lancet*, *368*, 1260-1269.

García-Moreno, C., Zimmerman, C., Morris-Gehring, A., Heise, L., Amin, A., Abrahams, N., ... & Watts, C. (2015). Addressing violence against women: a call to action. *The Lancet*, *385*, 1685-1695.

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2011). Performance of Velicer's minimum average partial factor retention method with categorical variables. *Educational and Psychological Measurement*, *71*(3), 551-570.

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2016). Are fit indices really fit to estimate the number of factors with categorical variables? Some cautionary findings via Monte Carlo simulation. *Psychological methods*, *21*(1), 93.

Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, *70*, 491.

Glick, P., Fiske, S. T., Mladinic, A., Saiz, J. L., Abrams, D., Masser, B., ... & Annetje, B. (2000). Beyond prejudice as simple antipathy: hostile and benevolent sexism across cultures. *Journal of Personality and Social Psychology*, *79*, 763.

Glick, P., Sakalli–Ugurlu, N., Ferreira, M. C., & Aguiar de Souza, M. (2002). Ambivalent sexism and attitudes toward wife abuse in Turkey and Brazil. *Psychology of Women Quarterly*, *26*, 292-297.

Gracia, E. (2014a). Public attitudes toward partner violence against women. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 4325–4328). Dordrecht, Netherlands: Springer.

Gracia, E. (2014b). Intimate partner violence against women and victim-blaming attitudes among Europeans. *Bulletin of the World Health Organization*, *92*, 380-381.

Gracia, E., & Herrero, J. (2006a). Public attitudes toward reporting partner violence against women and reporting behavior. *Journal of Marriage and Family*, *68*, 759-768.

Gracia, E., & Herrero, J. (2006b). Acceptability of domestic violence against women in the European Union: A multilevel analysis. *Journal of Epidemiology and Community Health*, *60*, 123-129.

Gracia, E., & Lila, M. (2015). *Attitudes towards violence against women in the EU*. Luxembourg: Publication Office of the European Union.

Gracia, E., & Merlo, J. (2016). Intimate partner violence against women and the Nordic paradox. *Social Science & Medicine*, *157*, 27-30.

Gracia, E., & Tomás, J. M. (2014). Correlates of victim-blaming attitudes regarding partner violence against women among the Spanish general population. *Violence Against Women, 20*, 26–41.

Gracia, E., García, F., & Lila, M. (2008). Police involvement in cases of intimate partner violence against women: The influence of perceived severity and personal responsibility. *Violence Against Women*, *14*, 697-714.

Gracia, E., García, F., & Lila, M. (2009). Public responses to intimate partner violence against women: The influence of perceived severity and personal responsibility. *The Spanish Journal of Psychology*, *12*, 648-656.

Gracia, E., García, F., & Lila, M. (2011). Police attitudes toward policing partner violence against women: Do they correspond to different psychosocial profiles? *Journal of Interpersonal Violence*, *26*, 189-207.

Gracia, E., García, F., & Lila, M. (2014). Male police officers' law enforcement preferences in cases of intimate partner violence versus non-intimate interpersonal violence: do sexist attitudes and empathy matter? *Criminal Justice and Behavior*, *41*, 1195-1213.

Gracia, E., Rodriguez, C. M., & Lila, M. (2015). Preliminary evaluation of an analog procedure to assess acceptability of intimate partner violence against women: The Partner Violence Acceptability Movie Task. *Frontiers in Psychology*, *6:1567*.

Herrera, M. C., Expósito, F., & Moya, M. (2012). Negative reactions of men to the loss of power in gender relations: Lilith vs. Eve. *The European Journal of Psychology Applied to Legal Context*, *4*, 17-42.

Herrero, J., Rodríguez, F. J., & Torres, A. (2017). Acceptability of partner violence in 51 societies: The role of sexism and attitudes toward violence in social relationships. *Violence Against Women*, *23*, 351-367.

Hu L. T., & Bentler P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.

Jewkes, R., Flood, M., & Lang, J. (2015). From work with men and boys to changes of social norms and reduction of inequities in gender relations: A conceptual shift in prevention of violence against women and girls. *The Lancet*, *385*, 1580-1589.

Jose, A., Olino, T. M., & O'leary, K. D. (2012). Item response theory analysis of intimate-partner violence in a community sample. *Journal of Family Psychology*, *26*, 198-205.

Lila, M., Gracia, E., & Catalá-Miñana, A. (2017). More likely to dropout, but what if they don't? Partner violence offenders with alcohol abuse problems completing batterer intervention programs. *Journal of Interpersonal Violence.* http://dx.doi.org/10.1177/0886260517699952

Lila, M., Gracia, E., & García, F. (2013). Ambivalent sexism, empathy and law enforcement attitudes towards partner violence against women among male police officers. *Psychology, Crime & Law*, *19*, 907-919.

Lila, M., Gracia, E., & Herrero, J. (2012). Asunción de responsabilidad en hombres maltratadores: Influencia de la autoestima, la personalidad narcisista y la personalidad antisocial [Responsibility assumption among male batterers: self-esteem, narcissistic and antisocial personality influence]. *Revista Latinoamericana de Psicología, 44,* 99-108.

Lila, M., Gracia, E., Murgui, S. (2013). Psychological adjustment and victim-blaming among intimate partner violence offenders: The role of social support and stressful life events. *The European Journal of Psychology Applied to Legal Context, 5,* 147-153.

Lila, M., Oliver, A., Catalá-Miñana, A., Galiana, L., & Gracia, E. (2014). The intimate partner violence responsibility attribution scale (IPVRAS). *The European Journal of Psychology Applied to Legal Context*, *6*, 29-36.

Lila, M., Oliver, A., Galiana, L., & Gracia, E. (2013). Predicting success indicators of an intervention programme for convicted intimate-partner violence offenders: The Contexto Programme. *The European Journal of Psychology Applied to Legal Context, 5*, 73–95.

Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavior Research Methods*, *38*, 88-91.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*, 130-149.

Maydeu-Olivares, A., & Garcia-Forero, C. (2010). Goodness-of-fit testing. *International Encyclopedia of Education*, *7*, 190-196.

Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika*, *71*, 713.

Miles, J., & Shevlin, M. (2001). *Applying Regression and Correlation: A Guide for Students and Researchers*. Sage: London.

Muehlenhard, C. L., & Kimes, L. A. (1999). The social construction of violence: The case of sexual and domestic violence. *Personality and Social Psychology Review, 3*, 234-245.

Muthén, B., & Kaplan D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38*, 171-189.

Muthén, B., & Kaplan D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the

model. *British Journal of Mathematical and Statistical Psychology, 45*, 19-30.

Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide*. Los Angeles, CA: Author.

Reichenheim, M. E., Klein, R., & Moraes, C. L. (2007). Assessing the physical violence component of the Revised Conflict Tactics Scales when used in heterosexual couples: An item response theory analysis. *Cadernos de Saúde Pública*, *23*, 53-62.

Revelle, W. (2016) *psych: Procedures for personality and psychological research*. Northwestern University, Evanston, Illinois, USA. Retrieved from: http://CRAN.R-project.org/package=psych Version = 1.5.4

R Core Team. (2016). R: A language and environment for statistical computing. *R foundation for Statistical Computing*. Retrieved from https://www.R-project.org/

Ruiz-Hernández, J. A., García-Jiménez, J. J., Llor-Esteban, B., & Godoy-Fernández, C. (2015). Risk factors for intimate partner violence in prison inmates. *The European Journal of Psychology Applied to Legal Context*, *7*, 41-49.

Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores.* New York, NY: Psychometric Society.

Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, *29*, 304-321.

Schwab-Reese, L. M., & Renner, L. M. (2017). Attitudinal acceptance of and experiences with intimate partner violence among rural adults. *Journal of Family Violence*, *32*, 115-123.

Stith, S. M., Smith, D. B., Penn, C. E., Ward, D. B., & Tritt, D. (2004). Intimate partner physical abuse perpetration and victimization risk factors: A meta-analytic review. *Aggression and Violent Behavior*, *10*, 65-98.

Stöckl, H., Devries, K., Rotstein, A., Abrahams, N., Campbell, J., Watts, C., & García-Moreno, C. (2013). The global prevalence of intimate partner homicide: A systematic review. *The Lancet*, *382*, 859-865.

Sugarman, D. B., & Frankel, S. L. (1996). Patriarchal ideology and wife-assault: A meta-analytic review. *Journal of Family Violence*, *11*, 13-40.

Taylor, C. A., & Sorenson, S. B. (2005). Community-based norms about intimate partner violence: Putting attributions of fault and responsibility into context. *Sex Roles*, *53*, 573-589.

Thornton, L., Batterham, P. J., Fassnacht, D. B., Kay-Lambkin, F., Calear, A. L., & Hunt, S. (2016). Recruiting for health, medical or psychosocial research using Facebook: Systematic review. *Internet Interventions*, *4*, 72-81.

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality Assessment of Ordered Polytomous Items with Parallel Analysis. *Psychological Methods, 16*, 209-220.

Topolovec-Vranic, J., & Natarajan, K. (2016). The Use of Social Media in Recruitment for Medical Research Studies: A Scoping Review. *Journal of Medical Internet Research*, *18,* e286.

Tsai, A. C., Kakuhikire, B., Perkins, J. M., Vořechovská, D., McDonough, A. Q., Ogburn, E. L., ... & Bangsberg, D. R. (2017). Measuring personal beliefs and perceived norms about intimate partner violence: Population-based survey experiment in rural Uganda. *PLoS Medicine*, *14*, e1002303.

Valor-Segura, I., Expósito, F., & Moya, M. (2011). Victim blaming and exoneration of the perpetrator in domestic violence: The role of beliefs in a just world and ambivalent sexism. *The Spanish Journal of Psychology*, *14*, 195-206.

Vargas, V., Lila, M., & Catalá-Miñana, A. (2015). ¿Influyen las diferencias culturales en los resultados de los programas de intervención con maltratadores? Un estudio con agresores españoles y latinoamericanos [Do cultural differences influence batterer intervention program outcomes? A study with Spanish and Latin American offenders]. *Psychosocial Intervention*, *24*, 41-47.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41*, 321-327.

Waltermaurer, E. (2012). Public justification of intimate partner violence: A review of the literature. *Trauma, Violence, & Abuse, 13*, 167-175.

Wang, L. (2016). Factors influencing attitude toward intimate partner violence. *Aggression and Violent Behavior*, *29*, 72-78.

Wong, J. Y., Tiwari, A., Fong, D. Y., & Bullock, L. (2016). A cross-cultural understanding of depression among abused women. *Violence Against Women*, *22*, 1371-1396.

Worden, A. P., & Carlson, B. E. (2005). Attitudes and beliefs about domestic violence: Results of a public opinion survey: II. Beliefs about causes. *Journal of Interpersonal Violence*, *20*, 1219-1243.

World Health Organization (2002). *World report on violence and health*. Geneva: World Health Organization.

World Health Organization. (2013). *Global and regional estimates of violence against women: prevalence and health effects of intimate partner violence and non-partner sexual violence*. Geneva: World Health Organization.

Yount, K. M., Halim, N., Hynes, M., & Hillman, E. (2011). Response effects to attitudinal questions about domestic violence against women: A comparative perspective. *Social Science Research, 40*, 873–884.

Yount, K. M., VanderEnde, K., Zureick-Brown, S., Minh, T. H., Schuler, S. R., & Anh, H. T. (2014). Measuring attitudes about women's recourse after

exposure to intimate partner violence: The ATT-RECOURSE scale. *Journal of Interpersonal Violence, 29,* 1579-1605.

Appendix 1. A-IPVAW Scale.

I think it is acceptable for a man …

a-ipvaw1	to shout at his partner if she is constantly nagging/arguing
a-ipvaw2	to shout at his partner if she is not treating him with respect
a-ipvaw3	to set limits on how his partner dresses
a-ipvaw4	to set limits on where his partner goes
a-ipvaw5	to push someone into having sex if she has been flirting with him all night
a-ipvaw6	to control his partner's mobile phone
a-ipvaw7	to push someone into having sex if she has been dating him
a-ipvaw8	to threaten to leave his partner in order to achieve something he wants
a-ipvaw9	to hit his partner if she has been unfaithful
a-ipvaw10	to hit his partner if she is constantly nagging/arguing
a-ipvaw11	to push someone into having sex if he has spent a lot of money on her
a-ipvaw12	to hit his partner if she is not treating him with respect
a-ipvaw13	to prevent his partner from seeing family and friends
a-ipvaw14	not to allow his partner to work or study
a-ipvaw15	to tell his partner what she can or cannot do
a-ipvaw16	to throw/smash objects during an argument
a-ipvaw17	to record his partner with a mobile phone or video camera, or take pictures of her without her knowledge
a-ipvaw18	to send messages or images of his partner without her permission
a-ipvaw19	to threaten his partner with hurting her or others if she leaves him
a-ipvaw20	to constantly reproach his partner for the mistakes she has made during an argument

Note. a-ipvaw: acceptability of intimate partner violence against women.

Scoring the A-IPVAW: How to obtain the acceptability estimates

We would like to discourage using the sum of the items to generate a raw score for the A-IPVAW scale. Instead we would recommend generating the acceptability estimates following one of these two methods:

(1) Factor Scores: to obtain the factor scores, conduct a weighted sum of the items by the factor loadings presented in Figure 2. To do so, the factor loading of each item is multiplied by the score for each item before summing.

(2) IRT Scores: to obtain the IRT scores, estimate an IRT model by fixing the item parameters to the values presented in Table 3, and generate the person parameter θ estimates for each respondent. In appendix 2 we include an R script with the code to obtain the acceptability estimates with this method using the *mirt* library.

Appendix 2. R code to obtain the A-IPVAW scale IRT scores.

```
#################################
#### AIPVAW IRT Scores Script ####
#################################

library(mirt)

aipvaw_data <- read.table("data path and format")
# insert path and extension of the data (e.g., "C:/Users/Documents/R/AIP
VAW/AIPVAW_data.dat")
# aipvaw_data must be a matrix or data.frame
# with respondents on the rows and items on the columns

a_AIPVAW <- c(2.29, 2.25, 2.82, 2.36, 1.96, 2.65, 2.98, 2.41, 2.25, 3.20,
              2.44, 2.99, 3.66, 3.15, 2.59, 1.95, 1.71, 2.04, 3.60, 2.11)
b1_AIPVAW <- c(0.52, 0.53, 1.68, 1.49, 1.87, 1.67, 2.25, 1.85, 2.32, 2.44,
               2.40, 2.39, 2.57, 2.46, 1.75, 1.55, 1.87, 1.93, 2.44, 1.00)
b2_AIPVAW <-c(1.95, 2.00, 2.86, 2.87, 3.33, 2.88, 3.08, 2.91, 3.45, 3.26,
              3.48, 3.42, 3.34, 3.25, 2.84, 2.95, 3.23, 3.10, 3.25, 2.67)

AIPVAW_param <- mirt(aipvaw_data, 1, itemtype = "graded", pars = "values")
AIPVAW_param$est <- FALSE
AIPVAW_param$value[aipv_pars$name=="a1"] <- a_AIPVAW
AIPVAW_param$value[aipv_pars$name=="d1"] <- b1_AIPVAW * -a_AIPVAW
AIPVAW_param$value[aipv_pars$name=="d2"] <- b2_AIPVAW * -a_AIPVAW

AIPVAW_IRT <- mirt(AIPVAW_data, 1, itemtype = "graded", pars =
AIPVAW_param)
IRTScores <- fscores(AIPVAW_IRT, method = "EAP", full.scores = T)

write.table(IRTScores, "IRTScores.dat", col.names = FALSE, row.names =
FALSE)
# return a .dat file with the IRT Scores for each respondent
```

Study 5

Assessing Victim Blaming Attitudes towards cases of Intimate Partner Violence against Women: development and validation of the VB-IPVAW scale

Assessing Victim Blaming Attitudes towards cases of Intimate Partner Violence against Women: development and validation of the VB-IPVAW scale[*]

Manuel Martín-Fernández, Enrique Gracia, & Marisol Lila
Department of Social Psychology, University of Valencia, Spain

Abstract

Intimate partner violence against women (IPVAW) is the violence most commonly suffered by women and constitutes a serious public health problem of global proportions. Public attitudes towards IPVAW can influence the social context in which this type of violence occurs. Victim blaming attitudes are among those attitudes reflecting public tolerance and acceptability of IPVAW, and are often used to explain or justify IPVAW. In this study we develop and validate a new instrument assessing victim blaming attitudes towards cases of IPVAW. A sample of 1800 participants was collected through social media, and a second sample of 50 IPVAW offenders was used for validity purposes. Through a cross-validation approach and fitting and item response theory model to the data, we found that the latent structure of the instrument was one-dimensional and particularly informative for medium and high levels of victim blaming attitudes. Differential item functioning analysis showed that the items parameters did not differ by gender. We found, in addition, that (a) our measure was strongly related to acceptability and perceived severity of

IPVAW, and also to ambivalent sexism, (b) men presented higher levels of victim blaming attitudes than women, and (c) IPVAW offenders showed higher levels of victim blaming attitudes than men from the general population. A five-item short version of the scale was also presented for those studies in which application time or space are limited. Our findings support that this new scale is a precise and valid measure to assess victim blaming attitudes towards cases of IPVAW.

Keywords: Victim Blaming, Attitudes, Intimate Partner Violence, Item Response Theory, Measurement, Violence against Women

Acknowledgments

Intimate partner violence against women (IPVAW) is the violence most commonly suffered by women (Devries et al., 2013; Garcia-Moreno et al., 2006; Stockl et al., 2013) and constitutes a serious public health problem of global proportions (Ali & Naylor, 2013; WHO 2013), with important consequences for women's physical and psychological well-being (Campbell, 2002; Craparo, Gori, Petruccelli, Cannella, & Simonelli, 2014; Ellsberg, Jansen, Heise, Watts, & Garcia-Moreno, 2008). The estimated prevalence of IPVAW in high-income countries is 23.2% (WHO, 2013). According to the European Union Agency for Fundamental Rights survey (2014), in the European Union, lifetime prevalence of IPVAW is 22% (ranging across countries from 13% to 32%).

Public attitudes towards IPVAW influence the social context in which this type of violence occurs (Carlson & Worden, 2005; Copp et al., 2016; Flood & Pease, 2009; Gracia & Lila, 2015; Waltermaurer, 2012). In terms of Gracia and Lila (2015), IPVAW "is a complex phenomenon that needs to be understood within the wider social context and within the social and cultural norms that permeate it. Public attitudes and responses regarding violence against women reflect these norms and play an important role in shaping the social climate in which the violence occurs" (p. 13). Researchers increasingly acknowledge the importance of paying attention to attitudes towards IPVAW, as they are linked, for example, to IPVAW incidence, victim's help-seeking, or public and law enforcement responses (Browning, 2002; Faramarzi, Esmailzadeh, & Mosavi, 2005; Gracia & Herrero, 2006a; Gracia, Herrero, Lila, & Fuente, 2009; Gracia, Garcia, & Lila, 2008, 2011; Rizo & Macy, 2011; West & Wandrei, 2002).

Victim blaming attitudes are among those attitudes reflecting public tolerance and acceptability of IPVAW, and are often used to explain or justify IPVAW (Gracia & Tomás, 2014; Gracia, 2014; WHO, 2002). Victim blaming attitudes influence not only public responses and willingness to intervene in

known cases of IPVAW but also perpetrators and victims' responses. Victim blaming attitudes among those surrounding the victims may not only foster and facilitate perpetrators' behaviors, but also make more difficult for the victims to disclose the violence, and to seek and receive help form both informal and formal sources of help (Ansara & Hindin, 2010; Gracia & Herrero, 2006b; Gracia, García, & Lila, 2009; Gracia et al, in press; Gracia & Tomás, 2014; Liang, Goodman, Tummala-Narra, & Weintraub, 2005; Martín-Fernández, Gracia, & Lila, 2018; Valor-Segura, Exposito, & Moya, 2011; Voith, 2017; West & Wandrei, 2002).

The availability of reliable and valid measures of victim blaming attitudes in cases of IPVAW is key for research and intervention purposes (Gracia & Lila, 2015; Powell & Webster, 2018; Muehlenhard & Kimes, 1999). Previous research has addressed the measurement of victim blaming attitudes towards cases of IPVAW, however some of these measures have some drawbacks and/or are limited in a number of issues. Some instruments assessing victim blaming in cases of IPVAW addressed only physical assault as violence inside married couples, ignoring other important types of violence such as psychological or emotional violence (Jackson et al., 1994; Petretic-Jackson et al., 1994). Other available measures were validated only in samples of college students and are difficult to generalize to large populations (Fox & Cox, 2011; Scott & Strauss, 2007). Other instruments are based only in vignettes describing a single case (Koepke et al, 2014; Vidal-Fernández & Megías, 2014; Yamawaki et al., 2009), or are exploratory studies that need further research to establish their psychometric properties in larger and more representative samples (Fox & Cox, 2011; Yun & Vonk, 2011). Other instruments included subscales assessing victim blaming attitudes in cases of IPVAW, but were design to be used with IPVAW male offenders (Henning & Holdfold, 2006; Henning, Jones, & Holdfold, 2006; Lila, Oliver, Catalá-Miñana,

Galiana, & Gracia, 2015). Finally, large population survey data on victim blaming attitudes in cases of IPVAW are not usually based on measurement instruments with adequate reliability and validity evidences (Gracia & Lila, 2015).

Psychometrically sound measures are clearly still needed to assess victim blaming attitudes in cases of IPVAW both in research setting and for large population surveys.

The present study

This study aims to fill this gap in the literature providing a reliable, and valid self-reported measure of victim blaming attitudes in cases of IPVAW, the VB-IPVAW. To this end, we aim to develop a scale to measure victim blaming attitudes based on the pool of items that Gracia and Lila (2015) found in their review. We sought to adapt and validate this measure following a cross-validation approach, and fitting then an item response theory (IRT) model. IRT models have been increasingly used for personality and attitudinal measures, since they allow researchers to improve their psychological instruments studying the quality and suitability of individual items (Glockner-Rist & Herbert Hoijtink, 2003). In this regard, IRT allow researchers to test whether there is an effect in the item responses due to pertaining to different groups (e.g., gender). In addition, a short version of the VB-IPVAW would also be assembled for cases in which time and/or space are limited (e.g., large demographical surveys). We will assess the validity of this new measure of victim blaming attitudes exploring its relationships with other related constructs also linked to IPVAW, such as acceptability of IPVAW, perceived severity of IPVAW, sexist attitudes, and socio-demographical characteristics—i.e., gender and age differences—(Bryant & Spencer, 2003; Capezza & Arriaga, 2008; Gracia & Tomás, 2014; Keopke, et al., 2014; Lila et al., 2015; Scott & Straus, 2007; Taylor & Sorenson, 2005; Yun & Vonk, 2011). In addition, for

validity purposes we will compare the scores in the scale between male respondents from the general Spanish population and a sample of offenders court-mandated to an intervention program for IPVAW batterers. Male offenders are expected to show higher levels of victim blaming attitudes as they tend to use them to justify their behavior (Lila, Gracia, & Murgui, 2013; Lila et al., 2015).

Method

Sample

Data was collected using an online sampling through social media and e-mail snowballing. Previous studies have shown that these sampling methods are effective and cost-efficient (for systematic reviews see Thornton et al., 2016; Topolovec-Vranic & Natarajan, 2016). We recruited a total pool of 2698 respondents (67.6% of the respondents were women). We decided to balance the sample by gender, selecting a random sample that maintained a similar ratio of male and female participants.

The final sample was composed of 1800 participants (92.7% Spanish nationality), aged from 18 to 75 years old (M_{age} = 34.24, SD_{age} = 14.41) of which 52.8% were women. We divided the sample in two subsamples of 900 participants with similar ratios of gender, age, nationality, and educational level categories. The sample socio-demographic information is displayed in Table 1.

In addition, we recruited a second sample of 50 male IPVAW offenders. These offenders were court-mandated to attend an intervention program. The mean age of the sample was 39.84, ranging from 21 to 69 years old. Most of the offenders had completed compulsory secondary education (86%).

Table 1. Socio-demographics of the general sample (N = 1800).

	N	%
Sex		
Women	950	52.8
Men	850	47.2
Age		
18-24	724	40.2
25-54	877	48.7
55+	199	11.1
Nationality		
Spanish	1665	92.7
Immigrant	135	7.3
Educational level		
Compulsory	276	15.3
Upper Secondary	502	27.9
University: Undergraduate	394	21.9
University: Postgraduate	628	34.9

Instruments

Victim blaming attitudes in cases of IPVAW (VB-IPVAW). A pool of 60 items regarding victim blaming attitudes was drawn from a review of different European surveys on violence against women (Gracia & Lila, 2015). These items were translated to English from their original language by European experts on the field of IPVAW who provided the survey data for the review. A panel of six experts on IPVAW was asked to assess the relevance of each item in the pool (Lynn, 1986; Polit & Beck, 2006). To this end, the experts rated the relevance of the items on a 5-point Likert-type scale (i.e., "Is this item relevant to measure victim blaming attitudes in cases of IPVAW?"; 1 = Strongly Disagree, 5 = Strongly Agree). We selected those items with an average rating of 4 (i.e., the agree category) or above to compose a measure of 13 items. The items were translated to Spanish by the same authors of the review.

Respondents had to rate their level of agreement with the statements of the items on a 4-point Likert-type scale (1 = Strongly disagree, 4 = Strongly agree). The complete VB-IPVAW is shown in Appendix 1.

Acceptability of IPVAW (A-IPVAW; Martín-Fernández et al., 2018). The A-IPVAW scale is composed of 20 items in which respondents have to rate how acceptable are different men's behaviors against their female partners (e.g., "It is acceptable for a man to hit his partner if she has been unfaithful"). The response format is a 3 point Likert-type scale (0 = "Not acceptable", 1 = "Somewhat acceptable", 2 = "Acceptable"). This instrument was cross-validated in the general Spanish population, and it showed validity evidences based on its internal structure, and also based on its relationship with other variables, such as perceived severity of IPVAW or ambivalent sexism (Martín-Fernández et al., 2018). The A-IPVAW showed a good internal consistency in the general sample (Cronbach's α = .89).

Perceived severity of IPVAW (PS-IPVAW; Gracia, García, & Lila, 2008, 2009, 2011). The PS-IPVAW scale posits eight IPVAW scenarios (e.g., "A couple is having a quarrel; he insults her and threatens to beat her up"). Respondents had to rate the severity of each scenario (ranging from 1 = "No severe at all", to 10 = "Extremely severe"). This scale presents adequate psychometric properties and has been validated in the general Spanish population, and also with police officers and IPVAW offenders (Gracia, García, & Lila, 2008, 2009). The scale has been related not only to attitudes towards IPVAW, such as victim blaming attitudes and acceptability of IPVAW (Gracia & Tomás, 2014, Martín-Fernández et al., 2018), but to sexism, personal responsibility, and empathy (Lila, Gracia, & García, 2013; Vargas, Lila, & Catalá-Miñana, 2015). This instrument showed a good internal consistency in the general sample of this study (Cronbach's α = .89).

Ambivalent Sexism Inventory (ASI; Glick & Fiske, 1996). We used the Spanish version of the ASI (Expósito, Moya, & Glick, 1998). This instrument includes two subscales, hostile and benevolent sexism, each composed of 11 items. Hostile sexism is conceptualized as attitudes of discrimination and prejudice against women based on the assumption of women's inferiority (e.g., "Women are too easily offended"). Benevolent sexism reflects men's views of women as week and needing protection (e.g., "Women should be cherished and protected by men"). The ASI has been adapted and validated in more than twenty countries (Glick et al, 2000; Glick, Sakall, Urgurlu, Ferreira, & Aguilar de Souza, 2002), and has also shown to be strongly related to IPVAW responsibility attribution, with attitudes toward intervention in IPVAW cases among police officers, and with acceptability of IPVAW (Lila et al., 2013; Lila et al., 2015; Martín-Fernández et al., 2018). The internal consistency of both subscales was good in the general sample (Cronbach's α = .88 and .89 for hostile and benevolent sexism, respectively).

Procedure

The online survey included the VB-IPVAW scale, the PS-IPVAW scale, and the ASI. We maintained the survey opened for a recruitment period of four weeks, from November 2016 to December 2016. A message providing information about the study and calling for participation was posted in different social media groups. Informed consent information was supplied and implied through participation in the on-line survey. Participation was anonymous.

Data Analyses

The following analyses were conducted in order to evaluate the psychometric properties of the VB-IPVAW. Descriptive statistics of the items, corrected item-test correlations, and internal consistency were computed for the whole sample. The latent structure of the scale was assessed following a

cross-validation approach by splitting the general sample in two subsamples, each of 900 participants. An exploratory factor analysis (EFA) was carried out in order to identify a latent variable model for the scale items in the first subsample. This model was then replicated in the second subsample using a confirmatory factor analysis (CFA).

