Correspondence Between Self-Report Measures and Clinician Assessments of Psychopathology in Female Intimate Partner Violence Survivors

Meghan W. Cody,1 Judiann M. Jones,2 Matthew J. Woodward,2 Catherine A. Simmons,2 and J. Gayle Beck2

Abstract
Intimate partner violence (IPV) has potentially severe and long-lasting mental health consequences for survivors, including elevated symptoms and diagnoses of posttraumatic stress disorder (PTSD), depression, and generalized anxiety disorder (GAD). The current study examined the relationship between three self-report measures of psychological distress and ratings obtained from the corresponding clinician-administered measures in women seeking assessment for mental health problems following IPV (N = 185). PTSD symptoms were assessed using the self-report Impact of Event Scale–Revised (IES-R) and the interview-based Clinician-Administered PTSD Scale (CAPS). Depression symptoms were assessed using the self-report Beck Depression Inventory–II (BDI-II) and the depressive disorders sections from the clinician-administered Anxiety Disorders Interview Schedule–IV

1Mercer University School of Medicine, Macon, GA, USA
2University of Memphis, TN, USA

Corresponding Author:
Meghan W. Cody, Department of Psychiatry and Behavioral Sciences, Mercer University, 655 First Street, Macon, GA 31201, USA.
Email: cody_m@mercer.edu
Anxiety symptoms were assessed using the self-report Beck Anxiety Inventory (BAI) and the clinician-administered GAD section from the ADIS-IV. Results indicated that psychological distress was prevalent in the sample, with 27% receiving a PTSD diagnosis, 40% diagnosed with a depressive disorder, and 55% meeting criteria for GAD. Although each self-report measure was significantly and positively correlated with its corresponding clinician-administered measure, rates of diagnostic concordance were mixed. The BDI-II showed a high degree of agreement with the ADIS-IV depression section, but the IES-R and the CAPS were discordant at classifying PTSD. The BAI had acceptable sensitivity but poor specificity in relation to the ADIS-IV GAD section. These findings suggest that multiple assessment modalities should be considered when rating symptoms and estimating the prevalence of diagnoses among survivors of IPV.

Keywords
intimate partner violence, PTSD, depression, GAD, receiver operating characteristics

Intimate partner violence (IPV) consists of physical, sexual, and/or emotional abuse in an intimate relationship and can have a variety of negative physical, psychological, and social aftereffects (Tjaden & Thoennes, 2000). In particular, mental health consequences of IPV can be serious and long lasting for survivors (Plichta, 2004). One meta-analysis evaluating both interview and symptom checklist measures estimated that 64% of female IPV survivors have clinically significant symptoms of posttraumatic stress disorder (PTSD) and 48% have significant symptoms of major depressive disorder (MDD; Golding, 1999). Additional research has highlighted elevated rates of clinician-assessed generalized anxiety disorder (GAD) in this population, as well (J. G. Beck et al., 2014). However, previous work with IPV survivors has often neglected to account for variance in reported mental health symptoms due to differences in assessment modalities (Woodward et al., 2013). Lack of concordance among self-reports, collateral reports, and clinician assessments has been well-documented in other populations (e.g., Achenbach, Krukowski, Dumenci, & Ivanova, 2005; Bradley, Hilsenroth, Guarnaccia, & Westen, 2007; Chopra, Sullivan, Feldman, Landes, & Beck, 2008; Huprich & Bornstein, 2007) but has yet to be examined in IPV samples. The current project investigates the correspondence between self-reported posttraumatic stress, depression, and anxiety symptoms with clinician-rated PTSD, depressive disorders, and GAD, respectively, in a help-seeking sample of female IPV survivors.
Due to the chronic, repetitive nature of IPV, symptoms of psychological disorders may be especially likely to develop among survivors (Golding, 1999). In one community sample of 413 female IPV survivors, Mechanic, Weaver, and Resick (2008) found that the majority reported clinically significant symptoms of PTSD and depression on self-report questionnaires, with 39.6% reporting severe PTS symptoms, and 31.0% reporting severe depressive symptoms. Furthermore, IPV interacts with other psychosocial, economic, and environmental stressors such that it is associated with especially high levels of general psychological distress among certain groups, including African American women or those of lower socioeconomic status (Kaslow et al., 2010). In addition to self-reported psychopathology symptoms and general psychological distress, other studies using structured clinical interviews have identified high rates of mental disorders among IPV survivors (Housekamp & Foy, 1991). For example, a study using a diagnostic clinical interview with welfare recipients who experienced IPV identified 38.4% with PTSD, 44.6% with depression, and 13.4% with GAD (Tolman & Rosen, 2001). Few investigations, however, have examined both self-reported symptoms and clinician-assessed psychological disorders within the same IPV sample, despite evidence that outcomes may differ between the two assessment modalities (Woodward et al., 2013). Method variance, or variance attributed to differences in measurement rather than differences in the construct of interest, is the result of multiple biases related to rater effects, item characteristic effects, item context effects, and measurement context effects (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). The current study was conducted to examine the correspondence between two different assessment methods (self-report and clinician interview) of anxiety, depression, and PTSD symptoms. The current report focuses on GAD, IPV-related PTSD, and depressive disorders because previous work has found them to be the most common mental health diagnoses among female IPV survivors (J. G. Beck et al., 2014).

