

The Revised Conflict Tactics Scales (CTS2):

**A review of the properties, reliability, and validity of the CTS2 as a measure of partner abuse in
community and clinical samples**

Harriet Chapman¹, Steven M. Gillespie²

*¹Centre for Forensic and Criminological Psychology, School of Psychology, University of
Birmingham, Birmingham, UK*

*²Department of Psychological Sciences, Institute of Psychology, Health and Society, University
of Liverpool, UK*

Chapman, H., & Gillespie, S. M. (2018). The Revised Conflict Tactics Scales (CTS2): A review
of the properties, reliability, and validity of the CTS2 as a measure of partner abuse in
community and clinical samples. *Aggression and Violent Behavior, 44*, 27-35.

<https://doi.org/10.1016/j.avb.2018.10.006>

The Revised Conflict Tactics Scales (CTS2):

A review of the properties, reliability, and validity of the CTS2 as a measure of partner abuse in community and clinical samples

Abstract

The purpose of this review is to explore the Revised Conflict Tactics Scales (CTS2), focusing on its research uses for assessing IPV. An overview of the CTS2 is presented, including the area that it assesses, the purpose of the tool, and its content. The tool's attempts to measure the purported construct are evaluated, drawing upon principles of reliability, validity and appropriate norms. The CTS2 has been found to be a reliable and valid instrument to measure IPV across different populations and across different cultures. However, there are some concerns regarding the internal consistency of the sexual coercion scale among female samples, and explorations of the factor structure of the CTS2 have yielded inconsistent models. While some statistical properties of the scale have been established for different populations, more research is needed to ascertain the validity and reliability of the CTS2 in varied clinical and forensic settings. The strengths and limitations of the CTS2 considered in the review have important implications for interpreting findings of studies investigating IPV using only this tool.

Key words: CTS2; review; intimate partner violence; properties; reliability; validity

The Revised Conflict Tactics Scale (CTS2):

A review of the properties, reliability, and validity of the CTS2 as a measure of partner abuse in community and clinical samples

1. Introduction

Straus and Gelles famously coined “the marriage license as a hitting license” following the discovery that violence between intimates is perpetrated at a far more frequent rate than violence occurring outside the context of an intimate relationship (Straus & Gelles, 1986). Intimate Partner Violence (IPV), referring to physical violence, sexual violence, stalking and psychological aggression by a current or former intimate partner (e.g., Breiding, Basile, Smith, Black & Mahendra, 2015), was first brought to public attention in the 1970s (see Dobash & Dobash, 1979), and in the ensuing years it has emerged as a serious public health concern (Cohen, Schulz, Liu, Halassa, Waldinger, 2015).

To demonstrate the scale of the problem, IPV has a higher rate of repeat victimisation compared with any other crime (Kershaw et al., 2000), and prosecutions for IPV in the year 2015-2016 rose by 9% from the preceding year (Violence against Women and Girls Crime Report, 2015- 2016). Although official statistics paint a picture of IPV as a societal problem of men’s violence against women, with an estimated 1,028,000 female victims and 489,000 male victims of IPV (including non-physical abuse, threats, force, sexual assault or stalking) aged between 16 and 59 in England and Wales for the year ending March 2016 (Office for National Statistics, February 2017), there is a growing evidence base to suggest that male and female acts of IPV occur at approximately equal rates (Dixon & Graham-Kevan, 2011; Esquivel-Santoveña & Dixon, 2012). In fact, a meta-analysis, based primarily on studies with young samples in the United States of America, found that females physically aggress against their intimate partners slightly more often than males (Archer, 2000); a finding that has been replicated in later studies of similar samples (e.g., Straus, 2008). Thus, although initially surprising (Straus, 1979), the finding of gender symmetry in

rates of IPV has found consistent support, and more recent findings highlight shared characteristics of both male and female perpetrators of IPV (Brzozowski, Gillespie, Dixon, & Mitchell, 2018; Simmons, Lehmann, Cobb, & Fowler, 2005).

Although both sexes sustain injuries from IPV, male-female perpetrated IPV is proposed to do more serious damage than same-sex or female-male perpetrated IPV due to the imbalance in size and strength (Archer, 2000; Cascardi, Langhinrichsen & Vivian, 1992; Dobash & Dobash, 2004; Felson, 2006; Stets & Straus, 1990). Thus, although there is symmetry in rates of IPV, this may not extend to symmetry in severity of injury. Perhaps the most poignant statistic of this nature is that two women are killed each week by a current or former partner in the UK (Office for National Statistics, February 2015), while on average three women are killed by their partner each day in the United States (American Psychological Association, 2017), making IPV the primary cause of violent death among females (Catalano, 2013; Fox & Zawitz, 2007). It is therefore important that any measure of the nature, frequency, and context of IPV is a reliable and valid measure in both male and female samples.

The consequences of IPV are well-documented, with deleterious effects recognized for all family members, including the intergenerational transmission of violence (e.g., Black, Sussman & Unger, 2010; Carter, Weithorn & Behrman, 1999; Smith, Ireland, Park, Elwyn & Thornberry, 2011; Stover, Meadows & Kaufman, 2009). IPV victimisation is linked with increased risk for posttraumatic stress disorder (PTSD), depression, substance misuse and physical health problems in female victims (Campbell & Lewandowski, 1997; Cascardi et al., 1992; Coker, Smith, Bethea, King & McKeown, 2000). IPV also brings great financial costs to society and was estimated to have cost the British economy £16 billion in 2008 (Walby, 2009). Furthermore, the average lifetime cost of witnessing IPV in childhood was estimated to be over \$50,000 per victim in the United States in the year 2016 (Holmes, Richter, Votruba, Berg & Bender; 2018), demonstrating the longer-term economic burden of IPV.

One of the most commonly used measures of IPV is the revised Conflict Tactics Scales (CTS2; Straus, Hamby, Boney-McCoy & Sugarman, 1996). This paper reviews the CTS2 in terms of its scientific properties, focusing on its uses as a tool for assessing IPV perpetration. While others have reviewed the development of the CTS and CTS2, highlighted comparisons with similar measures, or commented on specific psychometric properties or the performance on the CTS2 in particular samples (see Jones, Browne, & Chou, 2017), this review will focus more closely on the psychometric properties of the CTS2. Thus, we will review findings on reliability (internal, cross-cultural, inter-rater, test-retest), validity (face, concurrent, predictive, content, construct), factor structure, and normative reference groups.

The studies presented were identified through an electronic search using Web of Knowledge. Search terms included “intimate partner violence”, “partner abuse”, “measurement”, “conflict tactics scales”, “validity”, “reliability”, “factor structure”, and “properties”. In addition to this, citation searches and searches of reference lists were also carried out looking specifically for papers that examined the psychometric properties of the CTS2 in relation to partner violence. Only articles written in English were examined. Studies were included if they evaluated any psychometric property of the CTS2 as a measure of IPV perpetration. It is noted that other versions of the CTS2 exist that are beyond the scope of this review (e.g., the sibling version, see Fernandes, Relva, Rocha & Alarcão, 2016; Relva, Fernandes & Costa, 2013) and as a measure of parent-child violence, see Straus, Hamby, Finkelhor, Moore & Runyan, 1998).

First, an overview of the CTS2 is presented, including the area that it assesses, the purpose of the tool, and its content. Second, the tool’s attempts to measure the purported construct are evaluated, drawing upon principles of reliability, validity and appropriate norms.

1.1. Overview of the area assessed by Conflict Tactics Scales (CTS)

Although conflict denotes negative ideation such as “war” and “aggression” (Ledlow, 2009), conflict theorists regard conflict as a natural part of relationships (Adams, 1965; Coser, 1967; Dahrendorf, 1959).

Such theorists maintain it is not the conflict itself that poses a threat to relationships but the tactics that are employed to resolve it (i.e., “conflict tactics”) (e.g., Coser, 1967; Dahrendorf, 1959; Fincham & Beach, 1999). In intimate relationships, conflict managed constructively can help to foster satisfaction and intimacy (Fincham & Beach, 1999; Simmel, 1920, cited in Lilly, Cullen & Ball, 2014, p.180). However, conflict managed through physical or psychological aggression can have a detrimental impact on victims and relationships.