Before conducting the EFA, we tested the suitability of the dataset using the Bartlett's sphericity test and the Kaiser-Meyer-Olkin (KMO) statistic. Then a parallel analysis based on minimum rank factor analysis using polychoric correlations was computed (Timmerman & Lorenzo-Seva, 2011). This method has shown a good performance testing the number of factors to extract for a categorical EFA (Garrido, Abad, & Ponsoda, 2013; 2016). The parallel analysis utilizes Monte Carlo simulation to generate randomized datasets similar to the empirical dataset. Ranked factor analysis is used to compute the percentage of variance explained by a series of different factorial models (i.e., one-factor model, two-factor model, three-factor model, etc...) in the random datasets; the mean and the 95th percentile are obtained. When the percentage of explained variance by a given factorial model in the empirical data is below the percentage expected for that same model in the simulated datasets, then that model is adding more factors than needed. Thus we looked for the minimum number of factors needed to explain more variance in the empirical data than in the random datasets.

We conducted an EFA using weighted least squares with adjusted means and variances (WLSMV) as estimation method, since it is more robust for ordinal and categorical data (Asparouhov & Muthén, 2010). Model fit was assessed using a combination of fit indices: the comparative fit index (CFI), the Tucker Lewis index (TLI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). CFI and TLI values greater than .95 are indicative of good model fit (Hu & Bentler, 1999),

whereas SRMR values lower than .08, and RMSEA values lower than .06 are considered good fitting models (MacCallum, Browne, & Sugawara, 1996).

We replicated the results yielded by the EFA conducting a CFA. The CFA model was estimated using again WLSMV as estimation method, and model fit was evaluated using the same fit indices and their cutoffs (CFI & TLI ≥ .95; RMSEA ≤ .06).

Once the latent structure of the scale was evaluated, we fitted an item response theory (IRT) model for the whole sample. IRT includes improved factor scores estimates and does not assume measurement precision to be constant. These two aspects of IRT allow researchers to identify which levels of the latent construct are better assessed by their instruments (Chernyshenko, Stark, Chan, Drasgow, & Williams, 2001). Given the ordinal nature of the data, the graded response model was selected (Samejima, 1969). The model was estimated using the *MHRM* algorithm (Cai, 2010), and model fit was assessed with the same combination of fit indices and their aforementioned cut-off values (CFI & TLI ≥ .95; RMSEA ≤ .06). To compute these indices, the Maydeu's M_2 statistic for ordinal variables was computed instead of other approximations of the chi-square for ordinal variables, since it was particularly developed to assess the overall fit for IRT models (Maydeu-Olivares & Garcia-Forero, 2010; Maydeu-Olivares & Joe, 2006). The test information function was obtained and the IRT scores (i.e., the person parameters of the model) were used for the validity analyses.

To establish whether the VB-IPVAW is invariant across gender, we conduct a differential item functioning (DIF) analysis for polytomous data using the logistic regression method (French & Miller 1996; Choi, Gibbons, & Crane, 2011). DIF occur when the probability of endorsement of an item category is not the same for male and female respondents with an equivalent IRT score, meaning that men and women answer the item differentially. If DIF

is detected for an item, then the item parameters should be recalibrated for each subgroup in order to obtain a comparable IRT score.

We assessed the VB-IPVAW scale validity for the whole sample. To this end, we correlated first the IRT scores (i.e., victim blaming attitudes estimates) with the A-IPVAW scores, the PS-IPVAW scores, and with the hostile and benevolent sexism scores from the ASI. Finally, we compared the VB-IPVAW scores between age groups and gender for the general sample, and between men from the general sample and the sample of male offenders.

Finally, we assembled a shortened version of the scale using the most informative items to measure higher levels of the latent trait, using automated test assembly (Diao & van der Linden, 2011). Through this procedure, the minimum number of items that meet different criteria imposed by the researchers are selected to be included in the shortened version. We selected those items that measure more accurately higher levels of victim blaming attitudes.

Descriptive statistics, classical internal consistency, and IRT analyses were computed with the statistical package R (R Core Team, 2017). In particular, we used the *psych* (Revelle, 2016), the *mirt* (Chalmers, 2012), the *lordif* (Choi, Gibbons, & Crane, 2011), and the *lpSolveApi* libraries (Konis, 2014). The parallel analysis was conducted with the *factor* package (Lorenzo-Seva & Ferrando, 2006), whereas the EFA and CFA analysis were carried out with *Mplus 7.1* (Muthén & Muthén, 2010).

Results

Descriptive analyses and reliability

The mean, standard deviation, range, skew and kurtosis statistics, and item-total corrected correlations are displayed in Table 2. All items present a mean centered in the lower category (i.e., "strongly disagree"), with standard

Table 2. VB-IPVAW items descriptive statistics.

	M	*SD*	*Min*	*Max*	*Skew*	*Kurtosis*	$r_{item\text{-}total}$
vb-ipvaw1	1.38	0.64	1	4	1.61(.02)	1.94(.02)	0.60
vb-ipvaw2	1.23	0.49	1	4	2.23(.01)	5.56(.01)	0.74
vb-ipvaw3	1.22	0.55	1	4	2.80(.01)	8.32(.01)	0.59
vb-ipvaw4	1.21	0.47	1	4	2.58(.01)	8.07(.01)	0.71
vb-ipvaw5	1.19	0.45	1	4	2.72(.01)	8.94(.01)	0.73
vb-ipvaw6	1.21	0.49	1	4	2.63(.01)	7.97(.01)	0.64
vb-ipvaw7	1.15	0.41	1	4	3.14(.01)	12.09(.01)	0.69
vb-ipvaw8	1.73	0.77	1	4	0.83(.02)	0.10(.02)	0.37
vb-ipvaw9	1.25	0.55	1	4	2.45(.01)	6.41(.01)	0.63
vb-ipvaw10	1.20	0.50	1	4	2.80(.01)	8.87(.01)	0.68
vb-ipvaw11	1.29	0.55	1	4	2.11(.01)	5.04(.01)	0.58
vb-ipvaw12	1.12	0.38	1	4	3.64(.01)	16.10(.01)	0.61
vb-ipvaw13	2.19	1.09	1	4	0.30(.01)	-1.34(.01)	0.08

Note: *M* = Mean, *SD* = Standard Deviation, *Min* = Minimum, *Max* = Maximum, $r_{item\text{-}total}$ = item-total corrected correlation. In brackets: the standard error for the *skew* and *kurtosis* statistics. vb-ipvaw: VB-IPVAW item.

deviations around .50, positively skewed and with high values of kurtosis. This implies that most of the respondents disagree with the statements. Regarding the item-total corrected correlations, all items were strongly related with the scale raw scores, except the last item which was removed from the scale for this reason. The internal consistency of the scale was good (Cronbach's α = .89).

Exploratory Factor Analysis

We carried out and EFA with the first subsample. Bartlett's sphericity test was significant ($p < .001$) and the KMO index was acceptable (KMO = .939), indicating that the matrix was suitable to perform a factor analysis. The parallel analysis based on minimum rank factor analysis showed that a one-factor solution accounted for the 74.2% of the variance, above the expected 20.8% for the simulated datasets. However, a two-factor solution accounted
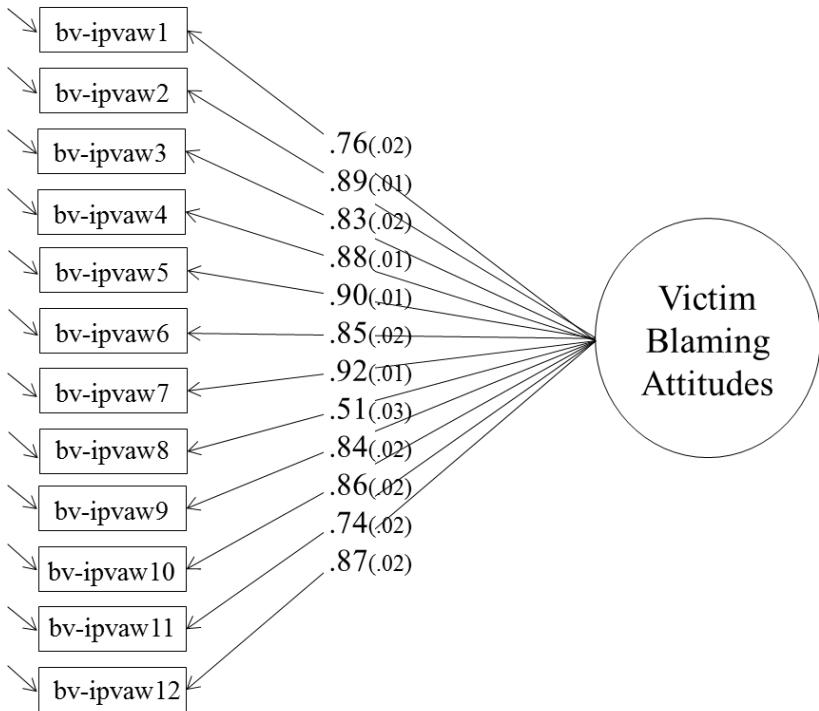
Figure 1. VB-IPVAW one-factor model.

only for the 6.9% of the variance, below the expected 14.5% for the simulated datasets. Thus, a one-factor solution was considered for the EFA. We extracted then one factor using WLSMV as the estimation method using the polychoric correlation matrix. The model converged normally and the model fitted very well the data (CFI = .99, TLI = .98, RMSEA = .051, SRMS = .038).

Confirmatory Factor Analysis

We conducted a CFA with the second subsample. To this end we specified a one-factor model and estimated it using the WLSMV method. As shown in Figure 1, all standardized loadings were greater than .70, with standard estimation errors around .02. The comparative fit indices of the model were good (CFI = .99, TLI = .99), with well-fitted residuals (RMSEA = .051), replicating the EFA results in a different subsample and yielding an

196

Table 3. VB-IPVAW scale IRT item parameters.

	a	b_1	b_2	b_3
vb-ipvaw1	2.05	0.82	2.07	3.47
vb-ipvaw2	3.68	1.03	2.31	3.00
vb-ipvaw3	2.39	1.34	2.33	2.98
vb-ipvaw4	3.43	1.15	2.46	2.90
vb-ipvaw5	3.72	1.28	2.56	3.03
vb-ipvaw6	2.49	1.24	2.55	3.20
vb-ipvaw7	3.36	1.40	2.64	3.12
vb-ipvaw8	0.95	-0.18	2.22	4.31
vb-ipvaw9	2.49	1.17	2.30	3.04
vb-ipvaw10	3.16	1.25	2.30	2.85
vb-ipvaw11	1.93	1.03	2.67	3.45
vb-ipvaw12	3.05	1.59	2.75	3.29

Note: a: discrimination parameter; b_k: threshold parameters.
vb-ipvaw: VB-IPVAW item.

overall good fit to the data. We kept the one-factor solution as the latent structure of the VB-IPVAW.

Item Response Theory

Once the dimensionality of the VB-IPVAW was determined, we used the full sample to fit the IRT graded response model (Samejima, 1969).

Item parameters are displayed in Table 3. The threshold parameters (b_1, b_2, and b_3) are in the same metric than the IRT scores (i.e., estimates of victim blaming attitudes), indicating the point in the latent trait continuum where the probability of endorsement between two adjacent categories is .50 for any respondent with an IRT score equal to the threshold parameter value. Respondents with IRT scores lower than the b_1 parameter would endorse more likely the lowest category (i.e., "Strongly disagree"), whereas those respondents with IRT scores higher than the b_3 parameter would tend to endorse the upper category (i.e., "Strongly agree"). Those respondents with IRT scores between the b_1 and b_2 parameters would endorse more likely the second category (i.e., "Somewhat disagree"), and those respondents with IRT

scores comprised between the b_2 and b_3 parameters would more likely endorse the third category (i.e., "Somewhat agree"). The b_1 threshold parameters were in general moderate (i.e., around 1), whereas the b_2 and b_3 parameters present high (i.e., above 2) and very high values (i.e., above 3), indicating that the instrument is sampling moderate to high levels of victim blaming attitudes.

The discrimination parameters (a), in turn, provide information about the accuracy of each item. In particular, the greater this parameter is, the less likely it is that a given respondent will endorse a category above or below their IRT score. The discrimination parameters of the VB-IPVAW were high, with values above 2 for almost all the items, with the exception of item 8, which present a moderate a value.

To assess differential item functioning (DIF), we conduct a series of logistic regression models (e.g., Choi, Gibbons, & Crane, 2011). These models tested whether there is an effect by pertaining to each group (i.e., men or women) in the latent trait continuum (i.e., estimates of victim blaming attitudes), and whether this effect is constant (uniform), or it varies across the continuum (non-uniform). These models were compared though a χ^2 test; Nagelkerke's pseudo-R^2 was also computed to assess the size of the DIF effect. This pseudo-R^2 indicates the improvement from the base model (i.e., no DIF model) to the fitted model (i.e., uniform or non-uniform DIF model). We found uniform DIF on items 1, 2 and 8 ($p < .001$, $R^2_{Nagelkerke}$ = 0.010, 0.007, and 0.018, respectively), and non-uniform DIF on the item 3 ($p = .002$, $R^2_{Nagelkerke}$ = 0.010), all with small values of the Nagelkerke's pseudo-R^2. This means that the adding the DIF effect to the model improved the fit of the model less than 2%, which could be considered a negligible effect size for this statistic (Choi, Gibbons, & Crane, 2011; Cohen, 1988).

The Figure 2 shows the test information function and the standard

Victim Blaming Attitudes
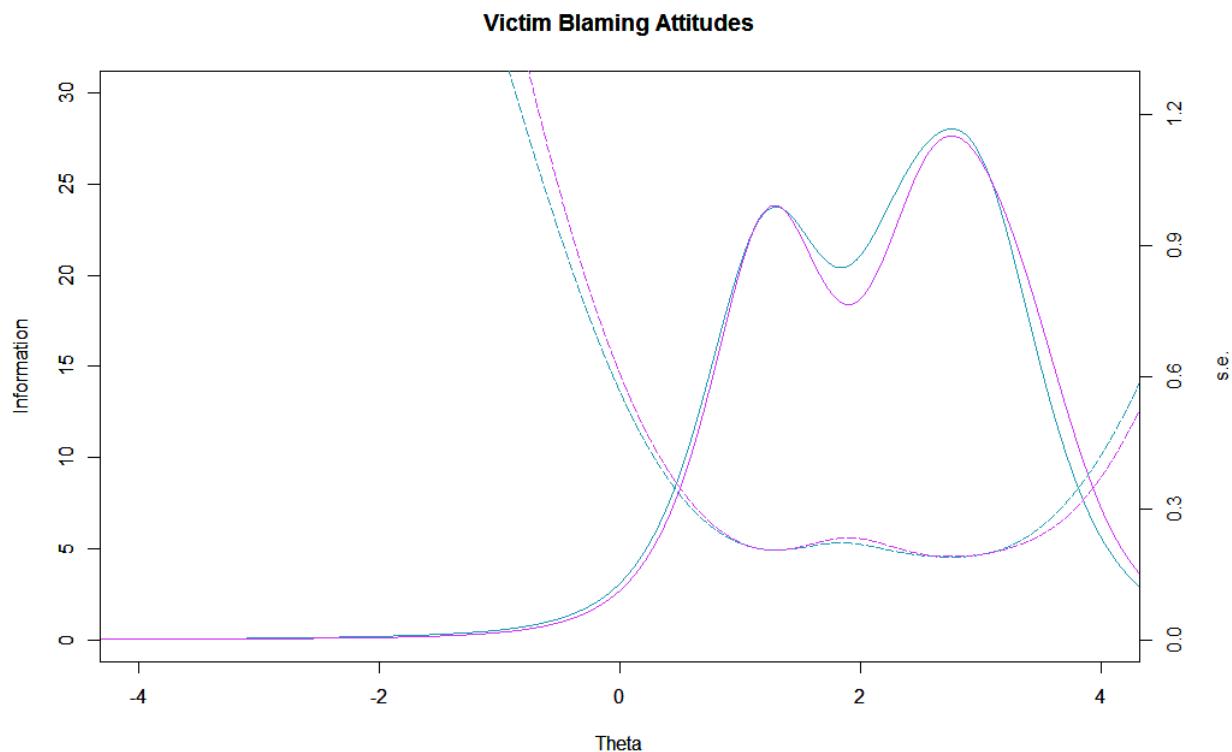
Figure 2. Test Information Function.

Theta: IRT scores of the scale (i.e., victim blaming attitudes estimates). *Information*: accuracy of the measure over the latent trait continuum (i.e., Theta). s.e.: Standard error of estimation. The pink continuous and discontinuous lines represent to the test information function and s.e. for women respectively, whereas the blue lines correspond to the test information function and s.e. for men.

error of estimation (s.e.) for men and women when the flagged items are taken into account. The VB-IPVAW scale resulted especially informative for moderate, high, and very high levels of the latent trait continuum (i.e., estimates of victim blaming attitudes), especially among men. In the same line, the s.e. showed the precision of the scale for the latent trait levels; the lower the s.e. result, the higher is the accuracy of the scale for a given latent trait level. In particular, s.e. values below 0.3 are equivalent to a Cronbach's α of .91 or higher, and s.e. values between 0.5 and 0.3 are equivalent to an α around .75

and .90 for their respective IRT score. The test information function resulted very similar for men and women.

The overall fit of the model was tested using the ordinal version of the M_2 statistic. This statistic could be used to compute an approximation of the most common fit indices from the factor analysis (CFI, TLI, and RMSEA). The model showed a good fit to the data when the DIF was taken into account, M_2 (92) = 373.50, $p < .001$, CFI = .98, TLI = .98, RMSEA = .04.

Validity Analyses

We used the IRT scores (i.e., "estimates of victim blaming attitudes") for validity analyses, since the items of the scale are not tau-equivalent (i.e., equally discriminative). IRT scores were on logistic metric, with an expected mean value of 0 and a standard deviation of 1 (see Chalmers, 2012).

The correlations between the VB-IPVAW and the variables measuring related constructs were in the expected direction (see Table 4). Estimates of victim blaming attitudes were positively related to the A-IPVAW scores, to both subscales of the ASI, in particular with the hostile sexism subscale; and negatively related to PS-IPVAW scores. Therefore, those participants scoring higher on the VB-IPVAW tended to score higher on acceptability of IPVAW, hostile and benevolent sexism, and tended to perceived the cases described in the PS-IPVAW as less severe.

Table 4. VB-IPVAW scale correlation to other relevant IPVAW variables.

	A-IPVAW	Hostile Sexism	Benevolent Sexism	PS-IPVAW
VB-IPVAW	.41**	.45**	.38**	-.36**
A-IPVAW		.44**	.34**	-.47**
Hostile Sexism			.81**	-.39**
Benevolent Sexism				-.30**

Note: **: $p < .001$. VB-IPVAW: Victim blaming attitudes towards cases of intimate partner violence against women. A-IPVAW: Acceptability of intimate partner violence against women scale. PS-IPVAW: Perceived severity of intimate partner violence against women scale.

We then compared the VB-IPVAW scores by gender and age. We found significant differences between men (M = 0.23, SD = 0.94) and women (M = -0.09, SD = 0.83), t(1711.9) = 7.63, p < .001, d = .36, with a moderate effect size. We also found significant differences between male respondents from the general sample and those from the offenders sample (M = 0.70, SD = 1.35), t(52.48) = -3.31, p = .002, d = .44, with a moderate effect size. There were significant differences between age groups, F(2) = 5.48, p = .004, η^2 = .006; however the effect of age on the VB-IPVAW scores could be considered negligible, since the size effect was below the low cut-off value of .01 for the partial eta-squared (Miles & Shevlin, 2001).

Short Version

To build the short version of the scale we computed an automatic test assembly algorithm. Two criteria were used: finding which items were more informative to assess moderate and high levels of the IRT scores for men and women (i.e., $\hat{\theta} \geq 0$), and using the minimum number of items to ensure that those levels are measured accurately (i.e., s.e. ≤ 0.5). The items 1, 2, 6, 9, and 11 were selected.

The internal consistency of the resulting short version was good (Cronbach's α = .77), and the correlation between the full and the short version was strong (r = .95). The correlations between the short version and the validity measures were in the same direction than the full VB-IPVAW (r = .46 with acceptability of IPVAW, r = -.36 with perceived severity of IPVAW, r = .36 with hostile sexism, and r = .33 with benevolent sexism).

Discussion

In this paper we set out to develop and validate a new tool to assess victim blaming attitudes towards IPVAW. Taken together, our findings provide strong evidences for the reliability and validity of the VB-IPVAW and its short

version. According to the test development standards, our measure presents validity evidences based on its content, its internal structure, and its relations to other variables (APA, AERA, & NCME, 2014). Content validity of the VB-IPVAW was evaluated through a careful selection of items from European surveys (Gracia & Lila, 2015), and through a panel of experts who ensured the actual relevance of the items to tap the key aspects of the construct. Regarding the internal structure of the scale, our findings support that a single dimension is sufficient to account for the variability of respondents' victim blaming attitudes towards cases of IPVAW, presenting a good fit to the data and high internal consistency in two different samples.

Using item response theory (IRT) to study the psychometric properties of the VB-IPVAW is one of the main strengths of this paper, since to the best of our knowledge this is the first time that this approach is used to address the assessment of victim blaming attitudes towards cases of IPVAW. One of the main advantages of IRT is that it provides improved factor scores that can be used to evaluate which levels on the latent trait continuum (i.e., victim blaming estimates) are better measured through the test information function. Unlike Cronbach's α, that assumes that the internal consistency of an instrument is constant for the entire latent trait, the information function provides a dynamic approach that allows assessing the precision of the scale over the different levels of the latent trait continuum. Our measure is particularly informative for moderately high and very high levels of victim blaming attitudes. The VB-IPVAW can detect respondents with high levels of victim blaming attitudes and discriminate among them with high accuracy. On the other hand, the precision of the scale is lower for respondents with low and very low levels of victim blaming attitudes.

Differential item functioning (DIF) was also assessed to establish whether any of the items of the VB-IPVAW was a potential indicator of item

bias (Sireci & Ríos, 2013). We found four items that showed DIF between male and female respondents. Nevertheless, the effect size of these discrepancies could be considered negligible and these results should be taken with caution. Given the low improvement of the models considering DIF in comparison with the non-DIF model, we recommend to compute the IRT scores of the VB-IPVAW using the same set of parameters for both male and female respondents. To this end, we provide in the Appendix 2 an R code to compute the IRT scores, in both the full and short versions of the scale.

Regarding the validity evidences based on the relation of the VB-IPVAW with other variables, we found that victim blaming attitudes towards cases of IPVAW are strongly related to the acceptability of IPVAW, and strongly and negatively related to the perceived severity of IPVAW. These relationships are consistent with previous research (Taylor & Sorenson, 2005; Witte, Schroeder, & Lhor, 2006). We also found a strong relationship between victim blaming attitudes and both hostile and benevolent sexism. In this line, previous studies have also found that individuals showing sexist attitudes tend to blame more the victims for IPVAW (Capezza & Arriaga, 2008; Gracia, Garcia, & Lila 2014; Lila et al., 2015; Judicibus & McCabe, 2001; Scott & Strauss 2007; Valor-Segura et al., 2011; Vidal-Fernández & Megías, 2014).

With regard to gender differences, our findings show that female respondents tend to present lower victim blaming attitudes towards IPVAW than males, which is congruent with previous research (Bryant & Spencer, 2003; Flood & Pease, 2009; Gracia et al., 2015; Langhinrichsen-Rohling, Shlien-Dellinger, Huss, & Kramer, 2004; Scott & Strauss, 2007; Vidal-Fernández & Megías, 2014). In addition, we also found that IPVAW male offenders tend to show higher levels of victim blaming attitudes than men from the general sample (Gracia et al., 2015; Lila et al., 2015; Lila, Gracia, & Murgui, 2013). This result highlights that the VB-IPVAW is especially informative for those

respondents with higher levels of victim blaming attitudes towards cases of IPVAW. Our measure can thus be used to evaluate attitudinal changes during and after the intervention with IPVAW offenders, as well as a screening tool to detect and discriminate between those individuals that are more prone to blame the victims for the IPVAW (Carbajosa, Catalá-Miñana, Lila, & Gracia, 2017; Lila, Gracia, & Catalá-Miñana, 2017; Lila, Oliver, Galiana, & Gracia, 2013).

In this study we have also developed a short 5-item version of the VB-IPVAW that can be useful when space and/or time limitations are an issue (e.g., large demographical surveys). Large scale surveys tend to use single items or a limited set of items evaluating IPVAW attitudes with unknown reliability or validity (Gracia & Lila, 2015). Short versions, on the other hand, may have limited reliability and validity, which makes it particularly important to ensure that short versions of questionnaires maintain adequate psychometric properties (Smith et al., 2000; Stanton et al., 2002). Our results showed that the short version of the VB-IPVAW has a high internal consistency and an adequate validity (i.e., it is strongly related to acceptability, perceived severity of IPVAW, and ambivalent sexism in the same direction that in the full version). Although further research is needed using different samples to ensure the validity of this short version (Goetz et al., 2013), the first analyses with the short version of VB-IPVAW are promising and suggest that it is as an adequate tool to assess victim blaming attitudes with a limited set of informative items. We recommend, however, using the long version of the scale when possible.

This study is not without limitations. The VB-IPVAW was developed in the Spanish cultural setting, and further studies are needed to adapt and generalize our findings to other cultures (Gracia & Lila, 2015; Boira et al., 2016). The sampling method is another limitation, since online sampling has some tradeoffs that may limit the generalizability of this study (Thornton et al., 2016; Topolovec-Vranic & Natarajan, 2016). Although this method is effective

and cost-efficient, allowing researchers to obtain large sample sizes, however, self-selection bias could be an issue, as those participants more motivated may be more willing to participate in the study. In addition, it is difficult to verify the socio-demographical characteristics facilitated by the respondents. The socio-demographic variables of the sample, nonetheless, are in concordance to other internet-based demographic studies conducted in Spain with internet users (Acebes Arribas, 2016). In addition, further research is needed to address the gender invariance of the VB-IPVAW through both IRT and factorial invariance methods, as DIF is only an initial step to assess item bias (Gómez-Benito, Sireci, Padilla, Hidalgo, & Benítez, 2018).

The development of the VB-IPVAW is a step forward in the study of attitudes toward IPVAW, allowing researchers to advance their knowledge about their conceptualization, measurement, prevalence, and the social factors that may influence these attitudes in order to improve prevention and intervention strategies (Powell & Webster, 2018). Addressing attitudes towards IPVAW is becoming a central issue in research and population surveys and, in this regard, both versions of the VB-IPVAW provide psychometrically sound instruments to fill this need.

References

Acebes Arribas, B. (2016). *Estudio de medios de comunicación on line*. Madrid: IAB.

Ali, P. A., & Naylor, P. B. (2013). Intimate partner violence: A narrative review of the feminist, social and ecological explanations for its causation. *Aggression and Violent Behavior*, *18*, 611-619.

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education (2014).

Standards for educational and psychological testing. Washington, DC: American Psychological Association.

Ansara, D. L., & Hindin, M. J. (2010). Formal and informal help-seeking associated with women's and men's experience of intimate partner violence in Canada. *Social Science and Medicine, 70*, 1011-1018.

Asparouhov, T., & Muthén, B. (2010). *Weighted least squares estimation with missing data.* Retrieved from:
http://www.statmodel.com/download/GstrucMissingRevision.pdf

Boira, S., Carbajosa, P., & Méndez, R. (2016). Miedo, conformidad y silencio: la violencia en las relaciones de pareja en áreas rurales de Ecuador [Fear, conformity and silence. Intimate partner violence in rural areas of Ecuador]. *Psychosocial Intervention*, *25*, 9-17.

Bryant, S. A., & Spencer, G. A. (2003). University students' attitudes about attributing blame in domestic violence. *Journal of Family Violence*, *18*, 369-376.

Browning, C. R. (2002). The span of collective efficacy: Extending social disorganization theory to partner violence. *Journal of Marriage and Family*, *64*, 833-850.

Cai, L. (2010). High-dimensional exploratory item factor analysis by a Metropolis–Hastings Robbins–Monro algorithm. *Psychometrika*, *75*, 33-57.

Campbell, J. C. (2002). Health consequences of intimate partner violence. *The Lancet, 359,* 1331-1336.

Capezza, N. M., & Arriaga, X. B. (2008). Factors associated with acceptance of psychological aggression against women. *Violence Against Women*, *14*, 612-633.

Carbajosa, P., Catalá-Miñana, A., Lila, M., & Gracia, E. (2017). Differences in treatment adherence, program completion, and recidivism among

batterer subtypes. *The European Journal of Psychology Applied to Legal Context, 9,* 93-101.

Carlson, B. E., & Worden, A. P. (2005). Attitudes and beliefs about domestic violence: Results of a public opinion survey I. Definitions of domestic violence, criminal domestic violence, and prevalence. *Journal of Interpersonal Violence*, *20*, 1197-1218.