A large body of work examining psychopathology after IPV has focused on PTSD symptoms (e.g., J. G. Beck et al., 2011; Wilson et al., 2011). Early models of PTSD proposed a two-factor structure, with intrusion and avoidance symptoms representing the core of the disorder (Elklit & Shevlin, 2007). Based on this model, Horowitz and colleagues created the Impact of Event Scale (IES; Horowitz, Wilner, & Alvarez, 1979) to assess intrusion and avoidance following a traumatic event. A cluster of hyperarousal symptoms was added to the PTSD diagnosis in the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-IV-TR*; American Psychiatric Association [APA], 2000), and the avoidance cluster was split into effortful avoidance and numbing symptoms for the manual’s fifth edition (*DSM-5*;
Data collection for the current study was completed in 2013, so PTSD symptoms and diagnoses were assessed using the *DSM-IV-TR*’s three-factor model. Accordingly, symptoms in the domains of intrusion, avoidance, and hyperarousal were assessed using the revised IES, which had added a third subscale to its original two (IES-R; Weiss & Marmar, 1997). Creamer, Bell, and Failla (2003) compared mean item cut scores on the IES-R with the recommended total cut score of 50 on the PTSD Checklist (PCL; Weathers, Litz, Herman, Huska, & Keane, 1993) and found that the IES-R had excellent classification power in predicting scores greater than 50 on the PCL, with 91% sensitivity and 82% specificity. However, Shakespeare-Finch and Armstrong (2010) found that sexual assault survivors reported higher IES-R scores than motor vehicle accident and bereavement survivors, with 96.9% exceeding Creamer et al.’s cut score of 33 (mean item score of 1.5). They theorized that the intentional, interpersonal nature of sexual assault heightened PTSD symptoms, suggesting that IPV survivors such as those examined in the present study may also self-report greater symptoms than other trauma survivors. To date, no study has reported the relationship between IES-R scores and clinician-rated PTSD in the same IPV sample.

Similarly, little is known about how self-reported depression symptoms compare with clinician-assessed depressive disorders among IPV survivors. The Beck Depression Inventory–II (BDI-II; A. T. Beck, Steer, & Brown, 1996) is one of the most widely used measures of depressive symptomatology and has been described as “the gold standard of self-rating scales” (Cusin, Yang, Yeung, & Fava, 2010). The BDI-II includes 21 items, each rated on a scale of 0 to 3, that broadly sample depression symptoms as described in the *DSM-IV-TR*. The authors of the BDI-II recommend the following interpretation of scores: 0 to 13 for minimal depression, 14 to 19 for mild depression, 20 to 28 for moderate depression, and 29 to 63 for severe depression, with a cut score of 17 recommended for identifying symptoms likely to meet diagnostic criteria for MDD. Studies using the BDI-II to characterize depressive symptomatology in female IPV survivors have found mean scores ranging from approximately 15 (for volunteers selected from IPV-related police reports; DePrince, Labus, Belknap, Buckingham, & Gover, 2012) to 25 (for participants recruited from a domestic violence advocacy agency; Torres et al., 2013). These studies suggest that mean BDI-II scores for IPV survivors fall in the range of mild to moderate depression.

Finally, although high rates of GAD have been found in trauma-exposed samples (Grant, Beck, Marques, Palyo, & Clapp, 2008), few studies have examined the relationship between diagnosed GAD and self-reported anxiety symptoms in these or other clinical populations. The Beck Anxiety Inventory (BAI; A. T. Beck & Steer, 1993), a widely used measure designed to assess
somatic and cognitive symptoms of anxiety, was administered in the current study. Recommended cut scores are 0 to 7 for minimal anxiety, 8 to 15 for mild anxiety, 16 to 25 for moderate anxiety, and 26 to 63 for severe anxiety (A. T. Beck & Steer, 1993). A BAI score of 13 was able to classify survivors of motor vehicle accidents as having any mental health disorder with 81% sensitivity and 71% specificity 18 months post-accident (Silove et al., 2003). Research using receiver operating characteristics (ROC) analysis (Kabacoff, Segal, Hersen, & Van Hasselt, 1997) to detect anxiety disorders in older adults has shown the BAI to be an inadequate measure, perhaps due to the high rate of somatic complaints among this population. Other work using the BAI to identify specific anxiety disorders in individuals seeking treatment found it to have acceptable sensitivity and specificity at detecting GAD (Leyfer, Ruberg, & Woodruff-Borden, 2006). BAI norms for IPV-exposed samples are not available, and the research on its relationship with clinician diagnoses of GAD in general is inconsistent.

