Measuring Acceptability of Intimate Partner Violence Against Women: Development and Validation of the A-IPVAW Scale

Manuel Martín-Fernández, Enrique Gracia, Miriam Marco, Viviana Vargas, Faraj A. Santirso, and Marisol Lila

University of Valencia, Spain

A B S T R A C T

Intimate partner violence against women (IPVAW) is a major social and public health problem of global proportions. Public attitudes toward IPVAW shape the social environment in which such violence takes place, and attitudes of acceptability of IPVAW are considered a risk factor to actual IPVAW. The aim of this study was to develop and validate a scale measuring acceptability of IPVAW (A-IPVAW). To this end, a sample of 1,800 respondents was recruited via social media. A second sample of 50 IPVAW offenders was used for concurrent validity analyses. Following a cross-validation approach and using item response theory analyses, we found that the latent structure of the scale was one-dimensional and very informative for high and very high levels of acceptability of IPVAW. Regarding criterion-related validity, we found that (a) our measure was related to perceived severity of IPVAW and ambivalent sexism, (b) men showed higher levels of acceptability than women, and (c) IPVAW offenders reported higher levels of acceptability than men from the general population. Taken together, our results provide evidence that the A-IPVAW is a reliable and valid instrument to assess acceptability of IPVAW.

RESUMEN

La medida de la aceptabilidad de la violencia de pareja contra la mujer: desarrollo y validación de la escala A-IPVAW

La violencia de pareja contra la mujer (VPM) es un grave problema social y de salud pública con repercusión mundial. Las actitudes públicas hacia la VPM contribuyen al contexto social en el que esta violencia tiene lugar, y las actitudes de aceptabilidad de la VPM son un factor de riesgo en la comisión de este tipo de violencia. El objetivo de este estudio fue el desarrollo y validación de una escala de medida de la aceptabilidad de la VPM (A-IPVAW, Acceptability of Intimate Partner Violence against Women scale). Con este fin, se obtuvo una muestra de 1800 participantes reclutados a través de las redes sociales. Se utilizó una segunda muestra de 50 hombres agresores de VPM para los análisis de validez concurrente. Mediante una validación cruzada (cross-validation) y utilizando análisis basados en la Teoría de Respuesta al Ítem, se estableció que la estructura latente de la escala era unidimensional y muy informativa para los niveles altos y muy altos de aceptabilidad de la VPM. Con respecto a la validez de criterio se encontró que (a) esta medida se relacionaba con la gravedad percibida de la VPM y el sexismo ambivalente, (b) los hombres mostraban niveles más altos de aceptabilidad que las mujeres y (c) los hombres agresores de VPM mostraban niveles más altos de aceptabilidad que los hombres de la muestra general. En conjunto, estos resultados muestran que la A-IPVAW es un instrumento fiable y válido para evaluar la aceptabilidad de la VPM.
attitudes and responses regarding violence against women reflect these norms and are key to better understanding its root causes and, therefore, developing more effective intervention measures (Gracia & Lila, 2015; Gracia & Merlo, 2016). Public attitudes toward IPVAV can shape the social environment in which such violence takes place, contributing to either foster or discourage IPVAV in societies (Brownlow, 2002; Campbell & Manganello, 2006; Frye, 2007; Gracia, 2014a). These attitudes have been related to incidence and reporting rates, public and professional responses, and the victims’ own responses (Carlson & Worden, 2005; Frye, 2007; Gracia, Garcia, & Lila, 2008, 2011, 2014; Gracia & Herrera, 2006b).