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*, 1-29.

Chernyshenko, O. S., Stark, S., Chan, K. Y., Drasgow, F., & Williams, B. (2001). Fitting item response theory models to two personality inventories: Issues and insights. *Multivariate Behavioral Research*, *36*, 523-562.

Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software*, *39*, 1-30.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hilsdale: *Lawrence Earlbaum Associates*.

Copp, J. E., Giordano, P. C., Longmore, M. A., & Manning, W. D. (2016). The development of attitudes toward intimate partner violence: An examination of key correlates among a sample of young adults. *Journal of Interpersonal Violence*. Retrieved from: http://journals.sagepub.com/doi/pdf/10.1177/0886260516651311

Craparo, G., Gori, A., Petruccelli, I., Cannella, V., & Simonelli, C. (2014). Intimate partner violence: relationships between alexithymia, depression, attachment styles, and coping strategies of battered women. *The Journal of Sexual Medicine*, *11*, 1484-1494.

De Judicibus, M., & McCabe, M. P. (2001). Blaming the target of sexual harassment: Impact of gender role, sexist attitudes, and work role. *Sex Roles*, *44*, 401-417.

Devries, K. M., Mak, J. Y., Garcia-Moreno, C., Petzold, M., Child, J. C., Falder, G., ... & Pallitto, C. (2013). The global prevalence of intimate partner violence against women. *Science*, *340*, 1527-1528.

Diao, Q., & van der Linden, W. J. (2011). Automated test assembly using lp solve version 5.5 in R. *Applied Psychological Measurement*, *35*, 398-409.

Ellsberg, M., Jansen, H. A., Heise, L., Watts, C. H., & Garcia-Moreno, C. (2008). Intimate partner violence and women's physical and mental health in the WHO multi-country study on women's health and domestic violence: an observational study. *The Lancet*, *371*, 1165-1172.

Expósito, F., Moya, M. C., & Glick, P. (1998). Sexismo ambivalente: medición y correlatos [Ambivalent sexism: measurement and correlates]. *Revista de Psicología Social*, *13*, 159-169.

Faramarzi, M., Esmailzadeh, S., & Mosavi, S. (2005). Prevalence and determinants of intimate partner violence in Babol City, Islamic Republic of Iran. *Eastern Mediterranean Health Journal*, *11*, 870-879.

Flood, M., & Pease, B. (2009). Factors influencing attitudes to violence against women. *Trauma, Violence, & Abuse*, *10*, 125-142.

Fox, K. A., & Cook, C. L. (2011). Is knowledge power? The effects of a victimology course on victim blaming. *Journal of Interpersonal Violence*, *26*, 3407-3427.

French, A. W., & Miller, T. R. (1996). Logistic regression and its use in detecting differential item functioning in polytomous items. *Journal of Educational Measurement*, *33*, 315-332.

Garcia-Moreno, C., Jansen, H. A., Ellsberg, M., Heise, L., & Watts, C. H. (2006). Prevalence of intimate partner violence: findings from the WHO multi-country study on women's health and domestic violence. *The Lancet*, *368*, 1260-1269.

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2013). A new look at Horn's parallel analysis with ordinal variables. *Psychological Methods*, *18*, 454-474.

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2016). Are fit indices really fit to estimate the number of factors with categorical variables? Some cautionary findings via Monte Carlo simulation. *Psychological Methods*, *21*, 93.

Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, *70*, 491.

Glick, P., Fiske, S. T., Mladinic, A., Saiz, J. L., Abrams, D., Masser, B., ... & Annetje, B. (2000). Beyond prejudice as simple antipathy: hostile and benevolent sexism across cultures. *Journal of Personality and Social Psychology*, *79*, 763.

Glick, P., Sakalli–Ugurlu, N., Ferreira, M. C., & Aguiar de Souza, M. (2002). Ambivalent sexism and attitudes toward wife abuse in Turkey and Brazil. *Psychology of Women Quarterly*, *26*, 292-297.

Glockner-Rist, A., & Hoijtink, H. (2003). The best of both worlds: Factor analysis of dichotomous data using item response theory and structural equation modeling. *Structural Equation Modeling*, *10*, 544-565.

Goetz, C., Coste, J., Lemetayer, F., Rat, A. C., Montel, S., Recchia, S., ... & Guillemin, F. (2013). Item reduction based on rigorous methodological guidelines is necessary to maintain validity when shortening composite measurement scales. *Journal of Clinical Epidemiology*, *66*, 710-718.

Gómez-Benito, J., Sireci, S., Padilla, J. L., Hidalgo, M. D., & Benítez, I. (2018). Differential Item Functioning: Beyond validity evidence based on internal structure. *Psicothema*, *30*, 104-109.

Gracia, E. (2014). Intimate partner violence against women and victim-blaming attitudes among Europeans. *Bulletin of the World Health Organization*, *92*, 380-381.

Gracia, E., García, F., & Lila, M. (2008). Police involvement in cases of intimate partner violence against women: The influence of perceived severity and personal responsibility. *Violence Against Women*, *14*, 697-714.

Gracia, E., García, F., & Lila, M. (2009). Public responses to intimate partner violence against women: The influence of perceived severity and personal responsibility. *The Spanish Journal of Psychology*, *12*, 648-656.

Gracia, E., García, F., & Lila, M. (2011). Police attitudes toward policing partner violence against women: Do they correspond to different psychosocial profiles? *Journal of Interpersonal Violence*, *26*, 189-207.

Gracia, E., & Herrero, J. (2006a). Public attitudes toward reporting partner violence against women and reporting behavior. *Journal of Marriage and Family*, *68*, 759-768.

Gracia, E., & Herrero, J. (2006b). Acceptability of domestic violence against women in the European Union: A multilevel analysis. *Journal of Epidemiology and Community Health*, *60*, 123-129.

Gracia, E., Herrero Olaizola, J. B., Lila, M., & Fuente Barrera, A. (2009). Perceived neighborhood social disorder and attitudes toward domestic violence against women among Latin-American immigrants. *European Journal of Psychology Applied to Legal Context*, *1*, 25-43.

Gracia, E., & Lila, M. (2015). *Attitudes towards violence against women in the EU*. Luxembourg: Publication Office of the European Union.

Gracia, E., & Tomás, J. M. (2014). Correlates of victim-blaming attitudes regarding partner violence against women among the Spanish general population. *Violence Against Women*, *20*, 26-41.

Henning, K., & Holdford, R. (2006). Minimization, denial, and victim blaming by batterers: How much does the truth matter?. *Criminal Justice and Behavior*, *33*, 110-130.

Henning, K., Jones, A. R., & Holdford, R. (2005). "I didn't do it, but if I did I had a good reason": Minimization, denial, and attributions of blame among male and female domestic violence offenders. *Journal of Family Violence*, *20*, 131-139.

Hu L. T., & Bentler P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Jackson, T. L., Dienst, R. D., Efird, T. L., Mobley, B. D., Schroeder, D. A., Hout, A. D., . . . , & LaBine, A. L, (1994). The Violence Attitudes Scale (VAS). In L. VandeCreek, S. Knapp & T. L. Jackson (Eds.), *Innovations in clinical practice: A source book* (Vol. 13, pp. 279-291). Sarasota, FL: Professional Resource Press/Professional Resource Exchange.

Koepke, S., Eyssel, F., & Bohner, G. (2014). "She Deserved It" Effects of Sexism Norms, Type of Violence, and Victim's Pre-Assault Behavior on Blame Attributions Toward Female Victims and Approval of the Aggressor's Behavior. *Violence Against Women*, *20*, 446-464.

Konis, K. (2016). lpSolveAPI: R Interface to 'lp_solve' Version 5.5.2.0. R package version 5.5.2.0-17. Retrieved from: https://CRAN.R-project.org/package=lpSolveAPI

Langhinrichsen-Rohling, J., Shlien-Dellinger, R. K., Huss, M. T., & Kramer, V. L. (2004). Attributions about perpetrators and victims of interpersonal abuse: Results from an analogue study. *Journal of Interpersonal Violence*, *19*, 484-498.

Liang, B., Goodman, L., Tummala-Narra, P., & Weintraub, S. (2005). A theoretical framework for understanding help-seeking processes

among survivors of intimate partner violence. *American Journal of Community Psychology*, *36*, 71-84.

Lila, M., Gracia, E., & García, F. (2013). Ambivalent sexism, empathy and law enforcement attitudes towards partner violence against women among male police officers. *Psychology, Crime & Law*, *19*, 907-919.

Lila, M., Gracia, E., & Murgui, S. (2013). Psychological adjustment and victim-blaming among intimate partner violence offenders: The role of social support and stressful life events. *The European Journal of Psychology Applied to Legal Context*, *5*, 147-153.

Lila, M., Oliver, A., Catalá-Miñana, A., Galiana, L., & Gracia, E. (2014). The intimate partner violence responsibility attribution scale (IPVRAS). *The European Journal of Psychology Applied to Legal Context*, *6*, 29-36.

Lila, M., Oliver, A., Galiana, L., & Gracia, E. (2013). Predicting success indicators of an intervention programme for convicted intimate-partner violence offenders: The Contexto Programme. *The European Journal of Psychology Applied to Legal Context*, *5*, 73-95.

Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavior Research Methods*, *38*, 88-91.

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35*, 382-385.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*, 130-149.

Martín-Fernández, M., Gracia, E., Marco, M., Vargas, V., Santirso, F. A., & Lila, M. (2018). Measuring acceptability of intimate partner violence against

women: Development and validation of the A-IPVAW scale. *The European Journal of Psychology Applied to Legal Context*, *10*, 26-34.

Maydeu-Olivares, A., & Garcia-Forero, C. (2010). Goodness-of-fit testing. *International Encyclopedia of Education*, *7*, 190-196.

Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psycho-metrika*, *71*, 713.

Miles, J., & Shevlin, M. (2001). *Applying Regression and Correlation: A Guide for Students and Researchers*. Sage: London.

Muehlenhard, C. L., & Kimes, L. A. (1999). The social construction of violence: The case of sexual and domestic violence. *Personality and Social Psychology Review*, *3*, 234-245.

Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide*. Los Angeles, CA: Author.

Petretic-Jackson, P., Sandberg, G., & Jackson, T. (1994). The Domestic Violence Blame scale. In L. VandeCreek, S. Knapp & T. Jackson (Eds.), *Innovations in clinical practice: A source book* (Vol. 13, pp. 101-118). Sarasota, FL: Professional Resource Press/ Professional Resource Exchange.

Polit, D. F., & Beck, C. T. (2006). The content validity index: are you sure you know what's being reported? Critique and recom-mendations. *Research in Nursing & Health*, *29*, 489-497.

Powell, A., & Webster, K. (2018). Cultures of gendered violence: An integrative review of measures of attitudinal support for violence against women. *Australian & New Zealand Journal of Criminology, 51,* 40-57.

R Core Team. (2017). R: A language and environment for statistical computing. *R foundation for Statistical Computing*. Retrieved from: https://www.R-project.org/

Revelle, W. (2016) *psych: Procedures for personality and psychological research*. Northwestern University, Evanston, Illinois, USA. Retrieved from: http://CRAN.R-project.org/package=psych Version = 1.5.4 .

Rizo, C. F., & Macy, R. J. (2011). Help seeking and barriers of Hispanic partner systematic review of the literature. *Aggression and Violent Behavior, 6,* 250-264.

Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores.* New York, NY: Psychometric Society.

Scott, K., & Straus, M. (2007). Denial, minimization, partner blaming, and intimate aggression in dating partners. *Journal of Interpersonal Violence*, *22*(7), 851-871.

Sireci, S. G., & Rios, J. A. (2013). Decisions that make a difference in detecting differential item functioning. *Educational Research and Evaluation*, *19*(2-3), 170-187.

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment, 12*, 102-111.

Stanton, J. M., Sinar, E. F., Balzer, W. K., & Smith, P. C. (2002). Issues and strategies for reducing the length of self-report scales. *Personnel Psychology, 55*, 167-194.

Stöckl, H., Devries, K., Rotstein, A., Abrahams, N., Campbell, J., Watts, C., & García-Moreno, C. (2013). The global prevalence of intimate partner homicide: A systematic review. *The Lancet*, *382*, 859-865.

Taylor, C. A., & Sorenson, S. B. (2005). Community-based norms about intimate attributions of fault and responsibility into context. *Sex Roles, 53*, 573-589.

Thornton, L., Batterham, P. J., Fassnacht, D. B., Kay-Lambkin, F., Calear, A. L., & Hunt, S. (2016). Recruiting for health, medical or psychosocial research using Facebook: Systematic review. *Internet Interventions*, *4*, 72-81.

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality Assessment of Ordered Polytomous Items with Parallel Analysis. *Psychological Methods, 16*, 209-220.

Topolovec-Vranic, J., & Natarajan, K. (2016). The Use of Social Media in Recruitment for Medical Research Studies: A Scoping Review. *Journal of Medical Internet Research*, *18,* e286.

Valor-Segura, I., Expósito, F., & Moya, M. (2011). Victim blaming and exoneration of the perpetrator in domestic violence: The role of beliefs in a just world and ambivalent sexism. *The Spanish Journal of Psychology*, *14*, 195-206.

Vargas, V., Lila, M., & Catalá-Miñana, A. (2015). ¿Influyen las diferencias culturales en los resultados de los programas de intervención con maltratadores?: un estudio con agresores españoles y latinoamericanos [Do cultural differences influence batterer intervention program outcomes? A study with Spanish and Latin American offenders]. *Psychosocial Intervention*, *24*, 41-47.

Vidal-Fernández, A., & Megías, J. L. (2014). Attributions of Blame to Battered Women when they are perceived as Feminists or as "Difficult to Deal With". *The Spanish Journal of Psychology*, *17,* E21.

Voith, L. A. (2017). Understanding the relation between neighborhoods and intimate partner violence: An integrative review. *Trauma, Violence, & Abuse*.

Waltermaurer, E. (2012). Public justification of intimate partner violence: a review of the literature. *Trauma, Violence, & Abuse, 13*, 167-175.

West, A., & Wandrei, M. L. (2002). Intimate partner violence a model for predicting interventions by informal helpers. *Journal of Interpersonal Violence*, *17*, 972-986.

Witte, T. H., Schroeder, D. A., & Lohr, J. M. (2006). Blame for intimate partner violence: An attributional analysis. *Journal of Social and Clinical Psychology*, *25*, 647-667.

World Health Organization. (2013). *Global and regional estimates of violence against women: prevalence and health effects of intimate partner violence and non-partner sexual violence.* Geneva: World Health Organization.

Yamawaki, N., Ostenson, J., & Brown, C. R. (2009). The functions of gender role traditionality, ambivalent sexism, injury, and frequency of assault on domestic violence perception: A study between Japanese and American college students. *Violence Against Women*, *15*, 1126-1142.

Yun, S. H., & Vonk, M. E. (2011). Development and initial validation of the Intimate Violence Responsibility Scale (IVRS). *Research on Social Work Practice*, *21*, 562-571.

Appendix 1. Victim blaming attitudes towards cases of intimate partner violence against women scale

vb-ipvaw1	Men are violent toward their partners because they make them jealous [Los hombres son violentos con sus parejas porque estas les provocan celos]
vb-ipvaw2	Men are violent toward their partners because women provoke them [Los hombres son violentos con sus parejas porque las mujeres les provocan]
vb-ipvaw3	Men are violent toward their partners because it is necessary to control them [Los hombres son violentos con sus parejas porque es necesario controlarlas]
vb-ipvaw4	Men are violent toward their partners because women are difficult to understand [Los hombres son violentos con sus parejas porque las mujeres son difíciles de comprender]
vb-ipvaw5	Men are violent toward their partners because women are not patient enough with them [Los hombres son violentos con sus parejas porque las mujeres no tienen la suficiente paciencia con ellos]
vb-ipvaw6	Men are violent toward their partners because that makes them attractive to women [Los hombres son violentos con sus parejas porque ser así es atractivo para las mujeres]
vb-ipvaw7	Men are violent toward their partners because women like it [Los hombres son violentos con sus parejas porque a las mujeres les gusta]
vb-ipvaw8	Women file false complaints to obtain economic benefits and hurt their partners [Las mujeres realizan denuncias falsas para conseguir beneficios económicos o herir a su pareja]
vb-ipvaw9	Men will change his violent behaviour towards his partner if she becomes more obedient [Los hombres cambiarían la conducta violenta hacia su pareja si ellas fueran más obedientes]
vb-ipvaw10	Women could avoid being beaten by their male partners if they knew when to stop talking

	[Las mujeres podrían evitar la violencia de sus parejas si ellas supieran cuando callarse]
vb-ipvaw11	If a woman is mistreated by her partner and does not leave him, that means that she does not dislike the situation
	[Si una mujer es maltratada por su pareja y no le deja, significa que no le desagrada la situación]
vb-ipvaw12	It is justified for a man to beat his partner if she decides to leave him
	[Está justificado que un hombre pegue a su pareja si esta decide dejarle]

vb-ipvaw: VB-IPVAW item.

VB-IPVAW scale scores: computing victim blaming attitudes estimates

Instead of using the raw sum of the items to compute the VB-IPVAW scale scores, we would recommend generating the victim blaming attitudes estimates following one of these two methods:

(1) Factor Scores: If the sample size is small (N < 100), conduct a weighted sum of the items by the factor loadings presented in Figure 1. In this manner the factor loading of each item is multiplied by the score for each item before summing. Alternatively, if the sample size is larger, conduct a new factor analysis replicating the one-factor model and compute the factor scores for whole sample.

(2) IRT Scores: to obtain the person parameter estimates (i.e., θ) for each respondent, estimate an IRT model by either fixing the item parameters to the values presented in Table 3 (for small sample sizes), or re-estimating the item parameters for the new sample (for larger sample sizes). To this end, we provide an R script with the code to compute the VB-IPVAW scale IRT scores using the *mirt* library.

Appendix 2. R script for computing victim blaming attitudes estimates.

```
##################################
#### AIPVAW IRT Scores Script ####
##################################

library(mirt)

VBIPVAW_data <- read.table("data path and format")
# insert path and extension of the data (e.g., "C:/Users/Documents/R/VB-IPV
AW/my_data.dat")
# my_data must be a matrix or data.frame with respondents on the rows and
items on the columns

# IRT model with fixed items parameters (small sample sizes):

a_VBIPVAW <- c(2.05, 3.68, 2.39, 3.43, 3.72, 2.49,
               3.36, 0.95, 2.49, 3.16, 1.93, 3.05)
d1_VBIPVAW <- c(-1.69, -3.81, -3.21, -3.94, -4.77, -3.09,
               -4.71, 0.17, -2.90, -3.93, -2.00, -4.86)
d2_VBIPVAW <-c(-4.26, -8.50, -5.57, -8.45, -9.54, -6.36,
               -8.85, -2.10, -5.72, -7.26, -5.17, -8.38)
d3_VBIPVAW <-c(-7.12, -11.06, -7.12, -9.97, -11.25, -7.97,
               -10.47, -4.18, -7.55, -8.98, -6.68, -10.03)

VBIPVAW_param <- mirt(VBIPVAW_data, 1, itemtype = "graded", pars =
"values")
VBIPVAW_param$est <- FALSE
VBIPVAW_param$value[VBIPVAW_param$name=="a1"] <- a_VBIPVAW
VBIPVAW_param$value[VBIPVAW_param$name=="d1"] <- d1_VBIPVAW
VBIPVAW_param$value[VBIPVAW_param$name=="d2"] <- d2_VBIPVAW
VBIPVAW_param$value[VBIPVAW_param$name=="d3"] <- d3_VBIPVAW

VBIPVAW_IRT <- mirt(VBIPVAW_data, 1, itemtype = "graded", method = "MHRM",
pars = VBIPVAW_param)
IRTScores <- fscores(VBIPVAW_IRT, method = "EAP", full.scores = T)

write.table(IRTScores, "IRTScores.dat", col.names = FALSE, row.names =
FALSE)
# return a .dat file with the IRT Scores for each respondent

# IRT graded model with free parameters (large sample sizes):

VBIPVAW_IRT <- mirt(VBIPVAW_data, 1, itemtype = "graded", method = "MHRM")
IRTScores <- fscores(VBIPVAW_IRT, method = "EAP", full.scores = T)
write.table(IRTScores, "IRTScores.dat", col.names = FALSE, row.names =
FALSE)
```

Study 6

The Willingness to Intervene in Cases of Intimate Partner Violence Against Women (WI-IPVAW) Scale: Development and Validation of the Long and Short Versions

The Willingness to Intervene in Cases of Intimate Partner Violence Against Women (WI-IPVAW) Scale: Development and Validation of the Long and Short Versions[*]

Enrique Gracia, Manuel Martín-Fernández, Miriam Marco, Faraj A. Santirso, Viviana Vargas, & Marisol Lila
Department of Social Psychology, University of Valencia, Valencia, Spain

Abstract

The willingness to intervene when one becomes aware of a case of intimate partner violence against women (IPVAW) reflects the level of tolerance and acceptance of this type of violence. Increasing the likelihood that people will intervene to help victims of IPVAW is also a target for prevention strategies aiming to increase the informal social control of IPVAW. In this study, we present the development and validation of the Willingness to Intervene in Cases of Intimate Partner Violence (WI-IPVAW) scale. We present data for both the long and short versions of the scale. We analyzed the latent structure, the reliability and validity of the WI-IPVAW across 4 samples (N = 1648). Exploratory and confirmatory factor analyses supported a bifactor model with a general non-specific factor expressing the willingness to intervene in cases of IPVAW, and three specific factors reflecting different intervention preferences: a preference for setting the law enforcement in motion ('calling the cops' factor), a preference for personal intervention ('personal involvement' factor), and a preference for non-intervention ('not my business'

factor). Configural, metric, and partial scalar invariance across genders was supported. Two short versions of the scale of nine and six items were constructed based on quantitative and qualitative criteria. The long and short versions of the WI-IPVAW demonstrated high reliability, as well as construct validity as they were strongly related to the acceptability of IPVAW, victim blaming attitudes, perceived severity of IPVAW, and hostile sexism. These results provide strong support for the reliability and validity of both the long and short versions of the WI-IPVAW scale. The long and the two short versions are psychometrically sound instruments to analyze the willingness to intervene in cases of IPVAW in different settings and with different research needs (e.g., long versions for clinical and research settings, and short versions for large population surveys). The WI-IPVAW is also a useful instrument to assess prevention policies and public education campaigns aiming to promote a more responsive social environment in cases of IPVAW, thus contributing to the deterrence and reduction of this major social and public health problem.

The World Health Organization defined intimate partner against women (IPVAW) as a "global public health problem of epidemic proportions" (WHO, 2013, p. 7). IPVAW has profound consequences not only for the physical and psychological health of their victims, but also for the well-being of their children, and the wider society (e.g., Campbell 2002, Devries et al., 2011; Ellsberg et al., 2008; Guedes et al., 2016; WHO, 2013). IPVAW is considered the most common form of violence suffered by women (Devries et al., 2013; Garcia-Moreno et al., 2006; Stöckl et al., 2013). In high-income countries, the estimated prevalence of IPVAW is 23.2%, and the percentage of IPVAW homicides 41.2% (WHO, 2013). In Europe, a survey among the 28 European Union (EU) Member States estimated that an average of 22% of European women had been victims of physical and/or sexual violence by their partners since the age of 15, with a lifetime prevalence across countries ranging from 13% to 32% (European Union Agency for Fundamental Rights, 2014). In Spain, where this study was conducted, according to different sources, IPVAW lifetime prevalence is around 13%, among the lowest in the EU (European Union Agency for Fundamental Rights, 2014; Gracia & Merlo, 2016; Ministerio de Sanidad, Servicios Sociales e Igualdad, 2015; Vives-Cases et al., 2011).

Public attitudes towards IPVAW, shape the social context in which IPVAW takes place, and play an important role in perpetuating the levels of this type of violence in our societies (Carlson y Worden, 2005; Copp, et al., 2016; Flood & Pease, 2009; Gracia et al., 2018; Powell & Webster; 2018; Stith et al., 2004; Waltermaurer, 2012). The public willingness to intervene when one becomes aware of a case of IPVAW reflects the level of tolerance and acceptance of this type of violence, and can contribute either to deter or facilitate this type of violence (Browning, 2002; Emery et al., 2011; Gracia & Herrero, 2006; Jewkes et al., 2015; WHO, 2013; Wright & Benson, 2010). In the

current study, we set out to develop a scale measuring public willingness to intervene in cases of IPVAW.

A reason for the study of the willingness to act in cases of IPVAW is that, despite being still a largely unreported offense, IPVAW is, at the same time, widely known by the social environment surrounding the victims (Gracia, 2004, Taylor & Sorenson, 2005; Taylor et al., 2016). For example, in a survey across the 28 European Union member states, nearly 23% of respondents reported to know a woman among their family members or friends who had been victim of IPVAW, 17% reported to know women in their immediate in their area or neighborhood, and 9% knew a woman where they worked or study (European Commission, 2016). Those who are aware of IPVAW incidents are in a position of doing something to help the victims and stop de violence (e.g., offering help, taking personal action, or setting the law in motion), but also they can choose not to get involved, to ignore the situation, and do nothing (Banyard & Moynihan, 2011; Taylor at al., 2016). Therefore, whether or not those who are aware of this violence are willing to intervene, is a matter of no small importance.

Attitudes of non-intervention among the social circle surrounding the victims may facilitate or reinforce the perpetrators' behavior, but also may inhibit victims' disclosure, making more difficult to seek help and escape the violence. On the other hand, pro-intervention attitudes (e.g., reporting to the authorities or direct intervention) among those aware of this violence, can have a protective effect for victims, and may inhibit or deter IPVAW, by increasing the social and legal costs for perpetrators (Gracia, 2014; Koepsell et al., 2006; McDonnell et al., 2011; Voith, 2017). Whether or not those who are aware of IPVAW incidents are willing to intervene is also particularly relevant if we take into account that victims tend to seek help among informal sources of help (friends, family, neighbors coworkers, etc.) rather than from formal

sources such as the police (Ansara & Hindin, 2010; Liang et al., 2015; McDonnell et al. 2011; McCart & Smith, 2010; Wee et al., 2016). Also, pro-intervention attitudes among these potential informal sources of help, when shared collectively, can contribute to shape local social norms that can help deter this type of violence (Powell & Webster; 2018; Voith, 2017; Wee et al., 2016). As Voith (2017) noted in her review "the protective effects of pro-IPV-intervention norms in a community are 2-fold, in that community members will directly intervene if they witness IPV and perpetrators are less likely to continue the use of violence against their partners as a result of social pressure" (p. 4).

Another reason to study and appropriately measure public willingness to act in cases of IPVAW is that indications are that non-intervention attitudes are still quite prevalent. In this regard, a report on attitudes toward violence against women in the EU (Gracia & Lila, 2015) suggested that attitudes favoring non-intervention were still prevalent. For example, data available from surveys in different countries, indicated that a sizable number of respondents preferred no to get involved even if they were aware of a case of violence against women ("not my business", or "is a private matter" were among the reason for not intervening). For instance, across the EU (Eurobarometer, 2016), the most common reason for those who knew victims of domestic violence but did not speak about it to anyone was that it was none of their business (26%). 'Lack of proof' (18%), 'not wanting to create trouble' (16%), 'concerned about negative consequences or retaliation' (11%), 'did not know who to speak to' (8%), and 'it was not serious enough' (6%), were among other reasons. In Spain, where the present study was conducted, most of the officially reported cases of IPVAW are made by the victims, and only around 4% of the reports to the authorities are made by family members or other third parties (Consejo General del Poder Judicial, 2016). In this regard, increasing

the likelihood that people will intervene to help victims of IPVAW becomes a target for prevention strategies aiming to translate public awareness of this social problem into a greater sense of personal responsibility and involvement, thus contributing to the informal social control of IPVAW (Gracia et al., 2009).

Present Study

Drawing from the above, it becomes apparent the need to advance our knowledge on the public willingness to intervene in cases of IPVAW and related key issues such as the prevalence of pro- or non-intervention attitudes, intervention preferences, its correlates or determinants, or the assessment of the effectiveness of interventions targeting these attitudes. The availability of reliable and valid instruments measuring public willingness to intervene in cases of IPVAW is key to conduct this type of research. Although some measurement instruments have been developed to examine willingness to help in cases of violence, most of this research has been conducted in the context of bystander intervention behavior in cases of dating violence, and sexual harassment or rape situations (Banyard, 2008; Banyard & Moynihan, 2011; Banyard, Moynihan, Cares, & Warner, 2014; Branch, Richard, & Dretsch, 2013; McMahon et al 2014; Stein, 2007). Other studies assessing willingness to intervene have limited generalizability as they use small non-community samples (e.g., college students), and others instruments report low reliabilities (Baldry & Pagliaro, 2014; Baldry et al., 2015; Cinquegrana et al., 2018). Also, large population survey data on public attitudes towards intervention in cases of IPVAW are not usually based on measurement instruments with adequate reliability and validity, or are based on single items (Gracia & Lila, 2015). Clearly, there is still a need for psychometrically sound instruments measuring willingness to intervene in cases of IPVAW, appropriate for use with community samples, and suitable for large-scale surveys.