1.2. History of CTS development

Underpinned by the theoretical framework of conflict theory (Adams, 1965; Coser, 1967; Dahrendorf, 1959; Straus, 1979), the CTS (Straus, 1979) was designed to be an objective instrument providing data on the prevalence and chronicity of tactics employed by partners to resolve conflict in a dating, cohabiting or marital relationship. The first version of the scales was developed through data obtained from questionnaires administered to university students about assaults between their parents (Straus, 1974).

In 1996, the scales were revised in light of critiques and on the basis of research and clinical work (CTS2; Straus et al., 1996). They were further developed using data derived through the National Family Violence Survey (Straus, 1979) and through questionnaires administered to students about their own partner violence. The primary changes included: clarifying the wording of items; increasing the number of items for enhanced content validity and reliability; replacing some items; improving the distinction between minor and severe acts; including scales to measure sexual coercion and injury (i.e., consequence); replacing the weakest original scale (reasoning) with negotiation; and interspersing the order of questions to reduce response sets (Straus et al., 1996).

1.3. Application of the CTS2

The scales have been used in multiple contexts, including evidence-based initiatives to inform IPV treatment and policy (Dixon & Graham-Kevan, 2011), in therapy to assist disclosure (O’Leary &

Murphy, 1992), and in correctional settings to monitor behaviour and treatment progress (Straus, 1993). The CTS2 is also used as the primary measure of IPV in research (Capaldi et al., 2012; Schwartz, 2000), and was found by Thompson and colleagues to be the only measure capturing the extent of IPV through its measurement of multiple forms of perpetration and victimisation (Thompson, Basile, Hertz & Sitterle, 2006).

The scales have been used on participants from diverse cultural backgrounds, including African Americans and Hispanic Americans, in over 20 different countries (see Straus et al., 1996), and they have been translated into at least 15 different languages, including Chinese, Hebrew and Zulu (Straus, 2004). Consequently, understandings of IPV are largely based on the CTS/CTS2, and data derived therefrom play an important part in research, policy and practice. It is therefore important that the scales' theoretical bases and psychometric properties are scrutinised for understanding the validity and reliability of conclusions drawn from IPV research utilising the CTS2.

2. Overview of measure

The CTS2 is a 39-item, self-report questionnaire devised to assess the nature and frequency of tactics used by partners to manage conflict in an intimate relationship. It is designed to be understood by persons with a sixth-grade reading ability, and its simple format allows it to be fully administered in 10-15 minutes (Straus et al., 1996).

Using a Likert scale, respondents are asked to rate how frequently they have perpetrated a stated act, and how frequently they have been victim of that act by their partner over a specific referent period. By measuring both perpetration and victimisation, the scales are purported to take into account reciprocal violence and recognise that violence perpetration might be dependent on violence victimisation (Feld & Straus, 1989). Moreover, the scales allow for an estimate of the prevalence of 'Dyadic Concordance Types' (DCTS) (Straus, 2015), and in particular, whether partner violence within heterosexual relationships is male only, female only, both, or neither. Using data from 11 nations, Straus (2015) found

that, of the couples that reported physical violence, the predominant pattern was that this was reciprocal (42-43%). Cuenca and colleagues examined dyadic report rates across three of the scales of the Spanish version of the CTS2 and found similar prevalence rates for men and women for physical aggression (16.8% vs 17.6%) and psychological aggression (80.7% vs 81.4%) but not for sexual aggression (26.8% vs 16.1%) (Cuenca, Graña & Redondo, 2015). Identifying DCTS can be helpful in furthering understandings of the causes and consequences of IPV and may help to inform clinical intervention through elucidating treatment needs.

The standard referent period is 'during the past year' but this can be modified according to its application purposes (Straus et al., 1996). The CTS2 can also be adapted to measuring IPV occurring during a particular event or situation; although this does not then allow overall prevalence rates to be inferred (Straus et al., 1996).

2.1. Response format

The response options are: 1 = once; 2 = twice; 3 = 3-5 times; 4 = 6-10 times; 5 = 11-20 times; 6 = more than 20 times; 7 = not in referent period but happened before; 0 = never. The scores can be recoded using the mid-points of items to create a measure of frequency. As with the referent period, adaptations for scoring are also possible. The authors of the scales note that a quantitative response format is preferable since different meanings can be inferred from terms such as "*often*" and "*frequently*", thereby reducing the validity of between- respondent comparisons (Straus et al., 1996; Straus, 2012). Scoring lends itself to the measurement of two variables: a prevalence variable and a chronicity variable. The prevalence score enables a researcher to say, for example, that a certain percentage of a group had physically assaulted a partner, or to look at differences in psychological or other characteristics between those have and have not aggressed against an intimate partner (e.g., Brzozowski et al., 2018; Theobald, Farrington, Coid, & Piquero, 2016). The chronicity variable, on the other hand, might be more valuable for measuring treatment efficacy, and this variable allows for one to examine relationships between dimensional

constructs of interest (e.g., the “Big Five” personality traits) and chronicity of partner aggression, even where count data refers to infrequently occurring events (see Hines & Saudino, 2008).

The CTS2 has scales developed to measure three tactics that are frequently employed in conflicts between partners: physical aggression, verbal aggression and negotiation. It also has supplementary scales derived to measure injury and sexual coercion. The various scales can be selected to suit the purpose at hand.

Each of the scales contain two subscales, minor and severe behaviours, and emotional and cognitive scales for negotiation. Information about the scales is provided in Table 1.

The flexible application of the CTS2, and the resultant nature and number of variables that can be measured, allows relevant data to be obtained depending on particular measurement aims (Straus, 2012).

However, such adaptations make it difficult to compare across studies and caution is therefore necessary when applying psychometric properties of the standard format to its derivatives. Straus et al. (1996) note, however, that the CTS1 was found to be robust to modifications (e.g., Pan et al., 1994), and to produce findings that were comparable with the original format.

Table 1. Information about the scales and items comprising the CTS2

Scale	Development and definition	Items
Negotiation	Replaced 'Reasoning' scale of CTS1. Examines frequency of discussion tactics employed to settle conflict, and level of concern showed to their partner through cognitive and emotional techniques.	6 items: 3 emotional e.g., " <i>I showed my partner I cared even though we disagreed</i> "; 3 cognitive e.g., " <i>I explained my side of a disagreement to my partner</i> "
Psychological aggression	Replaced 'Verbal aggression' scale. Asks about frequency of tactics used that cause psychological distress to partner, including verbal and non-verbal aggressive acts.	8 items: 4 minor e.g., " <i>I shouted or yelled at my partner</i> "; 4 severe e.g., " <i>I threatened to hit or throw something at my partner</i> "
Physical assault	Asks about physical aggression and physical tactics used.	12 items: 5 minor e.g., " <i>I slapped my partner</i> "; 7 severe e.g., " <i>I choked my partner</i> "
Sexual coercion	Asks about the frequency of behaviour intended to coerce partner to engage in unwanted sexual activity.	7 items: 3 minor e.g., " <i>I insisted on sex when my partner did not want to (but did not use physical force)</i> "; 4 severe e.g., " <i>I used force (like hitting, holding down, or using a weapon) to make my partner have sex</i> "
Injury	Asks about physical injury received/inflicted due to physical conflict. Severity is assessed by asking about the need for medical attention.	6 items: 2 minor " <i>I had a sprain, bruise, or small cut because of a fight with my partner</i> "; 4 severe e.g., " <i>I had a broken bone from a fight with my partner</i> "

Psychometric properties of the CTS2

Although Straus et al. (1996) argued that the evidence supporting the validity of the CTS1 could apply to the CTS2 due to their conceptual and methodological similarities, for the purpose of this critique only the properties of the CTS2 are examined due to the number of alterations that were made when the scales were revised (as previously highlighted).

2.2. Instrument characteristics

The standard level of measurement for the CTS2 is ordinal level data, which provides units for analysis. The severity variable can be used to classify respondents into mutually exclusive categories of 'no violence', 'minor only', or 'severe'. Thus, it can be used as a nominal variable or a three-level ordinal scale (Straus & Douglas, 2004).

