

Evaluación de

Versión revisada de la

Conflict Tactics Scales

∞CTS-2∞

RESUMEN DE LA VALORACIÓN DEL TEST

Descripción general

Característica	Descripción
Nombre del test	Versión revisada de la Conflict Tactics Scales (CTS-2)
Autor	Murray A. Straus, Sherry L. Hamby, Sue Boney-McCoy, y David B. Sugarman
Autor de la adaptación española	Adaptación: Ismael Loinaz Validación: Ismael Loinaz, Enrique Echeburúa, Margarita Ortiz-Tallo y Pedro J. Amor
Variable(s)	Violencia de pareja, Violencia de género, Violencia doméstica, Resolución de conflictos en la pareja
Áreas de aplicación	Psicología clínica Psicología forense Salud general y bienestar Investigación
Soporte	Papel y lápiz

Valoración general

Característica	Valoración	Puntuación
Materiales y documentación	No evaluado	--
Fundamentación teórica	Adecuada	3
Adaptación	Adecuada	3
Análisis de ítems	Buena	4
Validez: contenido	Adecuada con carencias- Adecuada	2.5
Validez: relación con otras variables	Adecuada con carencias- Adecuada	2.6
Validez: estructura interna	Adecuada con carencias	2
Validez: análisis del DIF	--	--
Fiabilidad: equivalencia	--	--
Fiabilidad: consistencia interna	Adecuada	3
Fiabilidad: estabilidad	Adecuada con carencias	2
Fiabilidad: TRI	--	--
Fiabilidad inter-jueces	--	--
Baremos e interpretación de puntuaciones	Adecuada	3

Comentarios generales

La CTS-2 es una herramienta muy popular y accesible que permite evaluar un fenómeno complejo y de alta relevancia social como es la violencia en el seno de la pareja. La escala está compuesta por 78 ítems de respuesta graduada, 39 para la evaluación de la propia conducta y 39 para la de la conducta de la pareja. Fue desarrollada originariamente en inglés (Straus et al., 1996) y se han propuesto distintas adaptaciones al español (p.ej., Corral y Calvete, 2006; Graña et al., 2013). La presente revisión se centra en la evidencia disponible para la versión en español de la CTS-2 propuesta por Loinaz y colaboradores (Loinaz, 2009; Loinaz, Echeburúa, et al., 2012), aunque inevitablemente se hace referencia también a la versión original de Straus et al. (1996).

El equipo de Loinaz estudió, en una muestra de agresores condenados por violencia contra la pareja, el comportamiento psicométrico de los 39 ítems diseñados para valorar la propia conducta. Entre los puntos fuertes de la escala, destacan que es relativamente fácil de administrar y que puede ser cumplimentada por los participantes en cuestión de unos pocos minutos, facilitando su inclusión en baterías amplias de evaluación. También es conveniente señalar que ofrece una visión relativamente completa de los comportamientos de agresión dentro de la pareja, ya que permite evaluar hasta cinco dimensiones diferentes (i.e., actos de violencia física, psicológica y sexual concretos, estrategias de resolución de conflictos, y posibles consecuencias de tales conflictos) y obtener hasta tres tipos de puntuaciones: frecuencia, prevalencia y severidad de la violencia. Además, es un instrumento versátil, ya que puede adaptarse a diversos períodos de tiempo y obtener información de un único individuo o de los dos miembros de la pareja. Esta riqueza en términos de información ofrecida, la facilidad y claridad en su administración (p.ej., instrucciones claras), la facilidad de computación e interpretación de sus puntuaciones y los niveles por lo general adecuados de los índices de consistencia interna (alfa ≥ 0.70 ; Media ≈ 0.80), ha llevado a que la CTS-2 sea una medida de referencia en psicología forense-criminal, clínica y aplicada. Esto queda ampliamente demostrado por el elevado número de contribuciones científicas que han empleado la CTS-2 para discriminar entre hombres con diferentes tipos de delitos de agresión (p.ej., Bueso-lzquierdo et al., 2016), valorar la calidad de intervenciones experimentales en el campo de la violencia entre grupos (p.ej., Lila et al., 2018), estudiar la conexión con variables de personalidad socialmente aversivas (p.ej., Redondo et al., 2019), etc. También hay que destacar que, especialmente en la diferenciación o comparación de grupos, la CTS-2 parece discriminar de forma coherente y satisfactoria entre poblaciones con diferentes características (véase el apartado sobre validez), lo que sin duda es una fortaleza de la herramienta. Por último, haber establecido, como se hace en el artículo de Loinaz, Echeburúa, et al. (2012), una clasificación en deciles para poder interpretar individualmente las puntuaciones de hombres maltratadores en personas hispanohablantes supone un salto importante en la utilidad de la medida en el sector clínico y aplicado. Considerando las dificultades para acceder a la población objetivo de la evaluación, es importante resaltar este punto como otra fortaleza.

En lo que respecta a las debilidades o aspectos de mejora de la CTS-2, se destacan las más importantes. Primero, la falta de una definición clara del dominio conductual del test y de una visión crítica-experta sobre los elementos o contenidos relevantes del constructo de interés que, a su vez, podría justificar la no inclusión de algunos indicadores en la operativización actual de la medida. Segundo, la ausencia de información sobre el proceso de adaptación de los ítems al contexto español. No se describe de forma detallada el proceso de traducción y retro-traducción, ni posibles adaptaciones a fin de facilitar la equivalencia del constructo evaluado en diferentes contextos culturales. Tercero, no se realiza, en ninguna de las dos versiones idiomáticas, un estudio en profundidad de la calidad de los ítems utilizados. En el estudio de Straus et al. (1996) se informa de la correlación ítem-test corregida, pero no por cada escala, y tampoco proporciona datos de la distribución de las puntuaciones ni de la carga factorial de los ítems (de esto último sí se informa en el estudio de Loinaz, Echeburúa, et al., 2012). Cuarto, en relación a la dimensionalidad de la CTS-2, parece no existir, al menos con muestras españolas, un estudio exhaustivo de la estructura interna de la adaptación de Loinaz (2009). Aunque Loinaz, Echeburúa, et al. (2012) reconocen que, en los 39 ítems de evaluación de la propia conducta, algunos ítems presentan cargas secundarias

importantes en otros factores y podrían ser descartados, estos ítems siguen siendo utilizados en estudios posteriores que emplean la CTS-2. Por otro lado, la escala de Daños, que presenta también limitaciones importantes en términos de fiabilidad (alfa < 0.60 en el artículo de Loinaz, Echeburúa, et al., 2012), sigue empleándose (aunque con mucha menor frecuencia que otras subescalas). El impacto de todas estas limitaciones se agrava cuando se emplea la puntuación global de la medida, dado que no existen estudios que respaldan su uso. Tampoco se ha localizado ningún estudio que legitime empíricamente, en términos de estructura interna, la división entre comportamientos de gravedad menor o gravedad mayor. Sería conveniente confirmar la adecuación de este criterio teórico. Se constata que muchos investigadores/as optan por utilizar solo algunas de las escalas de la CTS-2 (p.ej., Llor-Esteban et al., 2016; Ruiz-Hernández et al., 2015; Vargas et al., 2017), ignorando el resto, lo que, en cierto modo, refleja parte las limitaciones descritas con anterioridad.