In this study, we present the development and validation of the Willingness to Intervene in Cases of Intimate Partner Violence (WI-IPVAW) scale. We present data for both the long and short versions of the scale, which allows to analyze the willingness to intervene in cases of IPVAW in different settings and with different research needs (e.g., long versions for clinical and research settings, and short versions for large population surveys). By using advanced statistical analyses, we will address important issues such as social desirability, measurement invariance, and will ensure that the shortened versions of the WI-IPVAW scale retain high quality psychometric properties. For validity purposes, we will explore the relationship between the long and short versions of the WI-IPVAW scale and other relevant constructs regarding attitudes towards IPVAW such as the acceptability of IPVAW, victim blaming attitudes, perceived severity of IPVAW, and hostile sexism (Flood & Pease, 2009; Gracia & Herrero, 2006; Gracia, 2014; Herrero, Rodríguez, & Torres, 2017; Lila, Gracia, & García, 2013; Martín-Fernández et al., 2018; Taylor & Sorenson, 2005). Gender, age and education differences in willingness to intervene in cases of IPVAW will be also explored (Carlson & Worden, 2005; Fincham et al., 2008; Flood & Pease, 2009; Gracia et al., 2009, 2015).

Method

Participants

Four samples were recruited for the current study. The first one was an incidental sample used to conduct a pilot study. It was composed by 148 Valencia University undergraduates who participated for course credit (31 males and 117 females), aged 19-32 years old (M = 21.29; SD = 2.60). The second, third, and fourth samples were recruited through online sampling. Online sampling is an effective and cost-efficient sampling method (Thornton et al., 2016; Topolovec-Vranic & Natarajan, 2016). A total pool of 2698

Table 1. Socio-demographics

	Sample 1	Sample 2	Sample 3	Sample 4
Gender				
Male	117	231	510	94
Female	31	269	490	106
Age				
18-24	131	214	243	108
25-34	14	83	311	30
35-54	2	141	347	53
55+	1	62	99	9
Nationality				
Spanish	128	429	869	186
Inmigrant	20	61	131	14
Education				
Compulsory	0	65	143	25
Upper Secondary	0	88	191	38
Undergraduate	135	190	321	89
Postgraduate	13	157	345	48

responses was collected. We equilibrated these samples by gender and removed those participants who omitted socio-demographic information. The socio-demographic characteristics of the samples are shown in Table 1.

The second sample consisted of 500 participants (231 males and 269 females), aged 18-80 (M = 33.83; SD = 14.77), and was used to study the psychometric properties of the scale. The third sample consisted of 1000 participants (490 males and 510 females), aged 18-82 (M = 35.40; SD = 13.46). This sample was used to test different levels of measurement invariance and to conduct the criterion-related validity analyses. The fourth sample consisted

of 200 participants (94 males and 106 females), aged 18-71 (M = 29.39; SD = 11.82), and was used to assemble two short versions of the scale.

Instruments

The *Willingness to Intervene in cases of IPVAW* (WI-IPVAW). The development of the WI-IPVAW was based on an initial pool of 96 items. The development of these pool of items was based on a review of European surveys addressing attitudes towards intervention in cases of violence against women (Gracia & Lila, 2015), and in previous research addressing public attitudes, and response preferences in cases of IPVAW (Gracia & Herrero, 2006; Gracia et al, 2009). The item development and selection process was also informed by literature pointing to the different scenarios other than the privacy of the home were IPVAW also takes place, and it is witnessed by third parties (Banyard & Moynihan, 2011; Hamby et al., 2015; Taylor et al., 2016). This initial pool of items presented hypothetical scenarios describing IPVAW situations, taking place in different places, and that could be witnessed by the respondent, or disclosed to him/her by the victim (e.g., next door apartment, staircase or common areas in buildings, street, shops, bars, etc.). These scenarios included different types of IPVAW behaviors (e.g., physical aggression, insults, threats, violent arguments, fights, etc.), and different types of potential responses or involvement (i.e., calling the police, scold or reprehend the aggressor, protecting the woman victim, ignoring the situation, doing nothing, etc.). After the initial pool of items was developed, they were reviewed by a panel of six experts on IPVAW to establish construct representativeness, and clarity (Beck & Gable, 2001; Delgado-Rico, et al., 2012). The experts had to rate the representativeness (i.e., whether the item is suitable to measure the willingness to intervene in cases of IPVAW), and the clarity (i.e., the extent to which the item is concise) of the items on a 7-point Likert-type scale (1 = "Very unrepresentative/unclear"; 7 = "Very representative/clear"). An item was

considered representative and/or clear if the average score in the expert ratings were above 5 on a Likert-type scale of seven points (i.e., the "somewhat representative/clear" category).

After this review, 31 items were selected. Respondents had to rate their perceived likelihood to intervene in the hypothetical scenario described in each item on a 6-point Likert-type scale (1= "Not at all likely", 6 = "Extremely likely"). The final version of the WI-IPVAW scale is shown in Appendix 1.

Acceptability of IPVAW (A-IPVAW; Martín-Fernández, et al., 2018a,b). The short form of the A-IPVAW scale was used in this study. This instrument is composed of eight items tapping attitudes of acceptability of IPVAW (e.g., It is acceptable for a man "to shout his partner if she is continuously arguing and nagging him"). Respondents had to rate on a 3-point Likert-type scale how acceptable where the men's behaviors against their female partners (0 = "Not acceptable", 1 = "Somewhat acceptable", 2 = "Acceptable"). The A-IPVAW scale was cross-validated in the general Spanish population, and also with IPVAW male offenders. The scale used in this study has showed both internal and external validity, as it has been related to perceived severity of IPVAW and ambivalent sexism (Martín-Fernández, et al., 2018b). This instrument showed a fair internal consistency across Samples 2, 3, and 4 (Cronbach's α = .75, .72, .68, respectively).

Victim Blaming attitudes towards IPVAW (VB-IPVAW; Martín-Fernández, et al., 2017). This instrument is composed of five items assessing the tendency to blame the victims of IPVAW (e.g., "A man will change his behavior towards her partner if she becomes more obedient"). Respondents had to rate their level of agreement with each statement in a 4-point Likert-type scale (1 = "Strongly disagree", 4 = "Strongly agree"). This instrument has demostrated validity evidence based on its relationships with other variables such as the acceptability and perceived severity of IPVAW, and ambivalent

sexism (Martín-Fernández et al., 2018b). It also presented a high internal consistency in Samples 2, 3, and 4 (Cronbach's α = .81, .84, .83, respectively).

Perceived Severity of IPVAW (PS-IPVAW; Gracia, et al., 2009, 2011). This scale presents eight IPVAW scenarios (e.g., "During an argument, a man hits his partner and afterwards he asks for her forgiveness) and respondents had to assess the severity of each scenario (ranging from 1, "Not severe at all", to 10, "Extremely severe"). The PS-IPVAW scale had been validated in the general Spanish population, and with police officers and IPVAW male offenders as well, presenting adequate psychometric properties. It also has been related to sexism, empathy, personal responsibility, and IPVAW victim blaming attitudes (Gracia, et al., 2009; Gracia & Tomás, 2014; Lila, et al., 2013; Vargas et al., 2015). The scale showed a good internal consistency in Samples 2, 3, and 4 (Cronbach's α = .83, .85, .87, respectively).

Ambivalent Sexism Inventory Short Version (ASI; Glick & Fiske, 1996; Rollero, et al., 2014). The reduced hostile sexism subscale was used for the current study, composed of six items assessing attitudes of prejudice and discrimination against women based on the assumption of women's inferiority and their differences from men (e.g., "Women seek to gain power by getting control over men"). The Spanish version of the items was used (Expósito, Moya, & Glick, 1998). The complete ambivalent sexism inventory has been validated in more than twenty countries (Glick et al, 2000; Glick, Sakall, Urgurlu, Ferreira, & Aguilar de Souza, 2002), and the hostile sexism subscale has showed strong relationships with attitudes toward intervention in IPVAW cases among police officers, IPVAW responsibility attribution, and to acceptability of IPVAW (Lila et al., 2013; Lila, Oliver, Catalá-Miñana et al., 2014; Martín-Fernández, et.al, 2018a). It presented a good internal consistency in Samples 2, 3, and 4 (Cronbach's α =.89, .88, .87, respectively).

Balanced Inventory of Desirable Responding short form (BIDR-16; Hart et al., 2015). The Impression Management subscale was used for the pilot study. This subscale is composed of eight items evaluating the tendency of participants to provide overestimated self-descriptions to create a socially desirable image. (e.g., "I never cover up mistakes"), and presented a moderate reliability on the first sample (Cronbach's α = .68).

Procedure

Two online forms were constructed to collect the data. The first form included the WI-IPVAW, the BIDR items of the Impression Management subscale, and a set of socio-demographical questions (i.e., gender, age, nationality, and study level). This form was used only for Sample 1. Participants were informed about the objectives of the study and gave their informed consent. The second form included the WI-IPVAW, the PS-IPAVAW, and the short forms of the A-IPVAW, VB-IPVAW, and the Hostile Sexism subscale form the ASI, and the same socio-demographical questions. After the participants gave their informed consent, they completed the online form. Participants received no payment. The data were collected from October 2017 to December 2017.

Data Analysis

A pilot study was conducted first using the sample of college students (Sample 1) in order to explore the psychometric properties of the WI-IPVAW and the effect of social desirability on the items. To this end, the descriptive statistics and the item-test corrected correlations were computed, and the internal consistency of the scale was evaluated by means of Cronbach's α. The latent structure of the scale was also assessed carrying out an exploratory factor analysis (EFA). Before conducting the EFA, the suitability of the data matrix was tested, computing the Bartlett's sphericity test and the KMO statistic. To determine the number of factors to extract, a parallel analysis

based on minimum rank factor analysis was conducted (Timmerman & Lorenzo-Seva, 2011). An EFA was then performed using the polychoric correlation matrix and weighted least squared means and variances adjusted estimation method (WLSMV), as this procedure is especially recommended for categorical data (Asparouhov & Muthén, 2010; Muthén & Kaplan, 1985, 1992). The fit of the model was assessed using the CFI, TLI, SRMR and RMSEA fit indices. CFI and TLI values ≥ .95 are indicative of very good fit, and values between .90-.95 indicate minimally acceptable model fit (Bentler, 1995; Hu & Bentler, 1999). RMSEA values ≤ .06, and ≤.08, indicate very good and acceptable fit, respectively, and SRMR values ≤ .08 are considered for well-fitting models (MacCallum et al., 1996). Once the latent structure of the scale was established, the social desirability of each item was evaluated. To do so, a confirmatory factor analysis (CFA) was conducted adding a social desirability factor to the EFA model. All the items of the BIDR and the WA-IPVAW scale were constrained to load onto this social desirability factor, using the BIDR items as social desirability markers (Ferrando, 2005; 2008). To make the model identifiable, the loadings of the BIDR were fixed to the same value. If a WI-IPVAW item loading on the social desirability factor is greater than the BIDR loadings, then we considered that the item is biased by social desirability. Those items were removed from the scale.

A larger sample (Sample 2) was used to study further the psychometric properties of the WI-IPVAW scale and to cross-validate the factorial model. The descriptive statistics, the item-test correlations, and the Cronbach's α were again computed. A CFA was carried out using the WLSMV estimation method. Several nested models were compared. Model fit was evaluated by using the same combination of fit indices and the same cut-offs.

Measurement invariance across genders was as well evaluated in an independent sample (Sample 3). To this end, several levels of group invariance

were tested, conducting and comparing a series of multi-group CFAs. Configural, metric, scalar and strict invariance models were estimated using the WLSMV estimation method (Milfont & Fisher, 2010). Configural invariance tests whether men and women conceptualize the construct in the same manner, estimating the same factorial model for each group and allowing the structural parameters (i.e., loadings, thresholds, and item variances) to vary across groups. The metric invariance model constraints the item loadings to have the same value for both groups, testing if men and women interpret the items in the same manner. The scalar invariance model fixes the threshold parameters to the same value across groups, establishing whether the latent construct yields the same score in the items for men and women. The strict invariance model assesses whether the measurement error is equal in each group, constraining the variances of the observed variables (i.e., the items) to have the same values across groups. The models where compared following the guidelines of Cheung and Rensvold (2002), computing the change in CFI (ΔCFI) and RMSEA (ΔRMSEA) to test which of the invariance models is better supported by the data. A change in the CFI (ΔCFI) and in the RMSEA (ΔRMSEA) ≤ .010 and ≤ .015, respectively, support the more restrictive model (i.e., being the configural model the most flexible model and the strict invariance the most restrictive). However, these criteria were proposed for models estimated with maximum likelihood estimation for continuous variables and, given that we are using weighted least squares estimation for categorical data, we decided to also apply a corrected chi-square difference testing (DIFFTEST; Asparouhov et al., 2006). If the fit indices comparisons and the DIFFTEST yield a similar result, then that invariance levels is assumed.

The validity of the scale was assessed comparing it with related variables, such as acceptability of IPVAW, attitudes of victim blaming in cases of IPVAW, perceived severity of IPVAW, and hostile sexism. Socio-

demographic comparisons were also made, testing differences across gender, age, and study levels groups.

Finally, two short versions of the WI-IPVAW scale of nine and five items were assembled following the Goetz et al. (2013) recommendations. First, the most relevant items were selected attending to the internal consistency, the previous factorial models, and to the assessments of the panel of experts. The psychometric properties of the shortened scales were then studied and compared with the original WI-IPVAW scale using a different sample (Sample 4).

All analyses were computed using the statistical package R (R Core Team, 2017) and the *psych* library (Revelle, 2016). EFA, CFA, and multi-group CFAs where conducted with the *MPlus 7.1* package (Muthén & Muthén, 2010).

Results

Pilot Study: factor structure and social desirability

The psychometric properties, the latent structure and the effect of social desirability on the WI-IPVAW items were explored conducting a pilot study with the Sample 1. Descriptive statistics revealed that most of the items were slightly displaced to the right, with means around 3-5 (e.g., "somewhat likely", "quite likely", "very likely"), and moderate negative skew (around -0.50). This means that the participants tended to select the upper categories of the scale. The overall internal consistency of the scale was very high (Cronbach's α = .93), showing a strong relation between the score on the scale and the items, with item-test corrected correlations around .50. Deleting any item did not improve the scale internal consistency.

Before conducting an EFA, the suitability of the matrix for factor analysis was tested. The Bartlett's sphericity test resulted significant (χ^2 = 2505.8, *df* = 465, *p* < .001) and the Kaiser-Meyer-Olkin statistic was good (KMO

= .88), indicating that the data are adequate for an EFA. The parallel analysis based on minimum rank factor analysis using the polychoric correlation matrix revealed that three factors should be extracted, since adding more factors did not contribute to explain more variance in our data than in a random dataset. A three factor model was thus estimated using WLSMV with the oblique OBLIMIN rotation. The model converged normally, and showed an acceptable fit (χ^2 = 2505.8, df = 465; CFI = .94; TLI = .92; RMSEA = .068; SRMR = .069). Although the CFI and the TLI were below the .95 cutt-off, they are not below .90, and the RMSEA and SRMR suggested that the model was well-fitted. The items were grouped in three factors. The first factor groups all the items that involve setting the law in motion by calling or reporting to the police the IPVAW incident (i.e., "calling the cops" factor), the second factor groups all items implying ignoring the situation or doing nothing (i.e., "not my business factor"), and the third factor groups all items in which the respondents personally intervene to stop the situation (i.e., "personal involvement" factor). All the items presented factor loadings above .30 in their factor, and only three items presented cross-loadings in more than one factor. In these three cases the loading on the main factor is above .50 and near to .30 in the secondary factor, indicating that the items are more related to the main factor (i.e., "personal involvement" factor on the first case, and "calling the cops" factor in the other two cases). The correlation between the "calling the cops" and the "personal involvement" factors was positive (r = .29), whereas the correlations between the "not my business" factor and the "calling the cops" and the "personal involvement" factors were negative (r = -.55 and r = -.28, respectively).

In order to test the extent in which the items of the scale were affected by social desirability bias a CFA was conducted. The CFA model posited the three previous content factors (i.e., "calling the cops", "not my business",

Table 2. Confirmatory factor analysis with social desirability markers (Sample 1)

	Calling the cops		No my business		Personal involvement		Social Desirability	
Item 1			0.63	(.06)			-0.32	(.08)
Item 2					0.63	(.05)	0.20	(.08)
Item 3					0.68	(.05)	0.34	(.09)
Item 4	0.70	(.06)					0.01	(.12)
Item 5	0.82	(.06)					-0.21	(.13)
Item 6					0.76	(.04)	0.01	(.08)
Item 7			0.68	(.05)			-0.29	(.08)
Item 8	0.69	(.06)					-0.17	(.10)
Item 9					0.77	(.04)	0.20	(.09)
Item 10	0.64	(.05)					0.15	(.09)
Item 11					0.78	(.05)	0.33	(.10)
Item 12					0.82	(.04)	0.19	(.09)
Item 13	0.79	(.05)					-0.08	(.11)
Item 14					0.53	(.06)	0.24	(.08)
Item 15			0.82	(.04)			-0.15	(.09)
Item 16			0.67	(.05)			-0.25	(.08)
Item 17	0.88	(.06)					-0.28	(.15)
Item 18	0.66	(.06)					0.35	(.08)
Item 19					0.73	(.04)	0.28	(.08)
Item 20					0.71	(.05)	0.01	(.08)
Item 21	0.81	(.04)					-0.15	(.11)
Item 22	0.66	(.05)					0.31	(.09)
Item 23					0.77	(.04)	0.21	(.09)
Item 24			0.61	(.06)			-0.29	(.08)
Item 25			0.68	(.05)			-0.01	(.09)
Item 26	0.81	(.04)					-0.17	(.11)
Item 27			0.81	(.04)			-0.15	(.09)
Item 28	0.55	(.06)					0.32	(.09)
Item 29	0.65	(.07)					**0.41**	(.09)
Item 30	0.47	(.07)					**0.38**	(.08)
Item 31	0.46	(.07)					**0.47**	(.08)
BIDR1-8							0.37	(.03)

Note: each cell contains the factor loadings (s.e. in brackets). BIDR 1-8: items from the impression management subscale of the Balanced Inventory of Desirable Response Short Form.

"personal involvement") and a new factor of social desirability. The content factors were allowed to correlate between them, whereas the social desirability factor is not correlated with any content factor. The WI-IPVAW items loaded on their main factor and also on the social desirability factor. The BIDR items were used as social desirability markers and only loaded on the social desirability factor. In addition, the BIDR items were constrained to have the same factor loadings on this factor. The model was estimated using WLSMV, converged normally, and showed and adequate fit (χ^2 = 1130, df = 837; CFI = .93; TLI = .92; RMSEA = .049). The factor loadings are shown on table 2.

The items 8, 26, and 27 (e.g., "If a man insults his partner in the street, I will say him something to call its attention"; "If a man strongly grabs his partner's arm on the street, forcing her to go with him, I will call the police; "If a newcomer couple of my building argues and yell constantly, I will immediately call the police") presented factor loadings on the social desirability factor higher than the markers (λ = .37), and thus were removed from the scale. Ferrando (2005) recommend removing those items that present factor loadings above |.30|; however we decided to apply a more conservative criterion (i.e., removing only items that had factor loadings above the markers loadings on the social desirability factor), since the internal consistency of the BIDR was moderate in the pilot study.

Descriptive Analyses and reliability

Sample 2 was used to assess the psychometric properties of the scale. Descriptive statistics and item-test corrected correlations can be found on Table 3. The descriptive statistics were in the same line as in the pilot study, with items slightly displaced to the right. The item means were around 4, with a standard deviation around 1, meaning that the respondents tended to endorse de upper intermediate categories (e.g., "somewhat likely", "quite

Table 3. Descriptive statistics of the WI-IPVAW items (Sample 2)

	M	*SD*	*Min*	*Max*	*Skew*	*Kurtosis*	*s.e.*	$r_{item\text{-}test}$
Item 1	3.07	1.31	1	6	0.28	-0.55	0.06	0.47
Item 2	4.10	1.40	1	6	-0.36	-0.78	0.06	0.63
Item 3	3.52	1.44	1	6	0.04	-0.97	0.06	0.65
Item 4	5.56	0.89	1	6	-2.54	7.18	0.04	0.46
Item 5	5.70	0.77	1	6	-3.48	14.31	0.03	0.43
Item 6	4.11	1.55	1	6	-0.49	-0.84	0.07	0.54
Item 7	2.87	1.35	1	6	0.31	-0.70	0.06	0.50
Item 8	5.38	1.05	1	6	-1.94	3.66	0.05	0.56
Item 9	3.77	1.42	1	6	-0.13	-0.81	0.06	0.71
Item 10	4.90	1.33	1	6	-1.21	0.70	0.06	0.60
Item 11	3.20	1.55	1	6	0.21	-1.02	0.07	0.66
Item 12	4.15	1.43	1	6	-0.39	-0.75	0.06	0.66
Item 13	5.55	0.90	1	6	-2.40	6.11	0.04	0.50
Item 14	3.14	1.60	1	6	0.26	-1.02	0.07	0.58
Item 15	2.24	1.25	1	6	0.94	0.25	0.06	0.53
Item 16	3.14	1.50	1	6	0.27	-0.95	0.07	0.55
Item 17	5.66	0.82	1	6	-3.03	10.26	0.04	0.43
Item 18	5.03	1.34	1	6	-1.41	1.20	0.06	0.57
Item 19	3.59	1.55	1	6	-0.03	-1.09	0.07	0.62
Item 20	4.43	1.48	1	6	-0.71	-0.49	0.07	0.60
Item 21	5.44	1.05	1	6	-2.21	4.68	0.05	0.44
Item 22	4.67	1.50	1	6	-0.99	-0.05	0.07	0.68
Item 23	3.71	1.47	1	6	-0.16	-0.91	0.07	0.68
Item 24	2.90	1.46	1	6	0.46	-0.74	0.07	0.50
Item 25	2.71	1.48	1	6	0.54	-0.74	0.07	0.41
Item 26	5.42	1.03	1	6	-2.21	5.17	0.05	0.51
Item 27	2.83	1.48	1	6	0.52	-0.68	0.07	0.50
Item 28	4.50	1.55	1	6	-0.78	-0.49	0.07	0.63

Note: *M*: mean. *SD*: standard deviation. *Min*: minimum. *Max*: maximum. *s.e.:* standard error for the Skew and Kurtosis statistics. $r_{item\text{-}test}$: item-test corrected correlation.

likely", "very likely"). The skew statistics were moderate and negative for many of the items, and some of them also presented high values of kurtosis,

indicating that the items were not normally distributed. The item-test corrected correlations presented values above .40, indicating a strong relationship between the items and the total score of the scale. The overall internal consistency of the scale was again very good (Cronbach's α = .94), and the internal consistency of each factor was also good (Cronbach's α = .88, .84, and .92 for the "calling the cops", "not my business", and "personal involvement" factors respectively).

Confirmatory Factor Analysis

Three models were estimated using the Sample 2 in order to test the factor structure of the WI-IPVAW. The first model was a one-factor model in which all items loaded onto a general factor of "willingness to intervene in cases of IPVAW". The second model was the three-factor model found on the pilot study, with three correlated factors differentiated by the responses to the scenarios described by the WI-IPVAW items (i.e., "calling the cops", "not my business", and "personal involvement"). The third model was a bifactor model with three specific factors reflecting different intervention preferences—as in the previous three-factor model— and a general, non-specific factor, of "willingness to intervene". This general factor accounts for all the shared elements of the specific factors. The specific factors account only for the core elements of their items, in this case the type of response to the scenarios described by the items. To this end, all the items loaded on their specific factor and also on the general factor. The factors are orthogonal, so they are not correlated. All models were estimated using WLSMV and the polychoric correlation matrix. All models converged normally.

The fit indices of the models are shown in Table 4. The one-factor model showed a poor fit to the data, presenting fit indices too far away from their cut offs. The three-factor model showed an acceptable RMSEA and a minimally acceptable CFI and TLI, which could be kept as the latent structure

of the scale. However, adding a general dimension of "willingness to intervene" to the model improved substantially the fit of the model to the data. Thus, we decided to retain the bifactor model.

The loadings of the bifactor model are displayed in Table 5. All the loadings for the specific factors were significant, with values above .30 in all the items except for the items 2 and 3, whose loadings were around .20. The loadings of the general factor were all significant with values above |.40|. Note that the "not my business" item loadings were negative in the general factor, pointing out that agreeing with these items yielded a lower score on the general "willingness to intervene" factor. Overall, the loadings in the general factor were higher than in the specific factor. Furthermore, the percentage of common explained variance of the general "willingness to intervene" factor was 56.85%, whereas the specific "calling the cops" factor explained the 23.16%, the "personal involvement" the 11.04%, and the "not my business" the 8.95% of the common explained variance.

Measurement Invariance

Once the bifactor model was kept as the latent structure of the scale, the measurement invariance of the scale was tested across genders using Sample 3.

Table 4. CFA fit Indices (Sample 2)

	χ^2	*df*	CFI	TLI	RMSEA
Model					
One-Factor	3658.43	350	0.79	0.77	.137[.133; .142]
Three-Factor	1264.39	347	0.94	0.93	.073[.068; .077]
Bifactor	1052.62	322	0.95	0.95	.067[.063; .072]

Note: CFI: comparative fit index. TLI: Tucker-Lewis index.
RMSEA: Root Mean Square Error of Approximation (95% CI in square brackets).

Table 5. CFA item loadings in the bifactor model (Sample 2)

	Calling the cops	No my business	Personal involvement	Willingness to Intervene
Item 1		.46 (.04)		-0.48 (.04)
Item 2			0.19 (.05)	0.73 (.03)
Item 3			0.20 (.06)	0.75 (.03)
Item 4	0.61 (.04)			0.47 (.05)
Item 5	0.65 (.04)			0.52 (.05)
Item 6			0.69 (.04)	0.50 (.05)
Item 7		0.41 (.05)		-0.53 (.04)
Item 8	0.59 (.04)			0.57 (.04)
Item 9			0.38 (.05)	0.75 (.03)
Item 10	0.49 (.04)			0.58 (.04)
Item 11			0.46 (.04)	0.67 (.03)
Item 12			0.41 (.04)	0.70 (.03)
Item 13	0.68 (.04)			0.53 (.05)
Item 14			0.46 (.04)	0.57 (.04)
Item 15		0.44 (.05)		-0.62 (.04)
Item 16		0.51 (.04)		-0.56 (.04)
Item 17	0.74 (.04)			0.45 (.05)
Item 18	0.48 (.04)			0.56 (.04)
Item 19			0.37 (.05)	0.65 (.04)
Item 20			0.60 (.04)	0.59 (.04)
Item 21	0.71 (.03)			0.41 (.05)
Item 22	0.47 (.03)			0.67 (.03)
Item 23			0.30 (.05)	0.73 (.03)
Item 24		0.53 (.04)		-0.50 (.04)
Item 25		0.44 (.04)		-0.46 (.04)
Item 26	0.65 (.03)			0.50 (.04)
Item 27		0.46 (.05)		-0.55 (.04)
Item 28	0.40 (.03)			0.62 (.03)

Note: each cell contains the factor loadings (s.e. in brackets).

244

Table 6. Measurment Invariance fit indices (Sample 3)

	χ^2	*df*	CFI	TLI	RMSEA
Configural Model	1881.65	594	.951	.943	.066 [.063; .069]
Metric Invariance Model	1194.39	648	.979	.978	.041 [.037; .045]
Scalar Invariance Model	1410.71	776	.976	.978	.040 [.037; .044]
Partial Scalar Invariance Model	1355.42	766	.978	.980	.039 [.036; .043]
Strict Invariance Model	1658.43	739	.965	.967	.050 [.047; .053]

Note: CFI: comparative fit index. TLI: Tucker-Lewis index. RMSEA: Root Mean Square Error of Approximation (95% CI in square brackets).

A stepwise approach was used, testing first the configural invariance, and then comparing it with the metric, scalar, and strict invariance models. The fit indices of the models and the model comparisons are showed in Table 6 and Table 7.

The configural model fitted well the data, meaning that men and women conceptualize the latent construct in the same manner, and was used as a base line for the model comparisons. Then it was compared with the metric invariance model, which constrained the factor loadings to be equivalent across groups, and we found that both CFI and RMSEA indices improved once the factor loadings were constrained. The DIFFTEST also showed that these improvement were marginally significant (p = .02). This is most likely due to the reduction in the number of parameters to estimate, making the model more parsimonious, and it is not an unusual phenomenon when conducting measurement invariance analysis with categorical data (e.g., Brummelman et al., 2015; Mejías et al., 2017). Given the improvement in model fit and the reduction in the number of parameters to estimate, the metric invariance was supported.