The instrument is designed for self-report. Notwithstanding the benefit of ease of administration, the validity of the CTS2 is undermined by biases inherent in self-reporting, including memory bias and response bias (Junger-Tas & Marshall, 1999). Specifically, respondents must be both willing to and able to admit to the act (i.e., they must be honest and they must be able to recall correctly).

As the CTS2 includes behaviours that are socially unacceptable and punishable by law, respondents may not be willing to respond truthfully. This is particularly pertinent in community samples, who may be more keen to present in a prosocial manner, and could therefore be expected to have greatest impact on the reporting of more serious acts in this population (see Junger-Tas & Marshall, 1999). For incarcerated populations, however, the reliability and validity of self-report data have been found to be higher than police data (Chaiken & Chaiken, 1982). Thus, the impact of social desirability effects on reporting IPV may vary according to the sample studied. That said, a meta-analysis exploring IPV reporting and social desirability found relatively small negative associations between the two constructs (Sugarman & Hoating, 1997) and, consistent with this, Straus (2004) found little correlation between social desirability

scores and the CTS2. Indeed, threats posed to validity through impression management can be managed through anonymous reporting in research, although this would not be feasible in a therapy or correctional context.

There is evidence to suggest that less severe acts may be more easily forgotten, and that more frequently perpetrated offences are more likely to be underreported (Junger-Tas & Marshall, 1999); something which Straus himself acknowledges as an issue with the scales (Straus, 2012). This latter finding has implications for incarcerated populations, who could be assumed to have perpetrated acts with greatest frequency. Although limiting the referent period to a shorter time-frame could help to reduce inaccuracies caused by memory biases, this does not avoid problems with memory for temporal sequencing (Junger-Tas & Marshall, 1999). As such, respondents might incorrectly judge acts to fall inside or outside the referent period, serving to distort data obtained. Indeed, memory has been found to be particularly problematic when respondents are asked to recall quantitative facts (Junger-Tas & Marshall, 1999), which further threatens the validity of the chronicity variable of the CTS2.

2.3. Reliability

3.2.1 Internal Reliability

Internal reliability refers to the extent to which items within a particular scale measure the same construct. In the pilot study of the CTS2, the scales were found to have good internal consistency, with coefficients as high or higher than those reported for the CTS1 (Straus et al., 1996). Specifically, coefficients of the scales ranged from .79 (psychological aggression) to .95 (injury). Straus et al. (1996) note that the psychological aggression scale has the lowest internal consistency because some items were selected to increase the diversity of its content as opposed to enhancing its internal consistency. Thus, the authors note that the scale attempts to balance internal consistency with representation of different forms of psychological aggression. Across 41 papers published up until 2005, the mean reliability coefficient was found to be .77 (Straus, 2005). However, alpha coefficients as low as .34 were reported in samples where

behaviours were altogether absent, such as attacking a partner with a knife or gun. In a recent review of 34 methodologically diverse studies, Cronbach's alpha estimates for the various scales of the CTS2 were found to be high, with the lowest reliability coefficients found for the sexual coercion scale (Costa & Barros, 2016).

Good coefficients have been found for the psychological aggression and physical aggression scales in a sample of pregnant women (ranging .73 - .77) (Hellmuth, Gordon, Stuart & Moore, 2013) and in a community sample of males and females (ranging from .74 - .88) (O'Leary & Williams, 2006). In the latter study, a good coefficient was found for injury reported by women by their partners (.74), but injury of the man by the woman had lower internal consistency (.47). Low coefficients were also found for the sexual coercion scale (.17 - .42). A coefficient as low as .18 has been reported for the sexual coercion scale in a community sample of females (Yun, 2010). However, this coefficient was yielded for the *minor* sexual coercion scale, with the *severe* sexual coercion scale yielding a coefficient of .77.

In terms of legal samples, Shorey and colleagues reported an alpha coefficient of .78 for the psychological aggression scale and .80 for the physical aggression scale using a sample of men and women who had been referred by the court for batterer programmes (Shorey, Ninnemann, Elmquist, Labrecque & Zucosky, 2012). The internal consistencies of individual scales have been found to be even greater among an incarcerated population of Spanish males (.80 - .83). However, the injury scale showed low reliability in this sample (.59) (Loinaz, Echeburua, Ortiz-Tallo & Amor, 2012). Lucente, Fals-Stewart, Richards and Goscha (2001) reported high levels of consistency across each scale using a population of incarcerated females (> 0.74), with the exception of the sexual coercion subscale in this sample (.34). Similarly, Tuomi Jones, Ji, Beck and Beck (2002) found moderate to excellent reliability for each of the scales using a sample of incarcerated women (.62- .91), again with the lowest coefficient found for the sexual coercion scale.

Thus, overall, the scales have demonstrated good internal consistency in a variety of samples. However, given the low coefficients reported for the sexual coercion scale in female samples and, to a lesser extent, the injury scale in male samples, these scales may benefit from being supplemented with interviews to provide context and further information pertaining to the items when used with the respective samples (O'Leary & Williams, 2006).

3.2.2 Cross-cultural reliability

The CTS2 has good cross-cultural reliability, at least among college samples. Straus (2004) pooled data from 33 studies across 17 countries representing every major world region except Africa and found that even the lowest coefficient surpassed "good" reliability (.70). Small differences were found between male and female students, with the scales having slightly higher reliability for male students (.78- .93 compared to .72- .87). More recently, cross-cultural reliability has been confirmed using other populations, including Spanish IPV men (Loinaz et al., 2012) and an Italian female community sample (Signorelli, Arcidiacono, Musumeci, Nuovo & Aguglia, 2014). However, a greater evidence base is still needed to gain a better picture of cross-cultural reliability in varied clinical and forensic settings given potential nuances in patterns of responding across cultures and ethnic groups.

3.2.3. Inter-rater agreement

Inter-partner agreement scores have been found to generally range from just short of medium to large, depending on the strategy used to determine reliability, the scale being assessed (Vega & O'Leary, 2007; O'Leary & Williams, 2006) and the sample reporting (e.g., Caetano, Schafer, Field & Nelson, 2002). Perhaps not surprisingly, higher partner-agreement has been found on more objective and specific items, as defined by the extent to which an independent observer could identify the act as having happened or not (Simpson & Christensen, 2005).

Cuenda and colleagues found that reports of psychological aggression were similar across perpetrators' and victims' reports but that the same was not true for physical aggression or sexual coercion. This would suggest that dyads may be less reliable in reporting their involvement in these latter forms of aggression, at least within a community sample (Cuenda et al., 2015). The finding of lower agreement on acts of sexual coercion is consistent with earlier research that found low agreement for sexual coercion and injury in a community sample (O'Leary & Williams, 2006.) However, in this study, dyadic agreement on physical aggression was found to be similar to dyadic agreement on positive relationship acts. As such, O'Leary and Williams highlight the importance of considering dyadic concordance among the scales in light of partner rates of agreement for other behaviours.

Cuenda and colleagues highlighted that a number of factors can influence the concordance of dyadic report, including social desirability and relationship satisfaction (Cuenda et al., 2015). To overcome the challenges of inter-partner agreement ratings, some studies have used a method known as *maximum dyadic report*, whereby the higher report of the dyad is utilised (see O'Leary & Williams, 2006; Cuenda et al., 2015). Inter-rater concordance is particularly relevant for clinical and legal purposes, and the need to take account of contextual factors becomes critical in these settings. However, it should be noted that the findings of inter-rater agreement from community samples cannot not be reliably generalised to clinical and legal samples.

3.2.4 Test-retest reliability

Test-retest reliability refers to the extent to which scores are consistent over time. Among a sample of convicted IPV offenders, Vega and O'Leary (2007) found strong stability for physical assault ($r = .76$), injury ($r = .70$), psychological aggression ($r = .69$) and negotiation ($r = .60$). However, they found low stability for the sexual coercion scale ($r = .30$). Using a slightly adapted version of the CTS2, Goodman et al. (1999) examined temporal consistency in the reports of participants with mental illness and found good consistency for the scales assessed (physical assault, sexual coercion and injury) across both genders

(79-90% for women and 62-81% for men). There has been minimal research exploring test-retest reliability for the CTS2, and thus Straus, Hamby and Warren (2003) recommend drawing comparisons between those of an individual and that observed in a similar group.