En definitiva, la valoración global de la CTS-2 y, más concretamente, de su versión en español de Loinaz (2009), podría definirse como aceptable dado el contexto de medición (i.e., complejidad de la población objetivo), pero con un amplio margen de mejora. Resulta imperativo reflejar mejor las bases conceptuales de la medida, realizar un estudio pormenorizado de la calidad de los ítems, profundizar en las evidencias de fiabilidad (p.ej., estabilidad temporal de la medida), etc. Pero, sobre todo, situaríamos como prioridad en el estudio de la CTS-2, la necesidad de escrutar de forma exhaustiva la dimensionalidad del instrumento. Sin esto, todas las interpretaciones derivadas de las puntuaciones del test pueden quedar en entredicho, ya que la ausencia de una estructura factorial clara afecta de forma notoria al uso de las puntuaciones del test. Todos estos aspectos deberían estudiarse en ambas partes de la CTS-2, así como para las diferentes puntuaciones (p.ej., frecuencia o severidad). Finalmente, disponer de baremos y puntos de corte para población española, en ambas partes del test y en diversos contextos de aplicación (p.ej., población general), sería no solo un aporte fundamental al uso del test en el contexto español, sino también una contribución capital al campo de estudio de la violencia. No obstante, estas limitaciones psicométricas no restan utilidad al uso de la herramienta como autoinforme para cuantificar la presencia de las conductas que en ella se recogen y que pueden formar parte tanto de las dinámicas de pareja como de la realidad de víctimas y agresores.

Referencias de los estudios utilizados en la elaboración del informe

- Amaoui, S., Marín-Morales, A., Martín-Pérez, C., Pérez-García, M., & Verdejo-Román, J. (2022). Social mentalizing in male perpetrators of intimate partner violence against women is associated with resting-state functional connectivity of the Crus II. *Journal of Psychiatric Research*, *150*, 264-271. <https://doi.org/10.1016/j.jpsychires.2022.03.044>
- Amaoui, S., Martín-Pérez, C., Marín-Morales, A., Bueso-Izquierdo, N., García-León, M. J., Pérez-García, M., & Verdejo-Román, J. (2022). Resting-state functional connectivity and socioemotional processes in male perpetrators of intimate partner violence against women. *Scientific Reports*, *12*(1). <https://doi.org/10.1038/s41598-022-14181-2>
- Boira, S. & Jodrá, P. (2013). Typology of men convicted of intimate partner violence in psychological treatment in a community setting. *Revista Latinoamericana de Psicología*, *45*(2), 289-303.
- Bueso-Izquierdo, N., Hidalgo-Ruzzante, N., Burneo-Garcés, C., & Pérez-García, M. (2015). Procesamiento emocional en maltratadores de género mediante el Test de Expresiones Faciales de Ekman y la Tarea Stroop Emocional. *Revista Latinoamericana de Psicología*, *47*(2), 102-110. <https://doi.org/10.1016/j.rlp.2015.02.001>
- Bueso-Izquierdo, N., Verdejo-Román, J., Contreras-Rodríguez, O., Carmona-Perera, M., Pérez-García, M., & Hidalgo-Ruzzante, N. (2016). Are batterers different from other criminals? An fMRI study. *Social Cognitive and Affective Neuroscience*, *11*(5), 852-862. <https://doi.org/10.1093/scan/nsw020>
- 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*(2), 93-101. <https://doi.org/10.1016/j.ejpal.2017.04.001>
- Daugherty, J. C., Marañón-Murcia, M., Hidalgo-Ruzzante, N., Bueso-Izquierdo, N., Jiménez-González, P., Gómez-Medialdea, P., & Pérez-García, M. (2019). Severity of neurocognitive impairment in

- women who have experienced intimate partner violence in Spain. *Journal of Forensic Psychiatry & Psychology*, 30(2), 322-340. <https://doi.org/10.1080/14789949.2018.1546886>
- Dobash, E. R., & Dobash, R. P. (1984). The nature and antecedents of violent events. *British Journal of Criminology*, 24(3), 269-288.
- Espejo-Navarro, A.L., & Valdivia-Ramírez, D.M. (2023). *Propiedades psicométricas de la Conflict Tactics Scale en estudiantes universitarios de Lima*. Universidad Peruana de Ciencias Aplicadas. Extraído de: <https://repositorioacademico.upc.edu.pe/handle/10757/658452>
- Gallego Rodríguez, C. , & Fernández-González, L. (2019). ¿Se relaciona el consumo de pornografía con la violencia hacia la pareja?: El papel moderador de las actitudes hacia la mujer y la violencia. *Psicología Conductual*, 27(3), 431-454.
- Graña, J. L., Redondo, N., Muñoz-Rivas, M. J., & Cantos, A. L. (2014). Subtypes of batterers in treatment: Empirical support for a distinction between type I, type II and type III. *Plos One*, 9(10), e110651. <https://doi.org/10.1371/journal.pone.0110651>
- 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(4), 309-320. <https://doi.org/10.1037/ccp0000291>
- Llor-Esteban, B., García-Jiménez, J. J., Ruiz-Hernández, J. A., & Godoy-Fernández, C. (2016). Profile of partner aggressors as a function of risk of recidivism. *International Journal of Clinical and Health Psychology*, 16(1), 39-46. <https://doi.org/10.1016/j.ijchp.2015.05.004>
- Loinaz, I. (2009). *Aproximación teórica y empírica al estudio de las tipologías de agresores de pareja. Análisis descriptivo y variables e instrumentos de evaluación en el centro penitenciario Brians-2*. Madrid: Ministerio del Interior, Secretaría General Técnica.
- Loinaz, I., Echeburúa, E., Ortiz-Tallo, M., & Amor, P. J. (2012). Propiedades psicométricas de la Conflict Tactics Scales (CTS-2) en una muestra española de agresores de pareja. *Psicothema*, 24(1), 142-148.
- Loinaz, I., Ortiz-Tallo, M., & Ferragut, M. (2012). MCMI-III Grossman personality facets among partner-violent men in prison. *International Journal of Clinical and Health Psychology*, 12(3), 389-404.
- Marín-Morales, A., Bueso-Izquierdo, N., Hidalgo-Ruzzante, N., Pérez-García, M., Catena-Martínez, A., & Verdejo-Román, J. (2022). Would you allow your wife to dress in a miniskirt to the party? Batterers do not activate default mode network during moral decisions about intimate partner violence. *Journal of Interpersonal Violence*, 37(3-4), NP1463-NP1488. <https://doi.org/10.1177/0886260520926494>
- Marín-Morales, A., Pérez-García, M., Catena-Martínez, A., & Verdejo-Román, J. (2021). Emotional regulation in male batterers when faced with pictures of intimate partner violence: Do they have a problem with suppressing or experiencing emotions? *Journal of Interpersonal Violence*, 37(11-12), NP10271-NP10295. <https://doi.org/10.1177/0886260520985484>
- Muñoz-Sánchez, S.Y. (2018). *Violencia de pareja y resolución de conflictos en relaciones LGTBI en Bogotá*. Universidad Nacional de Colombia. Extraído de: <https://repositorio.unal.edu.co/handle/unal/69077>
- Navarro, S., Frías-Ibáñez, A., & Palma, C. (2020). Clinical differences and associations between women with borderline personality and their partners: an exploratory study. *Journal of Psychopathology*, 2, 264-272. <https://doi.org/10.36148/2284-0249-407>
- Puente-Martínez, A., Ubillos-Landa, S., & Páez, D. (2022). Problem-focused coping strategies used by victims of gender violence across the stages of change. *Violence Against Women*, 28(14), 3331-3351. <https://doi.org/10.1177/10778012211054866>
- Redondo, N., Cantos, A. L., Graña, J. L., Muñoz-Rivas, M. J., & O'Leary, K. D. (2019). Treatment-induced changes in undercontrolled and overcontrolled anger subtypes of perpetrators of intimate partner violence and 5-year recidivism. *Criminal Justice and Behavior*, 46(12), 1700-1718. <https://doi.org/10.1177/0093854819879201>
- Redondo, N. & Graña, J. I. (2015). Consumo de alcohol, sustancias ilegales y violencia hacia la pareja en una muestra de maltratadores en tratamiento psicológico. *Adicciones*, 27(1), 27-36. <https://doi.org/10.20882/adicciones.191>