The scalar invariance model, which besides the factor loading also constrained the item thresholds to be equal across gender, was compared with the metric one. Although the reduction in the CFI and RMSEA fit indices were

Table 7. Measurement Invariance model comparisons (Sample 3)

	ΔCFI	ΔRMSEA	DIFFTEST	df	p
Configural Model					
Metric Invariance Model	-.028	.025	77.50	54	.020
Scalar Invariance Model	.003	.001	280.69	128	.000
Partial Scalar Invariance Model	-.002	.001	144.20	118	.051
Strict Invariance Model	.013	-.110	61.87	30	.001

Note: ΔCFI: change in CFI. ΔRMSEA: Change in RMESEA. DIFFTEST: Robust chi square difference testing, *df*: degree of freedom of the DIFTEST, *p:* p-value of the DIFTEST.

between the cut offs established by Cheung and Rensvold (2002), the DIFFTEST resulted significant ($p < .001$). The modification indices where then used to identify potential items to be unconstrained and test the partial scalar invariance. The thresholds of two items (item 6 and item 21) were allowed to vary across groups and we found that the partial invariance model did not differ from the metric one ($p = .051$). The partial scalar invariance model was thus supported.

Finally, the strict invariance was tested, constraining the item variances to be equal across groups and comparing it with the partial invariance model. We found that the CFI decreased below the ΔCFI = .01 cut off and the DIFFTEST was significant. Thus the strict invariance could not be supported.

Validity Analyses

The Sample 3 was also used to conduct validity analyses. The correlations of the WI-IPVAW factorial scores with other related constructs are showed in Table 8. The general factor "Willingness to Intervene" was negatively related to acceptability of IPVAW, attitudes of victim blaming, and hostile sexism, implying that those respondents with higher scores on this factor tend to present lower levels of attitudes of acceptability, tend to blame

Table 8. WI-IPVAW relations to other variables (sample 3)

	Acceptability	Victim blaming	Percieved severity	Hostile sexism
Calling the cops	-0.13*	-0.21*	0.23*	-0.15*
No my business	0.12*	0.11*	-0.11*	0.22*
Personal involvement	0.03	0.02	-0.12*	0.06
Willingness to Intervene	-0.23*	-0.19*	0.25*	-0.20*

Note:: *p* < .01.

less the victims of IPVAW, and show lower levels of sexist attitudes. On the other hand, the general factor was positively related with the perceived severity of IPVAW (those with higher scores on willingness to intervene tend to perceive IPVAW situations as more severe). Regarding the specific factors, the "calling the cops" factor showed a similar relation with these variables, although they were more moderated, whereas the "not my business" factor presented the opposite tendency: it was positively related with acceptability of IPVAW, attitudes of victim blaming, and hostile sexism, and negatively related with perceived severity of IPVAW. The "personal involvement" factor only presented a significant and negative relation to perceived severity.

A series of ANOVA were conducted with each factor to test differences across gender, age, and study level using the factor scores of the partial scalar invariance model. Regarding the general factor "willingness to intervene", significant differences were found between genders, $F(1) = 23.53$, $p < .001$, $\eta^2 = .023$, with a small effect size, presenting women higher values on this factor than men; marginal differences between age groups, $F(3) = 3.09$, $p = .026$, $\eta^2 = .009$; and no differences regarding the study level, $F(3) = 1.30$, $p = .274$, $\eta^2 = .004$. The effect sizes of age and study levels were considered negligible, since they were below the .01 cut-off for small size effects (Miles & Shevlin, 2001).

Significant differences were also found in the specific "calling the cops" factor by gender, $F(1) = 21.24$, $p < .001$, $\eta^2 = .021$, and age, $F(3) = 3.73$, $p = .011$,

$\eta^2 = .011$, both with a small effect size. Women showed higher scores on this factor than men, and so did the respondents of the upper age categories (i.e., "35-54" and "55+") in comparison with the lower category (i.e., "18-24"). The study level had not significant effect on this factor, $F(3) = 0.89$, $p = .444$, $\eta^2 = .002$. We found significant differences for the specific factor "not my business" by gender, $F(1) = 5.45$, $p = .020$, $\eta^2 = .005$, although the effect size was considered negligible. No differences were found in this factor regarding age, $F(3) = 2.27$, $p = .079$, $\eta^2 = .006$, or study level, $F(3) = 2.27$, $p = .079$, $\eta^2 = .002$.

Regarding the specific factor "personal involvement", significant differences were again found between genders, $F(1) = 85.00$, $p < .001$, $\eta^2 = .079$, with a medium effect size, and age groups, $F(3) = 5.08$, $p = .002$, $\eta^2 = .015$, with a small effect size. Men showed higher scores on this factor than women, and respondents of the upper age categories (i.e., "35-54" and "55+") presented higher scores than respondents of the lower age categories (i.e., "18-24" and "25-34"). Again, the study level showed no effect on this factor, $F(3) = 1.53$, $p = .197$, $\eta^2 = .005$.

WI-IPVAW Shortened forms

A combination of quantitative and qualitative criteria was used to decide which items should compose the shortened versions of the scale (see Table 9). We included those items that presented low loadings (i.e., below .20) on the social desirability factor used on the pilot study, with medium or high loadings (i.e., between .20-.50, and above .50, respectively) on their specific and general factor, and that were invariant across genders. In addition to these criteria, the assessment of the panel of experts regarding the representativeness and clarity of each item was also used.

Nine-item version of the WI-IPVAW scale. To ensure the content coverage of each specific factor (Smith et al, 2000), three items of each specific factor were selected to compose a 9-item version of the WI-IPVAW scale. In

Table 9. Criteria to reduce the WI-IPVAW

	Specific factor	SD factor loading	Specific factor loading	General factor loading	Invariant across gender	Expert Ratings
Item 1	Not my business	Medium	Medium	Medium	Yes	CR
Item 2	Personal involvement	Low	Low	High	Yes	CR
Item 3	Personal involvement	Medium	Medium	High	Yes	CR
Item 4	Calling the cops	Low	High	Medium	Yes	
Item 5	Calling the cops	Low	High	High	Yes	R
Item 6	Personal involvement	Low	High	Medium	No	C
Item 7	Not my business	Low	Medium	High		CR
Item 8	Calling the cops	Low	High	High	Yes	CR
Item 9	Personal involvement	Low	Medium	High	Yes	CR
Item 10	Calling the cops	Low	Medium	High	Yes	CR
Item 11	Personal involvement	Medium	Medium	High	Yes	CR
Item 12	Personal involvement	Low	Medium	High	Yes	CR
Item 13	Calling the cops	Low	High	High	Yes	C
Item 14	Personal involvement	Low	Medium	High	Yes	R
Item 15	Not my business	Low	Medium	High	Yes	CR
Item 16	Not my business	Low	High	High	Yes	CR
Item 17	Calling the cops	Low	High	Medium	Yes	
Item 18	Calling the cops	Medium	Medium	High	Yes	CR
Item 19	Personal involvement	Low	Medium	High	Yes	C
Item 20	Personal involvement	Low	High	High	No	CR
Item 21	Calling the cops	Low	High	Medium	Yes	R
Item 22	Calling the cops	Medium	Medium	High	Yes	CR
Item 23	Personal involvement	Low	High	Medium	Yes	
Item 24	Not my business	Low			Yes	CR
Item 25	Not my business	Low	Medium	Medium	Yes	CR
Item 26	Calling the cops	Low	High	Medium	Yes	
Item 27	Not my business	Low	Medium	High	Yes	C
Item 28	Calling the cops	Medium	Medium	High	Yes	CR

Note: Exper ratings: items rated as clear (C) and/or representative (R) by the panel of experts.

particular, the items 2, 9, 10, 11, 13, 16, 17, 29, and 30 were selected. Although the item 2 presented a low loading in the "calling the cops" factor, it was

selected as it fulfills the other criteria and the loading on the specific factor were close enough to the .20 cut off for medium loadings (i.e., λ = .19). Then, the Sample 4 was used to study the psychometric properties of the nine-item version of the scale. The internal consistency of this version of the scale was adequate (Cronbach's α = .77), and the item-test corrected correlations were above .30 for all items except for the item 29, which was .28. The factor structure of the nine-item version showed and excellent fit to the data when the bifactor model was fitted using WLSMV estimation with the polychoric correlation matrix (χ^2(18) = 33.01, CFI = .98, TLI = .97, RMSEA = .065 [90% CI .027; .099]). Regarding the validity evidences based on its relationships with other constructs, correlations are showed in Table 10 and are in the same direction than with the complete WI-IPVAW version. Finally, the correlation between the nine-item version and the complete scale was very strong, r = .92, t(198) = 32.81, p < .001, suggesting that both versions provided similar assessments.

Five-item version of the WI-IPVAW scale. For those circumstances in which the space is very limited (e.g., socio-demographical surveys), a shorter version of the scale was assembled focusing on the general factor. To this end, two items for the "calling the cops" and "personal involvement" factors and one item of the "not my business" factor were selected. These items were those that present higher factor loadings in the general "willingness to intervene" factor on the nine-item version, in particular the items 9, 10, 11, 13, and 30. Sample 4 was used to study the psychometric properties of this version of the scale. The internal consistency of the scale was again fair (Cronbach's α = .73), and the item-test corrected correlations were above .30 for all items except for the item 30 in this case, which was .27. A one-factor model was fitted to the five-item version of the scale since there were less than three items

Table 10. WI-IPVAW short forms relations to other variables (Sample 4).

	Acceptability	Victim blaming	Percieved severity	Hostile sexism
Nine-item version				
Calling the cops	-0.17*	-0.20*	0.33*	-0.10*
No my business	0.10*	-0.07	0.05	0.13*
Personal involvement	-0.04	0.04	-0.04	0.02
Willingness to Intervene	-0.20*	-0.24*	0.29*	-0.16*
Five-item version				
Willingness to Intervene	-0.23*	-0.29*	0.29*	-0.16*

Note: *: $p < .01$.

per specific factor, using WLSMV estimation. The model fitted reasonably well to the data ($\chi^2(5) = 30.44$, CFI = .96, TLI = .92, RMSEA = .150 [90% CI .099; .207]), although the residuals were below the .08 cut off for a well-fitted model. The correlations between the "willingness to intervene" factor and the criterion-related variables were again in the same direction than with the complete version of the scale (see Table 10). The correlation between the five-item version and the complete version of the scale was high, $r = .86$, $t(198) = 24$, $p < .001$, however it was smaller than with the nine-item version.

Discussion

In this paper, we described the development and psychometric properties of the long and short forms of the WI-IPVAW, a set of new self-report questionnaires assessing willingness to intervene in cases of IPVAW. Taken together, our results provided strong support for the reliability and validity (including content, internal structure, and relationships with related constructs) of both the long and short versions of the WI-IPVAW scale.

Content validity of the WI-IPVAW was assessed during the scale development process using the rating of a panel of experts, to ensure that the items were capturing adequately the different aspects of the construct. One of the advantages of the WI-IPVAW is that it also takes into account different community settings (next door house, streets, bars, etc.) where IPVAW can occur, as well as different types of IPV violence (e.g., verbal, threats, physical violence) in diverse situations and with different degrees of severity. The WI-IPVAW includes also a variety of potential responses to different IPVAW scenarios (e.g., talking to victims, personal involvement, calling the police, etc.). Tapping situation-specific responses across different settings provides greater ecological validity to this measure, and also facilitates future research on situational correlates of this type of attitudes (Banyard, 2008; Banyard et al. 2011; Carlo and Randall, 2002; Copp et al., 2016). Moreover, the effect of social desirability bias was controlled in a pilot study through a confirmatory factor analysis using social desirability markers (Ferrando, 2005; 2008). This analytical approach is one of the major strengths of the present study, because it allowed to identify and remove those items with higher loadings on the social desirability factor from the scale.

Regarding the internal structure of the scale, our results supported a bifactor model as the latent structure of the scale, as it was the one that best fitted the data in comparison with other models. In this model, each item loaded on one specific factor and also onto a general factor. This general factor (i.e., "willingness to intervene intervene in cases of IPVAW") captures the common variance of all items, reflecting the shared elements of the measured construct. On the other hand, the specific factors (i.e., "calling the cops", "personal involvement", and "not my business") represent the remaining unique variance not attributable to the general factor. The model is orthogonal and thus the factors are uncorrelated, meaning that the general factor is

assumed to be independent of the specific factors, as well as the specific factors are assumed to be distinctive and independent to each other (e.g., Chen, et al., 2006; Gibbins et al., 2012). Our results emphasize, in addition, the relevance of the general factor, since most of the loadings presented higher values on the general factor than on their respective specific factor. Moreover, the general factor accounted for the largest proportion of the common explained variance, the 56.85%. The "calling the cops" factor accounted for almost the half of the remaining common variance, the 23.16%, whereas the "personal involvement" and "not my business" specific factors explained the rest, 11.04% and 8.95% respectively.

We also studied the extent to which the scores on the WI-IPVAW were invariant across genders. A partial scalar invariance model was supported, meaning that both men and women conceptualize the underlying latent structure in the same manner (configural invariance), that the scale unit is the same, and thus the items are interpreted similarly by men and women (metric invariance), and that the thresholds of the items are the same for both genders, being the factorial scores comparable across gender groups (scalar invariance). However, the threshold parameters of two items (items 6 and 21) were allowed to vary across groups, implying that men and women do not share the same distribution on these items. To obtain comparable scores for men and women in the general "willingness to intervene" factor and in the specific factors, researchers and practitioners could remove the items 6 and 21 from the scale. We recommend, however, using the invariant items as anchor items and treat these two items differently for each gender. To this end we provide a Mplus syntax to compute this model on Appendix 2.

Regarding validity analyses based on the relationships of the WI-IPVAW with other variables, we found that the general factor (i.e., 'willingness to intervene in cases of IPVAW) was significantly associated with a set of

relevant variables linked to IPVAW. Thus, as expected, respondents with higher scores on the WI-IPVAW (i.e., those more willing to intervene), perceived IPVAW situations as more severe, find IPVAW less acceptable, have less victim-blaming attitudes, and score lower in hostile sexism. This supports the idea that the willingness to intervene in cases of IPVAW reflects the personal level of tolerance and acceptance of this type of violence, and suggests that attitudes towards intervention in cases of IPVAW are also linked to attitudes justifying IPVAW, such as victim-blaming, and to hostility about women (Gracia et al., 2014; Glick et al., 2002; Herrero et al., 2017; Ivert et al., 2017; Taylor & Sorenson, 2005; Valor-Segura et al., 2011). With respect to the specific factors, both 'calling the cops' and 'not my business', were related as expected (i.e., in a positive way the first one and in a negatively way the second one) with the same set of variables. For example, those scoring high in the 'not my business' tended to perceive IPVAW as less severe and more acceptable, and scored higher in both victim-blaming atitudes and hostile sexism. Interestingly, the 'personal involvement' factor was related only, in a negative way, with the perceived severity of IPVAW, suggesting that the more severe an IPVAW situation is perceived other intervention preferences are favored, as greater personal costs or negative consequences can be involved. For example, as Gracia et al. (2009) observed, reporting to the police incidents of IPVAW is more likely among those who tend to perceive them as more severe.

In this study, we also developed two shortened versions of the WI-IPVAW scale. The parent WI-IPVAW scale is a relatively lengthy questionnaire (28 items) The length of questionnaires often prevents their inclusion in population surveys where space is scant and expensive, or in studies where time is an issue, but there is nevertheless interest in addressing other important constructs linked to IPVAW, such as public attitudes. Large-scale surveys tend to resort to single items addressing these attitudes or use a set of

questions with unknown reliability or validity (Gracia & Lila, 2015; Richins, 2004). On the other hand, shortened versions can have the drawback of limited reliability and validity, which makes it particularly important to ensure that short versions of questionnaires retain its psychometric soundness (Kovacs, 2017; Smith et al., 2000; Stanton et al., 2002). As Smith et al. (2000) pointed out, it is important the rigorous application of psychometric principles in validating short forms. In the present study, two short versions of the parent WI-IPVAW scale of nine and six items were constructed based on quantitative and qualitative criteria (Goetz et al., 2013), supporting the adequate transfer of validity from the parent form of the WI-IPVAW to the two short forms. The long and short versions of the WI-IPVAW demonstrated high reliability, as well as construct validity as they were strongly related to the acceptability of IPVAW, victim blaming attitudes, perceived severity of IPVAW, and hostile sexism. Although some loss of reliability is inevitable, our results provide strong empirical support for the high quality of their psychometric properties of the short versions of the WI-IPVAW scale. When research or survey needs (large-scale surveys, limited space or time, etc.) require the use of short forms, our results support that both the 9- and the 5-item short forms are reliable and valid alternatives to the most comprehensive and broader assessment of the willingness to intervene in cases of IPVAW provided by the long version of the WI-IPVAW (both reduced versions presented a high correlation with the parent WI-IPVAW scale). For example, the 9-item WI-IPVAW short scale showed not only adequate reliability, but also allows the meaningful assessment of both the general non-specific factor expressing the willingness to intervene in cases of IPVAW, and the three specific factors reflecting different intervention preferences (by incorporating three items of each of the specific factors of the original scale the adequate representation of the construct is ensured). On the other hand, the 5-item WI-IPVAW short scale is

particularly recommended for the reliable and valid assessment of the general "willingness to intervene" factor, when space and/or time constraints are an issue, but, nevertheless this construct is a target for research or policy making. The five-item version only mapped the general factor as there were not enough items to preserve the original latent structure of the scale. The scores on the general factor of the five-item version presented a similar pattern when related to acceptability of IPVAW, attitudes of victim blaming, perceived severity, and hostile sexism.

This study is not without limitations. Although social desirability was controlled on the pilot study using the procedure proposed by Ferrando (2005), the items used as social desirability markers presented a mediocre reliability, and thus these results should be taken with caution. Regarding the measurement invariance, although we a partial scalar invariance level for the WI-IPVAW across genders was supported, further research is needed to establish whether this instrument is also invariant across age and educational level groups.

As for practical implications, addressing attitudes that supports IPVAW, such as the willingness to intervene in cases of IPVAW, and advancing in their conceptualization, measurement, prevalence, and determinants is key to monitor social changes in this type of attitudes and to better inform prevention and intervention strategies (Powell & Webster, 2018). Public willingness to intervene in cases of IPVAW reflects the level of tolerance and acceptability of IPVAW, and these attitudes, when collectively held at different levels of aggregation (e.g. social groups, neighborhoods, communities, countries) can create a social climate that can help legitimize or deter this type of violence (Browning, 2002; Emery et al., 2011; Heise, 2011; Heise & Fulu, 2014; Heise & Kotsadam, 2015; Voith, 2017; Wright & Benson, 2010). In this regard, the different versions of the WI-IPVAW, in particular the short

versions, more appropriate for survey type research, can be used to assess pro- or non-intervention norms at different aggregation levels, such as neighborhoods or communities, when they become key targets for social and community intervention strategies addressing the prevalence of IPVAW and its correlates, such as public attitudes (Gracia, 2014; Gracia et al., 2015; Voith, 2017). In terms of Klein et al. (1997), "we need to educate people to recognize that they have a role in helping battered women and to teach them that their behavior matters and showed them how to get involved" (p. 90). In this regard, as Gracia et al. (2009) proposed, to increase feelings of social and personal responsibility regarding high prevalence of IPVAW in our societies, public education efforts need to promote attitudes that reinforce the helping role of the social circle that surrounds the victims. Increasing the likelihood of public intervention to help IPVAW victims, not only among the general public, but also among different professional groups (social services, health, law enforcement, etc.) can contribute to the deterrence and reduction of this major social and public health problem (Gracia et al, 2014; Ferrer PI, 2016; Ossorio, 2016). The WI-IPVAW can therefore become a useful instrument to assess prevention policies and public education campaigns aiming to promote a more responsive social environment in cases of IPVAW, thus contributing to the deterrence and reduction of this major social and public health problem.

Ethics Statement

This study was performed in accordance with the Declaration of Helsinki. Informed consent information was supplied and implied through participation in the on-line survey. The study and protocol were reviewed and approved by the University of Valencia Ethics Committee.

Author Contributions

EG conceived the study and supervised the writing of the manuscript. MM-F designed the analytic strategy, contributed to developing materials, conducted the statistical analysis, and wrote the methods and results sections of the manuscript. MM contributed to developing materials, data collection, and writing some parts of the manuscript. FS contributed to developing materials, data collection, and writing some parts of the manuscript. VV contributed to developing materials, data collection, and writing some parts of the manuscript. ML coordinated the data collection and contributed to the writing of the manuscript.

Funding

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supplementary Material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01146/full#supplementary-material

References

Ansara, D. L., and Hindin, M. J. (2010). Formal and informal help-seeking associated with women's and men's experience of intimate partner violence in Canada. *Soc. Sci. Med.* 70, 1011–1018. doi: 10.1016/j.socscimed.2009.12.009

Asparouhov, T., and Muthén, B. (2010). *Weighted Least Squares Estimation With Missing Data.* Available at: http://www.statmodel.com/download/GstrucMissingRevision.pdf

Asparouhov, T., Muthén, B., and Muthén, B. O. (2006). Robust chi square difference testing with mean and variance adjusted test statistics. *Matrix* 1, 1–6. doi: 10.1177/0013164416683754

Baldry, A. C., Pacilli, M. G., and Pagliaro, S. (2015). She's not a person.. she's just a woman! Infra-Humanization and intimate partner violence. *J. Interpers. Violence* 30, 1567–1582. doi: 10.1177/ 0886260514540801

Baldry, A. C., and Pagliaro, S. (2014). Helping victims of intimate partner violence: the influence of group norms among lay people and the police. *Psychol. Violence* 4, 334–347. doi: 10.1037/a0034844

Banyard, V., and Moynihan, M. M. (2011). Variation in bystander behavior related to sexual and intimate partner violence prevention: correlates in a sample of college students. *Psychol. Violence* 1, 287–301. doi: 10.1037/a0023544

Banyard, V. L. (2008). Measurement and correlates of prosocial bystander behavior: the case of interpersonal violence. *Violence Vict.* 23, 83–97. doi: 10.1891/0886-6708.23.1.83

Banyard, V. L., Moynihan, M. M., Cares, A. C., and Warner, R. (2014). How do we know if it works? Measuring outcomes in bystander-focused abuse prevention on campuses. *Psychol. Violence* 4, 101–115. doi: 10.1037/a0033470

Beck, C. T., and Gable, R. K. (2001). Ensuring content validity: illustration of the process. *J. Nurs. Meas.* 9, 201–215.

Bentler, P. M. (1995). *EQS Program Manual.* Encino, CA: Multivariate Software Inc.

Boira, S., Carbajosa, P., and Méndez, R. (2016). Miedo, conformidad y silencio: la violencia en las relaciones de pareja en áreas rurales de Ecuador [Fear, conformity and silence. Intimate partner violence in rural areas of Ecuador]. *Psychosoc. Interv.* 25, 9–17. doi: 10.1016/j.psi.2015.07.008

Branch, K. A., Richards, T. N., and Dretsch, E. C. (2013). An exploratory analysis of college students' response and reporting behavior regarding intimate partner violence victimization and perpetration among their friends. *J. Interpers. Violence* 28, 3386–3399. doi: 10.1177/0886260513504494

Browning, C. R. (2002). The span of collective efficacy: extending social disorganization theory to partner violence. *J. Marriage Fam.* 64, 833–850. doi: 10.1111/j.1741-3737.2002.00833.x

Brummelman, E., Thomaes, S., Nelemans, S. A., Orobio de Castro, B., and Bushman, B. J. (2015). My child is God's gift to humanity: development and validation of the parental overvaluation scale (POS). *J. Pers. Soc. Psychol.* 108, 665–679. doi: 10.1037/pspp0000012

Campbell, J. C. (2002). Health consequences of intimate partner violence. *Lancet* 359, 1331–1336. doi: 10.1016/S0140-6736(02)08336-8

Carlo, G., and Randall, B. A. (2002). The development of a measure of prosocial behaviors for late adolescents. *J. Youth Adolesc.* 31, 31–44. doi: 10.1023/A:1014033032440

Carlson, B. E., and Worden, A. P. (2005). Attitudes and beliefs about domestic violence: results of a public opinion survey: I. Definitions of domestic violence, criminal domestic violence, and prevalence. *J. Interpers. Violence* 20, 1197–1218. doi: 10.1177/0886260505278530

Chen, F. F., West, S. G., and Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behav. Res.* 41, 189–225. doi: 10.1207/s15327906mbr4102_;5

Cheung, G. W., and Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct. Equat. Model.* 9, 233–255. doi: 10.1097/NNR.0b013e3182544750

Cinquegrana, V., Baldry, A. C., and Pagliaro, S. (2018). Intimate partner violence and bystanders' helping behaviour: an experimental study. *J. Aggress. Conflict Peace Res.* 10, 24–35.

Consejo General del Poder Judicial (2016). *La Violencia Sobre la Mujer en la Estadística Judicial: Anual 2016 [Violence Against Women in Judicial Statistics: YEAR 2016].*Available at: http://www.poder judicial.es/cgpj/es/Temas/Violencia-domestica-y-de-genero/Actividad-del-Observatorio/Datos-estadisticos/La-violencia-sobre-la-mujer-en-la-estadistica-judicial--Anual-2016

Copp, J. E., Giordano, P. C., Longmore, M. A., and Manning, W. D. (2016). The development of attitudes toward intimate partner violence: an examination of key correlates among a sample of young adults. *J.*

Interpers. Violence. doi: 10.1177/0886260516651311 [Epub ahead of print].

Delgado-Rico, E., Carretero-Dios, H., and Ruch, W. (2012). Content validity evidences in test development: an applied perspective. *Int. J. Clin. Health Psychol.* 12, 449–459. doi: 10.1111/jan.12741

Devries, K. M., Mak, J. Y., García-Moreno, C., Petzold, M., Child, J. C., Falder, G., et al. (2013). The global prevalence of intimate partner violence against women. *Science* 340, 1527–1528. doi: 10.1126/science.1240937

Devries, K. M., Watts, C., Yoshihama, M., Kiss, L., Schraiber, L. B., Deyessa, N., et al. (2011). Violence against women is strongly associated with suicide attempts: evidence from the WHO multi-country study on women's health and domestic violence against women. *Soc. Sci. Med.* 73, 79–86. doi: 10.1016/ j.socscimed.2011.05.006

Ellsberg, M., Jansen, H. A., Heise, L., Watts, C. H., and Garcia-Moreno, C. (2008). Intimate partner violence and women's physical and mental health in the WHO multi-country study on women's health and domestic violence: an observational study. *Lancet* 371, 1165–1172. doi: 10.1016/S0140-6736 (08)60522-X

Emery, C. R., Jolley, J. M., and Wu, S. (2011). Desistance from intimate partner violence: the role of legal cynicism, collective efficacy, and social disorganization in Chicago neighborhoods. *Am. J. Community Psychol.* 48, 373–383. doi: 10.1007/s10464-010-9362-5

European Commission (2016). *Gender Based Violence (Special Eurobaromenter 449 – Wave EB85.3 – TNS Opinion & Social).* Brussels: European Commission.

European Union Agency for Fundamental Rights (2014). *Violence Against Women: An EU-Wide Survey.* Luxembourg: Publications Office of the European Union.