2.4. Validity

3.3.1 Face validity

Face validity refers to the extent to which a test appears to measure what it is supposed to measure. It includes the wording of items and how this might affect responding. To improve the scales' face validity, the authors of the CTS2 amended the wording of the original CTS, making it more explicit. Moreover, the CTS2 balances face validity with demand characteristics by interspersing the order of the questions, which helps to reduce response sets (Dahlstrom, Brooks & Peterson, 1990). Indeed, Ramirez and Straus (2006) found that presenting questions in a slightly modified order to university students resulted in higher disclosure rates for physical assault, injury and sexual coercion, as compared to presenting questions in a sequential order, possibly owing to the redundancy of response sets.

3.3.2 Concurrent validity

Concurrent validity is indicated by the correlation of the CTS2 with other measures of the five constructs (i.e., negotiation, physical assault, psychological aggression, sexual coercion and injury). Few studies have examined the concurrent validity of the CTS2. In 2004, all of the five studies that had examined the concurrent validity of the CTS2 found that the scales correlated with other measures of roughly the same constructs (Straus, 2004).

Comparison of the CTS2 scores with official sources can be used to provide a reference point for concurrent validity (Erickson & Empey, 1963). The CTS2 has been linked to increased disclosure rates relative to other measures of IPV, such as the National Crime Victimization Survey (Straus, 2007), which could be a testament to the sensitivity of the measure. Indeed, a sensitive measure (i.e., its ability to detect

the presence of a phenomenon) is particularly crucial for self-report measures of undesirable behaviour (Straus, 2007).

Due to the largely held view of the CTS as the “gold standard” for IPV measurement (e.g., Kraanen, Vedal, Scholing & Emmelkamp, 2013), the CTS2 itself has been used to assess the construct validity of other measures. For example, significant correlations have been found between the CTS2 and the Jellinek Inventory for assessing Partner Violence (J- IPV; Kraanen et al., 2013), a screening tool for measuring victimisation and perpetration of IPV over the past year. Significant correlations have also been found between the scales of the CTS2 and the Abusive Behaviours Checklist (ABC; Beck & Beck, 1998), the Abusive Behaviour Inventory (ABI; Shepard & Campbell, 1992), and the Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER, Kropp, Hart & Belfrage, 2005) (Au, Cheung, Kropp, Yuk-Chung, Lam & Sung, 2008; Tuomi Jones et al., 2002; Zink, Klesges, Levin & Putman, 2007;). However, it is not possible to make inferences about the concurrent validity of the CTS2 based on these correlations without making assumptions about the validity of the newer tests.

3.3.3. Predictive validity

The CTS2 was not devised to be a predictive measure (Straus et al., 2003). However, it measures past behaviour and past behaviour is often said to be the best predictor of future behaviour (see Ouellette & Wood, 1998). Indeed, the Spousal Assault Risk Assessment (SARA; Kropp, Hart, Webster & Eaves, 1995) identifies a number of risk factors for IPV perpetration, and, whilst this includes many factors measured by the CTS2 (such as physical and sexual assault), it also includes factors which are not (e.g., employment problems). This suggests that the CTS2 may not be an adequate predictor of risk for future IPV perpetration when used in isolation.

Researchers have suggested that the psychological aggression scale of the CTS2 can be a useful predictive tool for using alongside clinical assessment for the prediction of physical violence (Salis, Salwen & O’Leary, 2014). In particular, Salis and colleagues found that when individuals scored at the

80th percentile on the psychological aggression scale of the CTS2 there was a 70% chance that a man was physically aggressive and this rose to 85% chance for a female. Furthermore, scoring in the 80th percentile or higher was predictive of physical aggression one year later.

3.3.4 Content validity

Content validity refers to the extent to which a measure reflects the entire construct that it is purporting to examine. The CTS2 was derived to assess all aspects of IPV, thereby covering physical and psychological abuse and sexual coercion. However, for practicality, the CTS2 can only include a sample of a limitless number of violent behaviours (Straus, 2007). Although the questions themselves can be assumed to be valid since they ask about behaviour, this does not mean that the strategy used to select the items (i.e., qualitative interviews, suggestions and reviews) ensured that a sufficient sample of acts were represented in the measure (Straus, 2007).

Indeed, in the measure's development, the authors included items that they considered to be "inappropriate" (Straus et al., 2003, p. 8) and only included acts that are common to all couples (pet-related aggression, for example, is not included). It is noted, however, that Dobash & Dobash (1984) derived a list of violent acts from qualitative methods that showed considerable overlap with the CTS items. According to Straus (1990), this provided evidence for an adequate coverage of violent acts represented in the CTS.

3.3.5 Construct validity

Construct validity refers to the extent to which a test measures what it is proposed to measure. It can be examined by exploring the instrument's correlation with other variables that are known, or expected, to be theoretically associated with the construct purportedly measured (Campbell & Fiske, 1959; Straus et al., 1996). Although the overarching construct of the CTS is conflict tactics, the scales assess four different

areas of tactics, and thus construct validity of the CTS2 is best understood by examining each of the scales (Straus et al., 1999).

First, a core premise of conflict theory is that inequality between persons increases the risk of violence, which is used as a means for the dominant person to maintain their position or for the subordinate person to balance power (Cosser, 1967; Dahrendorf, 1959). In line with this, positive correlations have been found between dominance and physical assault scores, irrespective of gender (Straus 2004). Second, consistent with empirical evidence indicating that childhood physical abuse is a risk factor for later violence (e.g., Widom, 1989), Straus (2004) found that corporal punishment in childhood correlated with physical assault perpetration scores.

In terms of correlations that would be expected between subscale scores, Straus and colleagues found that high physical assault perpetration rates correlate with high injury rates, and, in accordance with gender differences in size and strength (e.g., Archer, 2000), a higher correlation between these scores was found for males. Psychological aggression and physical assault have been found to be highly correlated (Straus et al., 1996; Salis et al., 2014) and Straus et al. (1996) also found a low correlation between negotiation and sexual coercion and injury.

While Straus and colleagues have argued that the quantitative and objective stance of the CTS2 is a strength as it limits minimisation, denial or distorted cognitions (e.g., Straus et al., 2003), others have argued that motives and meaning are vital for making sense of behaviour, and thus determining whether the behaviour constitutes IPV (e.g., Kimmel, 2002). In support of the latter, research employing interview data found that 58% of female undergraduates were miscategorized using the CTS2 due to their having engaged in mock violence; which, objectively, would meet the criteria of the behaviour-based CTS2 (Lehrner & Allen, 2014). Furthermore, one study found that 78% of 'victims' considered every act of physical aggression they received from their partner as evidence of their partner "playing around" (Jouriles, Platt, & McDonald, 2009), and, similarly, another study found that over one third of female

victims of violent acts did not consider themselves to have experienced “physical abuse”, as a “victim of violence” or as a “battered woman” (Hamby & Gray-Little, 2000).

Lehrner and Allen (2014) argue that the validity of interview data obtained can be inferred from participants’ willingness to report incidents of intentional violence, as well as from research finding that play violence is frequent in dating relationships (e.g., Perry & Fromouth, 2005; Ryan & Mohr, 2005).

Although aggressive play has been found to be related to risk of violence by intimate partners (Gonzalez-Mendez & Hernandez-Cabrera, 2009), Lehrner and Allen (2014) argue that playful violence in this context is not evidence of a conflict tactic, which the CTS2 purports to assess. Such findings raise concerns that the CTS2’s focus on behaviour without context can inflate rates of IPV in this population (i.e., female undergraduates) (Lehrner & Allen, 2014). Indeed, although the CTS2’s instructions specify acts used during an “argument”, it is clear that respondents do not reliably distinguish between acts that occurred inside and outside of conflict (Lehrner & Allen, 2014).