Attitudes of acceptability of IPVAV have been linked to the perpetration of this type of violence (Copp, Giordano, Longmore, & Manning, 2016; Gracia, Rodriguez, & Lila, 2015; Sugarman & Frankel, 1996). High levels of acceptability of IPVAV can lead to the perception of this type of behavior as normative, increasing the risk of men perpetrating IPVAV and of this type of violence being justified by victims and their social circles (Waltermauer, 2012). In this regard, attitudes of acceptability of IPVAV have been considered as a risk factor of actual IPVAV (Archer & Graham-Kevan, 2003; Capaldi, Knoble, Shortt, & Kim, 2012; Flood & Please, 2009; Lila, Gracia, & Murgui, 2013; Stith, Smith, Penn, Ward, & Tritt, 2004; WHO, 2002). Attitudes toward IPVAV are hence one of the main targets for intervention and prevention strategies (Garcia-Moreno et al., 2015; Gracia & Lila, 2015; Jewkes, Flood, & Lang, 2015; Lila, Gracia, & Herrera, 2012).

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

The Present Study

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

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

Method

Sample

Online sampling recruitment was used for the current study. Data were collected through social media and e-mail snowballing. These methods have proven to be effective and cost-efficient in previous studies (for systematic reviews see Thornton et al., 2016; Topolovec-Vranic & Natarajan, 2016). A total pool of 2,698 responses was collected, most of which were from women (67.6% of the respondents). Although previous studies have found a similar proportion of participation in social media by gender (Thornton et al., 2016), we opted to use a representative sample balanced by gender. To do so, we selected a random sample that maintained a similar proportion of male and female participants.

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

Table 1. Socio-demographic Variables of the General Sample (N = 1800)

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

A second sample of 50 male offenders court-ordered to an intervention program for IPVAV was selected for validity purposes. These offenders had a suspended sentence conditioned to their attending an intervention program. This sample had a mean age of...
Instruments

Acceptability of IPVAW scale (A-IPVAW). A pool of 102 items tapping acceptability of IPVAW was drawn from a recent review of European surveys on violence against women (Gracia & Lila, 2015). The items drawn from the pool were translated to English from their original language by European experts on the field of IPVAW who provided the survey data for the review. For this study, a panel of six experts on IPVAW rated the relevance of each item on a 5-point Likert-type scale (i.e., “Is this item relevant to measure attitudes of acceptability of IPVAW?”; 1 = strongly disagree, 5 = strongly agree). Those items rated with a 4 (i.e., the agree category) or more by the six experts were selected to compose a twenty-item measure of acceptability of IPVAW. The items were translated to Spanish by the same authors of the review of the European surveys. Respondents had to rate how acceptable were a range of men’s behaviors against their female partners on a 3-point Likert-type scale (0 = not acceptable, 1 = somewhat acceptable, 2 = acceptable). A fourth category, very acceptable, was merged with the acceptable category since it was selected by almost no respondents in both the general and offender samples. In particular, the frequencies of the upper category in both samples were quite similar. The items reflected physical violence (e.g., it is acceptable for a man “to hit his partner if she has been unfaithful”), coercion or verbal violence (e.g., it is acceptable for a man “to threaten his partner with hurting her or others if she leaves him”), and emotional violence such as controlling behaviors (e.g., it is acceptable for a man “to set limits on how his others if she leaves him”), and emotional violence such as controlling behaviors (e.g., it is acceptable for a man “to set limits on how his

Procedure

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

Data Analyses

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

To decide the number of factors needed for extraction, a parallel analysis based on minimum rank factor analysis with polychoric correlations (Timmerman & Lorenzo-Seva, 2011) and the minimum average partial (MAP) criterion (Velicer, 1976) was performed on the first subsample. These methods have shown a good performance determining the number of latent dimensions with categorical data (Garrido, Abad, & Ponsoda, 2011, 2016). Bartlett’s sphericity test and the Kaiser-Meyer-Olkin (KMO) statistic were computed to evaluate the suitability of the data for an EFA. Given the categorical nature of the data, the estimation method used was weighted least squares (WLSMV), since it is more robust for ordinal and categorical data (Asparouhov & Muthén, 2010). The polychoric correlation matrix was used to conduct the EFA (Muthén & Kaplan, 1985, 1992). Model fit was assessed in terms of relative comparative fit, with the CFI and TLI indices. Model residuals were also evaluated with the SRMS and the RMSEA statistics. CFI and TLI values ≥ .95 are indicative of good model fit (Hu & Bentler, 1999), whereas SRMR values ≤ .08, and RMSEA values ≤ .06 are considered good fitting models (McCallum, Browne, & Sugawara, 1996). Estimated chi-square values and their df were also provided, although this statistic tends to be affected by large sample sizes.