- Ruiz-Hernández, J. A., García-Jiménez, 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(1), 41-49. <https://doi.org/10.1016/j.ejpal.2014.11.003>
- Straus, M. A. (2007). Conflict Tactics Scales. En N. A. Jackson (Ed.). *Encyclopedia of domestic violence* (pp. 190–197). New York, Routledge: Taylor & Francis Group.
- Straus, M. A., Hamby, S. L., Boney-McCoy, S. U. E., & Sugarman, D. B. (1996). The revised conflict tactics scales (CTS2) development and preliminary psychometric data. *Journal of Family Issues*, 17(3), 283-316. <https://doi.org/10.1177/019251396017003001>
- 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 <https://doi.org/10.5944/ap.14.2.20753>
- Vega, E. M., & O'Leary, K. D. (2007). Test–retest reliability of the revised Conflict Tactics Scales (CTS2). *Journal of Family Violence*, 22(8), 703–708. <https://doi.org/10.1007/s10896-007-9118-7>
- Verdejo-Román, J., Bueso-Izquierdo, N., Daugherty, J. C., Pérez-García, M., & Hidalgo-Ruzzante, N. (2019). Structural brain differences in emotional processing and regulation areas between male batterers and other criminals: A preliminary study. *Social Neuroscience*, 14(4), 390-397. <https://doi.org/10.1080/17470919.2018.1481882>

Referencias de estudios que analizan las propiedades psicométricas de una versión española de la CTS-2 diferente a la versión evaluada en el informe (Loinaz, 2009; Loinaz, Echeburúa, et al., 2012)

- Calvete, E., Corral, S. & Estévez, A. (2007). Factor structure and validity of the Revised Conflict Tactics Scale for Spanish women. *Violence Against Women*, 13, 1072-1087. [evaluación de la conducta de la pareja; mujeres involucradas en violencia en la pareja]
- Corral, S. & Calvete, E. (2006). Evaluación de la violencia en las relaciones de pareja mediante las Escalas de tácticas para conflictos: estructura factorial y diferencias de género en jóvenes. *Psicología Conductual*, 14, 215-233. [evaluación de la propia conducta y de la conducta de la pareja; estudiantes universitarios]
- Graña, J. L., Andreu, J. M., Peña, M., & Rodríguez-Biezma, M. J. (2013). Validez factorial y fiabilidad de la "Escala de tácticas para el conflicto revisada" (Revised Conflict Tactics Scales, CTS2) en población adulta española. *Psicología Conductual*, 21(3), 525-543. [evaluación de la propia conducta y de la conducta de la pareja; población adulta]
- Medina-Ariza, J.; Barberet, R. y Lago, M. J. (Noviembre, 1998). Validity and reliability of the CTS-II: Another sledgehammer?. Ponencia presentada en el *Annual Meeting of the American Society of Criminology*, Washington, DC. [evaluación de la conducta de la pareja; estudiantes universitarios]

ANÁLISIS DETALLADO DE LA PRUEBA

1. DESCRIPCIÓN GENERAL DEL TEST

1.1. Nombre del test:

Versión revisada de la Conflict Tactics Scales (CTS-2)

1.2. Nombre del test en su versión original:

Revised Conflict Tactics Scales (CTS2)

1.3. Autor/es del test original:

Murray A. Straus, Sherry L. Hamby, Sue Boney-McCoy, y David B. Sugarman

1.4. Autor/es de la adaptación española:

Ismael Loinaz es autor de la adaptación e Ismael Loinaz, Enrique Echeburúa, Margarita Ortiz-Tallo y Pedro J. Amor son autores del estudio de validación

1.5. Editor del test en su versión original:

Las escalas CTS se encuentran comercializadas por WPS. La CTS-2 se publicó en: Straus, M.A., Hamby, S., Boney-McCoy, S., & Sugarman, D. (1996). The Revised Conflict Tactics Scales: Development and preliminary psychometric data. *Journal of Family Issues*, 17, 283-316.

1.6. Editor de la adaptación española:

Se trata de un test no comercializado. Traducción publicada en: Loinaz, I. (2009). *Aproximación teórica y empírica al estudio de las tipologías de agresores de pareja. Análisis descriptivo y variables e instrumentos de evaluación en el centro penitenciario Brians-2*. Madrid: Ministerio del Interior, Secretaría General Técnica. Validación en hombres condenados publicada en: Loinaz, I., Echeburúa, E., Ortiz-Tallo, M., y Amor, P. J. (2012). Propiedades psicométricas de la Conflict Tactics Scales (CTS-2) en una muestra española de agresores de pareja. *Psicothema*, 24(1), 142-148.

1.7. Fecha de publicación del test original:

1996

1.8. Fecha de publicación del test en su adaptación española:

2009 (la versión sometida a revisión)

1.9. Fecha de la última revisión del test:

2012

1.10. Área general de la/s variable/s que pretende medir el test:

Violencia de pareja, Violencia de género, Violencia doméstica, Resolución de conflictos en la pareja.

1.11. Breve descripción de la/s variable/s que pretende medir el test:

La CTS-2 es una escala autoadministrada que pretende medir el grado en que los miembros de una pareja se ven envueltos en ataques físicos, psicológicos o sexuales en el marco de su relación (i.e., indicios de violencia de pareja). Asimismo, también evalúa el uso de estrategias de razonamiento y negociación a la hora de resolver sus conflictos y la presencia de lesiones como consecuencia de posibles peleas entre los miembros de la pareja.