The current study uses data gathered in a psychodiagnostic assessment clinic to evaluate the relationship between self-reported symptoms on the IES-R, BDI-II, and BAI with clinician-assessed PTSD, depressive disorders, and GAD, respectively, in women exposed to IPV. We examine each of these measures in turn, calculating internal consistency using Cronbach’s alpha, as well as bivariate correlations and diagnostic concordance between self-reported and clinician-assessed ratings. Although universal standards of interpretation for Cronbach’s alpha do not exist (Lance, Butts, & Michels, 2006), scales intended for individual assessment and clinical decision making should have reliability coefficients of at least .80 (“moderately high” or “good” reliability) and preferably above .90 (“high” or “excellent” reliability; Sattler, 2008). Past research has found each measure used in the current study to have good psychometric properties in a range of samples; the current report will examine internal consistency of the measures when used in an IPV-exposed sample. It is hypothesized that Cronbach’s alpha will exceed .90 for each measure, indicating high internal consistency reliability. In addition, because little is known about the relationship of self-report scores to clinician assessments in this sample, we will examine this issue. To date, no ROC analyses have been reported for any of these measures in IPV survivors, despite the high prevalence of depression and anxiety in this population. In addition to the range of mental health disorders that IPV survivors are likely to experience, they commonly report a high level of general distress (Kaslow et al., 2010), so we hypothesize that the optimal cut score on each measure will be substantially higher in our sample compared with non-IPV samples. Such a result would indicate that self-report measures normed on non-IPV samples may provide an inflated rate of false positives when those cut scores
are used to screen IPV survivors. Self-report measures may be more sensitive to general psychological distress, compared with diagnoses made by clinicians, who are better able to identify and parse out disorder-specific symptoms (Woodward et al., 2013).

Method

Participants

Women who sought assessment and possible treatment for mental health conditions following IPV were screened at a university-based research clinic. Participants were recruited from flyers posted at local businesses and college campuses, presentations by study staff at health fairs, and through referrals from health care providers (e.g., gynecology clinics) and community agencies (e.g., the county family safety center). Individuals participated in exchange for a thorough psychological evaluation conducted by doctoral level psychologists or advanced psychology or social work graduate students, feedback, and treatment referrals as indicated.

Data were included in the current analyses if the reported IPV met DSM-IV-TR Criterion A for PTSD (APA, 2000). Out of 215 women who consented and completed the initial interview, 17 described abuse that did not satisfy Criterion A. Data from participants who demonstrated cognitive impairment (n = 5), psychotic symptoms (n = 6), or extremely inconsistent responding (n = 2) were also excluded because of concerns about validity. The full sample for the present study thus included 185 women. Due to attrition over multiple study visits for the full assessment, which could take up to 6 hr, the sample size for each analysis varied slightly.¹

Participants’ ages ranged from 18 to 75 years (M = 36.33, SD = 11.71). Thirty-one women (16.8%) reported still being involved in a romantic relationship with their most recent abusive partner (each denied immediate safety concerns, and the interviewer assisted with formulating a plan to prevent escalation of violence). Of those no longer involved with an abusive partner, the mean time since separation was 3.73 years (SD = 5.52). Participants reported an average of 3.77 additional non-IPV stressful life events (such as serious accidents or childhood abuse, SD = 2.30). See Table 1 for other demographic characteristics.

Measures

IPV characteristics. Description of the IPV was elicited by the Domestic Violence Interview (DVI), a semi-structured interview adapted from Blanchard
Table 1. Sample Characteristics.