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

Once the scale dimensionality was assessed, an item response theory (IRT) model was fitted to the whole sample. IRT models have some major advantages over CFA: (a) IRT includes improved factor scores estimates, and (b) does not assume measurement precision to be constant, allowing researchers to identify which factor scores of the latent construct are better assessed by their psychological instruments (Chen, 2007). Samejima’s graded response model (1969) was selected given the ordinal nature of the data. Then the model was estimated using the MHRM algorithm (Cai, 2010), and model fit was evaluated with the same cut-off values for the same fit indices (CFI & TLI ≥ .95, RMSEA ≤ .06). To compute these indices, the Maydeu’s M2 statistic for ordinal variables was used instead of other approximations of the chi-square for ordinal variables, since it was specifically developed to assess the overall fit for IRT models (Maydeu-Olivares & García-Forero, 2010; Maydeu-Olivares & Joe, 2006). The test information function was provided and the IRT scores—the person parameters of the model—were used as an acceptability estimate for the subsequent validity analyses.

Finally, the A-IPVAW scale validity was evaluated for the whole sample. First, acceptability estimates (i.e., IRT factor scores) were correlated with the PS-IPVAW scores and with the hostile and benevolent sexism scores from the ASI. Then comparisons were made of the A-IPVAW scale scores between men and women and age groups from the general sample, and between men from the general sample and the sample of male batters.
Descriptive statistics, classical internal consistency, and IRT analyses were conducted with the statistical package R (R Core Team, 2016). The psych (Revelle, 2016) and the mirt (Chalmers, 2012) libraries were used for this purpose. Bartlett’s sphericity test, the KMO statistic, parallel analysis, and the MAP criterion were computed with the factor package (Lorenzo-Seva & Ferrando, 2006), whereas the EFA and CFA analysis were conducted with Mplus 7.1 (Muthén & Muthén, 2010).

Results

Descriptive Statistics and Internal Consistency

The item descriptive analyses and item-total correlations are shown in Table 2. All items presented mean values close to 0, with standard deviations around 0.30. In addition, the skew and kurtosis indices showed that the item distributions were strongly displaced to the left, displaying a leptokurtic distribution. Taken together, the descriptive analyses reveal that respondents tended to choose the not acceptable category for almost all items. Regarding the item-total corrected correlations, all items presented values above .40, indicating that the items were strongly related to the measured construct. The internal consistency of the scale was good (Cronbach’s α = .89).

Confirmaotry Factor Analysis

A CFA was conducted with the second subsample. A one-factor model was posited and estimated with the WLSMV method. As displayed in Figure 1, all standardized loadings were greater than .70 with standard estimation errors around .03. Again, the chi-square test was significant, χ²(170) = 502.94, p < .001, and the comparative fit indices of the model were above the cut-off values (CFI = .96, TLI = .96), and the model residuals were fair (SRMR = .080, RMSEA = .047). These fit indices indicated that the model fitted well the data.