Tanto la versión original en inglés (Straus et al., 1996) como la adaptación al español revisada en este informe (Loinaz, 2009) consta de 78 ítems, 39 para evaluar la propia conducta (i.e., estrategias de negociación y conductas violentas que quién responde ha ejercido sobre su pareja, y daños sufridos) y 39 para evaluar la conducta de la pareja (i.e., estrategias de negociación y conductas violentas que quién responde ha recibido de su pareja, y daños ocasionados a la pareja). Los ítems constan de 8 alternativas de respuesta. Los valores del 0 al 6 corresponden a la frecuencia con la que puede haber ocurrido una conducta en el último año (desde 0 "nunca ha ocurrido" hasta 6 "más de 20 veces"), mientras que el valor 7 se asigna a la categoría de respuesta "nunca el año pasado, pero sí antes". Aunque el test está diseñado tomando como periodo de referencia los últimos 12 meses de la relación de pareja, se contempla que pueda modificarse el periodo de referencia en las instrucciones a fin de adaptarse a las distintas necesidades del evaluador/a (p.ej., desde que la relación de pareja comenzó, antes de iniciar un tratamiento o terapia de pareja, en los últimos seis meses, etc.). En este último caso, también se modifica la categoría de respuesta de la opción 7 (p.ej., "nunca en los últimos seis meses, pero sí antes").

La CTS-2 subsume un total de cinco dimensiones, a saber: estrategias de negociación (p.ej., 'Mostré a mi pareja que me preocupaba por ella incluso cuando no estábamos de acuerdo'), violencia o agresión psicológica (p.ej., 'Salí dando un portazo de la habitación o de la casa durante una discusión'), violencia o agresión física (p.ej., 'Le retorcí el brazo o tiré del pelo a mi pareja'), coacción sexual (p.ej., 'Usé la fuerza [como golpear, inmovilizar o usar un arma] para obligar a mi pareja a practicar sexo oral o anal conmigo'), y daños o lesiones (p.ej., 'Sentí dolor físico que duró hasta el día siguiente a causa de una pelea con mi pareja'). La dimensión de negociación puede a su vez dividirse en dos escalas fundamentadas en el tipo de estrategia: "estrategias cognitivas" y "estrategias emocionales". El resto de dimensiones de la CTS-2 pueden también dividirse en función de la gravedad del comportamiento evaluado, obteniéndose puntuaciones para dos escalas: "gravedad menor" y "gravedad severa".

El test permite obtener diferentes puntuaciones: (a) prevalencia: presencia o ausencia de la conducta; (b) frecuencia: número de veces que ha sucedido una conducta dentro de cada escala en el período de referencia; y (c) severidad de la violencia física.

1.12. Áreas de aplicación:

Psicología clínica, Psicología forense, Salud general y bienestar, Investigación

1.13. Formato de los ítems:

Respuesta graduada / Tipo Likert

1.14. Número de ítems:

Total: 78 ítems

Parte dedicada a la evaluación de la propia conducta: 39 ítems

Parte dedicada a la evaluación de la conducta de la pareja: 39 ítems

Cada parte incluye:

- Escala de Negociación: 6 ítems
 - Estrategias cognitivas: 3 ítems
 - Estrategias emocionales: 3 ítems
- Escala de Agresión psicológica: 8 ítems
 - Gravedad menor: 4 ítems
 - Gravedad severa: 4 ítems
- Escala de Agresión física: 12 ítems
 - Gravedad menor: 5 ítems
 - Gravedad severa: 7 ítems
- Escala de Coacción sexual: 7 ítems
 - Gravedad menor: 3 ítems
 - Gravedad severa: 4 ítems
- Escala de Daños: 6 ítems
 - Gravedad menor: 2 ítems
 - Gravedad severa: 4 ítems

1.15. Soporte:

Papel y lápiz

1.16. Cualificación requerida para el uso del test de acuerdo con la documentación aportada:

Sería recomendable tener conocimientos en victimología y/o criminología. También comprender las dimensiones y marcos temporales de la herramienta para poder ajustar el formato a los objetivos de la evaluación e interpretar los resultados de forma adecuada.

1.17. Descripción de las poblaciones a las que el test es aplicable:

En el estudio de Straus y colaboradores (1996) se indica que la CTS-2 es aplicable a personas adultas (con al menos 18 años de edad) que han tenido una relación sexo-afectiva de al menos un mes durante el último año y con un nivel educativo mínimo equivalente al 6º grado para entender el contenido de los ítems. Este mismo estudio anticipó que los ítems de la CTS-2 deberían funcionar adecuadamente en personas de diversas minorías étnicas, dado que comparte similitudes de contenido con versiones anteriores que se administraron con éxito (i.e., CTS-1). Adicionalmente, menciona que el instrumento ha sido utilizado en al menos 20 países y en sujetos pertenecientes a colectivos culturales diversos (p.ej., afroamericanos, hispanoamericanos).

El uso de la versión española de la CTS-2 (Loinaz, 2009), validada únicamente para perpetradores (39 ítems de evaluación de la propia conducta) (Loinaz et al., 2012) aunque en el estudio se contó con un grupo control formado por hombres adultos de la población general, también se limita a personas mayores de edad. Si bien se aplicó a hombres condenados a prisión por haber ejercido violencia contra la pareja tanto de nacionalidad española como extranjeros, no se especifica el nivel de competencia lingüística mínimo para responder satisfactoriamente al test. Tampoco el nivel educativo mínimo para poder garantizar un comportamiento adecuado de las puntuaciones del test, aunque asumimos que debería ser similar al de la versión original de Straus et al. (1996). Algunos de los estudios que han empleado la CTS-2 validada por Loinaz et al. (2012) sugieren que esta versión también puede utilizarse en mujeres de población general y mujeres víctimas de violencia de género (p.ej., Daugherty et al., 2019; Puente-Martínez et al., 2022), hombres condenados por otro tipo de delitos (p.ej., Amaoui, Marín-Morales et al., 2022; Bueso-Izquierdo et al., 2015), población clínica o psicopatológica (p.ej., mujeres con un diagnóstico de trastorno límite de la personalidad [Navarro et al., 2020]), y población de otros países

hispanohablantes (p.ej., Perú [Espejo-Navarro y Valdivia-Ramírez, 2023] o Colombia [Muñoz-Sánchez, 2018]), entre otras poblaciones.