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>93</td>
<td>50.3</td>
</tr>
<tr>
<td>African American</td>
<td>65</td>
<td>35.1</td>
</tr>
<tr>
<td>Hispanic</td>
<td>5</td>
<td>2.7</td>
</tr>
<tr>
<td>Asian</td>
<td>3</td>
<td>1.6</td>
</tr>
<tr>
<td>Other</td>
<td>15</td>
<td>8.1</td>
</tr>
<tr>
<td>Did not report</td>
<td>4</td>
<td>2.2</td>
</tr>
<tr>
<td><strong>Education level</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school or less</td>
<td>20</td>
<td>10.8</td>
</tr>
<tr>
<td>Some college</td>
<td>89</td>
<td>48.1</td>
</tr>
<tr>
<td>Undergraduate degree</td>
<td>39</td>
<td>21.1</td>
</tr>
<tr>
<td>Some graduate school</td>
<td>9</td>
<td>4.9</td>
</tr>
<tr>
<td>Graduate degree</td>
<td>26</td>
<td>14.1</td>
</tr>
<tr>
<td>Did not report</td>
<td>2</td>
<td>1.1</td>
</tr>
<tr>
<td><strong>Annual household income</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Below US$20,000</td>
<td>76</td>
<td>41.0</td>
</tr>
<tr>
<td>US$20,000-US$40,000</td>
<td>32</td>
<td>17.3</td>
</tr>
<tr>
<td>US$40,000-US$60,000</td>
<td>18</td>
<td>9.7</td>
</tr>
<tr>
<td>Above US$60,000</td>
<td>35</td>
<td>19.0</td>
</tr>
<tr>
<td>Did not report</td>
<td>24</td>
<td>13.0</td>
</tr>
<tr>
<td><strong>Employment status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-time</td>
<td>61</td>
<td>33.0</td>
</tr>
<tr>
<td>Part-time</td>
<td>47</td>
<td>25.4</td>
</tr>
<tr>
<td>Unemployed</td>
<td>52</td>
<td>28.1</td>
</tr>
<tr>
<td>Homemaker</td>
<td>8</td>
<td>4.3</td>
</tr>
<tr>
<td>Disability</td>
<td>10</td>
<td>5.4</td>
</tr>
<tr>
<td>Retired</td>
<td>3</td>
<td>1.6</td>
</tr>
<tr>
<td>Did not report</td>
<td>4</td>
<td>2.2</td>
</tr>
<tr>
<td><strong>Types of IPV experienced</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical only</td>
<td>2</td>
<td>1.1</td>
</tr>
<tr>
<td>Sexual only</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>Emotional only</td>
<td>11</td>
<td>5.9</td>
</tr>
<tr>
<td>Physical and sexual</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Physical and emotional</td>
<td>82</td>
<td>44.3</td>
</tr>
<tr>
<td>Sexual and emotional</td>
<td>7</td>
<td>3.8</td>
</tr>
<tr>
<td>Physical, sexual, and emotional</td>
<td>82</td>
<td>44.3</td>
</tr>
</tbody>
</table>

Note. IPV = intimate partner violence.
and Hickling (2004). The DVI was used to inquire about physical, sexual, and emotional abuse experienced in relationships with the most recent and the worst abusive intimate partner. Information about the nature of the abuse and the respondent’s emotional reactions were used to determine whether the IPV met Criterion A for PTSD. Table 1 shows the number of participants endorsing each form of abuse.

**Self-report measures.** PTSD, depression, and anxiety symptoms were measured using three self-report questionnaires. The IES-R (Weiss & Marmar, 1997) is a 22-item self-report questionnaire assessing distress over the past week from posttraumatic intrusion, avoidance, and hyperarousal symptoms. Items are rated on a scale of 0 (not at all) to 4 (extremely), and a mean item score is obtained, with higher scores indicating more severe symptoms. In the current study, the instructions were modified for respondents to answer “with respect to your abusive romantic relationships.” Past research showed the IES-R to have good internal consistency (alpha of .95) in a sample of motor vehicle accident survivors (J. G. Beck et al., 2008).

The BDI-II (A. T. Beck et al., 1996) is a 21-item inventory of depressive symptoms in the past 2 weeks. Items are scored on a 4-point scale, with higher ratings indicating more severe symptoms, and total scores ranging from 0 to 63. The BDI-II has excellent psychometric properties across a variety of clinical and non-clinical samples, with alphas ranging from .86 to .92 (Dozois, Dobson, & Ahnberg, 1998; Segal, Coolidge, Cahill, & O’Riley, 2008).

Self-reported anxiety symptoms were measured with the BAI (A. T. Beck & Steer, 1993), a 21-item scale that assesses cognitive, affective, and somatic symptoms of anxiety over the past week. Items are rated on a scale of 0 (not at all) to 3 (severely), for a total possible score of 63. The BAI has been shown to accurately distinguish between anxiety and depression when paired with the BDI (Hewitt & Norton, 1993).