Table 2. A-IPVAW Item Descriptive Statistics

<table>
<thead>
<tr>
<th>Item</th>
<th>M</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>Skew</th>
<th>Kurtosis</th>
<th>r_item-total</th>
</tr>
</thead>
<tbody>
<tr>
<td>a-ipvaw1</td>
<td>.42</td>
<td>.61</td>
<td>0.00</td>
<td>2.00</td>
<td>1.15</td>
<td>.26</td>
<td>.54</td>
</tr>
<tr>
<td>a-ipvaw2</td>
<td>.41</td>
<td>.61</td>
<td>0.00</td>
<td>2.00</td>
<td>1.17</td>
<td>.32</td>
<td>.53</td>
</tr>
<tr>
<td>a-ipvaw3</td>
<td>.10</td>
<td>.33</td>
<td>0.00</td>
<td>2.00</td>
<td>3.63</td>
<td>.15</td>
<td>.63</td>
</tr>
<tr>
<td>a-ipvaw4</td>
<td>.14</td>
<td>.39</td>
<td>0.00</td>
<td>2.00</td>
<td>2.80</td>
<td>.75</td>
<td>.59</td>
</tr>
<tr>
<td>a-ipvaw5</td>
<td>.10</td>
<td>.33</td>
<td>0.00</td>
<td>2.00</td>
<td>3.48</td>
<td>.12</td>
<td>.51</td>
</tr>
<tr>
<td>a-ipvaw6</td>
<td>.10</td>
<td>.34</td>
<td>0.00</td>
<td>2.00</td>
<td>3.52</td>
<td>.12</td>
<td>.62</td>
</tr>
<tr>
<td>a-ipvaw7</td>
<td>.09</td>
<td>.32</td>
<td>0.00</td>
<td>2.00</td>
<td>4.01</td>
<td>.16</td>
<td>.59</td>
</tr>
<tr>
<td>a-ipvaw8</td>
<td>.04</td>
<td>.22</td>
<td>0.00</td>
<td>2.00</td>
<td>6.63</td>
<td>.47</td>
<td>.56</td>
</tr>
<tr>
<td>a-ipvaw9</td>
<td>.02</td>
<td>.18</td>
<td>0.00</td>
<td>2.00</td>
<td>5.46</td>
<td>.34</td>
<td>.51</td>
</tr>
<tr>
<td>a-ipvaw10</td>
<td>.02</td>
<td>.18</td>
<td>0.00</td>
<td>2.00</td>
<td>8.12</td>
<td>.73</td>
<td>.56</td>
</tr>
<tr>
<td>a-ipvaw11</td>
<td>.04</td>
<td>.21</td>
<td>0.00</td>
<td>2.00</td>
<td>6.48</td>
<td>.45</td>
<td>.50</td>
</tr>
<tr>
<td>a-ipvaw12</td>
<td>.03</td>
<td>.18</td>
<td>0.00</td>
<td>2.00</td>
<td>7.47</td>
<td>.61</td>
<td>.56</td>
</tr>
<tr>
<td>a-ipvaw13</td>
<td>.03</td>
<td>.20</td>
<td>0.00</td>
<td>2.00</td>
<td>7.97</td>
<td>.67</td>
<td>.63</td>
</tr>
<tr>
<td>a-ipvaw14</td>
<td>.02</td>
<td>.18</td>
<td>0.00</td>
<td>2.00</td>
<td>8.34</td>
<td>.75</td>
<td>.55</td>
</tr>
<tr>
<td>a-ipvaw15</td>
<td>.09</td>
<td>.34</td>
<td>0.00</td>
<td>2.00</td>
<td>3.77</td>
<td>.14</td>
<td>.50</td>
</tr>
<tr>
<td>a-ipvaw16</td>
<td>.15</td>
<td>.41</td>
<td>0.00</td>
<td>2.00</td>
<td>2.74</td>
<td>.71</td>
<td>.53</td>
</tr>
<tr>
<td>a-ipvaw17</td>
<td>.12</td>
<td>.37</td>
<td>0.00</td>
<td>2.00</td>
<td>3.27</td>
<td>.10</td>
<td>.48</td>
</tr>
<tr>
<td>a-ipvaw18</td>
<td>.09</td>
<td>.33</td>
<td>0.00</td>
<td>2.00</td>
<td>3.86</td>
<td>.15</td>
<td>.54</td>
</tr>
<tr>
<td>a-ipvaw19</td>
<td>.02</td>
<td>.17</td>
<td>0.00</td>
<td>2.00</td>
<td>8.59</td>
<td>.80</td>
<td>.57</td>
</tr>
<tr>
<td>a-ipvaw20</td>
<td>.26</td>
<td>.49</td>
<td>0.00</td>
<td>2.00</td>
<td>1.71</td>
<td>.20</td>
<td>.55</td>
</tr>
</tbody>
</table>

Note. a-ipvaw: acceptability of intimate partner violence against women; r_item-total = item-total corrected correlation. In brackets the standard error for the skew and kurtosis statistics.