1.18. Existencia o no de diferentes formas del test y sus características:

La versión completa de la CTS-2 está compuesta por 78 ítems, 39 centrados en la evaluación de la propia conducta (i.e., estrategias de negociación utilizadas por uno mismo, actos de violencia ejercidos sobre la pareja, y daños y lesiones sufridos por uno mismo) y 39 en la evaluación de la conducta de la pareja (i.e., estrategias de negociación utilizadas por la pareja, actos de violencia sufridos por uno mismo, y daños y lesiones ejercidos a la pareja). Existe una correspondencia entre los ítems de evaluación de la propia conducta y de evaluación de la conducta de la pareja. Es decir, la conducta o lesión descrita por el ítem es la misma, lo que cambia es la perspectiva (i.e., la propia conducta o la conducta de la pareja). Por ejemplo: el ítem de evaluación de la propia conducta "Llamé gordo/a o feo/a a mi pareja" corresponde al ítem de evaluación de la conducta de la pareja "Mi pareja me llamó gordo/a o feo/a". Las dos partes pueden utilizarse separadamente. Por ejemplo, Calvete et al. (2007) únicamente administró la parte de evaluación de la conducta de la pareja, mientras que Loinaz et al. (2012) solo evaluó la parte de la propia conducta y no la correspondiente a lo que la pareja podía haber hecho, al no disponer de ningún dato para valorar la veracidad de dicha información.

Tanto la versión completa, como las dos partes de la CTS-2, pueden ser respondidas por ambos miembros de la pareja (datos diádicos) o únicamente por uno de ellos (datos individuales).

En el estudio de Straus et al. (1996), con el fin de reducir el tiempo de administración de la prueba, se menciona la posibilidad de administrar únicamente los ítems de las escalas de Agresión física, Violencia psicológica y Negociación (formato similar al de la CTS-1). No obstante, no se han identificado estudios que pongan a prueba el comportamiento psicométrico de esta versión reducida en muestras españolas.

En la mayoría de trabajos consultados para la elaboración de este informe, la administración del test fue en formato de papel y lápiz dónde la CTS-2 formó parte de una batería amplia de instrumentos de medida. Por último, cabe señalar que durante la revisión del test no se ha localizado ningún recurso que describa requisitos inusuales para una posible administración online de la CTS-2.

1.19. Procedimiento de corrección:

Manual

1.20. Puntuaciones:

Siguiendo las indicaciones de los autores originales (Straus et al., 1996), las puntuaciones directas que se obtienen con la CTS-2 son:

- Frecuencia: número de veces que una conducta ocurrió en cada escala en el período de referencia evaluado. La puntuación de cada ítem se obtiene asignando el número correspondiente en las primeras tres categorías ("Nunca" = 0, "una vez" = 1, "dos veces" = 2), el punto medio en las siguientes tres categorías ("entre tres y cinco veces" = 4, "de seis a diez veces" = 8, "de once a veinte veces" = 15), mientras que a la categoría "más de veinte veces" se les asigna el valor de 25. Para cada escala, la puntuación de 'frecuencia' se obtiene sumando las puntuaciones de los ítems que conforman la escala.
- Prevalencia: presencia o ausencia de la conducta en cada escala en el período de referencia evaluado. Para calcular la prevalencia, a la respuesta de "nunca" se le asigna un valor de 0 (ausencia), mientras que al resto de opciones de respuesta se les asigna un 1 (presencia). No obstante, las puntuaciones en los ítems no se suman, simplemente se considera si algunos de los comportamientos recogidos en la escala están presentes o no. De esta forma, la puntuación de prevalencia por escala presenta únicamente valores de 0 o 1, los cuales indican ausencia o presencia, respectivamente. Esta puntuación no se calcula en la subescala de Negociación. En la subescala de Agresión psicológica los autores originales advierten que emplear la prevalencia puede no ser útil, ya que casi la totalidad de las muestras suelen reportar al menos un incidente.
- Severidad: indica el grado de gravedad de la conducta. Se puede obtener para la escala Agresión física. Para obtener la puntuación, cada comportamiento tiene asignado un valor teórico de gravedad,

de forma que algunos ítems tienen un valor de 1 punto (ítems 7, 9, 17, 45 y 53; p.ej., ítem 7: "Tiré algo a mi pareja que le pudo hacer daño"), otros de 3 puntos (ítems 27 y 73; p.ej., ítem 27: "Golpeé a mi pareja con algo que le pudo hacer daño"), otros de 5 puntos (ítems 33, 37, 43 y 61; p.ej., ítem 33: "Estrangué [agarré por el cuello] a mi pareja") y, por último, un único ítem con un valor de 8 puntos (ítem 21: "He usado un cuchillo o arma contra mi pareja"). El sumatorio de estos valores permite calcular una puntuación directa de la 'severidad de la violencia física' ejercida.

1.21. Escalas utilizadas:

Puntuaciones directas, Centiles

1.22. Posibilidad de obtener informes automatizados:

No

1.23. Tiempo estimado para la aplicación del test:

En aplicación individual: 10-15 minutos.

En aplicación colectiva: No se establece diferenciación temporal en función del tipo de administración individual vs. colectiva.

1.24. Documentación aportada por el editor:

La versión española no se encuentra comercializada. La documentación ha sido aportada por la coordinación de la evaluación de test (i.e., artículos seleccionados sobre la escala original y la adaptada).

1.25. Precio de un juego completo de la prueba:

No procede. No se trata de un test comercializado.

1.26. Precio y número de ejemplares del paquete de cuadernillos:

No procede. No se trata de un test comercializado.

1.27. Precio y número de ejemplares del paquete de hojas de respuesta:

No procede. No se trata de un test comercializado.

1.28. Precio de la administración y/o corrección, y/o elaboración de informes por parte del editor:

No procede. No se trata de un test comercializado.

2. VALORACIÓN DE LAS CARACTERÍSTICAS DEL TEST

2.1. Aspectos generales:

Contenido	Valoración	Puntuación
2.1. Calidad de los materiales del test	No evaluado	--
2.2. Calidad de la documentación aportada	No evaluado	--
2.3. Fundamentación teórica	Adecuada	3
2.4. Adaptación del test	Adecuada	3
2.5. Desarrollo de los ítems del test	Adecuada	3
2.6. Calidad de las instrucciones para el participante	Excelente	5
2.7. Calidad de las instrucciones (administración, puntuación, interpretación)	Excelente	5
2.8. Facilidad para registrar las respuestas	Buena-Excelente	4.5
2.9. Bibliografía del manual	No evaluado	--
2.10. Datos sobre el análisis de los ítems	Buena	4

2.11. Validez:

2.11.1. Evidencias de validez de contenido:

Contenido	Valoración	Puntuación
2.11.1.1. Calidad de la representación del contenido o dominio	Adecuada	3
2.11.1.2. Consultas a expertos	Se ha consultado de manera informal a un pequeño número de expertos	2