**Clinician assessments.** Participants were administered the Anxiety Disorders Interview Schedule–IV (ADIS-IV; DiNardo, Brown, & Barlow, 1994) to assess for current anxiety, mood, and related disorders according to DSM-IV criteria. Based on the respondent’s answers, the interviewer assigned a clinical severity rating (CSR) for each disorder, rating functional interference and subjective distress created by that condition. The CSR ranges from 0 (none) to 8 (very severe); a CSR of 4 (definite interference or definite distress) or higher indicates that the symptoms meet the diagnostic threshold for a disorder. The ADIS-IV has adequate to high reliability for both dimensional CSRs (r = .72 for GAD and .69 for depressive disorders) and categorical diagnoses.
(κ = .67 for GAD and .72 for depressive disorders) over two independent administrations (Brown, DiNardo, Lehman, & Campbell, 2001). ADIS-IV interviews in the present study were videotaped, and a random selection of 23.2% (n = 43) were rated for reliability. Inter-rater diagnostic agreement was substantial for depression (κ = .77 for both MDD and dysthymic disorder) and GAD (κ = .86). Following Brown and colleagues (2001), MDD and dysthymia were collapsed into one category of depressive disorders (κ = .81).

The Clinician-Administered PTSD Scale (CAPS; Blake et al., 1995) is considered the current “gold standard” in assessment of PTSD due to its excellent reliability (inter-rater correlation coefficients greater than .90; κ approaching 1.00; internal consistency alphas .80-.90) and validity (Weathers, Keane, & Davidson, 2001). The CAPS is a semi-structured interview that assesses the frequency and intensity of PTSD symptoms over the past month on a scale of 0 (does not occur/is not distressing) to 4 (occurs nearly every day/is extremely distressing). A symptom counts toward a DSM-IV-TR diagnosis of PTSD if it has a rating of at least 1 (occurs once or twice/10% of the time) on the frequency scale and at least 2 (is moderately distressing) on the intensity scale (Weathers, Ruscio, & Keane, 1999). In the current study, the CAPS was administered by clinical psychologists or advanced graduate students in reference to the participant’s IPV. All assessors were thoroughly trained and supervised on the administration of the diagnostic interviews (see protocol in DiNardo, Moras, Barlow, Rapee, & Brown, 1993). Interviews were videotaped; of participants whose data are included in the current analysis, 44 (23.8%) were randomly selected for reliability checks. Inter-rater agreement on PTSD diagnosis was nearly perfect, with discrepant ratings for only one participant (κ = .94).

Procedure

Following informed consent, participants were interviewed with the DVI, the CAPS (anchored to their IPV), and the ADIS-IV. Participants were also asked to complete several questionnaires, including the IES-R, BDI-II, and BAI. Following the assessment, participants received feedback on their evaluation and referrals for additional services. All procedures were implemented in compliance with the regulations of the university’s Institutional Review Board.

Data Analyses

Measurement properties of the self-report questionnaires were assessed through calculation of Cronbach’s alpha as a measure of internal consistency.
(Cronbach’s alpha was also obtained for the frequency and intensity symptom ratings on the CAPS but could not be calculated for the single-item clinician ratings of depressive disorders and GAD on the ADIS-IV). Next, bivariate correlations were computed between each self-report measure and the clinician ratings obtained from the respective clinical interview. Finally, the three self-report symptom measures (the IES-R, BDI-II, and BAI) were compared with the clinician diagnoses made using structured interviews (the CAPS and ADIS-IV) by calculating ROC curves. An ROC curve graphs scores on a measure according to their true positive rate (Sensitivity) by their false positive rate \((1 – \text{Specificity})\), revealing an area under the curve (AUC) that shows the total rate of concordant classification. The AUC statistic represents the probability that a randomly selected person clinically diagnosed with a problem will have a higher score on a self-report measure than a randomly selected person not clinically diagnosed (Fawcett, 2006).

Next, cut scores were identified for each measure by choosing the point on the ROC graph that maximizes Youden’s index, the vertical distance between the point and the line of chance, calculated as \(J = \max(\text{Sensitivity} + \text{Specificity} – 1)\) and ranging from 0 to 1 (Kelly, Dunstan, Lloyd, & Fone, 2008; Youden, 1950). In the current study, each measure’s cut score was chosen by identifying the score with the largest \(J\) value, while also keeping sensitivity and specificity both above a minimum of 50%. Following Kessel and Zimmerman’s (1993) recommendations, \(2 \times 2\) tables were computed comparing the number of individuals classified as a case or non-case by the self-report test’s cut score and the clinical interview. Cell sizes were used to calculate positive and negative predictive values (PPV and NPV, the proportion of true positives and true negatives, respectively) and Cohen’s kappa to compare diagnostic agreement (Kessel & Zimmerman, 1993). Interpretation of kappa values follows recommendations from Landis and Koch (1977): 0-.20 poor to slight, .21-.40 fair, .41-.60 moderate, .61-.80 substantial, and .81-1 almost perfect.