Exploratory Factor Analysis

An EFA was conducted using the first subsample. Bartlett’s sphericity test was significant (p < .001), indicating that the items were dependent, and the KMO index was acceptable (KMO = .928). Therefore, the correlations of the items and the data matrix were suitable for carrying out a factor analysis. The parallel analysis based on minimum rank factor analysis showed that the first factor accounted for 44.6% of the explained variance, far above the average 19.6% expected for a data matrix of random responses, whereas a hypothetical second factor accounted only for 8.4%, below the average 8.5% expected for a randomly generated data matrix. In addition, the MAP criterion indicated that the differences between the one-factor and the two-factor solutions were negligible; a one-factor solution was therefore considered more suitable. As both tests yielded similar results, only one factor was extracted using WLSMV as the estimation method. The estimated model converged normally and showed a good fit to the data. Although the chi-square test was significant, χ²(170) = 485.74, p < .001, which a priori implies that the model did not fit well the data, the comparative fit indices of the model were above the cut-off values (CFI = .96, TLI = .96), and the model residuals were fair (SRMR = .080, RMSEA = .045). These fit indices indicated that the model fitted well the data.

Figure 1. CFA One-factor Model.

Note. Standardized factor loadings (with their standard error in brackets) are depicted in the diagram for each item of the scale; a-ipvaw: acceptability of intimate partner violence against women item.
**Item Response Theory**

Once the dimensionality of the A-IPVAW scale was delimited, the items were calibrated under Samejima’s (1969) graded response model for the full sample.

**Table 3. A-IPVAW Scale IRT Item Parameters**

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b1</th>
<th>b2</th>
</tr>
</thead>
<tbody>
<tr>
<td>a-ipvaw1</td>
<td>2.29</td>
<td>0.52</td>
<td>1.95</td>
</tr>
<tr>
<td>a-ipvaw2</td>
<td>2.25</td>
<td>0.53</td>
<td>2.00</td>
</tr>
<tr>
<td>a-ipvaw3</td>
<td>2.82</td>
<td>1.68</td>
<td>2.86</td>
</tr>
<tr>
<td>a-ipvaw4</td>
<td>2.36</td>
<td>1.49</td>
<td>2.87</td>
</tr>
<tr>
<td>a-ipvaw5</td>
<td>1.96</td>
<td>1.87</td>
<td>3.33</td>
</tr>
<tr>
<td>a-ipvaw6</td>
<td>2.65</td>
<td>1.67</td>
<td>2.88</td>
</tr>
<tr>
<td>a-ipvaw7</td>
<td>2.98</td>
<td>2.25</td>
<td>3.08</td>
</tr>
<tr>
<td>a-ipvaw8</td>
<td>2.41</td>
<td>1.85</td>
<td>2.91</td>
</tr>
<tr>
<td>a-ipvaw9</td>
<td>2.25</td>
<td>2.32</td>
<td>3.45</td>
</tr>
<tr>
<td>a-ipvaw10</td>
<td>3.20</td>
<td>2.44</td>
<td>3.26</td>
</tr>
<tr>
<td>a-ipvaw11</td>
<td>2.44</td>
<td>2.40</td>
<td>3.48</td>
</tr>
<tr>
<td>a-ipvaw12</td>
<td>2.99</td>
<td>2.39</td>
<td>3.42</td>
</tr>
<tr>
<td>a-ipvaw13</td>
<td>3.66</td>
<td>2.57</td>
<td>3.34</td>
</tr>
<tr>
<td>a-ipvaw14</td>
<td>3.15</td>
<td>2.46</td>
<td>3.25</td>
</tr>
<tr>
<td>a-ipvaw15</td>
<td>2.59</td>
<td>1.75</td>
<td>2.84</td>
</tr>
<tr>
<td>a-ipvaw16</td>
<td>1.95</td>
<td>1.55</td>
<td>2.95</td>
</tr>
<tr>
<td>a-ipvaw17</td>
<td>1.71</td>
<td>1.87</td>
<td>3.23</td>
</tr>
<tr>
<td>a-ipvaw18</td>
<td>2.04</td>
<td>1.93</td>
<td>3.10</td>
</tr>
<tr>
<td>a-ipvaw19</td>
<td>3.60</td>
<td>2.44</td>
<td>3.25</td>
</tr>
<tr>
<td>a-ipvaw20</td>
<td>2.11</td>
<td>1.00</td>
<td>2.67</td>
</tr>
</tbody>
</table>