2.11.2. Evidencias de validez basadas en la relación entre las puntuaciones del test y otras variables:

2.11.2.1. Evidencias de validez basadas en la relación entre las puntuaciones del test y otras variables:

Contenido	Valoración	Puntuación
2.11.2.1.1. Diseños empleados	Correlaciones con otros test Diferencias entre grupos Correlación de las puntuaciones de las escalas de Agresión física, Agresión psicológica y Coerción sexual con el tipo de delito principal por el que los agresores habían sido condenados	
2.11.2.1.2. Tamaño de las muestras	Un estudio con una muestra moderada ($200 \leq N \leq 500$) o varios estudios con muestras pequeñas ($N < 200$)	2
2.11.2.1.3. Procedimiento de selección de las muestras	Incidental. Para la versión original de Straus et al. (1996), cuya muestra de análisis estuvo compuesta por un total de 317 participantes, se buscaron voluntarios entre estudiantes de sociología o criminología que fueran mayores de edad y que tuvieran (o hubieran tenido en el último año) una relación de pareja heterosexual (i.e., entendiéndose como pareja quedar regularmente, vivir juntos o estar casados). En el caso de la validación de la versión en español de Loinaz et al. (2012), se describen dos procedimientos de carácter no probabilístico. Por un lado, con respecto a la muestra de 173 hombres agresores, se describe que la cumplimentación de la escala se dio entre los años 2008 y 2010 como parte de estudios de investigación más amplios realizados con muestras de diferentes tipologías de agresores (se entiende que estas personas fueron voluntarias en la investigación). Con respecto a los hombres de población general ($n= 108$), estos fueron reclutados a través de contactos personales de colaboradores/as de los/as investigadores/as para	

	aparearlos por edad y nacionalidad con el grupo de internos (se entiende, por tanto, que también eran voluntarios que se seleccionaron por conveniencia).	
2.11.2.1.4. Calidad de los tests empleados como criterio o marcador	Adecuada	3
2.11.2.1.5. Promedio de las correlaciones con otros tests que miden constructos similares	Adecuada con carencias ($0.35 \leq r < 0.50$)-Adecuada ($0.50 \leq r < 0.60$)	2.5
2.11.2.1.6. Promedio de las correlaciones con otros tests que miden constructos no relacionados	Inadecuada-Adecuada con carencias	1.5
2.11.2.1.7. Resultados de la matriz multirrasgo-multimétodo	--	--
2.11.2.1.8. Resultados de las diferencias intergrupo	Buena	4

2.11.2.2. Evidencias de validez basadas en la relación entre las puntuaciones del test y un criterio:

Contenido	Valoración	Puntuación
2.11.2.2.1. Criterios empleados	--	--
2.11.2.2.2. Calidad de los criterios empleados	--	--
2.11.2.2.3. Relación temporal entre test y criterio	--	--
2.11.2.2.4. Tamaño de las muestras	--	--
2.11.2.2.5. Procedimiento de selección de las muestras	--	--
2.11.2.2.6. Promedio de las correlaciones del test con los criterios	--	--

2.11.3. Evidencias de validez basadas en la estructura interna:

Contenido	Valoración	Puntuación
2.11.3.1. Resultados del análisis factorial	Adecuada con carencias	2
2.11.3.2. Funcionamiento diferencial de los ítems	--	--

2.11.4. Acomodaciones en la administración del test:

Contenido	Valoración	Puntuación
2.11.4. El manual del test informa sobre las acomodaciones en la administración del test		No

2.11.5. Comentarios generales sobre evidencias de validez:

Existe abundante literatura sobre las evidencias de validez de la CTS-2 en su versión original, mientras que para la versión española que nos ocupa, los estudios se centran principalmente en el análisis de la estructura dimensional y la relación con otras variables.

En lo que respecta a la **evidencia basada en el contenido**, se valora de forma global la información aportada para la versión original de la CTS-2 como adecuada. Si bien el trabajo realizado para la forma original del test desarrollada por Straus et al. (1996) delimita el marco teórico general del instrumento, basado en la "teoría del conflicto", la definición conceptual u operacional de los dominios del constructo es escueta y la justificación de la selección de las conductas específicas a incluir es limitada. Straus (2007) expone que la selección de los contenidos del test se basó en los resultados de entrevistas y sugerencias y revisiones propuestas por investigadores y clínicos expertos en violencia en la pareja. Sin embargo, él mismo reconoce que los actos de violencia incluidos en la CTS-2 son únicamente una muestra del universo de posibles conductas violentas. Argumenta a favor de la evidencia basada en el contenido el solapamiento que existe entre los ítems de la CTS-2 y la lista de conductas de violencia típicas elaborada por Dobash y Dobash (1984). Toda esta información es extrapolable a la versión en español de Loinaz (2009). El usuario del test tiene que leer el contenido de los ítems para hacerse una idea más precisa de los indicadores que pretende evaluar el test. Tampoco hay información sobre criterios formales de los ítems como son el grado de claridad, brevedad o ambigüedad de los indicadores utilizados en ninguna de las dos versiones. En la versión original de Straus et al. (1996) se describe un procedimiento de pilotaje por el cual estudiantes universitarios dieron retroalimentación sobre la redacción de los ítems, aunque no se reportan los detalles de los resultados. Por último, cabe señalar que para la versión española no se ha abordado explícitamente la posibilidad de incluir adaptaciones de contenido en los ítems relacionadas con la idiosincrasia de la cultura española al constructo objeto de medición. Sin embargo, tal y como advierte Loinaz et al. (2012) (p. 147), hay que tener en cuenta que los factores originales estructuran los ítems a partir de consideraciones culturales de la sociedad norteamericana y que, por ello, conductas que originalmente son consideradas físicas/psicológicas/sexuales, o de gravedad menor/severa, pueden no tener la misma consideración en nuestra sociedad, como ocurre en el caso de los ítems 25 (Llamé gorda o fea a mi pareja) y 29 (Destrocé algo que pertenecía a mi pareja).

Respecto a las **evidencias basadas en la relación entre las puntuaciones del test y otras variables**, es conveniente destacar que la gran mayoría de estudios identificados en los que se empleó la versión española de la CTS-2, utilizaron esta medida para establecer diferencias entre grupos (p.ej., agresores de violencia de género vs. hombres de la población general o vs. delincuentes por otros delitos), usan muestras relativamente pequeñas-moderadas y muestreos no probabilísticos por conveniencia para el acceso a la muestra. En este sentido, las evidencias sobre la capacidad de las puntuaciones de la CTS-2 para discriminar entre diferentes grupos poblaciones son extensas. En el estudio de validación de la versión en castellano (Loinaz et al., 2012), los autores pudieron diferenciar entre hombres condenados por delitos relacionados con la violencia de pareja y hombres adultos de la población general. En concreto, hubo diferencias entre ambas muestras en todas las escalas de la CTS-2 con la excepción de la Coacción sexual, con puntuaciones relativamente bajas en ambos grupos. De forma similar, algunos estudios han utilizado las puntuaciones de la CTS-2 para diferenciar entre presos condenados por un delito de violencia de género y presos condenados por otro tipo de delitos (p.ej., Amaoui, Martín-Pérez et al., 2022; Marín-Morales et al., 2022; Verdejo-Román et al., 2019), encontrando resultados consistentes (p.ej., diferencias en todas las escalas menos en la escala de Negociación). Otros estudios han realizado comparaciones similares entre hombres que comenten un delito de violencia de pareja y delincuentes comunes, pero analizando diferencias para una puntuación global de la CTS-2 o las puntuaciones aisladas en alguna de las escalas, como en la Violencia psicológica (i.e., obteniéndose en ambos casos puntuaciones más elevadas para el grupo de maltratadores condenados por violencia de género: Bueso-Izquierdo et al., 2015; Marín-Morales et al., 2021;