A related issue is that, by limiting IPV measurement to acts used in the context of an argument, the CTS2 has a limited coverage of IPV, failing to measure acts used to control a partner, for example (Dobash, Dobash, Wilson, & Daly, 1992; Foshee, Bauman, Linder, Rice & Wilcher, 2007; Sillito, 2012).

Furthermore, despite acknowledging the mutuality of violence, the CTS2 does not attempt to match acts to a particular event or determine the ordering of these events (Britton, 2011). Consequently, an act of self-defence would receive the same score as an act of instigated violence and thus its reputation as a measure of reciprocity has been classed as misleading (e.g., Britton, 2011; Krahe, Bieneck & Möller, 2000).

To summarise, the scores of the CTS2 cannot reliably distinguish between acts that reflect conflict tactics (deliberate or self-defence) or acts used outside of conflict (controlling or playful behaviour). Thus, by limiting contextual information, the CTS2 loses valuable information about the nature of behaviour, which ultimately impacts upon its discriminant validity. In response to these issues, the authors of the

CTS2 have argued that the purpose of the measure is to assess rates of IPV perpetration, not context, highlighting that the CTS2 is intended for use in conjunction with other tools to make sense of the scores (Straus et al., 2003). Furthermore, Straus (1990) has argued that measuring context alongside behaviour makes assumptions about a relationship.

At the very least, the discrepancies noted above highlight the need for a more clearly defined construct. They also demonstrate the complexity of defining and measuring IPV (Waltermaurer, 2005). With these caveats in mind, multi-modal methods incorporating interview data have been proposed for screening and assessment of IPV (see Lehrner & Allen, 2014).

Further factors threatening the construct validity of the CTS2 include that, as standard, it assesses conflict tactics employed in the last 12 months and only in the current relationship. Conflict tactics measured using these instructions give a limited overview of IPV and assume heterogeneity in behaviour across both time and relationships (Kimmel, 2002; Lehrner & Allen, 2014). Another factor, and one which Straus and colleagues have identified as a reason for differing IPV rates obtained in research (Straus et al., 2003), relates to how the instrument is presented, such as the instructions given. Indeed, Hamby and Finkelhor (2000) reported that advertising the CTS2 as a 'crime survey' rather than a 'family survey' affects reporting. This is perhaps not surprising given research demonstrates that wording can serve as a cue for recall or can bias memory (see Junger-Tus & Marshall, 1999).

Finally, research carried out by Loinaz et al. (2012) demonstrates convergent and discriminant validity of the CTS2. In terms of the former, the researchers found significant relationships between conviction of physical assault against a partner and the overall frequency of reported physical violence. What is more, the severity of reported physical assault was significantly related to conviction for physical assault. In terms of discriminant validity, the CTS2 differentiated the IPV men from the general population on all scales, except the sexual coercion scale, where there were no significant group differences. These findings raise doubts about the discriminant validity of the sexual coercion scale, and when added to concerns

around internal reliability (reviewed earlier), suggest the need for some caution when interpreting results related to this scale. Furthermore, only a *small* difference was found between offenders and non-offenders on the negotiation scale (as measured by Cohen's *d*).

2.5. Factor structure

The factor structure of the CTS2 has been largely explored using female samples. For example, Lucente and colleagues found support for the five-factor model using incarcerated female substance abusers (Lucente et al., 2001), while Signorelli and colleagues found evidence for a five-factor solution (corresponding to negotiation, violence, extreme violence, injury and sexual coercion) in an Italian female community sample (Signorelli, Arcidiacono, Musumeci, Nuovo, & Aguglia, 2014). Other studies have also demonstrated a five-factor solution among female samples using confirmatory factor analyses (Straus, 2004; Connelly, Newton & Adams, 2005).

Newton, Connelly & Landsverk (2001) found that a five-factor model using minor and severe categories for psychological aggression, physical aggression, and negotiation produced a better fit of the data than did a three-factor model combining minor and severe scales of psychological and physical aggression for a sample of high-risk postpartum women. Although including both severities in this way means that the scores are not biased by minor acts, even within each category (i.e., minor and severe) the items cover a range of severities (Newton et al., 2001). In line with this, factor analyses have frequently shown that the weapon items form a distinct factor from the other items (Straus, 1979, 1990).

In contrast, Anderson and Leigh (2010) found that, with the exception of Negotiation, the Perpetration items failed to conform to the intended factor structure among a sample of deaf female students. The researchers commented that this may have been attributable to a lack of variance in the data. However, it is also important to consider the specific sample used by Anderson and Leigh and that, therefore, nuances in response styles between population groups may indicate that caution is warranted in assuming the same

factor structure is applicable across different samples. In support of this, Connelly et al. (2005) reported a weaker factor structure in the Spanish version of the CTS2.

Other researchers have suggested that a four-factor solution may represent a better fitting model for the CTS2, although the scales forming this model have varied across research. For example, Tuomi Jones et al. (2002) combined psychological and physical scales to produce a four-factor solution in a sample of incarcerated women, whilst Loinaz et al. (2012) found that physical, sexual, psychological violence and negotiation explained 51.4% of the total variance in a sample of convicted IPV men. However, some items were found to load onto more than one factor, thereby demonstrating overlap between the types of violence. For example, "*I damaged something that belonged to my partner*" was reported to load onto two factors corresponding to physical and psychological aggression.

In summary, a five-factor model has been established in female clinical and community samples. However, this has not been consistently found among different populations, including males and female minority groups. Given that the factor structure has been shown to vary according to gender (Schafer, 1996) and in samples characterised by high levels of aggression (e.g., Tuomi Jones et al., 2002), further research is needed to form a better understanding extent to which the factor structure of the CTS2 is consistent across more diverse samples, including among males in the community and in secure settings, and when using the CTS2 in different cultures (Connelly et al., 2005).

2.6. Normative samples as a reference group

A normative sample is a group of people assumed to be representative of the larger population who may utilise the measure. The norm group's data are utilised as a reference for evaluating future scores obtained on the measure.

College students comprise the reference sample of the CTS2 and it is recommended that this sample is used to compare the data of other college students (Straus et al., 2003). However, the prevalence of IPV is

higher in college students than in married couples. For example, Stets & Straus (1989) found rates of 25-30% in the former relative to 16% for the latter, and incarcerated populations have yet higher perpetration rates (Tuomi Jones et al., 2002). Moreover, a review conducted in 2012 found that rates of IPV varied significantly as a function of the sample, with college students and legal samples reporting the highest rates of perpetration (Desnarias, Reeves, Nicholls, Telford & Fiebert, 2012). The authors acknowledge that more normative groups are needed to compare data obtained from respondents who fall outside the reference sample for the CTS2 (i.e., primarily white, middle-class, educated, young adults). Indeed, the CTS2 has frequently been administered to clinical populations, including men with alcohol use disorders (Panuzio et al., 2006), participants with mental illness (Goodman et al., 2009), incarcerated female substance abusers (Lucente et al., 2001), incarcerated IPV men (Loinaz et al., 2012), and pregnant and postpartum women (Hellmuth et al., 2013 and Netwon et al., 2001, respectively). Whilst the CTS2 has been used successfully with individuals from various countries and cultural backgrounds, further research is needed with more diverse reference groups to more effectively establish the scales' validity and reliability in these populations. For instance, Anderson and Leigh (2010) found that three scales did not evidence reliability and that the factor structure was not valid for perpetration items among deaf female students. Additionally, a weaker internal consistency and factor structure has been reported in the Spanish CTS-2 (Connelly et al., 2005). Cultural, economic and political factors may be expected to impact on the reporting of partner violence and these variables therefore play an important role in developing appropriate norms among different reference groups (Jones, Browne & Chou, 2017). As such, further sample populations are doubtlessly needed for normative comparisons to be facilitated.

Standardised scores are not provided for the CTS2 as it is not intended for diagnostic purposes. However, the authors do advise that a score of one or more on the physical scale warrants further exploration and intervention (Straus, 2007).