Note. a: discrimination parameter; b1: threshold parameters; a-ipvaw: acceptability of intimate partner violence against women.

Item parameter estimates are shown in Table 3. The threshold parameters (b1 and b2) indicate the point on the latent trait continuum (i.e., acceptability of IPV AW) where the probability of endorsement between two adjacent categories is .50 for any respondent with a person parameter 0 (i.e., acceptability estimates) equal to the threshold parameter value. Therefore, respondents with acceptability estimates lower than the b1 parameter would be more likely to endorse the lowest category (i.e., not acceptable), whereas those respondents with acceptability estimates higher than the b2 parameter would tend to endorse one of the other two categories. Those respondents would more likely endorse the intermediate category (i.e., somewhat acceptable) if their acceptability estimate was lower than the b1 parameter, and the upper category (i.e., acceptable) if their acceptability estimate was higher than the b2 parameter. In general, the threshold parameters were high, indicating that the test was sampling high (above 1) and very high levels (above 2) of acceptability of IPV AW.

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

The information function of the test, as depicted in Figure 2, showed that the A-IPVAW scale was especially informative for respondents with high and very high acceptability estimates. The standard error of estimation (SE) informed about the accuracy of the scale for the different latent trait levels; the lower the SE, the higher the precision of the scale for a given latent trait level. In particular, SE values below .5 are equivalent to a Cronbach’s α of .75 or higher; and SE values below .3 are equivalent to a Cronbach’s α of .91 or higher. This means that the test can estimate very accurately the attitudes toward acceptability of those respondents with moderate, high, and very high levels of acceptability of IPV AW, although it cannot discriminate well among respondents with low and very low levels of acceptability.

**Figure 2. Test Information Function.**

Note. Acceptability Estimate: IRT scores of the scale; Information: accuracy of the measure over the latent trait continuum (i.e., acceptability); SE: standard error of estimation (red line).

Overall fit of the model was evaluated with the ordinal version of the M2 statistic. This statistic works similarly to the χ2 statistic in the factor analysis framework. Thus, the same fit indices (CFI, TLI, and RMSEA) could be computed for the IRT analyses using the M2. Using this procedure, the model showed an adequate fit, M2(150) = 560.87, p < .001; CFI = .99, TLI = .98, and RMSEA = .036.