Ruiz-Hernández et al., 2015). También se ha mostrado que las puntuaciones en la CTS-2 permiten diferenciar entre maltratadores con distintas probabilidades de reincidencia en sus delitos (Llor-Esteban et al., 2016) o con expresiones diferenciales de rasgos de personalidad violentos-antisociales (Boira y Jodrá, 2013; Carbajosa et al. 2017; Graña et al., 2014; Loinaz, Ortiz-Tallo, et al., 2012; Redondo et al., 2019). Adicionalmente, estudios de investigación aplicada han encontrado diferencias en las puntuaciones de la CTS-2 entre grupos de maltratadores sometidos a diferentes tipos de intervención (Lila et al., 2018), avalando la sensibilidad de las puntuaciones de la CTS-2 a los efectos de la implementación de programas eficaces para reducir la violencia de pareja. También se han observado diferencias en las puntuaciones de la CTS-2 entre grupos con distintos niveles de consumo de alcohol o de drogas ilegales (Redondo y Graña, 2015). Por último, las puntuaciones en la CTS-2 también permiten discriminar en función de las estrategias de afrontamiento que utilizan mujeres víctimas de violencia de género (Puente-Martínez et al., 2022) y diferenciar entre mujeres víctimas de violencia de género y aquellas no víctimas (Daugherty et al., 2019). Por otro lado, no se han hallado diferencias concluyentes basadas en las puntuaciones de la CTS-2 entre mujeres con un trastorno límite de personalidad y sus parejas (Navarro et al., 2020). Tampoco se han observado diferencias en términos de perpetración de actos de violencia de pareja entre maltratadores españoles y maltratadores inmigrantes de habla hispana (Vargas et al., 2017). A pesar de todas estas evidencias, se evalúa el apartado como bueno, y no excelente, porque no siempre se establecen hipótesis claras de las diferencias grupales esperadas y porque en ocasiones se comparan ambos grupos únicamente en algunas puntuaciones (i.e., en la mayoría de casos se utiliza una puntuación global o la puntuación en alguna de las escalas).

En relación a la evidencia convergente y discriminante, el estudio de Straus et al. (1996) muestra que la correlación entre la escala de Agresión física y una escala de integración social fue de -0.29, en línea con lo que esperaban los autores. En el estudio de Loinaz, Echeburúa, et al. (2012) se correlacionan las puntuaciones de las escalas de la CTS-2 de Agresión física, Agresión psicológica y Coerción sexual con el tipo de delito por el que el agresor estaba condenado en el momento del estudio: maltrato físico o psicológico. Si bien coherentes teóricamente, las correlaciones únicamente fueron estadísticamente significativas, aunque de tamaño modesto (≈ 0.25), entre la 'frecuencia' y 'gravedad' de la Agresión física y la presencia de una condena por maltrato físico (lesiones, maltrato, tentativa de homicidio u homicidio consumado). Adicionalmente, destacar que existen otras investigaciones que han ampliado la información en términos de correlatos de la versión en español de la CTS-2. Por ejemplo, en el estudio de Rodríguez y Fernández-González (2019) encontraron que las escalas de agresión física, psicológica y sexual correlacionaron positivamente (y con una magnitud similar entre ellas) tanto con medidas de consumo de pornografía violenta y no violenta como con indicadores de actitudes neosexistas y de valoración de las mujeres como un objeto sexual. De igual forma, puntuaciones altas en estas escalas fueron indicativas de mayores niveles de justificación de la violencia, mayores creencias en mitos sobre la violación e, incluso, una mayor tendencia global a la agresión. Considerando lo expuesto anteriormente, se concluye que las evidencias de validez convergente, aunque limitadas porque solo valoran algunas de las escalas, parecen ser bastante sólidas y que las evidencias de validez discriminante son insuficientes para ser valoradas adecuadamente, dado que faltan hipótesis sobre la ausencia de relación entre variables.

En lo relativo a las **evidencias basadas en la estructura interna del test**, la CTS-2 muestra un amplio margen de mejora. Por un lado, el artículo de la versión original de Straus et al. (1996), aunque sugiere una estructura teórica de cinco factores, carece de un estudio empírico de la dimensionalidad de la CTS-2. Loinaz, Echeburúa et al. (2012) no consiguieron demostrar la estructura de cinco dimensiones propuesta por Straus et al. (1996) a nivel teórico. El ajuste obtenido utilizando un análisis factorial confirmatorio en una muestra de hombres maltratadores fue claramente insatisfactorio: $\chi^2(655) = 1902$; $p < 0.001$; GFI = 0.88; SMRS = 0.09; RMSEA = 0.11). A la vista de estos resultados, estos autores decidieron abordar la dimensionalidad de la CTS-2 utilizando un análisis factorial exploratorio que sugirió la existencia de cuatro factores (violencia física, sexual, psicológica y negociación) tras eliminar algunos ítems con cargas cruzadas. Estos resultados indican la presencia de un solapamiento importante entre algunos de los ítems de la CTS-2, al menos en el caso de hombres maltratadores. Conjuntamente, los cuatro factores explicaron el 51.4% de la varianza. Aunque diversas investigaciones utilizan una puntuación global de la CTS-2, no existen evidencias que apoyen este uso. Por último, cabe señalar que no existen hasta la fecha estudios que hayan analizado el posible funcionamiento diferencial de los ítems en función de variables sociodemográficas u otras variables de interés, ni tampoco ningún estudio de invarianza de

la versión española de la CTS-2.

2.12. Fiabilidad:

Contenido	Valoración	Puntuación
2.12.1. Datos aportados sobre fiabilidad	Varios coeficientes de fiabilidad (para cada escala o subescala) Coeficientes de fiabilidad para diferentes grupos de personas	

2.12.2. Equivalencia formas paralelas:

Contenido	Valoración	Puntuación
2.12.2.1. Tamaño de las muestras	--	--
2.12.2.2. Puesta a prueba de los supuestos de paralelismo	--	--
2.12.2.3. Promedio de coeficientes de equivalencia	--	--

2.12.3. Consistencia interna:

Contenido	Valoración	Puntuación
2.12.3.1. Tamaño de las muestras	Un estudio con una muestra moderada ($200 \leq N < 500$) o varios estudios con muestras pequeñas ($N < 200$)	2
2.12.3.2. Coeficientes de consistencia interna presentados	Coeficiente alfa o KR-20	
2.12.3.3. Promedio de coeficientes de consistencia	Buena ($0.80 \leq r < 0.85$)	4

2.12.4. Estabilidad (test-retest):

Contenido	Valoración	Puntuación
2.12.4.1. Tamaño de las muestras	Versión original: Un estudio con una muestra pequeña ($N < 100$) Versión española: No se aporta	1
2.12.4.2. Coeficientes de estabilidad	Versión original: Adecuada ($0.65 \leq r < 0.75$)	3