3. Conclusion

The CTS2 is a versatile tool that can be used in a variety of ways in a number of settings. Despite a relative dearth of research examining the psychometric properties of the CTS2 in comparison to the CTS1, it has generally been found to be a reliable and valid instrument to measure IPV in different populations and across cultures. However, there are some concerns regarding the internal consistency of the sexual coercion scale among female samples, and consistency among dyad report has been shown to be poor for reporting sexual coercion in community samples. It is further noteworthy that explorations of the factor structure of the CTS2 have yielded inconsistent models. It appears that research using more specific populations has produced greater variance in the aforementioned properties and further research using more diverse populations would therefore be helpful in enhancing understandings of the validity of the CTS2.

As some psychometric properties of the CTS2 have been found to vary according to gender and level of aggression in the sample, it is important not to over-extrapolate findings between different samples.

Although some statistical properties of the scale have indeed been established for different populations, including incarcerated female substance abusers and IPV men, more research is doubtlessly needed to ascertain the validity and reliability of the CTS2 in both community and incarcerated samples of males and females. This will be important for the scales' reputation as the "gold standard" of IPV measurement to be upheld, and is paramount given the measure's frequent use in identifying IPV.

IPV is a challenging phenomenon to define and measure (Waltermaurer, 2005). While the behaviourally-based CTS2 is able to objectively identify acts suggestive of IPV, it is clear that accuracy in identifying IPV perpetration extends beyond simple administration and requires an understanding of the context in which the behaviour took place (Waltermaurer, 2005). Although the CTS2 could be criticised for being reductionist in its sole focus on the presence of an act, and largely ignoring the context in which the act took place, it does afford measurement of the type, severity and frequency of a range of conflict tactics, and thereby enables some differentiation of individuals' IPV perpetration. This represents an important property of any IPV measure, especially given the considerable heterogeneity among individuals with a

history of IPV perpetration (see Holtzworth-Munroe & Meehan, 2004). Nonetheless, the need to gather both quantitative and qualitative data in IPV assessment is important in order to more accurately assess and make sense of the use of violence within relationships (Strauchler et al., 2004).

Referent periods used in IPV measurement have an important role in determining understandings of the incidence or prevalence of IPV. Thus, Waltermaurer (2005) notes that researchers utilising the CTS2 and selecting referent periods for IPV measurement need to be mindful of their goals, since the time-frames used to determine IPV can essentially measure different types of abuse (i.e., within or across relationships). Researchers also need to be aware that adaptations of the scales will limit comparability with other studies of IPV and influence the extent to which psychometric properties can be applied. Nevertheless, Waltermaurer (2005) highlights that instrument consistency should not be at the cost of adding to the growing understanding of IPV. The strengths and limitations of the CTS2 outlined in this review have important implications for interpreting findings of studies investigating IPV using only this tool.

It is important to highlight that the review carried out was not a systematic review and, while efforts were made to present a widespread and representative body of literature through database searches, citation searches and reference list searches, given the large body of work carried out using the CTS2, it is entirely possible that a number of studies reporting on the properties of the scale have not been cited as part of this review. Furthermore, non-English papers have not been represented in the review. Additionally, this review focused specifically on CTS2 for measuring IPV perpetration. The CTS2 has since been modified for the assessment of other forms of violence, including sibling violence (Fernandes et al., 2016; Relva et al., 2013) and caregiver violence (Straus et al., 1998), and the psychometric properties of such versions of the scales have not been examined as part of this review; and therefore, nor should the findings be extended to these versions. Given the size of the literature base, a systematic review would be a good way to access articles on the CTS2, and its derivatives, and to assess for the quality of the research studies in drawing conclusions.

References

- Adams, D. N. (1965). Coercion and consensus theories: Some unresolved issues. *American Journal of Sociology, 71*, 714-716.
- American Psychological Association, intimate partner violence facts and resources. Retrieved from:
<http://www.apa.org/topics/violence/intimate-partner-violence.pdf>
- Anderson, M. L., & Leigh, I. W. (2010). Internal consistency and factor structure of the Revised Conflict Tactics Scales in a sample of deaf female college students. *Journal of Family Violence, 25*(5), 475-483.
- Angenent, M. (1984). "Medewerking aan Enquetes over Niet-geregistreeerde Criminaliteit." *Tijdschrift voor Criminologie, 6*, 345-55.
- Archer, J. (2000). Sex differences in aggression between heterosexual partners: A meta-analytic review. *Psychological Bulletin, 126*, 651-680.
- Au, A., Cheung, G., Kropp, R., Yuk-Chung, C., Lam, G. L. T., & Sung, P. (2008). A preliminary validation of the Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER) in Hong Kong. *Journal of Family Violence, 23*(8), 727-735. doi:10.1007/s10896-008-9198-z
- Beck, M., & Beck, N. (1998). The Abusive Behaviour Checklist. Unpublished instrument.
- Black, D. S., Sussman, S., & Unger, J. B. (2010). A further look at the intergenerational transmission of violence: Witnessing interparental violence in emerging adulthood. *Journal of Interpersonal Violence, 25*(6), 1022-1042. doi:10.1177/0886260509340539.
- Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health, 27*(3), 281-291.

- Brzozowski, A., Gillespie, S. M., Dixon, L., & Mitchell, I. J. (2018). Cardiac autonomic function and psychological characteristics of heterosexual female perpetrators of intimate partner physical aggression. *Journal of Interpersonal Violence*. doi: 10.1177/0886260518775748.
- Caetano, R., Schafer, J., Field, C., & Nelson, S. M. (2002). Agreement on reports of intimate partner violence among white, black, and Hispanic couples in the United States. *Journal of Interpersonal Violence*, 17(12), 1308-1322.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105.
- Campbell, J. C. & Lewandowski, L. A. (1997). Mental and physical health effects of intimate partner violence on women and children. *Psychiatric Clinics of North America*, 20, 353–374.
- 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(2), 231-280.
- Carter, L. S., Weithorn, L. A., & Behrman, R. E. (1999). Domestic violence and children: Analysis and recommendations. *Future of Children*, 9, 4–20.
- Cascardi, M., Langhinrichsen, J., & Vivian, D. (1992). Marital aggression: Impact, injury and health correlates for husbands and wives. *Archives of Internal Medicine*, 152, 1178–1184.
doi:10.1001/archinte.152.6.1178
- Catalano, S. (2013). Intimate partner violence: Attributes of victimization, 1993–2011. Washington, DC: US Dept of Justice, Office of Justice Programs, Bureau of Justice Statistics.
- Chaiken, J. M., & Chaiken, M. R. (1982). *Varieties of criminal behaviour*. Santa Monica, California:Rand.

- Coker, A. L., Smith, P. H., Bethea, L., King, M., McKeown, R. E. (2000). Physical health consequences of physical and psychological intimate partner violence. *Archives of Family Medicine*, 9, 451-457. doi:10.1001/archfami.9.5.451.
- Connelly, C. D., Newton, R. R., & Aarons, G. A. (2005). A psychometric examination of English and Spanish versions of the revised Conflict Tactics Scales. *Journal of Interpersonal Violence*, 20(12), 1560–1579.
- Coser, L. A. (1967). *Continuities in the study of social conflict*. New York: Free Press.
- Costa, D., & Barros, H. (2016). Instruments to assess intimate partner violence: a scoping review of the literature. *Violence and Victims*, 31(4), 591 – 621. DOI:10.1891/0886-6708
- Cuenca, M. L., Graña, J. L., & Redondo, N. (2015). Differences in prevalence of partner aggression as the revised conflict tactics scale: Individual and dyadic report. *Behavioral Psychology/Psicología Conductual*, 23(1), 27 – 140.
- Dahlstrom, W. G., Brooks, J. D., & Peterson, C. D. (1990). The Beck Depression Inventory: Item order and the impact of response sets. *Journal of Personality Assessment*, 55, 224-233.
- Dahrendorf, R. (1959). *Class and class conflict in industrial society*. Palo Alto, CA: Stanford University Press.
- Dixon, L., & Graham-Kevan, N. (2011). Understanding the nature and etiology of intimate partner violence and implications for practice and policy. *Clinical Psychology Review*, 31, 1145-1155.
- Dobash, R. E., & Dobash, R. P. (1984). The nature and antecedents of violent events. *The British Journal of Criminology*, 24(3), 269-288.