Results

Using the CAPS for diagnosis, 50 participants out of the full sample of 185 (27%) met criteria for current IPV-related PTSD, according to the DSM-IV-TR. Of the 153 participants who also completed the ADIS-IV, 61 (40%) were diagnosed with at least one DSM-IV-TR depressive disorder. For GAD, 84 (55%) of the participants who completed the ADIS-IV met DSM-IV-TR diagnostic criteria. The relationship between the self-report measures and the diagnoses made via clinical interviews are presented in the figures as flowcharts of diagnostic concordance, as recommended by the Standards for Reporting of Diagnostic Accuracy (STARD) Initiative (Bossuyt et al., 2004).
Assessment of PTSD Symptoms

Both self-report and clinician-administered measures of PTSD symptoms showed evidence of high internal consistency, with Cronbach’s alpha of .93 for the IES-R and .92 for the CAPS. On the IES-R, mean item scores ranged from 0 to 3.86 ($M = 2.26, SD = 0.87$). CAPS total scores ranged from 0 to 87 ($M = 31.10, SD = 21.32$). IES-R and CAPS scores were significantly positively correlated, Pearson’s $r = .33, p < .01$.

Twenty-four participants who completed the CAPS did not complete the IES-R, leaving a sample of $n = 161$ for the ROC analysis. The ROC curve for predicting PTSD diagnosis with the IES-R revealed an AUC of .59 ($SE = .05, p = .091, 95\%$ confidence interval $[CI] = [.50, .68]$), indicating performance no better than chance. The mean item score of 2.34 had the largest Youden’s index ($J = .12$), with 60% sensitivity and 52% specificity. Figure 1 depicts a diagnostic flowchart, with the $2 \times 2$ table comparing results using an IES-R score of 2.34 with the CAPS diagnoses forming the bottom four squares. These cells revealed a PPV of .30, an NPV of .78, and overall kappa of .09 ($p = .195$).

Assessment of Depressive Symptoms

Twenty-one participants in the original sample did not complete the BDI-II. Of those who did, 143 also completed the ADIS-IV. Inter-item reliability was excellent for the BDI-II, $\alpha = .92$, and scores ranged from 0 to 58 ($M = 27.20, SD = 12.84$). CSRs for the depressive disorders on the ADIS-IV ranged from 0 to 7 for MDD ($M = 2.20, SD = 2.36$) and 0 to 6 for dysthymia ($M = 0.50, SD = 1.50$). BDI-II scores were highly correlated with CSRs for depressive disorders ($r = .63, p < .01$).

The BDI-II performed significantly better than the chance at classifying individuals with depressive disorders relative to the ADIS-IV (AUC = .81, $SE = .04, p < .001, 95\% CI = [.74, .88]$). A total score of 29.5 was the best point for classification ($J = .52$, with 75% sensitivity and 77% specificity), leading to an optimal cut score of 30. Figure 2 presents the flowchart with the $2 \times 2$ results comparing the BDI-II with the ADIS-IV depressive disorders diagnosis. Overall diagnostic concordance between the self-report and clinician-administered measures of depression was moderate (PPV = .69, NPV = .83, $\kappa = .53, p < .001$).

Assessment of Anxiety Symptoms

Fifty participants in the original sample did not complete the BAI, and of the remaining 135, all but 3 completed the ADIS-IV ($n = 132$ in this analysis).
Cronbach’s alpha for the BAI was .93, and total scores ranged from 0 to 55 ($M = 21.55$, $SD = 13.44$). CSRs for GAD on the ADIS-IV ranged from 0 to 7.
BAI scores and GAD CSRs were significantly correlated, $r = .28, p < .01$.

The ROC curve for predicting GAD diagnosis on the ADIS-IV with the BAI revealed better than chance performance ($AUC = .62, SE = .05, p = .017, 95\% \text{ CI} = [.52, .72]$). The best cut point, a total score of 14.5, had 79% sensitivity and 54% specificity ($J = .33$). The 2 × 2 classification table for a score of 15 is shown in Figure 3. These cells revealed fair diagnostic concordance between the self-report measure of anxiety and the clinician-made GAD diagnosis ($\text{PPV} = .69, \text{NPV} = .66, \kappa = .34, p < .001$).
Discussion

This study examined the correspondence between self-report measures and clinician assessments of three common mental health problems in a sample of help-seeking female IPV survivors. Self-reported PTSD symptoms on the IES-R were compared with clinician ratings of PTSD on the CAPS, self-reported depression symptoms on the BDI-II were compared with clinician ratings of depressive disorders on the ADIS-IV, and self-reported anxiety symptoms on the BAI were compared with clinician ratings of GAD on the ADIS-IV. Results indicated that all of the self-report measures
and clinician-administered assessments had good psychometric properties. Self-reports of PTSD and anxiety symptoms were moderately correlated with clinician ratings of PTSD and GAD, respectively. Self-reported depression was highly correlated with clinician-made ratings of depressive disorders. ROC analyses showed that the BDI-II classified participants with depressive disorders at a high degree of agreement with the ADIS-IV MDD and dysthymia sections. However, IES-R scores and the CAPS were discordant at classifying PTSD in this sample, with sensitivity and specificity no better than chance. The BAI performed in between these two extremes, in relation to the ADIS-IV GAD section, with acceptable sensitivity but poor specificity. Overall, these findings suggest that both self-report questionnaires and clinical interviews may be used to detect psychopathology in IPV survivors. However, although findings from these different assessment modalities were positively correlated, classification results were discordant for many participants, particularly when assessing PTSD. Clinicians may wish to use self-report measures for initial screening of IPV survivors but will need to confirm diagnoses of mental disorders through clinical interviews.