Expósito, F., Moya, M. C., and Glick, P. (1998). Sexismo ambivalente: medición y correlatos [Ambivalent sexism: measurement and correlates]. *Rev. Psicol. Soc.* 13, 159–169. doi: 10.1174/ 021347498760350641

Ferrando, P. J. (2005). Factor analytic procedures for assessing social desirability in binary items. *Multivariate Behav. Res.* 40, 331–349. doi: 10.1207/s15327906mbr4003_;3

Ferrando, P. J. (2008). The impact of social desirability bias on the EPQ-R item scores: an item response theory analysis. *Pers. Individ. Diff.* 44, 1784–1794. doi: 10.1016/j.paid.2008.02.005

Ferrer-Perez, V. A., Ferreiro-Basurto, V., Navarro-Guzmán, C., and Bosch-Fiol, E. (2016). Programas de intervención con maltratadores en España: la perspectiva de los/as profesionales [Batterer intervention programs in Spain: the professionals' perspective]. *Psychosoc. Interv.* 25, 159–168. doi: 10.1016/j.psi.2016.06.001

Fincham, F. D., Cui, M., Braithwaite, S., and Pasley, K. (2008). Attitudes toward intimate partner violence in dating relation-ships. *Psychol. Assess.* 20, 260–269. doi: 10.1037/1040-3590.20.3.260

Flood, M., and Pease, B. (2009). Factors influencing attitudes to violence against women. *Trauma Violence Abuse* 10, 125–142. doi: 10.1177/1524838009334131

Garcia-Moreno, C., Jansen, H. A., Ellsberg, M., Heise, L., and Watts, C. H. (2006). Prevalence of intimate partner violence: findings from the WHO multi-country study on women's health and domestic violence. *Lancet* 368, 1260–1269. doi: 10.1016/S0140-6736 (06)69523-8

Gibbins, C., Toplak, M. E., Flora, D. B., Weiss, M. D., and Tannock, R. (2012). Evidence for a general factor model of ADHD in adults. *J. Atten. Disord.* 16, 635–644. doi: 10.1177/1087054711416310

Glick, P., and Fiske, S. T. (1996). The ambivalent sexism inventory: differentiating hostile and benevolent sexism. *J. Pers. Soc. Psychol.* 70:491. doi: 10.1037/0022-3514.70.3.491

Glick, P., Fiske, S. T., Mladinic, A., Saiz, J. L., Abrams, D., Masser, B., et al. (2000). Beyond prejudice as simple antipathy: hostile and benevolent sexism across cultures. *J. Pers. Soc. Psychol.* 79, 763–775. doi: 10.1037/0022-3514.79.5.763

Glick, P., Sakalli–Ugurlu, N., Ferreira, M. C., and Aguiar de Souza, M. (2002). Ambivalent sexism and attitudes toward wife abuse in Turkey and Brazil. *Psychol. Women Q.* 26, 292–297. doi: 10.1111/1471-6402.t01-1-00068

Goetz, C., Coste, J., Lemetayer, F., Rat, A. C., Montel, S., Recchia, S., et al. (2013). Item reduction based on rigorous methodological guidelines is necessary to maintain validity when shortening composite measurement scales. *J. Clin. Epidemiol.* 66, 710–718. doi: 10.1016/j.jclinepi.2012.12.015

Gracia, E. (2004). Unreported cases of domestic violence against women: towards an epidemiology of social silence, tolerance, and inhibition. *J. Epidemiol. Community Health* 58, 536–537. doi: 10.1136/jech.2003.019604

Gracia, E. (2014). Intimate partner violence against women and victim-blaming among Europeans. *Bull. World Health Organ.* 92, 380–381. doi: 10.2471/BLT.13.131391

Gracia, E., García, F., and Lila, M. (2009). Public responses to intimate partner violence against women: the influence of perceived severity and personal responsibility. *Span. J. Psychol.* 12, 648–656. doi: 10.1017/S1138741600002018

Gracia, E., García, F., and Lila, M. (2011). Police attitudes toward policing partner violence against women: do they correspond to different psychosocial profiles? *J. Interpers. Violence* 26, 189–207. doi: 10.1177/0886260510362892

Gracia, E., García, F., and Lila, M. (2014). Male police officers' law enforcement preferences in cases of intimate partner violence versus non-intimate interpersonal violence: do sexist attitudes and empathy matter? *Crim. Justice Behav.* 41, 1195–1213. doi: 10.1177/0093854814541655

Gracia, E., and Herrero, J. (2006). Public attitudes toward reporting partner violence against women and reporting behavior. *J. Marriage Fam.* 68, 759–768. doi: 10.1111/j.1741-3737.2006.00288.x

Gracia, E., and Lila, M. (2015). *Attitudes Towards Violence Against Women in the EU.* Luxembourg: Publication Office of the European Union.

Gracia, E., López-Quílez, A., Marco, M., Lladosa, S., and Lila, M. (2015a). The spatial epidemiology of intimate partner violence: do neighborhoods matter? *Am. J. Epidemiol.* 182, 58–66. doi: 10.1093/aje/kwv016

Gracia, E., and Merlo, J. (2016). Intimate partner violence against women and the Nordic paradox. *Soc. Sci. Med.* 157, 27–30. doi: 10.1016/j.socscimed.2016.03.040

Gracia, E., Rodriguez, C. M., and Lila, M. (2015b). Preliminary evaluation of an analog procedure to assess acceptability of intimate partner violence against women: the partner violence acceptability movie task. *Front. Psychol.* 6:1567. doi: 10.3389/fpsyg.2015. 01567

Gracia, E., and Tomás, J. M. (2014). Correlates of victim-blaming attitudes regarding partner violence against women among the Spanish general population. *Violence Against Women* 20, 26–41. doi: 10.1177/1077801213520577

Grimm, P. (2010). "Social desirability bias," in *Wiley International Encyclopedia of Marketing*, eds J. Sheth and N. K. Malhotra (New York, NY: John Wiley & Sons).

Guedes, A., Bott, S., Garcia-Moreno, C., and Colombini, M. (2016). Bridging the gaps: a global review of intersections of violence against women and violence against children. *Global Health Action* 9:31516. doi: 10.3402/gha.v9.31516

Hamby, S., Weber, M. C., Grych, J., and Banyard, V. (2015). What difference do bystanders make? The association of bystander involvement with victim out-comes in a community sample. *Psychol. Violence* 6, 91–102. doi: 10.1037/a0039073

Hart, C. M., Ritchie, T. D., Hepper, E. G., and Gebauer, J. E. (2015). The balanced inventory of desirable responding short form (BIDR-16). *Sage Open* 5, 1–9. doi: 10.1177/2158244015621113

Heise, L. (1998). Violence against women: an integrated, ecological framework. *Violence Against Women* 4, 262–290. doi: 10.1177/1077801298004003002

Heise, L. (2011). *What Works to Prevent Partner Violence? An Evidence Overview.* London: STRIVE.

Heise, L. L., and Kotsadam, A. (2015). Cross-national and multilevel correlates of partner violence: an analysis of data from population-based surveys. *Lancet Glob. Health* 3, e332–e340. doi: 10.1016/S2214-109X(15)00013-3

Herrero, J., Rodríguez, F. J., and Torres, A. (2017). Acceptability of partner violence in 51 societies: the role of sexism and attitudes toward violence in social relationships. *Violence Against Women* 23, 351–367. doi: 10.1177/1077801216642870

Hu, L. T., and Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: conventional criteria versus new alternatives. *Struct.l Equat. Model.* 6, 1–55. doi: 10.1080/10705519909540118

Ivert, A. K., Merlo, J., and Gracia, E. (2018). Country of residence, gender equality and victim blaming attitudes about partner violence: a multilevel analysis in EU. *Eur. J. Public Health.* 28, 559–564. doi: 10.1093/eurpub/ckx138

Jewkes, R., Flood, M., and Lang, J. (2015). From work with men and boys to changes of social norms and reduction of inequities in gender relations: a conceptual shift in prevention of violence against women and girls. *Lancet* 385, 1580–1589. doi: 10.1016/S0140-6736(14)61683-4

Klein, E., Campbell, J., Soler, E., and Ghez, M. (1997). *Ending Domestic Violence: Changing Public Perceptions/Halting the Epidemic.* Thousand Oaks, CA: Sage.

Koepsell, J. K., Kernic, M. A., and Holt, V. L. (2006). Factors that influence battered women to leave their abusive relationships. *Violence Vict.* 21, 131–147. doi: 10.1891/vivi.21.2.131

Kovacs, C., Batinic, B., Stiglbauer, B., and Gnambs, T. (2017). Development of a shortened version of the latent and manifest benefits of work (LAMB) scale. *Eur. J. Psychol. Assess.* doi: 10.1027/1015-5759/a000434

Latané, B., and Darley, J. M. (1970). *The Unresponsive Bystander: Why Doesn't He Help?* New York, NY: Appleton-Century Crofts.

Liang, B., Goodman, L., Tummala-Narra, P., and Weintraub, S. (2005). A theoretical framework for understanding help-seeking processes among survivors of intimate partner violence. *Am. J. Commun. Psychol.* 36, 71–84. doi: 10.1007/s10464-005-6233-6

Lila, M., Gracia, E., and García, F. (2013). Ambivalent sexism, empathy and law enforcement attitudes towards partner violence against women among male police officers. *Psychol. Crime Law* 19, 907–919. doi: 10.1080/1068316X.2012.719619

Lila, M., Oliver, A., Catalá-Miñana, A., Galiana, L., and Gracia, E. (2014). The intimate partner violence responsibility attribution scale (IPVRAS). *Eur. J. Psychol. Appl. Legal Context* 6, 29–36. doi: 10.5093/ejpalc2014a4

López-Ossorio, J. J., González-Álvarez, J. L., and Andrés-Pueyo, A. (2016). Eficacia predictiva de la valoración policial del riesgo de la violencia de género [Predictive effectiveness of the police risk assessment in intimate partner violence]. *Psychosoc. Interv.* 25, 1–7. doi: 10.1016/j.psi.2015.10.002

MacCallum, R. C., Browne, M. W., and Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychol. Methods* 1, 130–149. doi: 10.1037/1082-989X.1.2.130

Marco, M., Gracia, E., and López-Quílez, A. (2018). The university campus environment as a protective factor for intimate partner violence against women: an exploratory study. *J. Commun. Psychol.* doi: 10.1002/jcop.21980

Martín-Fernández, M., Gracia, E., and Lila, M. (2018b). Assessing victim-blaming attitudes in cases of intimate partner violence against women: Development and validation of the VB-IPVAW scale. *Psychosoc. Interv.* doi: 10.5093/pi2018a18

Martín-Fernández, M., Gracia, E., Marco, M., Vargas, V., Santirso, F. A., and Lila, M. (2018a). Measuring acceptability of intimate partner violence

against women: development and validation of the A-IPVAW scale. *Eur. J. Psychol. Appl. Legal Context* 10, 26–34. doi: 10.5093/ejpalc2018a3

McCart, M. R., Smith, D., and Sawyer, G. K. (2010). Help seeking among victims of crime: a review of the empirical literature. *J. Trauma Stress* 23, 198–206. doi: 10.1002/jts.20509

McDonnell, K. A., Burke, J. G., Gielen, A. C., O'Campo, P., and Weidl, M. (2011). Women's perceptions of their community's social norms towards assisting women who have experienced intimate partner violence. *J. Urban Health* 88, 240–253. doi: 10.1007/s11524-011-9546-9

McMahon, S., Allen, C. T., Postmus, J. L., McMahon, S. M., Peterson, N. A., and Hoffman, M. L. (2014). Measuring bystander attitudes and behavior to prevent sexual violence. *J. Am. Coll. Health* 62, 58–66. doi: 10.1080/07448481.2013.849258

Megías, J. L., Toro-García, V., and Carretero-Dios, H. (2017). The acceptance of myths about intimate partner violence against women (AMIVAW) scale: development and validation in Spanish and English. *Psychol. Women Q.* 42, 44–61. doi: 10.1177/0361684317742638

Miles, J., and Shevlin, M. (2001). *Applying Regression and Correlation: A Guide for Students and Researchers.* London: Sage.

Milfont, T. L., and Fischer, R. (2010). Testing measurement invariance across groups: applications in crosscultural research. *Int. J. Psychol. Res.* 3, 111–121. doi: 10.21500/20112084.857

Ministerio de Sanidad, Servicios Sociales e Igualdad (2015). *Macroencuesta de Violencia Contra la Mujer 2015 [Survey on Violence Against Women 2015].* Madrid: Ministerio de Sanidad, Servicios Sociales e Igualdad.

Muthén, B., and Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *Br. J. Math. Stat. Psychol.* 38, 171–189. doi: 10.1111/j.2044-8317.1985.tb00832.x

Muthén, B., and Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: a note on the size of the model. *Br. J. Math. Stat. Psychol.* 45, 19–30. doi: 10.1111/j.2044-8317.1992.tb00975.x

Muthén, L. K., and Muthén, B. O. (2010). *Mplus User's Guide.* Los Angeles, CA: Author.

Powell, A., and Webster, K. (2018). Cultures of gendered violence: an integrative review of measures of attitudinal support for violence against women. *Austral. N. Z. J. Criminol.* 51, 40–57. doi: 10.1177/0004865816675669

R Core Team (2017). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.

Revelle, W. (2016). *psych: Procedures for Personality and Psychological Research.* Evanston, IL: Northwestern University.

Richins, M. L. (2004). The material values scale: measurement properties and development of a short form. *J. Consum. Res.* 31, 209–219. doi: 10.1086/383436

Rizo, C. F., and Macy, R. J. (2011). Help seeking and barriers of Hispanic partner systematic review of the literature. *Aggress. Violent Behav.* 6, 250–264. doi: 10.1016/j.avb.2011.03.004

Rollero, C., Glick, P., and Tartaglia, S. (2014). Psychometric properties of short versions of the ambivalent sexism inventory and ambivalence toward men inventory. *TPM. Test. Psychometr. Methodol. Appl. Psychol.* 21, 149–159.

Smith, G. T., McCarthy, D. M., and Anderson, K. G. (2000). On the sins of short-form development. *Psychol. Assess.* 12, 102–111. doi: 10.1037/1040-3590.12.1.102

Stanton, J. M., Sinar, E. F., Balzer, W. K., and Smith, P. C. (2002). Issues and strategies for reducing the length of self-report scales. *Pers. Psychol.* 55, 167–194. doi: 10.1111/j.1744-6570.2002.tb00108.x

Stein, J. L. (2007). Peer educators and close friends as predictors of male college students' willingness to prevent rape. *J. College Stud. Dev.* 48, 75–89. doi: 10.1353/csd.2007.0008

Stöckl, H., Devries, K., Rotstein, A., Abrahams, N., Campbell, J., Watts, C., et al. (2013). The global prevalence of intimate partner homicide: a systematic review. *Lancet* 382, 859–865. doi: 10.1016/S0140-6736(13)61030-2

Taylor, C. A., and Sorenson, S. B. (2005). Community-based norms about intimate attributions of fault and responsibility into context. *Sex Roles* 53, 573–589. doi: 10.1007/s11199-006-9155-3

Taylor, E., Banyard, V., Grych, J., and Hamby, S. (2016). Not all behind closed doors: examining bystander involvement in intimate partner violence. *J. Interpers. Violence.* doi: 10.1177/0886 260516673629 [Epub ahead of print].

Thornton, L., Batterham, P. J., Fassnacht, D. B., Kay-Lambkin, F., Calear, A. L., and Hunt, S. (2016). Recruiting for health, medical or psychosocial research using Facebook: systematic review. *Internet Intervent.* 4, 72–81. doi: 10.1016/j.invent. 2016.02.001

Timmerman, M. E., and Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychol. Methods* 16, 209–220. doi: 10.1037/a0023353

Topolovec-Vranic, J., and Natarajan, K. (2016). The use of social media in recruitment for medical research studies: a scoping review. *J. Med. Internet Res.* 18, e286. doi: 10.2196/jmir.5698

Touza-Garma, C. (2017). Influence of health personnel's attitudes and knowledge in the detection and reporting of elder abuse: an exploratory systematic review. *Psychosoc. Intervent.* 26, 73–91. doi: 10.1016/j.psi.2016.11.001

Vargas, V., Lila, M., and Catalá-Miñana, A. (2015). >Influyen las diferencias culturales en los resultados de los programas de intervención con maltratadores?: un estudio con agresores españoles y latinoamericanos [Do cultural differences influence batterer intervention program outcomes? A study with Spanish and Latin American offenders]. *Psychosoc. Interv.* 24, 41–47. doi: 10.1016/j.psi.2015.03.001

Vives-Cases, C., Torrubiano-Domínguez, J., Escribá-Agüir, V., Ruiz-Pérez, I., Montero-Piñar, M. I., and Gil-González, D. (2011). Social determinants and health effects of low and high severity intimate partner violence. *Ann. Epidemiol.* 21, 907–913. doi: 10.1016/j.annepidem.2011.02.003

Voith, L. A. (2017). Understanding the relation between neighborhoods and intimate partner violence: an integrative review. *Trauma Violence Abuse* doi: 10.1177/1524838017717744 [Epub ahead of print].

Waltermaurer, E. (2012). Public justification of intimate partner violence: a review of the literature. *Trauma Violence Abuse* 13, 167–175. doi: 10.1177/1524838012447699

Wee, S., Todd, M. J., Oshiro, M., Greene, E., and Frye, V. (2016). Modifiers of neighbors' bystander intervention in intimate partner violence: a concept mapping study. *Violence Gender* 3, 55–63. doi: 10.1089/vio.2015.0012

World Health Organization [WHO] (2002). *World Report on Violence and Health.* Geneva: World Health Organization.

World Health Organization [WHO] (2013). *Global and Regional Estimates of Violence Against Women: Prevalence and Health Effects of Intimate Partner Violence and Non-Partner Sexual Violence.* Geneva: World Health Organization.

Wright, E. M., and Benson, M. L. (2011). Clarifying the effects of neighborhood context on violence "behind closed doors". *Just. Q.* 28, 775–798. doi: 10.1080/07418825.2010.533687

274

Appendix 1. Items of the WI-IPVAW scale (original Spanish items).

1. If in a bar a couple that clearly has been drinking too much start insulting and pushing each other, I would ignore them and would mind my own business. (Si en un bar una pareja que claramente ha bebido mucho se insulta y empuja mutuamente, los ignoraría y continuaría con lo mío)

*2. If in the common area where I live I heard a man shouting violently at his partner I would intervene to stop the situation. (Si en la escalera de mi edificio un hombre le gritase violentamente a su pareja, intervendría para parar la situación)

3. If on the street a man insults his partner, I would say something to get his attention. (Si en la calle un hombre insulta a su pareja, le diría algo para llamarle la atención)

4. If in a shop a woman tells that she is afraid because her partner has threatened her, I would recommend her to call the police. (Si una mujer cuenta en una tienda que tiene miedo porque su pareja le ha amenazado, le recomendaría que llame a la policía)

5. If a woman knocks on the door of my house to ask for help because her husband threatens to hit her, I would call the police. (Si una mujer llamara a la puerta de mi casa para pedir ayuda porque su marido le amenaza con pegarle, llamaría a la policía)

6. If a woman in the street is running away from her partner, I would stop the man. (Si una mujer corre por la calle huyendo de su pareja, detendría al hombre)

7. In a supermarket, if a man insulted his wife, I would ignore the situation. (Si en un supermercado un hombre insultara a su mujer, ignoraría la situación)

+*9. If I would find out that a woman neighbor of mine has been beaten by her husband, I would recommend her to report it. (Si me enterara de que en mi finca una vecina ha sido golpeada por su marido, le recomendarían que lo denunciara)

+*10. In a bar, if a man started screaming at his partner, I would stand between them to help the woman. (Si un hombre comenzase a gritar a su pareja en un bar, me interpondría para ayudar a la mujer)

+*11. If I found out that in my neighborhood a woman is frequently beaten by her partner, but does not want to report it to the authorities, I would call the police. (Si me enterase de que en mi finca una mujer es golpeada frecuentemente por su pareja, pero no quiere denunciar los hechos, llamarían a la policía)

12. If in the street a man picks up his partner's cell phone and hits it on the ground, I would approach the man and reprehend him for what he did. (Si en la calle un hombre coge el móvil de su pareja y lo estrella contra el suelo, me acercaría al hombre y le llamaría la atención)

+*13. If where I live I overheard a man shouting and threatening his partner, I would goto try to protect the woman. (Si en el rellano de mi edificio se escuchasen fuertes gritos y amenazas de un hombre hacia su pareja, acudiría e intentaría proteger a la mujer)

14. If in a bakery in the neighborhood a woman came crying telling that her partner is following her and threatening her, I would call the police.(Si en una panadería del barrio una mujer entrase llorando y diciendo que su pareja la está siguiendo y amenazándole, llamaría a la policía)

15. If a couple in the neighborhood is arguing and screaming at dawn, I would go and knock on the door to see what happens. (Si una pareja de vecinos discuten y se gritan de madrugada, llamaría a la puerta para ver qué ocurre)

276

*16. If at a bus stop a couple starts arguing and the man pushes the woman, I would look the other way, ignoring the situation. (Si en la parada del autobús una pareja comenzara a discutir y el hombre empuja a la mujer, miraría para otro lado ignorando la situación)

*17. If on the street a young couple is shouting and insulting each other, I would ignore them. (Si en la calle una pareja joven se grita e insulta mutuamente, les ignoraría)

18. If in the staircase or common areas of the building where I live a woman is asking for help because her partner is hitting her, I would call the police. (Si en el rellano de la escalera una mujer está pidiendo ayuda porque su pareja le está pegando, llamarían a la policía)

19. If on the street a couple insults and threatens each other, starting to hit each other, I would call the police. (Si en la calle una pareja se insulta y amenaza mutuamente, llegando a las manos, llamaría a la policía)

20. If in a bar of the neighborhood breaks a strong argument between a couple, with both shouting and insulting each other, I would reprehend them. (Si en un bar del barrio comenzara una fuerte discusión entre una pareja y ambos se están gritando e insultando, les llamaría la atención)

21 in the street, if a man pushed his partner falling her to the ground, I would intervene trying to stop him. (Si un hombre en la calle empujara a su pareja y esta cae al suelo, intentaría detener al hombre)

22. If a woman neighbor is telling that her husband is threatening her because she wants to separate or divorce, I would recommend to go to the police. (Si una vecina contara que su marido le amenaza porque quiere separarse o divorciarse, le recomendaría que fuese a la policía).

23. If a woman has been beaten in a bar by her partner and does not want to report it to the authorities, I would call the police. (Si una mujer ha sido

golpeada en un bar por su pareja y no quiere denunciar, yo llamaría a la policía).

24. If a couple had a strong argument in a store of the neighborhood, I would say something to them to stop the situation. (Si una pareja tuviera una fuerte discusión en una tienda del barrio, les dirían algo para parar la situación)

25. If overheard strong arguments and shouting between a couple in the neighborhood, I would ignore it. (Si en casa de unos vecinos se oyen fuertes discusiones y gritos entre una pareja, lo ignoraría)

28. If in a bar a man hits his partner during an argument and immediately afterwards he asks for her forgiveness, I would ignore it. (Si en un bar un hombre pega a su pareja durante una discusión e inmediatamente después le pide perdón, lo ignoraría)

*29. If a woman tells in a bar that her partner has hitted her, I would recommend her to call the police. (Si una mujer cuenta en un bar que su pareja le ha golpeado, le recomendaría que llamara a la policía)

+*30. If a couple of immigrants or from another culture were fighting on the street, I would ignore it and keep walking. (Si una pareja de inmigrantes o de otra cultura se peleara en la calle, pasaría de largo)

31. If in a bar a couple with clear symptoms of having drunk too much start to push and hit each other, I would call the police. (Si en un bar una pareja con claros síntomas de haber bebido mucho se empuja y golpea mutuamente, llamaría a la policía)

*9-item WI-IPVAW short scale. + 5-item WI-IPVAW short scale

Appendix 2. Scoring the WI-IPVAW

The WI-IPVAW scale has one general factor (i.e., "willingness to intervene"), comprised by all the items, and three specific factors (i.e., "calling the cops", "not my business", and "personal involvement"), comprised by different subsets of items. Both the general and specific factors could be used to obtain complementary information. However, if one is interested on the common elements of the measured construct, then the general factor should be used. On the other hand, if one is more interested on the type of interventions proposed by the items, then one should look at the specific factors.

To compute the scores of the scale we would like to discourage using the summation of the direct scores of the items, as this procedure assumed that all items are equally important to measure the factors of the scale (i.e., that are tau-equivalent), and, moreover, it could increase the residual correlations between the factors. In addition, one should be cautious with the items 6 and 20 of the scale, which have different threshold parameters for men and women. We would recommend using one of the following procedures:

a) Estimating the factor scores using the invariant items as anchor items: all the items of the scale are used to compute the factor scores for men and women. In order to make the factor scores comparable across genders, the invariant items are used as anchor items, constraining the loadings and thresholds parameters to be equal across gender groups, whereas the thresholds of the non-invariant items are allowed to vary. We provide an Mplus syntaxis to estimate this model and obtain the factor scores for men and women in the same metric.

b) Estimating the factor scores using only the invariant items: alternatively, one can remove the non-invariant items and estimate the same factorial model for men and women.

TITLE: Multigroup CFA WI-IPVAW
DATA:
! choose data extension and format. It should include the items and one grouping variable
FILE IS wi-ipvaw.dat; ! insert data name

VARIABLE:
NAMES ARE gender wi1-wi28;
USEV ARE wi1-wi28;
CATEGORICAL = wi1-wi28;
GROUPING IS gender (0=Women 1=Men); ! insert grouping values according to data

ANALYSIS:
ESTIMATOR = WLSMV;
PARAMETERIZATION=THETA;

MODEL:

   ! factor loadings constrained to be equal across groups
   Call BY wi4* wi5 wi8 wi10 wi13 wi17 wi18 wi21 wi22 wi26 wi28 (lcall1-lcall11);
   Ignore BY wi1* wi7 wi15 wi16 wi24 wi25 wi27 (lign1-lign7);
   PInvol BY wi2* wi3 wi6 wi9 wi11 wi12 wi14 wi19 wi20 wi23 (lpinvol1-lpinvol10);
   WI By wi1-wi28*(lwi1-lwi28);

   ! factor variances fixed to 1 in the reference group
   Call@1;
   Ignore@1;
   PInvol@1;
   WI@1;

  ! correlations between factors fixed to 0
   Call WITH Ignore@0;
   Call WITH PInvol@0;
   Ignore WITH PInvol@0;
   WI WITH Call@0;
   WI WITH Ignore@0;
   WI WITH PInvol@0;

! residuals fixed to 1
wi1-wi28@1;

! latent means fixed to 0 in the reference group
[Call@0]
[Ignore@0]
[PInvol@0]
[WI@0]

MODEL Men:

! factor loadings constrained to be equal across groups
Call BY wi4* wi5 wi8 wi10 wi13 wi17 wi18 wi21 wi22 wi26 wi28 (lcall1-lcall11);
Ignore BY wi1* wi7 wi15 wi16 wi24 wi25 wi27 (lign1-lign7);
PInvol BY wi2* wi3 wi6 wi9 wi11 wi12 wi14 wi19 wi20 wi23 (lpinvol1-
lpinvol10);
WI By wi1-wi28*(lwi1-lwi28);

! factor variances free
Call*;
Ignore*;
PInvol*;
WI*;

! correlations between factors fixed to 0
Call WITH Ignore@0;
Call WITH PInvol@0;
Ignore WITH PInvol@0;
WI WITH Call@0;
WI WITH Ignore@0;
WI WITH PInvol @0;

! free thresholds for items 6 and 21
[wi6$1-wi6$5*];
[wi20$1-wi20$5*];

! residuals fixed to 1
wi1-wi28@1;

! free latent means
[Call*]

281

[Ignore*]
        [PInvol*]
        [WI*]

OUTPUT:
  STDYX;
SAVEDATA: FILE IS WI_fscores.dat; ! return a .dat file with the factor scores
SAVE IS fscores;
FORMAT IS free;

CAPITULO 5

Discusión General

Discusión General

En esta tesis doctoral se han llevado a cabo seis estudios para dar respuesta a dos problemas de medición relacionados con la VG. El primero de ellos tiene que ver con la comparabilidad de los datos disponibles sobre VG de las grandes encuestas poblacionales en la UE, en los que rara vez se realiza un análisis riguroso sobre la equivalencia de medida (e.g., invarianza factorial) de los instrumentos empleados en las mismas (Davidov, Meuleman, Cieciuch, Schmidt, & Billie, 2014). El segundo problema se trata de la falta de herramientas de evaluación en la literatura para medir de manera adecuada (i.e., fiable y válida) diferentes tipos de actitudes hacia la VG (Gracia & Lila, 2015), en concreto: actitudes de aceptabilidad de la VG, de culpabilización a las víctimas de VG, y de disposición a intervenir en casos de VG.

Se presenta a continuación una discusión conjunta de los estudios realizados para abordar cada uno de estos problemas, así como las fortalezas y limitaciones de los diferentes estudios. Se exponen a continuación las principales implicaciones y las futuras líneas de investigación derivadas de los resultados de la presente tesis doctoral, y se ofrece finalmente una conclusión general de la misma.

Comparabilidad de las medidas de violencia de género en la Unión Europea

En los dos primeros estudios de esta tesis se han efectuado un conjunto de análisis para asegurar la comparabilidad entre países de las principales medidas de VG de la encuesta poblacional del FRA. Para ello, se ha puesto a prueba la invarianza factorial de las medidas de VG física, sexual y psicológica,

y se ha examinado cómo se distribuyen los niveles de estos tres tipos de violencia entre los 28 estados miembros de la UE.

El primer conjunto de análisis estuvo orientado a evaluar las propiedades psicométricas (i.e., estructura latente, fiabilidad y validez) de las diferentes medidas de VG de la encuesta del FRA. En el primer estudio, se encontró que las medidas de VG física y sexual incluidas en la encuesta del FRA medían dos constructos diferenciados, si bien relacionados. En el segundo estudio, se halló que los diferentes ítems de VG psicológica (i.e., conducta controladora y abuso emocional), evaluaban un constructo unifactorial, a pesar de que en estudios anteriores se han encontrado dos factores en grandes encuestas con ítems similares (García-Moreno, Jansen, Ellsberg, Heise, & Watts, 2006; Heise, Pallitto, García-Moreno, & Clark, 2019). Aunque todavía existe cierto debate sobre la estructura latente de la VG psicológica, la correlación entre conducta controladora y abuso emocional, al igual que en el segundo estudio, suele ser elevada, lo que puede ser indicativo de que una estructura unifactorial es suficiente para dar cuenta de la variabilidad de este constructo (Beeble, Bybee, & Sullivan, 2007; Heise et al., 2019; Salis, Salwen, & O'Leary, 2014).