- Dobash, R. P. & Dobash, R. E. (2004). Women's violence to men in intimate relationships: Working on a Puzzle. *British Journal of Criminology*, 44, 324- 349. doi:10.1093/bjc/azh026.
- Erickson, M. L., & Empey, L. T. (1963). Court records, undetected delinquency and decision-making. *Journal of Criminal Law, Criminology and Police Science*, 54(4), 456-470.
- Esquivel-Santoveña, E. E., & Dixon, L. (2012). Investigating the true rate of physical intimate partner violence: A review of nationally representative surveys. *Aggression and Violent Behavior*, 17(3), 208-219.
- Feld, S. L., & Straus, M. A. (1989). Escalation and distance of wife assault in marriage. *Criminology*, 27(1), 141-162.
- Felson, R. B. (2006). Is violence against women about women or about violence? *Contexts*, 5(2), 21-25. doi:10.1525/ctx.2006.5.2.21.
- Fernandes, O. M., Relva, I. C., Rocha, M., & Alarcão, M. (2016). Construct validity study of the Revised Conflict Tactics Scales--Sibling Version/Estudo da validade de construto das Revised Conflict Tactics Scales--Versao Irmaos. *Motricidade*, 12(1), 69-83.
- Fincham, F. D., & Beach, S. R. (1999). Marital conflict: Implications for working with couples. *Annual Review of Psychology*, 50, 47-77.
- Fox, J. A., & Zawitz, M. W. (2007). *Homicide trends in the United States*. Washington, DC: Bureau of Justice Statistics.
- Goodman, L. A., Thompson, K. M., Weinfurt, K., Corl, S., Acker, P., Mueser, K. T., & Rosenberg, S. D. (1999). Reliability of reports of violent victimization and posttraumatic stress disorder among men and women with serious mental illness. *Journal of Traumatic Stress*, 12, 587-599.

- Hamby, S. L., & Finkelhor, D. (2000). The victimization of children: Recommendations for assessment and instrument development. *Journal of the American Academy of Child & Adolescent Psychiatry, 39*(7), 829-840.
- Hamby, S. L., & Gray-Little, B. (2000). Labelling partner violence: When do victims differentiate among acts? *Violence and Victims, 15*(2), 173-186.
- Hellmuth, J. C., Gordon, K. C., Stuart, G. L., & Moore, T. M. (2013). Risk factors for intimate partner violence during pregnancy and postpartum. *Archives of Women's Mental Health, 16*(1), 19-27.
- Hernández González, E., & González Méndez, R. (2009). Coerción sexual, compromiso y violencia en las relaciones de pareja de los universitarios. *Escritos de Psicología (Internet), 2*(3), 40-47.
- Hines, D. A., & Saudino, K. J. (2008). Personality and intimate partner aggression in dating relationships: the role of the "Big Five". *Aggressive Behavior, 34*(6), 593-604.
- Holtzworth-Munroe, A., & Meehan, J. C. (2004). Typologies of men who are maritally violent: Scientific and clinical implications. *Journal of Interpersonal Violence, 19*(12), 1369-1389.
- Holmes, M. R., Richter, F. G., Votruba, M. E., Berg, K. A., & Bender, A. E. (2018). Economic Burden of Child Exposure to Intimate Partner Violence in the United States. *Journal of Family Violence, 33*(4), 239 - 249.
- Jones, T. J., Browne, K., & Chou, S. (2017). A critique of the revised Conflict Tactics Scales-2 (CTS2). *Aggression and Violent Behavior, 37*, 83-90.
- Jouriles, E. N., Platt, C., & McDonald, R. (2009). Violence in adolescent dating relationships. *Prevention Researcher, 16*(1), 3-7.

- Junger-Tas, J., & Marshall, I. H. (1999). The self-report methodology in crime research. *Crime and Justice*, 25, 291-367.
- Kershaw, C., Budd, T., Kinshott, G., Mattinson, J., Mayhew, P., & Myhill, A. (2000). *The 2000 British Crime Survey England and Wales*. London: Home Office.
- Kimmel, M. S. (2002). Gender symmetry in domestic violence: A substantive and methodological research review. *Violence Against Women*, 8(11), 1332-1363.
- Kraanen, F. L., Vedel, E., Scholing, A., & Emmelkamp, P. M. (2013). Screening on perpetration and victimization of intimate partner violence (IPV): Two studies on the validity of an IPV screening instrument in patients in substance abuse treatment. *PLoS One*, 8(5): e63681.
doi:10.1371/journal.pone.0063681
- Kropp, P. R., Hart, S. D., & Belfrage, H. (2005). *Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER) user manual*. Vancouver, BC: ProActive ReSolutions.
- Kropp, P. R., Hart, S. D., Webster, C. D., & Eaves, D. (1995). *Manual for the Spousal Assault Risk Assessment Guide* (2nd ed.). Vancouver, BC: British Columbia Institute against Family Violence.
- Langhinrichsen-Rohling, J. (2005). Top 10 greatest “hits”. Important findings and future directions for intimate partner violence research. *Journal of Interpersonal Violence*, 20(1), 108-118.
- Ledlow, G. R. (2009). Conflict and interpersonal relationships. In J.A. Johnson (Ed.), *Health organizations: Theory, behaviour, and development* (pp. 149). Sudbury, Massachusetts: Jones and Bartlett Publishers.
- Lehrner, A., & Allen, N. E. (2014). Construct validity of the Conflict Tactics Scales: A mixed-method investigation of women’s intimate partner violence. *Psychology of Violence*, 4(4), 477-490.

Loinaz, I., Echeburúa, E., Ortiz-Tallo, M., & Amor, P. J. (2012). Psychometric properties of the Conflict Tactics Scales (CTS-2) in a Spanish sample of partner-violent men. *Psicothema*, 24(1), 142-148.

Lucente, S. W., Fals-Stewart, W., Richards, H. J., & Goscha, J. (2001). Factor structure and reliability of the Revised Conflict Tactics Scales for incarcerated female substance abusers. *Journal of Family Violence*, 16(4), 437-450.

Newton, R. R., Connelly, C. D., & Landsverk, J. A. (2001). An examination of measurement characteristics and factorial validity of the Revised Conflict Tactics Scale. *Educational and Psychological Measurement*, 61(2), 317-335.

Office for National Statistics (2015). Violence crime and sexual offences- Intimate personal violence and serious sexual assault. Retrieved from http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/dcp171776_394500.pdf.

Office for National Statistics (2017). Focus on Violent crime and sexual offences, February 2017, Appendix tables. Retrieved from <https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/datasets/appendixtablesfocusonviolentcrimeandsexualoffences>

O'Leary, K. D., & Murphy, C. (1992). Clinical issues in the assessment of spouse abuse. In R.T. Ammerman & M. Hersen (Eds.). *Assessment of family violence: A clinical and legal sourcebook* (pp.26-46). New York, John Wiley & Sons.

O'Leary, K. D., & Williams, M. C. (2006). Agreement about acts of aggression in marriage. *Family Psychology*, 20, 656-662.

- Ouellette, J. A., & Wood, W. (1998). Habit and intention in everyday life: The multiple processes by which past behaviour predicts future behaviour. *Psychological Bulletin*, *124*, 54-74.
- Pan, H. S., Neidig, P. H., & O'Leary, K. D. (1994). Male-female and aggressor-victim differences in the factor structure of the Modified Conflict Tactics Scale. *Journal of Interpersonal Violence*, *9*(3), 366-382.
- Panuzio, J., O'Farrell, T. J., Marshall, A. D., Murphy, C. M., Murphy, M., & Taft, C. T. (2006). Intimate partner aggression reporting concordance and correlates of agreement among men with alcohol use disorders and their female partners. *Assessment*, *13*(3), 266-279.
- Perry, A. R., & Fromuth, M. E. (2005). Courtship violence using couple data characteristics and perceptions. *Journal of Interpersonal Violence*, *20*(9), 1078-1095.
- Ramirez, I. L., & Straus, M. A. (2006). The effect of question order on disclosure of intimate partner violence: An experimental test using the Conflict Tactics Scales. *Journal of Family Violence*, *21*(1), 1-9.
- Relva, I. C., Fernandes, O. M., & Costa, R. (2013). Psychometric properties of revised conflict tactics scales: Portuguese sibling version (CTS2-SP). *Journal of Family Violence*, *28*(6), 577-585.
- Ryan, K. M., & Mohr, S. (2005). Gender differences in playful aggression during courtship in college students. *Sex Roles*, *53*(7), 591-601.
- Salis, K., Salwen, J., & O'Leary, K. (2014). The predictive utility of psychological aggression for intimate partner violence, *Partner Abuse*, *5*(1), 83-97.