Although the IES-R is often used as a diagnostic indicator of PTSD in studies of reactions to trauma, it was not designed as such (Weiss, 2004). Although past work outside of the IPV literature has shown it to have adequate correspondence with clinician-administered PTSD assessments (J. G. Beck et al., 2008), the results of the current study suggest that the IES-R shows weaker relationships with clinical interviews when used to screen IPV survivors. Even the most accurate IES-R cut score (item mean of 2.34, total of 51.48) had only 60% sensitivity and 52% specificity at predicting the CAPS diagnosis of PTSD. Note that this score is much higher than the total cut score of 33 recommended by Creamer and colleagues (2003), but is comparable with the mean score of 66.57 reported by 14 IPV survivors in another study (Lindgren & Renck, 2008). The IES-R’s low positive predictive value in the current sample (.30) indicates that 70% of participants identified as having probable PTSD were actually false positives, an unacceptable value for screening purposes. One possible explanation for the lack of concordance between the IES-R and the CAPS is that the CAPS was designed to assess the four-factor model of PTSD reflected by the DSM-IV-TR; whereas the IES-R assesses only three factors (Elklit & Shevlin, 2007). The current findings need replication in other trauma samples using diagnostic interviews but suggest that the IES-R and the CAPS measure different facets of psychological reactions to trauma. Using the IES-R to assess severity of symptom domains in addition to conducting a diagnostic interview to identify PTSD cases may
allow clinicians to better individualize treatment. However, this measure should not be used for screening purposes.

Symptoms self-reported on the BDI-II, in contrast, largely matched clinician ratings of depressive disorders, with 81% accurate discrimination relative to the ADIS-IV using a cut score of 30. This score is in the range of severe depression, according to guidelines published by A. T. Beck et al. (1996), and is elevated with respect to the recommended cut score of 17. Thus, on both the self-report measure and clinician-administered assessment, depression symptoms among IPV survivors were found to be significant (mean BDI-II score of 27) and prevalent (40% of the sample met diagnostic criteria for depressive disorders on the ADIS-IV). A large amount of research has suggested that assessment of depression in community health care settings (e.g., Eack, Singer, & Greeno, 2008), and especially among survivors of interpersonal violence (e.g., Houry, Kaslow, & Thompson, 2005), is well worth the effort. Routine depression screening of IPV survivors is therefore recommended, with the BDI-II showing comparable results to a more time-intensive clinical interview. When necessary to allocate scarce clinician resources to individuals most likely to have clinically significant depression, a cut score of 30 on the BDI-II will maximize sensitivity while minimizing false positives.

In the assessment of anxiety symptoms, a cut score of 15 on the BAI correctly classified the majority of the sample relative to the ADIS-IV diagnosis of GAD. Sensitivity was adequate, at 79%, but specificity was only 54%, indicating that many individuals scoring above the cut point on the BAI did not meet criteria for GAD. The moderate level of discordance between the BAI and the ADIS-IV GAD section may be a reflection of the fact that the BAI was not designed to indicate GAD specifically, but rather to reflect anxiety symptoms independent of depression (A. T. Beck & Steer, 1993). In a highly distressed, diagnostically complicated sample such as the current one of IPV survivors, the BAI alone seems to be sensitive to elevated distress but may not be specific to GAD. Other measures, such as the GAD Questionnaire–IV (GAD-Q-IV; Newman et al., 2002), may be more appropriate when the goal is detection of GAD in particular. The GAD-Q-IV, a nine-item self-report measure designed to assess the diagnostic criteria for GAD according to the DSM-IV-TR, has excellent sensitivity and specificity relative to the ADIS-IV (Newman et al., 2002). However, the current results support use of the BAI as a measure of clinically significant anxiety that may be used to identify individuals with any anxiety disorder (not limited to GAD), given its evidence of good reliability and validity in this IPV sample.