Con respecto a la consistencia interna, las medidas de VG física, sexual y psicológica de la encuesta del FRA mostraron valores de ω superiores a .70 en los 28 países, lo que indica que son capaces de medir con una precisión adecuada estos tres tipos de violencia. Las evidencias de validez basadas en la relación de estas tres medidas con otras variables, pusieron de manifiesto que las mujeres con experiencias de abuso infantil tienen una mayor probabilidad de mostrar mayores niveles de VG física, sexual y psicológica. Este resultado es consistente con investigaciones previas en las que se muestra que las mujeres que han sido victimizadas durante su infancia, tienen más riesgo de ser victimizadas como adultas (Appel & Holden, 1998; Capaldi, Knoble, Shortt, &

Kim, 2012; Hamby & Grych, 2013; Schumacher, Feldbau-Kohn, & Heyman, 2001; Till-Tentschert, 2017). Se encontró además que las mujeres con mayores niveles de VG física y sexual tienden a presentar menores niveles de salud percibida y menores ingresos, un resultado bien establecido en la literatura sobre VG (Coker, et al., 2002; Lövestad & Krantz, 2012; Nybergh, Taft, & Krantz, 2013; Reichel, 2017; Sokoloff & Dupont, 2005). En lo concerniente a las relaciones entre la VG física, la sexual y la psicológica, se encontró, como se esperaba, una estrecha relación entre estas tres formas de violencia. La concurrencia de diferentes tipos de VG no es un fenómeno extraño en la literatura, ya que un tipo de violencia puede derivar fácilmente en otras formas de VG (Hamby & Sugarman, 1999, Karakurt & Silver, 2013; Liles et al., 2012; Murphy & O'Leary, 1989).

El segundo conjunto de análisis, por su parte, estuvo orientado a poner a prueba la invarianza factorial de las medidas de VG física, sexual y psicológica entre los 28 países de la UE. Para ello, se llevó a cabo una serie de MG-CFA para evaluar los niveles de invarianza configural, métrica y escalar. A pesar de las dificultades que distintos estudios han puesto de manifiesto para establecer los niveles de invarianza métrica o escalar cuando se emplean múltiples grupos (Davidov et al., 2014; Davidov, Schmidt, & Schwarz, 2008; Muthen & Asparouhov, 2018), fue posible mantener el nivel de invarianza escalar en las tres medidas de VG entre todos los países de la UE. Este resultado puede deberse al tipo de preguntas incluidas en la encuesta del FRA, ya que tienen una naturaleza principalmente conductual (e.g., "¿Te ha abofeteado tu pareja?", "¿Te ha forzado tu pareja a mantener relaciones sexuales?", "¿Ha hecho tu pareja algo para asustarte o intimidarte a propósito?").

Tras establecer un modelo invariante en las medidas de VG física, sexual y psicológica, se compararon los niveles de estos tipos de violencia entre los 28 países de la UE. En lugar de obtener las prevalencias de estas tres

formas de VG utilizando las puntuaciones directas de los ítems de la encuesta del FRA, una aproximación que ignora la estructura latente de cada medida y la relevancia de cada ítem para evaluar su factor, se realizó un análisis de medias latentes, comparando así las medias de los factores de VG entre los diferentes países (Jang et al., 2017; Putnick & Bornstein, 2016; Vanderberg & Lance, 2000). Esta es una de las principales fortalezas de los tres primeros estudios de esta tesis, pues a diferencia de los porcentajes de las tasas de prevalencia, el análisis de medias latentes permite cuantificar la diferencia media en los niveles de VG física, sexual o psicológica entre los distintos países de la UE. Los resultados de este análisis mostraron que no hay diferencias en los tres tipos de VG entre muchos de los países de la UE, ya que los intervalos de confianza de sus medias factoriales se solapaban considerablemente, lo que parece apuntar a que los niveles de VG física, sexual y psicológica son bastante similares entre ellos. Sí que se encontraron, no obstante, diferencias sustantivas entre los países con mayores niveles de VG y aquellos países con menores niveles de la misma. En concreto, países como Dinamarca, Finlandia, Suecia y Reino Unido presentaron, en promedio, mayores niveles de VG física y sexual que Chipre, España y Polonia. En lo concerniente a la violencia psicológica, países como Latvia, Lituania, Finlandia y Suecia mostraron asimismo mayores niveles de este tipo de violencia que Croacia, España, Eslovenia e Irlanda. Este resultado apoya la idea de Gracia y Merlo (2016) sobre la paradoja nórdica, ya que Dinamarca, Finlandia y Suecia, a pesar de ser los países con mayores niveles de igualdad de género dentro de la UE, se cuentan también entre los países con mayores niveles de VG física, sexual y psicológica.

En el tercer estudio de esta tesis doctoral se profundizó más en la cuestión de la paradoja nórdica, examinando si este fenómeno es fruto de un sesgo de medida. Para ello se llevó a cabo un análisis de invarianza factorial y

un análisis DIF con las medidas de VG física y sexual de la encuesta del FRA entre España y Suecia, dos países que ejemplifican esta paradoja. Se consiguió establecer un nivel de invarianza escalar entre ambos países, y no se detectó DIF en ninguno de los ítems. Los resultados conjuntos de ambos análisis mostraron que el país no tiene efecto sobre las puntuaciones factoriales de estos tipos de violencia, lo que permite descartar la hipótesis de que la paradoja nórdica se deba a un sesgo en las medidas empleadas para evaluar la VG. Se compararon también en este estudio las medias latentes de los factores de VG física y sexual entre España y Suecia, y se encontró que el 89.1% y el 99.4% de las mujeres suecas presentan valores más altos que el promedio de las mujeres españolas en los factores de violencia física y sexual, respectivamente.

Estos tres primeros estudios presentan, no obstante, algunas limitaciones. La primera de ellas es la naturaleza transversal del diseño de la encuesta del FRA, lo que imposibilita evaluar la invarianza factorial entre diferentes periodos temporales, así como monitorizar si los niveles de VG física, sexual y psicológica cambian o permanecen constantes a lo largo del tiempo. La segunda limitación tiene que ver con el empleo de un método de estimación para datos categóricos (i.e., WLSMV) y la utilización de los índices de ajuste para evaluar la invarianza factorial de las medidas de VG física, sexual y psicológica de la encuesta del FRA. De acuerdo con Sass, Schmitt, y Marsh (2014), este procedimiento puede incrementar las tasas de error Tipo I, llevando a asumir que las medidas son invariantes cuando en realidad no lo son. Para tratar este problema, en los dos primeros estudios se utilizaron los puntos de corte propuestos por Meade et al. (2008), ya que estos tienden a funcionar igual de bien en tamaños muestrales grandes que cuando se emplean métodos de estimación basados en máxima verosimilitud (Sass et al., 2014). En el tercer estudio, por su parte, se utilizaron los índices de ajuste en

combinación con un análisis DIF, encontrando resultados similares mediante ambas aproximaciones. Una cuarta limitación de los tres primeros estudios son los amplios intervalos de confianza de las medidas de VG física, sexual y psicológica de la encuesta del FRA, lo que sugiere que estos instrumentos pueden mejorarse para incrementar la precisión de la estimación de los niveles de estos tipos de violencia. El sesgo de auto-selección es otra de las limitaciones, ya que inicialmente sólo el 42% de las encuestadas accedió a participar en la encuesta del FRA (FRA, 2014b). Por último, los resultados de los tres primeros estudios de esta tesis están limitados a la VG, y la pregunta de si la invarianza factorial puede mantenerse entre los países de la UE permanece sin respuesta cuando los agresores de VG no son las parejas de las mujeres que responden a la encuesta.

Tomados en conjunto, los resultados de los tres primeros estudios de esta tesis subrayaron la importancia de emplear métodos estadísticos capaces de dar cuenta de la comparabilidad entre países de las medidas incluidas en las grandes encuestas poblacionales. Los análisis de invarianza factorial mostraron que las medidas de VG física, sexual y psicológica de la encuesta del FRA son invariantes entre los diferentes países de la UE, descartando que las diferencias entre ellos se deban a un sesgo de medida. Esto permitió a su vez examinar de manera más apropiada y sofisticada cómo se distribuyen los niveles de estas formas de violencia en la UE. Estos tres estudios suponen un importante paso adelante en la evaluación de la VG en la investigación internacional, si bien todavía es necesaria más investigación para esclarecer el rol que el país puede ejercer en las diferencias de VG entre los diferentes países de la UE.

Medidas de actitudes públicas hacia la violencia de género

Investigaciones previas han puesto de manifiesto la importancia de las actitudes públicas hacia la VG para un mejor entendimiento de este fenómeno (Abramsky et al., 2011; Archer & Graham-Kevan, 2003; Copp et al., 2016; WHO, 2002). Sin embargo, la evaluación de estas actitudes en las grandes encuestas poblacionales tiende a basarse en ítems únicos y en instrumentos cuyas propiedades psicométricas (i.e., fiabilidad, estructura latente, y validez) no son conocidas, o bien no son adecuadas (Gracia & Lila, 2015). En los tres últimos estudios de la presente tesis doctoral se aborda directamente este vacío en la literatura, presentando el desarrollo y la validación de tres nuevas escalas para medir tres tipos de actitudes públicas hacia la VG: actitudes de aceptabilidad de la VG (A-IPVAW), de culpabilización a las víctimas de VG (VB-IPVAW), y de disposición a intervenir en casos de VG (WI-IPVAW).

Para asegurar la validez de contenido de las tres escalas, se partió de una selección de ítems identificados por Gracia y Lila (2015) en su revisión sobre las actitudes públicas hacia la VG en la UE y se utilizó un panel de expertos para evaluar la relevancia de cada ítem para medir cada constructo (i.e., aceptabilidad, culpabilización a las víctimas, disposición a intervenir). Se seleccionaron así aquellos ítems que mejor capturaban los aspectos clave de cada uno de estos constructos.

Con respecto a las evidencias de validez basadas en la estructura interna, se examinaron en profundidad la estructura latente y la consistencia interna de cada escala. Para ello se empleó un diseño de validación cruzada, en el que se exploró primero la estructura latente mediante un EFA en una muestra inicial, para replicar a continuación el modelo encontrado a través de un CFA en una segunda muestra. En los estudios cuarto y quinto, donde se desarrollan las escalas A-IPVAW y VB-IPVAW, se encontró una estructura unidimensional, en la que un solo factor (i.e., "actitudes de aceptabilidad", o

"actitudes de culpabilización") era suficiente para explicar la variabilidad de las respuestas a estas escalas. En el sexto estudio, por su parte, se encontró inicialmente una estructura latente de tres factores para el WI-IPVAW. Sin embargo, el CFA posterior puso de manifiesto que la estructura latente de esta última escala se explicaba mejor con un modelo *bifactor* en lugar de con un modelo de tres factores correlacionados. Este modelo plantea un factor general que captura toda la varianza común de los ítems (i.e., "disposición a intervenir"), y tres factores específicos que dependen del tipo de respuesta que platean los ítems (i.e., "llamar a la policía", "no es asunto mío", "intervención personal").

La consistencia interna de las tres escalas fue muy buena, con valores del α de Cronbach superiores a .80 para las tres medidas (y sus subescalas, en el caso del WI-IPVAW). En los estudios cuarto y quinto se estimó además un modelo de respuesta graduada de la teoría de respuesta al ítem (TRI) para las escalas A-IPVAW y VB-IPVAW (Samejima, 1969). La función de información resultante de estos análisis de la TRI, resaltaron que ambas medidas son especialmente informativas para valores medios y altos de ambos rasgos latentes, lo que las hace especialmente propicias para detectar y discriminar entre participantes con niveles medios y altos de actitudes de aceptabilidad de la VG y de culpabilización a las víctimas de VG. La precisión de estas medidas es, sin embargo, más baja para discriminar entre participantes con niveles bajos y muy bajos de estos tipos de actitudes.

Se examinó también en los estudios quinto y sexto de esta tesis la equivalencia de medida en función del sexo, a través de un análisis DIF para la escala VB-IPVAW, y mediante un análisis de invarianza factorial para la escala WI-IPVAW. No se encontró DIF en los ítems de la escala VB-IPVAW, por lo que es posible asumir que los participantes de ambos sexos no responden de manera diferencial a ninguno de sus ítems. En el caso de la escala WI-IPVAW

se consiguió establecer un nivel de invarianza escalar parcial, dejando libres los umbrales de dos ítems ("si una mujer huye de su pareja en la calle, detendría al hombre", "Si en la escalera del edificio donde vivo escuchase a un hombre gritando y amenazando a su pareja, acudiría e intentaría proteger a la mujer"). Si bien es posible obtener puntuaciones factoriales comparables entre ambos sexos en esta última escala utilizando como test de anclaje el resto de los ítems, se recomienda por simplicidad no utilizar los dos ítems no invariantes (Putnick & Bornsetin, 2016).

En lo concerniente a las evidencias de validez de cada escala basadas en sus relaciones con otras variables, se halló una estrecha relación entre las tres escalas presentadas en esta tesis. Asimismo, las correlaciones entre las puntuaciones factoriales de las escalas A-IPVAW y VB-IPVAW con sexismo hostil y benevolente fueron positivas y elevadas, indicando una relación entre el sexismo ambivalente y los constructos de estas dos nuevas herramientas. Se encontró también una fuerte relación entre las puntuaciones de la escala WI-IPVAW y sexismo hostil. Estos resultados son consistentes con estudios previos que relacionan actitudes de justificación de la VG con niveles elevados de sexismo (Glick et al., 2002; Herrera, Expósito & Moya, 2012; Herrero et al., 2017; Valor-Segura, Expósito, & Moya, 2011). Con respecto a la severidad percibida de la VG, las tres escalas se relacionaron negativamente con esta variable, indicando, como se esperaba, que los participantes con mayores niveles de estos tres tipos de actitudes suelen percibir como menos graves los casos de VG (Gracia & Herrero, 2006a; Muehlenhard & Kimes, 1999; Taylor & Sorenson, 2005; Witte, Schroeder, & Lohr, 2006).

Al comparar las puntuaciones en las tres escalas por sexo, se encontró en línea con investigaciones previas que las mujeres tienden a presentar actitudes hacia la VG más negativas que los hombres (Carlson & Worden, 2005; Fincham et al., 2008; Flood & Pease, 2009; Gracia & Herrero, 2006b). En los

estudios cuarto y quinto, se encontró además que agresores condenados por VG son más propensos a mostrar altos niveles de aceptabilidad de la VG y de culpabilización a las víctimas de VG que los hombres de la población general, lo que también es congruente con estudios previos (Archer & Graham-Kevan, 2003; Gracia, Rodriguez, & Lila, 2015; Lila, Gracia, & García, 2013). Esto subraya que ambas escalas son en efecto especialmente informativas para detectar aquellas personas con mayores niveles de estos tipos de actitudes hacia la VG.

Estos tres últimos estudios no están exentos de limitaciones. El método de muestreo online es una de las limitaciones de estos estudios, ya que esta estrategia tiene algunas dificultades que pueden afectar a la generalizabilidad de los resultados. A pesar de que este método de recogida de datos es eficiente y efectivo para recolectar grandes tamaños muestrales en poco tiempo, es más difícil verificar la información sociodemográfica proporcionada por los participantes (Thornton et al., 2016; Topolovec-Vranic & Natarajan, 2016). El sesgo de auto-selección puede suponer asimismo otro problema del muestro online, ya que los participantes más motivados por la temática de los estudios pueden estar más dispuestos a participar en los mismos. No obstante, las características sociodemográficas de las muestras recogidas en los tres últimos estudios de esta tesis doctoral son consistentes con otros estudios demográficos realizados con muestras de usuarios de internet (Acebes Arribas, 2016). Otra limitación es la ausencia de análisis de invarianza factorial para las escalas A-IPVAW y VB-IPVAW, ya que los análisis DIF son sólo un primer paso en el análisis del posible sesgo de los ítems de una escala (Gómez-Benito, Sireci, Padilla, Hidalgo, & Benítez, 2018). Por último, el efecto de la deseabilidad social sobre los ítems de las escalas A-IPVAW y VB-IPVAW debería de evaluarse apropiadamente en futuros estudios, bien mediante el

empleo de medidas implícitas de actitudes hacia la VG (Gracia et al., 2015), o bien a través de modelos factoriales que modelen este sesgo (Ferrando, 2005).

Finalmente, los estudios cuarto y quinto de la presente tesis doctoral proporcionan una versión breve de la escala VB-IPVAW, y dos versiones breves de la escala WI-IPVAW, desarrolladas siguiendo las directrices de Goetz et al. (2013). Estas versiones breves resultan especialmente adecuadas para utilizarse en aquellos estudios en los que el espacio (e.g., cuando se evalúan una gran cantidad de variables) o el tiempo de aplicación puedan ser un problema.

En conclusión, los tres últimos estudios de esta tesis presentan el desarrollo y la validación de tres medidas complementarias de actitudes públicas hacia la VG con una adecuada consistencia interna y con suficientes evidencias de validez para ser empleadas en contextos de investigación o de intervención, pues son especialmente informativas para discriminar con precisión entre aquellas personas con altos niveles de estos tipos de actitudes.

Implicaciones y futuras líneas de investigación

De los resultados obtenidos en la presente tesis doctoral se derivan las siguientes implicaciones. La primera de ellas pone de manifiesto la importancia de utilizar modelos y análisis estadísticos adecuados para asegurar la comparabilidad entre países de las medidas utilizadas en las grandes encuestas poblacionales. Evaluar la invarianza factorial o el DIF de cualquiera de estas medidas es un requisito previo fundamental para garantizar la validez de las comparaciones realizas en las grandes encuestas (Davidov et al., 2008; 2014). Sin embargo, los estudios analizando este tipo de cuestiones en la literatura sobre VG han sido más bien escasos hasta el momento, lo que constituye una de las principales aportaciones de esta tesis.

El empleo de modelos factoriales para estudiar las propiedades psicométricas (i.e., fiabilidad, estructura latente, validez) de las medidas utilizadas en una gran encuesta poblacional es otra de las principales aportaciones de esta tesis. Este tipo de modelos permite dotar de mayor significado a las comparaciones realizadas entre países. A diferencia de los porcentajes de las tasas de prevalencia, comparar las medias factoriales de VG física, sexual y psicológica a través de un análisis de medias latentes, permite determinar el tamaño del efecto de estas comparaciones, cuantificando así la diferencia entre los diferentes países de la UE en los niveles de estos tipos de violencia. De este modo es posible determinar qué porcentaje de individuos de un país presenta, en promedio, niveles más altos de un tipo de VG que en otro país (Milfont & Fisher, 2010; Putnick & Bornstein, 2016; Ruscio, 2008). Asimismo, también es posible establecer el grado de solapamiento entre los 28 países de la UE en cada tipo de VG.

Las nuevas escalas de actitudes hacia la VG presentadas en esta tesis —A-IPVAW, VB-IPVAW, WI-IPVAW— conforman otra de las principales aportaciones de la misma, pues permiten llevar a cabo una mejor evaluación de estos tres tipos de actitudes que con las medidas existentes hasta el momento. Estas tres nuevas medidas pueden emplearse tanto para profundizar en cuestiones de investigación sobre la VG, como en contextos de intervención con agresores de VG, ya sea para evaluar los cambios actitudinales de los participantes durante la intervención, o bien como una herramienta de detección y discriminación entre aquellos individuos con altos niveles de aceptación de la VG, de culpabilización a las víctimas de VG, o con bajos niveles de disposición a intervenir en casos de VG (Carbajosa, Catalá-Miñana, Lila, & Gracia, 2017; Ferrer-Perez, Ferreiro-Basurto, Navarro-Guzmán, & Bosch-Fiol, 2016; Lila, Gracia, & Catalá-Miñana, 2018; Powell & Webster, 2018). Además, el desarrollo de las versiones breves de las escalas

VB-IPVAW y WI-IPVAW hace posible incluir estas escalas en grandes estudios poblacionales donde el espacio es limitado. Dada la importancia que las actitudes públicas hacia la VG pueden llegar a tener para entender y explicar este fenómeno, y dada la escasa presencia de este tipo de medidas en las grandes encuestas, la potencial inclusión de estas medidas en encuestas poblacionales puede resultar de extrema utilidad para la investigación sobre la VG (Gracia & Lila, 2015).

Finalmente, los resultados encontrados en la presente tesis doctoral permiten la apertura de nuevas líneas de investigación, así como la continuación de las existentes abordando algunas de las limitaciones de los estudios que la componen. En primer lugar, si bien se ha establecido el nivel de invarianza escalar para las escalas de VG física, psicológica y sexual de la encuesta del FRA, todavía no se ha estudiado si estas medidas son invariantes cuando la violencia no es ejercida por la pareja de las participantes en esta encuesta (e.g., compañeros de trabajo, conocidos, desconocidos). Asimismo, también podría estudiarse la invarianza factorial de otras medidas incluidas por el FRA, como son los conjuntos de preguntas de acoso sexual (*sexual harrasment*) y de acoso (*stalking*). Para ello, además de los procedimientos de invarianza factorial analíticos —como los utilizados en esta tesis—, podrían emplearse métodos de estimación bayesiana para estudiar la invarianza por aproximación, sorteando así las potenciales dificultades de los métodos de estimación habituales para datos categóricos (Cieciuch, Davidov, Schmidt, Algesheimer, & Schwartz, 2014; Muthen & Asparouhov, 2013, 2018; Van de Schoot, Schmidt, De Beuckelaer, Lek, & Zondervan-Zwijnenburg, 2015). Otra línea de investigación que queda pendiente del cuarto estudio de esta tesis es el desarrollo de una versión breve de la escala A-IPVAW para su potencial inclusión en grandes encuestas poblacionales, así como el estudio en profundidad de la invarianza factorial de la misma. Finalmente, es necesaria

más investigación para adaptar las escalas A-IPVAW, VB-IPVAW y WI-IPVAW, y generalizar los resultados obtenidos a otras sociedades (Boira, Carbajosa, & Mendez, 2016; Gracia & Lila, 2015; Ivert, Merlo, & Gracia, 2018).

Conclusión

A lo largo del desarrollo de la presente tesis doctoral se ha tratado de dar respuesta a dos problemas de medición relacionados con la VG: la comparabilidad de las medidas de diferentes formas de esta violencia en una gran encuesta poblacional, y el desarrollo de medidas adecuadas —esto es, con suficiente precisión y evidencias de validez—, para evaluar tres aspectos diferenciados de las actitudes públicas hacia la VG, una de las variables clave en el estudio de este fenómeno. Los resultados obtenidos en los seis estudios que conforman el corpus principal de esta tesis, subrayan la importancia de utilizar metodologías adecuadas y actualizadas para abordar estos dos problemas. Los tres primeros estudios ponen de manifiesto las ventajas de emplear modelos factoriales para examinar la distribución de los niveles de VG entre los 28 países de la UE, ya que permiten dotar de mayor significado a las comparaciones realizadas entre ellos que los porcentajes de las tasas de prevalencia de este tipo de violencia. Por su parte, en los tres últimos estudios de esta tesis se proporcionan tres nuevas escalas para evaluar con suficientes garantías las actitudes de aceptabilidad de la VG, de culpabilización a las víctimas para la VG, y la disposición a intervenir en casos de VG. Estas escalas suponen un avance en el estudio de las actitudes públicas hacia la VG, pues posibilitan ampliar el conocimiento sobre las mismas ahondando en su conceptualización, evaluación, prevalencia y en los factores sociales que pueden influir en estas actitudes para mejorar las estrategias de prevención e intervención de la VG.

Conclusion

Two measurement problems related to IPVAW were addressed in the present thesis dissertation: the comparability of the measures assessing different forms of this violence used in a large survey across the EU, and the development of appropriate measures (i.e., reliable and valid), to evaluate three aspects of the public attitudes towards IPVAW, which are one of the key variables in the study of this type of violence. To this end, six studies were carried out. Taken together, the findings of this thesis dissertation underlined the importance of utilizing appropriate and up-to-date methodologies to address these issues. The three first studies showed the advantages of using factorial models to examine how the levels of IPVAW are distributed across all EU countries, allowing us to conduct more meaningful and refined cross-country comparisons than using the percentages of the prevalence rates. The three last studies of this dissertation provided three new scales to assess with enough accuracy and evidences of validity three aspects of the public attitudes towards IPVAW: the acceptability of IPVAW, victim-blaming attitudes in cases of IPVAW, and the willingness to intervene in cases of IPVAW. The development of these measures is a step forward in the study of attitudes towards IPVAW, allowing researchers to extend knowledge about their conceptualization, measurement, prevalence, and the social factors that may influence these attitudes in order to improve prevention and intervention strategies.

Referencias

Abramsky, T., Watts, C. H., Garcia-Moreno, C., Devries, K., Kiss, L., Ellsberg, M., & Heise, L. (2011). What factors are associated with recent intimate partner violence? Findings from the WHO multi-country study on women's health and domestic violence. *BMC Public Health*, *11*, 109-120.

Acebes Arribas, B. (2016). *Estudio de medios de comunicación on line*. Madrid: IAB.

Ali, P. A., & Naylor, P. B. (2013). Intimate partner violence: A narrative review of the feminist, social and ecological explanations for its causation. *Aggression and Violent Behavior*, *18*, 611-619.

American Educational Research Association, American Psychological Association & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Ansara, D. L., & Hindin, M. J. (2010). Formal and informal help-seeking associated with women's and men's experience of intimate partner violence in Canada. *Social Science and Medicine, 70*, 1011-1018

Appel, A. E., & Holden, G. W. (1998). The co-occurrence of spouse and physical child abuse: a review and appraisal. *Journal of Family Psychology*, *12*, 578-599.

Archer, J., & Graham-Kevan, N. (2003). Do beliefs about aggression predict physical aggression to partners? *Aggressive Behavior, 29,* 41-54.

Baldry, A. C., Pacilli, M. G., and Pagliaro, S. (2015). She's not a person.. she's just a woman! Infra-Humanization and intimate partner violence. *Journal of Interpersonal Violence,* 30, 1567–1582.

Baldry, A. C., & Pagliaro, S. (2014). Helping victims of intimate partner violence: the influence of group norms among lay people and the police. *Psychological Violence,* 4, 334–347.

Banyard, V., & Moynihan, M. M. (2011). Variation in bystander behavior related to sexual and intimate partner violence prevention: correlates in a sample of college students. *Psychological Violence,* 1, 287–301.

Banyard, V. L., Moynihan, M. M., Cares, A. C., & Warner, R. (2014). How do we know if it works? Measuring outcomes in bystander-focused abuse prevention on campuses. *Psychological Violence,* 4, 101–115.

Beeble, M. L., Bybee, D., & Sullivan, C. M. (2007). Abusive men's use of children to control their partners and ex-partners. *European Psychologist*, *12*, 54-61.

Beck, C. T., & Gable, R. K. (2001). Comparative analysis of the performance of the Postpartum Depression Screening Scale with two other depression instruments. *Nursing Research*, *50*, 242-250.

Boira, S., Carbajosa, P., & Méndez, R. (2016). Miedo, conformidad y silencio: la violencia en las relaciones de pareja en áreas rurales de Ecuador [Fear, conformity and silence. Intimate partner violence in rural areas of Ecuador]. *Psychosocial Intervention*, *25*, 9-17.

Branch, K. A., Richards, T. N., & Dretsch, E. C. (2013). An exploratory analysis of college students' response and reporting behavior regarding intimate partner violence victimization and perpetration among their friends. *Journal of Interpersonal Violence*, 28, 3386–3399.

Browning, C. R. (2002). The span of collective efficacy: Extending social disorganization theory to partner violence. *Journal of Marriage and Family*, *64*, 833-850.

Bryant, S. A., & Spencer, G. A. (2003). University students' attitudes about attributing blame in domestic violence. *Journal of Family Violence*, *18*, 369-376.

Cai, L. (2010). High-dimensional exploratory item factor analysis by a Metropolis-Hastings Robbins-Monro algorithm. *Psychometrika, 75*, 33-57.

Campbell, J. C. (2002). Health consequences of intimate partner violence. *The Lancet*, *359,* 1331-1336.

Capaldi, D. M., Knoble, N. B., Shortt, J. W., & Kim, H. K. (2012). A systematic review of risk factors for intimate partner violence. *Partner Abuse*, *3*, 231-280.

Carbajosa, P., Catalá-Miñana, A., Lila, M., & Gracia, E. (2017). Differences in treatment adherence, program completion, and recidivism among batterer subtypes. *The European Journal of Psychology Applied to Legal Context, 9,* 93-101.