- Schafer, J. (1996). Measuring spousal violence with the Conflict Tactics Scale notes on reliability and validity issues. *Journal of Interpersonal Violence, 11*(4), 572-585.
- Schwartz, M. (2000). Methodological issues in the use of survey data for measuring and characterizing violence against women. *Violence Against Women, 6*(8), 815-838.
- Shepard, M. F., & Campbell, J. A. (1992). The abusive behavior inventory: A measure of psychological and physical abuse. *Journal of Interpersonal Violence, 7*(3), 291-305.
- Shorey, R. C., Ninnemann, A., Elmquist, J., Labrecque, L., Zucosky, H., Febres, J., ... & Stuart, G. L. (2012). Arrest history and intimate partner violence perpetration in a sample of men and women arrested for domestic violence. *International Journal of Criminology and Sociology, 1*, 132 – 140.
- Signorelli, M. S., Arcidiacono, E., Musumeci, G., Di Nuovo, S., & Aguglia, E. (2014). Detecting domestic violence: Italian validation of Revised Conflict Tactics Scale (CTS-2). *Journal of Family Violence, 29*(4), 361-369.
- Simmons, C. A., Lehmann, P., Cobb, N., & Fowler, C. R. (2005). Personality profiles of women and men arrested for domestic violence: An analysis of similarities and differences. *Journal of Offender Rehabilitation, 41*(4), 63-81.
- Simpson, L. E., & Christensen, A. (2005). Spousal agreement regarding relationship aggression on the Conflict Tactics Scale-2. *Psychological Assessment, 17*(4), 423-432.
- Smith, C. A., Ireland, T. O., Park, A., Elwyn, L., & Thornberry, T. P. (2011). Intergenerational continuities and discontinuities in intimate partner violence: A two- generational prospective study. *Journal of Interpersonal Violence, 26*(18), 3720- 3752. doi: 10.1177/0886260511403751.

- 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*(2), 161-180.
doi:10.1007/BF01006627.
- Stets, J. E., & Straus, M. A. (1990). Gender differences in reporting marital violence and its medical and psychological consequences. In M. A. Straus & R. J. Gelles (Eds.), *Physical violence in American families: Risk factors and adaptations to violence in 8,145 families* (pp. 151-165). New Brunswick, NJ: Transaction Publishing.
- Stover, C. S., Meadows, A. L., & Kaufman, J. (2009). Interventions for intimate partner violence: Review and implications for evidence-based practice. *Professional Psychology: Research and Practice*, *40*(3), 223- 233. doi:10.1037/a0012718.
- Strauchler, O., Mccloskey, K., Malloy, K., Sitaker, M., Grigsby, N., & Gillig, P. (2004). Humiliation, manipulation, and control: Evidence of centrality in domestic violence against an adult partner. *Journal of Family Violence*. *19*, 339-346. 10.1007/s10896-004-0679-4.
- Straus, M. A. (1974). Leveling, civility, and violence in the family. *Journal of Marriage and the Family*, *36*, 13-29.
- Straus, M. A. (1979). Measuring intrafamily conflict and violence: The Conflict Tactics Scales. *Journal of Marriage and the Family*, *41*, 75-88.
- Straus, M. A. (1990). The Conflict Tactics Scales and its critics: An evaluation and new data on validity and reliability. In M. A. Straus & R. J. Gelles (Eds.), *Physical violence in American families: Risk factors and adaptations to violence in 8,145 families* (pp. 49-73). New Brunswick, NJ: Transaction Publications.

- Straus, M. A. (1993). Identifying offenders in criminal justice research on domestic assault. *American Behavioural Scientist*, 36(5), 587-600.
- Straus, M. A. (2004). Cross-cultural reliability and validity of the Revised Conflict Tactics Scales: A study of university student dating couples in 17 Nations. *Cross-Cultural Research*, 38, 407-432.
- Straus, M. A. (2005). Bibliography and tabular summary of publications on the revised Conflict Tactics Scales (CTS2 and CTSPC). Durham, NH: Family Research Laboratory, University of New Hampshire.
- Straus, M. A. (2007). Conflict Tactics Scales. In N. A. Jackson (Ed.), *Encyclopaedia of domestic violence* (pp. 190 – 197). New York, Routledge: Taylor & Francis Group.
- Straus, M. A. (2008). Dominance and symmetry in partner violence by male and female university students in 32 nations. *Children and Youth Services Review*, 30(3), 252-275.
- Straus, M. A. (2015). Dyadic concordance and discordance in family violence: A powerful and practical approach to research and practice. *Aggression and Violent Behavior*, 24, 83–94.
- Straus, M. A., & Douglas, E. M. (2004). A short form of the Revised Conflict Tactics Scales, and typologies for severity and mutuality. *Violence and Victims*, 19(5), 507-520.
- 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(3), 283-316.
- Straus, M. A., Hamby, S. L., Finkelhor, D., Moore, D. W., & Runyan, D. (1998). Identification of child maltreatment with the Parent-Child Conflict Tactics Scales: Development and psychometric data for a national sample of American parents. *Child Abuse & Neglect*, 22(4), 249-270.

Straus, M. A., Hamby, S. L., & Warren, W. L. (2003). *The Conflict Tactics Scale handbook*. Los Angeles, CA: Western Psychological Services.

Sugarman, D. B., & Hotaling, G. T. (1989). Dating violence: Prevalence, context and risk markers. In M. A. Pirog-Good & J. E. Stets (Eds.), *Violence in dating relationships: Emerging social issues* (pp. 3-32). New York: Praeger.

Theobald, D., Farrington, D. P., Coid, J. W., & Piquero, A. R. (2016). Are male perpetrators of intimate partner violence different from convicted violent offenders? Examination of psychopathic traits and life success in males from a community survey. *Journal of Interpersonal Violence, 31*, 1687-1718.

Thompson, M.P., Basile, K.C., Hertz, M.F., & Sitterle, D. (2006). *Measuring intimate partner violence victimization and perpetration: A compendium of assessment tools*. Atlanta (GA): Centers for Disease Control and Prevention, National Center for Injury Prevention and Control.

Tuomi Jones, N., Ji, P., Beck, M., & Beck, N. (2002). The reliability and validity of the Revised Conflict Tactics Scale (CTS2) in a female incarcerated population. *Journal of Family Issues, 23*, 441-457.

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.

Violence against Women and Girls Crime Report, 2015- 2016. Retrieved from:

https://www.cps.gov.uk/sites/default/files/documents/publications/cps_vawg_report_2016.pdf

Walby, S. (2009). *The Cost of Domestic Violence Update 2009*. Unpublished Manuscript, Lancaster University. Retrieved from www.lancs.ac.uk/fass/doc_library/sociology/Cost_of_domestic_violence_update.doc

Waltermaurer, E. (2005). Measuring intimate partner violence (IPV) you may only get what you ask for. *Journal of Interpersonal Violence, 20*(4), 501-506.

Widom, C.S. (1989). The cycle of violence, *Science, 244*, 160–166.

Yun, S. H. (2010). Factor structure and reliability of the Revised Conflict Tactics Scales' (CTS2) 10 factor model in a community-based female sample. *Journal of Interpersonal Violence, 26*(4), 719-744.

Zink, T., Klesges, L. M., Levin, L., & Putnam, F. (2007). Abuse Behaviour Inventory: Cutpoint, validity, and characterization of discrepancies. *Journal of Interpersonal Violence, 22*, 921-931.