A limitation of this study is that the self-report measures examined were selected due to their frequency of use in past research and for the breadth of
symptoms covered in each domain, not for their utility as diagnostic screening measures. Nevertheless, these are commonly used assessments, and the current findings contribute valuable information about how they operate in an understudied IPV sample relative to clinical interview measures. Although briefer self-report questionnaires are available that follow the diagnostic criteria more closely for each disorder, an understanding of each scale’s psychometric performance in an IPV sample is still necessary for accurate clinical interpretation and comparison with research findings. Other limitations of the current study include concerns about the generalizability of the sample, as random selection of IPV survivors is impossible. In addition, participants in this study were self-selected as seeking help for mental health conditions following IPV and may not be representative of the majority of this population. In addition to using different recruitment strategies, such as surveying individuals who have filed IPV-related police reports (i.e., those who are not seeking mental health intervention), future research would benefit from including alternative measures of psychopathology, such as measures specifically designed to be used as screening instruments (e.g., the PCL for PTSD). Larger sample sizes would also allow for examination of other mental health disorders that are less common in this population.

In conclusion, psychometrically validated self-report and clinician-administered measures both have the potential to identify clinically significant distress among IPV survivors, so an understanding of the correspondence between the two modalities is critical. In particular, structured clinical interviews are more time- and resource-intensive assessment procedures than self-report measures. Cut scores on these self-report measures that could indicate mental health diagnoses may be useful tools in assisting health care providers with task shifting, the redistribution of service delivery to individuals with varying degrees of training (Kazdin & Rabbitt, 2013). As such, self-report measures are recommended for the detection of IPV survivors with significant symptoms of psychopathology, but clinician-administered interviews are necessary to confirm specific diagnoses. In the current study, both the self-report and clinician-administered measures of PTSD, depression, and anxiety symptoms were found to be reliable measures among IPV survivors. For each type of symptom, self-report and clinician-administered measures were significantly and positively correlated with each other. However, findings for diagnostic concordance between the self-report measures and the clinical interviews were mixed. The BDI-II was found to closely correspond to diagnoses of depressive disorders made using the ADIS-IV. The BAI and the IES-R showed less concordance with diagnoses made using the ADIS-IV GAD section and the CAPS, respectively, suggesting that prevalence estimates of psychological disorders made using self-report measures may not be
directly comparable with structured clinical interviews, the gold standard. In sum, the current study examines the psychometric properties of self-report and clinician-administered measures of psychopathology among IPV survivors, an understudied population that exhibits high levels of general distress. Psychological assessment of these individuals will provide the most accurate picture of their social and emotional functioning when conducted in multiple modalities.

Acknowledgments
The authors would like to thank the Trauma Research and Recovery Lab, especially Riley Nicholson, Michelle Presley, Michayela Rosario, and Elizabeth Vega, for their valuable research assistance. The authors would also like to thank the study participants and staff, especially Náthali Blackwell and Leslie Lindsey, involved in the Athena Project at the University of Memphis.

Authors’ Note
This research was conducted at the University of Memphis, Department of Psychology (M. W. Cody, J. M. Jones, M. J. Woodward, and J. G. Beck) and Department of Social Work (C. A. Simmons). M. W. Cody is now in the Department of Psychiatry and Behavioral Sciences at Mercer University School of Medicine; J. M. Jones is now at the University of Alabama at Birmingham School of Medicine.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Support for this project was provided in part by the Lillian and Morrie Moss Chair of Excellence position held by Dr. Gayle Beck.

Note
1. Completers versus non-completers were not significantly different from each other on any demographic characteristics ($p > .05$ for all comparisons).

References


**Author Biographies**

**Meghan W. Cody**, PhD, is an assistant professor in Clinical Medical Psychology at Mercer University Health Sciences Center in Macon and Atlanta, GA. The work presented in this manuscript was completed while she was a postdoctoral fellow at the University of Memphis. Her research interests include cognitive-behavioral processes in trauma, anxiety, and medical coping.

**Judiann M. Jones**, MS, is completing her clinical internship at the University of Alabama-Birmingham. She will graduate this year with a PhD from the University of Memphis and begin fellowship at the Memphis VA Medical Center. Her research interests include interpersonal trauma and the psychological after-effects of experiencing abuse.

**Matthew J. Woodward**, MS, is a doctoral student at the University of Memphis. His research interests include the role of interpersonal processes in adjustment to trauma.

**Catherine A. Simmons**, PhD, LCSW, is an Associate Professor in the Department of Social work at the University of Memphis. Prior to joining academia, Dr. Simmons spent 10 years as a military social work officer leading family violence, substance abuse, and mental health programs. The author/editor of two books and over 40 professional papers, her research interests include strengths based social work practice, family violence, and issues pertaining to gender, trauma, and violence.

**J. Gayle Beck**, PhD, is the holder of the Lillian and Morrie Moss Chair of Excellence at the University of Memphis. Her research interests include understanding the broad spectrum of psychopathology following trauma exposure, particularly interpersonal trauma.