	Versión española: No se aporta	
--	--------------------------------	--

2.12.5. Cuantificación de la precisión mediante TRI:

Contenido	Valoración	Puntuación
2.12.5.1. Tamaño de las muestras	--	--
2.12.5.2. Coeficientes proporcionados	--	
2.12.5.3. Tamaño de los coeficientes	--	--

2.12.6. Fiabilidad inter-jueces:

Contenido	Valoración	Puntuación
2.12.6.1. Tipos de coeficientes presentados	--	
2.12.6.2. Promedio de los coeficientes	--	--

2.12.7. Comentarios generales sobre evidencias de fiabilidad:

En relación a la versión original en inglés, son numerosos los estudios que han estudiado la fiabilidad como consistencia interna mediante la obtención del coeficiente alfa de Cronbach. Puede consultarse un recopilatorio de los estudios y los correspondientes valores en: [CTS Sourcebook](#) (clicar en CTS39). En general, los valores indican una consistencia interna entre moderada y buena (0.77 en promedio), aunque existe una variabilidad considerable, dado que se encuentran valores entre inadecuados y excelentes dependiendo de la muestra, la parte evaluada (propia conducta o conducta de la pareja) y la puntuación (escalas o total). En cuanto a la estabilidad temporal, los pocos estudios existentes establecen que la fiabilidad test-retest es, en general, adecuada (Vega y O'Leary, 2007).

La fiabilidad en la versión española de la CTS-2 analizada en este informe (39 ítems de evaluación de la propia conducta; Loinaz, Echeburúa, et al., 2012) ha sido estudiada en términos de consistencia interna de las puntuaciones de 'frecuencia' en una muestra de 173 hombres encarcelados por violencia contra la pareja. Se obtienen los siguientes coeficientes en las distintas escalas: Negociación (0.83), Agresión psicológica (0.81), Agresión física (0.83), Coacción sexual (0.80) y Daños (0.59). Para la puntuación total, el coeficiente alfa de Cronbach es de 0.88. En este sentido, la fiabilidad como consistencia interna es, en general, buena para las escalas y excelente para la puntuación total. La baja consistencia interna para la escala de Daños puede ser ocasionada por el cambio de población objetivo de la medición (Straus et al. 1996 = universitarios; Loinaz, Echeburúa, et al. 2012 = hombres maltratadores) o por los problemas descritos para la versión española en términos de dimensionalidad (véase el apartado sobre validez en el presente informe). De hecho, en el mismo estudio, Loinaz, Echeburúa, et al. (2012) realizaron un estudio adicional de la consistencia interna de las puntuaciones de la CTS-2 para el modelo de cuatro dimensiones que defienden tras los análisis preliminares de dimensionalidad. Este modelo, como señalamos, no incluye la escala de Daños y excluye ítems con cargas factoriales sustanciales en más de un factor. Tras estas modificaciones, también encontraron índices de consistencia interna entre buenos y excelentes para todas las escalas de la CTS-2 y la puntuación total.

Algunos trabajos científicos que han empleado la versión española de la CTS-2 evaluada en este informe ofrecen datos de consistencia interna de la medida (p.ej., Gallego Rodríguez y Fernández-González, 2019; Lila et al., 2018). Los valores para las escalas suelen situarse en

torno a 0.70 – 0.85 en casi todas las muestras identificadas y para la puntuación total en torno a 0.80 – 0.85.

Finalmente, cabe señalar que otros estudios reportan coeficientes de consistencia interna para otras adaptaciones al español de la CTS-2. Así, por ejemplo, para la adaptación de Corral y Calvete (2006), los valores obtenidos en la evaluación de la conducta de la pareja en una muestra de 919 mujeres (Calvete et al., 2007), los coeficientes alfa de Cronbach oscilan entre 0.80 (escala de Coacción sexual) y 0.93 (escala de Agresión física). Mientras que para la adaptación de Graña et al. (2013), en una muestra de 4550 adultos españoles, los valores oscilan entre 0.63 (Coacción sexual) y 0.80 (Agresión física) en la evaluación de la conducta de la pareja y entre 0.62 (Coacción sexual) y 0.79 (Agresión física) en la evaluación de la propia conducta.

2.13. Baremación e interpretación de las puntuaciones:

2.13.1. Interpretación normativa de las puntuaciones:

Contenido	Valoración	Puntuación
2.13.1.1. Calidad de las normas	Un baremo adecuado para la población objetivo	3
2.13.1.2. Tamaño de las muestras	Suficiente ($150 \leq N < 300$)	2
2.13.1.3. Aplicación de estrategia de tipificación continua	No	
2.12.1.4. Procedimiento de selección de las muestras	Incidental. Para la elaboración de los baremos españoles de la versión de Loinaz et al. (2012), se ha utilizado un procedimiento de selección incidental. La muestra está formada por 173 hombres condenados a prisión por haber ejercido violencia contra la pareja. Concretamente, 127 provienen del Centro Penitenciario Brians-2, en Barcelona, y 46 del de Alhaurín de la Torre, en Málaga. La media de edad es de 38.14 años, y un 80.3% son de nacionalidad española. Los datos de la CTS-2 se recogieron en el marco de una investigación más amplia que tuvo lugar entre los años 2008 y 2010.	
2.12.1.5. Actualización de baremos	Buena (entre 10 y 14 años)	4

2.13.2. Interpretación referida a criterio:

Contenido	Valoración	Puntuación
2.13.2.1. Adecuación del establecimiento de los puntos de corte	--	--
2.13.2.2. Procedimiento empleado para fijar los puntos de corte	--	
2.13.2.3. Procedimiento de obtención del acuerdo inter-jueces	--	
2.13.2.4. Valor del coeficiente de acuerdo inter-jueces	--	--

2.13.3. Comentarios generales sobre baremación e interpretación de las puntuaciones:

En relación con la creación de baremos de las puntuaciones de la CTS-2, para la versión española del test propuesta por Loinaz (2009), se ofrece una tabla de transformación de las puntuaciones directas de 'frecuencia' a percentiles para los tres tipos de violencia de pareja estudiados (violencia psicológica, física y sexual) en la evaluación de la propia conducta (Loinaz, Echeburúa, et al., 2012). Asimismo, también se incluye esta misma información para la severidad de la violencia física.

Este baremo se ha elaborado a partir de una muestra de 173 hombres condenados por delitos de violencia hacia la pareja. Permite establecer la ubicación relativa de un agresor concreto en función de sus puntuaciones en la CTS-2, facilitando la interpretación individual de las puntuaciones obtenidas por un agresor. La naturaleza de la población y la dificultad de acceso a ella pueden justificar que el tamaño muestral no sea muy elevado.

Para la versión española propuesta por Loinaz (2009), no se han identificado baremos españoles para las puntuaciones obtenidas a partir de los 39 ítems de evaluación de la conducta de la pareja ni para otro tipo de muestra o contexto de evaluación.