Carlson, B. E., & Worden, A. P. (2005). Attitudes and beliefs about domestic violence: Results of a public opinion survey I. Definitions of domestic violence, criminal domestic violence, and prevalence. *Journal of Interpersonal Violence*, *20*, 1197-1218.

Capezza, N. M., & Arriaga, X. B. (2008). Factors associated with acceptance of psychological aggression against women. *Violence Against Women*, *14*, 612-633.

Carney, M. M., & Barner, J. R. (2012). Prevalence of partner abuse: Rates of emotional abuse and control. *Partner Abuse*, *3*, 286-335.

Cinquegrana, V., Baldry, A. C., & Pagliaro, S. (2018). Intimate partner violence and bystanders' helping behaviour: an experimental study. *Journal of Aggression, Conflict, and Peace Resesearch,* 10, 24–35

Cieciuch, J., Davidov, E., Schmidt, P., Algesheimer, R., & Schwartz, S. H. (2014). Comparing results of an exact *vs.* an approximate (Bayesian) measurement invariance test: a cross-country illustration with a scale to measure 19 human values. *Frontiers in Psychology*, *5*, 982.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*, 464-504.

Chernyshenko, O. S., Stark, S., Chan, K. Y., Drasgow, F., & Williams, B. (2001). Fitting item response theory models to two personality inventories: Issues and insights. *Multivariate Behavioral Research*, *36*, 523-562.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233-255.

Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software*, *39*, 1-30.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hilsdale: Lawrence Earlbaum Associates.

Coker, A. L., Smith, P. H., McKeown, R. E., & King, M. J. (2000). Frequency and correlates of intimate partner violence by type: physical, sexual, and psychological battering. *American Journal of Public Health*, *90*, 553-559.

Copp, J. E., Giordano, P. C., Longmore, M. A., & Manning, W. D. (2016). The development of attitudes toward intimate partner violence: An examination of key correlates among a sample of young adults. *Journal of Interpersonal Violence*. Retrieved from: http://journals.sagepub.com/doi/pdf/10.1177/088626 0516651311

Craparo, G., Gori, A., Petruccelli, I., Cannella, V., & Simonelli, C. (2014). Intimate partner violence: relationships between alexithymia, depression, attachment styles, and coping strategies of battered women. *The Journal of Sexual Medicine*, *11*, 1484-1494.

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, *40*, 55-75.

Davidov, E., Schmidt, P., & Schwartz, S. H. (2008). Bringing values back in: The adequacy of the European Social Survey to measure values in 20 countries. *Public Opinion Quarterly*, *72*, 420-445.

Delgado-Rico, E., Carrctero-Dios, H., & Ruch, W. (2012). Content validity evidences in test development: An applied perspective. *International Journal of Clinical and Health Psychology*, *12*, 449-459.

Devries, K. M., Mak, J. Y., Garcia-Moreno, C., Petzold, M., Child, J. C., Falder, G., ... & Pallitto, C. (2013). The global prevalence of intimate partner violence against women. *Science*, *340*, 1527-1528.

Eagly, A., & Chaiken, S. (1998). Attitude structure. In D. Gilbert, S. Fiske, y G. Lindsay (Eds.), *Handbook of social psychology* (pp. 269-322). New York: McGraw-Hill.

Eagly, A. H., & Chaiken, S. (2005). *Attitude Research in the 21st Century: The Current State of Knowledge*. New York: McGraw-Hill.

Eagly, A. H., & Chaiken, S. (2007). The advantages of an inclusive definition of attitude. *Social Cognition*, *25*, 582-602.

EIGE – European Institute for Gender Equality. (2017). *Gender Equality Index, 2017. Measuring gender equality in the European Union 2005-2015*. Luxembourg: Publication Office of European Union.

Ellsberg, M., Jansen, H. A., Heise, L., Watts, C. H., & Garcia-Moreno, C. (2008). Intimate partner violence and women's physical and mental health in

the WHO multi-country study on women's health and domestic violence: an observational study. *The Lancet*, *371*, 1165-1172.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, New Jersey: Lawrence Erlbaum.

European Commission (2016). *Gender Based Violence (Special Eurobaromenter 449 – Wave EB85.3 – TNS Opinion & Social).* Brussels: European Commission.

Fernández-González, L., Calvete, E., & Orue, I. (2017). Mujeres víctimas de violencia de género en centros de acogida: características sociodemográficas y del maltrato. *Psychosocial Intervention*, *26*, 9-17.

Ferrando, P. J. (2005). Factor analytic procedures for assessing social desirability in binary items. *Multivariate Behavioral Research*, *40*, 331-349.

Ferrer-Perez, V. A., Ferreiro-Basurto, V., Navarro-Guzmán, C., & Bosch-Fiol, E. (2016). Programas de intervención con maltratadores en España: la perspectiva de los/as profesionales. *Psychosocial Intervention*, *25*(3), 159-168.

Fincham, F. D., Cui, M., Braithwaite, S., & Pasley, K. (2008). Attitudes toward intimate partner violence in dating relationships. *Psychological Assessment*, *20*, 260-269.

Flood, M., & Pease, B. (2009). Factors influencing attitudes to violence against women. *Trauma, Violence, & Abuse*, *10*, 125-142.

Fox, K. A., & Cook, C. L. (2011). Is knowledge power? The effects of a victimology course on victim blaming. *Journal of Interpersonal Violence*, *26*, 3407-3427.

Follingstad, D. R., Coyne, S., & Gambone, L. (2005). A representative measure of psychological aggression and its severity. *Violence and Victims*, *20*, 25-38.

FRA – European Union Agency for the Fundamental Rights (2014a). *Violence Against Women: An EU-Wide Survey.* Luxembourg: Publications Office of the European Union.

FRA – European Union Agency for the Fundamental Rights (2014b). *Violence against women: an EU-wide survey: Survey methodology*. Luxembourg: Publications Office of the European Union.

French, A. W., & Miller, T. R. (1996). Logistic regression and its use in detecting differential item functioning in polytomous items. *Journal of Educational Measurement*, *33*, 315-332.

García-Moreno, C., Hegarty, K., d'Oliveira, A. F. L., Koziol-McLain, J., Colombini, M., & Feder, G. (2015). The health-systems response to violence against women. *The Lancet*, *385*, 1567-1579.

Garcia-Moreno, C., Jansen, H. A., Ellsberg, M., Heise, L., & Watts, C. H. (2005). WHO multi-country study on women's health and domestic violence against women. *Geneva: World Health Organization*, *204*, 1-18.

Garcia-Moreno, C., Jansen, H. A., Ellsberg, M., Heise, L., & Watts, C. H. (2006). Prevalence of intimate partner violence: findings from the WHO multi-country study on women's health and domestic violence. *The Lancet*, *368*, 1260-1269.

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2013). A new look at Horn's parallel analysis with ordinal variables. *Psychological Methods*, *18*, 454-474.

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2016). Are fit indices really fit to estimate the number of factors with categorical variables? Some cautionary findings via Monte Carlo simulation. *Psychological Methods*, *21*, 93.

Garrido-Macías, M., Valor-Segura, I., & Expósito, F. (2017). ¿Dejaría a mi pareja? Influencia de la gravedad de la transgresión, la satisfacción y el

compromiso en la toma de decisión. *Psychosocial Intervention*, *26*, 111-116.

Glick, P., Sakalli–Ugurlu, N., Ferreira, M. C., & Aguiar de Souza, M. (2002). Ambivalent sexism and attitudes toward wife abuse in Turkey and Brazil. *Psychology of Women Quarterly*, *26*, 292-297.

Goetz, C., Coste, J., Lemetayer, F., Rat, A. C., Montel, S., Recchia, S., ... & Guillemin, F. (2013). Item reduction based on rigorous methodological guidelines is necessary to maintain validity when shortening composite measurement scales. *Journal of Clinical Epidemiology*, *66*, 710-718.

Gómez-Benito, J., Sireci, S., Padilla, J. L., Hidalgo, M. D., & Benítez, I. (2018). Differential Item Functioning: Beyond validity evidence based on internal structure. *Psicothema*, *30*, 104-109.

Gracia, E. (2014). Intimate partner violence against women and victim-blaming attitudes among Europeans. *Bulletin of the World Health Organization*, *92*, 380-381.

Gracia, E., García, F., & Lila, M. (2008). Police involvement in cases of intimate partner violence against women: The influence of perceived severity and personal responsibility. *Violence Against Women*, *14*, 697-714.

Gracia, E., García, F., & Lila, M. (2011). Police attitudes toward policing partner violence against women: Do they correspond to different psychosocial profiles? *Journal of Interpersonal Violence*, *26*, 189-207.

Gracia, E., & Herrero, J. (2006a). Acceptability of domestic violence against women in the European Union: A multilevel analysis. *Journal of Epidemiology and Community Health*, *60*, 123-129.

Gracia, E., & Herrero, J. (2006b). Public attitudes toward reporting partner violence against women and reporting behavior. *Journal of Marriage and Family*, *68*, 759-768.

Gracia, E., Herrero, J., Lila, M., & Fuente, A. (2009). Percepción de desorden social en el vecindario y actitudes hacia la violencia doméstica contra la mujer en inmigrantes Latinoamericanos. *European Journal of Psychology Applied to Legal Contexts*, *1*, 25-43.

Gracia, E., & Lila, M. (2015). *Attitudes towards violence against women in the EU*. Luxembourg: Publication Office of the European Union.

Gracia, E., Martín-Fernández, M., Marco, M., Santirso, F. A., Vargas, V., & Lila, M. (2018). The Willingness to Intervene in Cases of Intimate Partner Violence Against Women (WI-IPVAW) scale: Development and validation of the long and short versions. *Frontiers in Psychology*, *9*, 1146.

Gracia, E., & Merlo, J. (2016). Intimate partner violence against women and the Nordic paradox. *Social Science and Medicine*, *157*, 27-30.

Gracia, E., Rodriguez, C. M., & Lila, M. (2015). Preliminary evaluation of an analog procedure to assess acceptability of intimate partner violence against women: The Partner Violence Acceptability Movie Task. *Frontiers in Psychology*, *6:1567*.

Gracia, E., Rodriguez, C. M., Martín-Fernández, M., & Lila, M. (2017). Acceptability of family violence: Underlying ties between intimate partner violence and child abuse. *Journal of Interpersonal Violence*, 0886260517707310.

Gracia, E., & Tomás, J. M. (2014). Correlates of victim-blaming attitudes regarding partner violence against women among the Spanish general population. *Violence Against Women*, *20*, 26-41.

Guedes, A., Bott, S., Garcia-Moreno, C., & Colombini, M. (2016). Bridging the gaps: a global review of intersections of violence against women and violence against children. *Global Health Action,* 9:31516. doi: 10.3402/gha.v9.31516

Hanel, P. H., & Mehler, D. M. (2019). Beyond reporting statistical significance: Identifying informative effect sizes to improve scientific communication. *Public Understanding of Science*, *28*, 468-485.

Hamberger, L. K., Larsen, S. E., & Lehrner, A. (2017). Coercive control in intimate partner violence. *Aggression and Violent Behavior*, *37*, 1-11.

Hamby, S., & Grych, J. (2013). *The web of violence: Exploring connections among different forms of interpersonal violence and abuse*. London: Springer.

Hamby, S. L., & Sugarman, D. B. (1999). Acts of psychological aggression against a partner and their relation to physical assault and gender. *Journal of Marriage and the Family*, *61*, 959-970.

Heise, L. L. (1998). Violence against women: An integrated, ecological framework. *Violence Against Women*, *4*, 262-290.

Heise, L. (2011). *What works to prevent partner violence? An evidence overview*. London: London School of Hygiene and Tropical Medicine. Retrieved from: http://strive.lshtm.ac.uk/resources/ what-works-prevent-partner-violence-evidence-overview

Heise, L. L., & Kotsadam, A. (2015). Cross-national and multilevel correlates of partner violence: an analysis of data from population-based surveys. *The Lancet Global Health*, 3, e332-e340.

Heise, L., Pallitto, C., García-Moreno, C., & Clark, C. J. (2019). Measuring psychological abuse by intimate partners: Constructing a cross-cultural indicator for the Sustainable Development Goals. *SSM-Population Health*, 100377.

Henning, K., & Holdford, R. (2006). Minimization, denial, and victim blaming by batterers: How much does the truth matter?. *Criminal Justice and Behavior*, *33*, 110-130.

Henning, K., Jones, A. R., & Holdford, R. (2005). "I didn't do it, but if I did I had a good reason": Minimization, denial, and attributions of blame among

male and female domestic violence offenders. *Journal of Family Violence*, *20*, 131-139.

Herrera, M. C., Expósito, F., & Moya, M. (2012). Negative reactions of men to the loss of power in gender relations: Lilith vs. Eve. *The European Journal of Psychology Applied to Legal Context*, *4*, 17-42.

Herrero, J., Rodríguez, F. J., & Torres, A. (2017). Acceptability of partner violence in 51 societies: The role of sexism and attitudes toward violence in social relationships. *Violence Against Women*, *23*, 351-367.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Ivert, A. K., Merlo, J., & Gracia, E. (2017). Country of residence, gender equality and victim blaming attitudes about partner violence: a multilevel analysis in EU. *The European Journal of Public Health*, *28*, 559-564.

Jackson, T. L., Dienst, R. D., Efird, T. L., Mobley, B. D., Schroeder, D. A., Hout, A. D., . . . , & LaBine, A. L, (1994). The Violence Attitudes Scale (VAS). In L. VandeCreek, S. Knapp & T. L. Jackson (Eds.), *Innovations in clinical practice: A source book* (Vol. 13, pp. 279-291). Sarasota, FL: Professional Resource Press/Professional Resource Exchange.

Jang, S., Kim, E. S., Cao, C., Allen, T. D., Cooper, C. L., Lapierre, L. M., et al. (2017). Measurement invariance of the satisfaction with life scale across 26 countries. *Journal of Cross-Cultural Psychology*, *48*, 560-576.

Jewkes, R., Flood, M., & Lang, J. (2015). From work with men and boys to changes of social norms and reduction of inequities in gender relations: a conceptual shift in prevention of violence against women and girls. *The Lancet*, *385*, 1580-1589.

Juarros-Basterretxea, J., Overall, N., Herrero, J., & Rodríguez-Díaz, F. J. (2019). Considering the Effect of Sexism on Psychological Intimate Partner

Violence: A Study with Imprisoned Men. *European Journal of Psychology Applied to Legal Context*, *11,* 61-69.

Karakurt, G., & Silver, K. E. (2013). Emotional abuse in intimate relationships: The role of gender and age. *Violence and victims*, *28*, 804-821.

Kaukinen, C. (2004). Status compatibility, physical violence, and emotional abuse in intimate relationships. *Journal of Marriage and Family*, *66*, 452-471.

Koepke, S., Eyssel, F., & Bohner, G. (2014). "She Deserved It" Effects of Sexism Norms, Type of Violence, and Victim's Pre-Assault Behavior on Blame Attributions Toward Female Victims and Approval of the Aggressor's Behavior. *Violence Against Women*, *20*, 446-464.

Koepsell, J. K., Kernic, M. A., & Holt, V. L. (2006). Factors that influence battered women to leave their abusive relationships. *Violence Victims, 21*, 131–147

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*, 936-949.

Liang, B., Goodman, L., Tummala-Narra, P., & Weintraub, S. (2005). A theoretical framework for understanding help-seeking processes among survivors of intimate partner violence. *American Journal of Community Psychology*, *36*, 71-84.

Lila, M., Gracia, E., & Catalá-Miñana, A. (2018). Individualized motivational plans in batterer intervention programs: A randomized clinical trial. *Journal of Consulting and Clinical Psychology*, *86*, 309-320.

Lila, M., Gracia, E., & García, F. (2013). Ambivalent sexism, empathy and law enforcement attitudes towards partner violence against women among male police officers. *Psychology, Crime & Law*, *19*, 907-919.

Lila, M., Oliver, A., Catalá-Miñana, A., Galiana, L., & Gracia, E. (2014). The intimate partner violence responsibility attribution scale (IPVRAS). *The European Journal of Psychology Applied to Legal Context*, *6*, 29-36.

Liles, S., Usita, P., Irvin, V. L., Hofstetter, C. R., Beeston, T., & Hovell, M. F. (2012). Prevalence and correlates of intimate partner violence among young, middle, and older women of Korean descent in California. *Journal of Family Violence*, *27*, 801-811.

Lubke, G. H., Dolan, C. V., Kelderman, H., & Mellenbergh, G. J. (2003). On the relationship between sources of within-and between-group differences and measurement invariance in the common factor model. *Intelligence*, *31*, 543-566.

Little, L., & Kaufman Kantor, G. (2002). Using ecological theory to understand intimate partner violence and child maltreatment. *Journal of Community Health Nursing*, *19*, 133-145.

Lövestad, S., & Krantz, G. (2012). Men's and women's exposure and perpetration of partner violence: an epidemiological study from Sweden. *BMC Public Health*, *12*, 945.

Lundgren, E., Heimer, G., Westerstrand, J., & Kalliokoski, A. M. (2001). *Mens violence against women in" equal" Sweden-a prevalence study*. Stockholm: Brottsoffermyndigheten and Uppsala Universitet.

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35*, 382-385.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*, 130-149.

Maio, G. R., Olson, J. M., Bernard, M. M., & Luke, M. A. (2006). Ideologies, values, attitudes, and behavior. In *Handbook of social psychology* (pp. 283-308). Springer: Boston.

Martín-Fernández, M., Gracia, E., & Lila, M. (2018). Assessing victim-blaming attitudes in cases of intimate partner violence against women: Development and validation of the VB-IPVAW scale. *Psychosocial Intervention*, *27*, 133-143.

Martín-Fernández, M., Gracia, E., Marco, M., Vargas, V., Santirso, F. A., & Lila, M. (2018). Measuring acceptability of intimate partner violence against women: Development and validation of the A-IPVAW scale. *European Journal of Psychology Applied to Legal Context*, *10*, 26-34.

Maydeu-Olivares, A., & Garcia-Forero, C. (2010). Goodness-of-fit testing. *International Encyclopedia of Education*, *7*, 190-196.

Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psycho-metrika*, *71*, 713.

McDonnell, K. A., Burke, J. G., Gielen, A. C., O'Campo, P., & Weidl, M. (2011). Women's perceptions of their community's social norms towards assisting women who have experienced intimate partner violence. *Journal of Urban Health, 88*, 240–253.

McMahon, S., Allen, C. T., Postmus, J. L., McMahon, S. M., Peterson, N. A., & Hoffman, M. L. (2014). Measuring bystander attitudes and behavior to prevent sexual violence. *Jorunal of American College Health, 62*, 58–66.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, *93*, 568-592.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, *18*, 5-11.

Miles, J., & Shevlin, M. (2001). *Applying Regression and Correlation: A Guide for Students and Researchers*. Sage: London.

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: applications in crosscultural research. *International Journal of Psychological Research, 3*, 111–121.

Muehlenhard, C. L., & Kimes, L. A. (1999). The social construction of violence: The case of sexual and domestic violence. *Personality and Social Psychology Review, 3*, 234-245.

Murphy, C. M., & O'Leary, K. D. (1989). Psychological aggression predicts physical aggression in early marriage. *Journal of Consulting and Clinical Psychology*, *57*, 579-582.

Muthén, B., & Asparouhov, T. (2018). Recent methods for the study of measurement invariance with many groups: alignment and random effects. *Sociological Methods and Research*, *47*, 637-664.

Muthén, B., & Asparouhov, T. (2013). BSEM measurement invariance analysis. *Mplus web notes*, *17*, 1-48.

Nybergh, L., Taft, C., Enander, V., & Krantz, G. (2013). Self-reported exposure to intimate partner violence among women and men in Sweden: results from a population-based survey. *BMC Public Health*, *13*, 845.

Petretic-Jackson, P., Sandberg, G., & Jackson, T. (1994). The Domestic Violence Blame scale. In L. VandeCreek, S. Knapp & T. Jackson (Eds.), *Innovations in clinical practice: A source book* (Vol. 13, pp. 101-118). Sarasota, FL: Professional Resource Press/ Professional Resource Exchange.

Polit, D. F., & Beck, C. T. (2006). The content validity index: are you sure you know what's being reported? Critique and recom- mendations. *Research in Nursing & Health*, *29*, 489-497.

Powell, A., & Webster, K. (2018). Cultures of gendered violence: An integrative review of measures of attitudinal support for violence against women. *Australian & New Zealand Journal of Criminology, 51,* 40-57.

Putnick, D.L., & Bornstein, M.H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*, 71-90.

Reichel, D. (2017). Determinants of intimate partner violence in Europe: The role of socioeconomic status, inequality, and partner behavior. *Journal of Interpersonal Violence*, *32*, 1853-1873.

Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*, 354-373.

Rizo, C. F., & Macy, R. J. (2011). Help seeking and barriers of Hispanic partner systematic review of the literature. *Aggression and Violent Behavior, 6,* 250-264.

Romans, S., Forte, T., Cohen, M. M., Du Mont, J., & Hyman, I. (2007). Who is most at risk for intimate partner violence? A Canadian population-based study. *Journal of Interpersonal Violence*, *22*, 1495-1514.

Ruscio, J. (2008). A probability-based measure of effect size: Robustness to base rates and other factors. *Psychological Methods*, *13*, 19-30.

Salis, K. L., Salwen, J., & O'Leary, K. D. (2014). The predictive utility of psychological aggression for intimate partner violence. *Partner Abuse*, *5*, 83-97.

Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores.* New York: Psychometric Society.

Sass, D. A., Schmitt, T. A., & Marsh, H. W. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. *Structural Equation Modeling*, *21*, 167-180.

Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, *29*, 304-321.

Schumacher, J. A., Feldbau-Kohn, S., Slep, A. M. S., & Heyman, R. E. (2001). Risk factors for male-to-female partner physical abuse. *Aggression and Violent Behavior*, *6*, 281-352.

Schumacher, J. A., & Leonard, K. E. (2005). Husbands' and wives' marital adjustment, verbal aggression, and physical aggression as longitudinal predictors of physical aggression in early marriage. *Journal of Consulting and Clinical Psychology*, *73*, 28.

Scott, K., & Straus, M. (2007). Denial, minimization, partner blaming, and intimate aggression in dating partners. *Journal of Interpersonal Violence*, *22*, 851-871.

Smith, S. G., Zhang, X., Basile, K. C., Merrick, M. T., Wang, J., Kresnow, M., et al. (2018). *The National Intimate Partner and Sexual Violence Survey (NISVS): 2015 Data Brief – Updated Release*. Atlanta, GA: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention.

Sokoloff, N. J., & Dupont, I. (2005). Domestic violence at the intersections of race, class, and gender: Challenges and contributions to understanding violence against marginalized women in diverse communities. *Violence Against Women*, *11*, 38-64.

Stein, J. L. (2007). Peer educators and close friends as predictors of male college students' willingness to prevent rape. *Journal of College Student Development., 48*, 75–89.

Stets, J. E., & Straus, M. A. (1989). The marriage license as a hitting license: A comparison of assaults in dating, cohabiting, and married couples. *Journal of Family Violence*, *4*, 161-180.

Straight, E. S., Harper, F. W., & Arias, I. (2003). The impact of partner psychological abuse on health behaviors and health status in college women. *Journal of Interpersonal Violence*, *18*, 1035-1054.

Straus, M.A. (1979). Measuring intrafamily conflict and violence: The Conflict Tactics Scales. *Journal of Marriage and Family, 41*, 75-88.

Straus, M. A., Hamby, S. L., Boney-McCoy, S., & Sugarman, D. B. (1996). The revised conflict tactics scales (CTS2) development and preliminary psychometric data. *Journal of Family Issues*, *17*, 283-316.

Sugarman, D. B., & Frankel, S. L. (1996). Patriarchal ideology and wife-assault: A meta-analytic review. *Journal of Family Violence*, *11*, 13-40.

Schwab-Reese, L. M., & Renner, L. M. (2017). Attitudinal acceptance of and experiences with intimate partner violence among rural adults. *Journal of Family Violence*, *32*, 115-123.

Taylor, E., Banyard, V., Grych, J., & Hamby, S. (2016). Not all behind closed doors: examining bystander involvement in intimate partner violence. *Journal of Interpersonal Violence.* doi: 10.1177/0886 260516673629

Taylor, C. A., & Sorenson, S. B. (2005). Community-based norms about intimate partner violence: Putting attributions of fault and responsibility into context. *Sex Roles*, *53*, 573-589.

Thornton, L., Batterham, P. J., Fassnacht, D. B., Kay-Lambkin, F., Calear, A. L., & Hunt, S. (2016). Recruiting for health, medical or psychosocial research using Facebook: Systematic review. *Internet Interventions*, *4*, 72-81.

Till-Tentschert, U. (2017). The relation between violence experienced in childhood and women's exposure to violence in later life: Evidence from Europe. *Journal of Interpersonal Violence*, *32*, 1874-1894.

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality Assessment of Ordered Polytomous Items with Parallel Analysis. *Psychological Methods, 16*, 209-220.

Tjaden, P., & Thoennes, N. (2000). Prevalence and consequences of male-to-female and female-to-male intimate partner violence as measured by the National Violence Against Women Survey. *Violence Against Women*, *6*, 142-161.

Topolovec-Vranic, J., & Natarajan, K. (2016). The Use of Social Media in Recruitment for Medical Research Studies: A Scoping Review. *Journal of Medical Internet Research*, *18,* e286.

Tsai, A. C., Kakuhikire, B., Perkins, J. M., Vořechovská, D., McDonough, A. Q., Ogburn, E. L., et al. (2017). Measuring personal beliefs and perceived norms about intimate partner violence: Population-based survey experiment in rural Uganda. *PLoS Medicine*, *14*, e1002303.

United Nations Development Program. (2017). *Human Development Reports. Gender Inequality Index. 2017*. Retrieved from: http://hdr.undp.org/en/content/ gender-inequality -index-gii

Valor-Segura, I., Expósito, F., & Moya, M. (2011). Victim blaming and exoneration of the perpetrator in domestic violence: The role of beliefs in a just world and ambivalent sexism. *The Spanish Journal of Psychology*, *14*, 195-206.

Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*, 486-492.

Van De Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijnenburg, M. (2015). Measurement invariance. *Frontiers in Psychology*, *6*, 1064.

Vandello, J. A., & Cohen, D. (2003). Male honor and female fidelity: implicit cultural scripts that perpetuate domestic violence. *Journal of Personality and Social Psychology*, *84*, 997-1010.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4-70.

Vargas, V., Lila, M., Catalá-Miñana, A., & Gracia, E. (2017). Españoles e Inmigrantes Latinoamericanos condenados por violencia de género:¿ Existe un perfil diferencial?. *Acción Psicológica*, *14*(2), 51-68.

Vidal-Fernández, A., & Megías, J. L. (2014). Attributions of Blame to Battered Women when they are perceived as Feminists or as "Difficult to Deal With". *The Spanish Journal of Psychology*, *17,* e21.

Voith, L. A. (2017). Understanding the relation between neighborhoods and intimate partner violence: An integrative review. *Trauma, Violence, & Abuse, 20,* 385-397.

Waltermaurer, E. (2012). Public justification of intimate partner violence: a review of the literature. *Trauma, Violence, & Abuse, 13*, 167-175.

Wang, L. (2016). Factors influencing attitude toward intimate partner violence. *Aggression and Violent Behavior*, *29*, 72-78.

West, A., & Wandrei, M. L. (2002). Intimate partner violence a model for predicting interventions by informal helpers. *Journal of Interpersonal Violence, 17*, 972-986.

WEF – World Economic Forum. (2017). *Global Gender Gap. 2017*. Retrieved from: https://www.weforum.org/reports/the-global-gender-gap-report-2017

WHO – World Health Organization. (2002). *World report on violence and health*. Geneva: World Health Organization.

WHO – World Health Organization. (2012). *Understanding and addressing violence against women: Intimate partner violence*. Geneva: World Health Organization.

WHO – World Health Organization. (2013). *Global and regional estimates of violence against women: prevalence and health effects of intimate partner violence and non-partner sexual violence*. Geneva: World Health Organization.

Witte, T. H., Schroeder, D. A., & Lohr, J. M. (2006). Blame for intimate partner violence: An attributional analysis. *Journal of Social and Clinical Psychology*, *25*, 647-667.

Yamawaki, N., Ostenson, J., & Brown, C. R. (2009). The functions of gender role traditionality, ambivalent sexism, injury, and frequency of assault on domestic violence perception: A study between Japanese and American college students. *Violence Against Women*, *15*, 1126-1142.

Yount, K. M., Halim, N., Hynes, M., & Hillman, E. (2011). Response effects to attitudinal questions about domestic violence against women: A comparative perspective. *Social Science Research, 40*, 873–884.

Yun, S. H., & Vonk, M. E. (2011). Development and initial validation of the Intimate Violence Responsibility Scale (IVRS). *Research on Social Work Practice*, *21*, 562-571.