**Validity Analyses**

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

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

**Table 4. A-IPVAW Scale Correlation to Other Variables**

<table>
<thead>
<tr>
<th></th>
<th>Hostile Sexism</th>
<th>Benevolent Sexism</th>
<th>PS-IPVAW</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-IPVAW</td>
<td>.44***</td>
<td>.34***</td>
<td>-.47***</td>
</tr>
<tr>
<td>Hostile Sexism</td>
<td>.81***</td>
<td>-39***</td>
<td></td>
</tr>
<tr>
<td>Benevolent Sexism</td>
<td>-30***</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. A-IPVAW: acceptability of intimate partner violence against women scale; PS-IPVAW: perceived severity of intimate partner violence against women scale. *** p < .001

When the A-IPVAW scores were compared by gender, significant differences were found between men (M = 0.27, SD = 0.95) and women (M = -.017, SD = 0.77), t(1637.5) = 10.82, p < .001, d = 0.51, with a moderate effect size. There were also significant
Discernible differences between men from the general sample and those from the offenders sample (M = 0.98, SD = 1.35), t(51.88) = -3.67, p < .001, d = 0.61, with a moderate effect size. Significant differences were found between age groups, F(2) = 3.49, p = .03, η² = .004; however, the effect size was below the cut-off value of .01 for the partial eta-squared, and therefore the effect of age on the A-IPVAW scores could be considered negligible (Miles & Shevlin, 2001).

Discussion

Attitudes of acceptability of IPVAW are an important risk factor of actual IPVAW (Abramsky et al., 2011; Archer & Graham-Kevan, 2003; Copp et al., 2016; Gracia et al., 2015; WHO, 2002). However, the measurement of these attitudes in surveys in western countries tends to be based on single items and not supported by reliable and valid instruments (Gracia & Lila, 2015). The aim of this study was to develop and validate the A-IPVAW scale, a twenty-item instrument to assess the acceptability of IPVAW among the general population. Taken together, our results provide evidence that the A-IPVAW is a reliable and valid instrument to assess acceptability of IPVAW.

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

The use of IRT to study the psychometric properties of the scale constitutes a major strength of this paper, since little research has applied this analytical framework to the study and measurement of IPVAW. Some previous research has applied IRT to the study of intimate partner violence, either by fitting an IRT model to an existing scale (e.g., Beck, Menke, & Figueredo, 2013; Jose, Olino, & O’Leary, 2012; Reichenheim, Klein, & Moraes, 2007) or by testing differential item functioning across gender (e.g., Edelen, McCaffrey, Marshall, & Jaycox, 2009; Yount et al., 2014); however, none of these studies addressed the acceptability of IPVAW. Moreover, IRT offers improved factor scores that can be used to assess which latent trait levels (i.e., acceptability estimates) are measured more accurately. The test information function is a dynamic approach to study the reliability of a psychometric instrument. Unlike Cronbach’s α, the IRT information function does not assume that the accuracy of the scale is constant across the entire latent trait (i.e., acceptability estimates), and thus some latent trait levels are more accurately measured than others. In particular, the A-IPVAW scale is especially informative (i.e., accurate) for moderate, high, and very high levels of acceptability of IPVAW. Our measure can detect individuals with high levels of acceptability and discriminate among them with high precision. On the other hand, the precision of the scale is lower for individuals with low and very low levels of acceptability.

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

Regarding gender differences, women showed lower scores in the A-IPVAW scale than men, which is also consistent with previous research (Carlson & Worden, 2005; Fincham et al., 2008; Flood & Pease, 2009; Gracia & Herrero, 2006b). Moreover, in line with the findings of Gracia et al. (2015), we found that convicted batterers are more prone to show higher levels of acceptability than men from the general population. Although this preliminary result should be taken with caution as the sample of batterers is small and somewhat limited, it highlights that the scale is indeed especially informative for those individuals with higher levels of acceptability of IPVAW, and thus at higher risk of committing IPVAW (Archer & Graham-Kevan, 2003; Gracia et al., 2015). In this regard, the A-IPVAW scale can be used as a forensic tool that can detect cases of higher risk of IPVAW perpetration, as it can differentiate accurately between participants with low and high acceptability estimates (Andreu-Rodríguez, Peña-Fernández, & Loza, 2016). It can also be used as an evaluation instrument for intervention programs with IPVAW perpetrators, monitoring attitudinal changes during and after the intervention (Carbajosa, Catalá-Miñana, Lila, & Gracia, 2017; Ferrer-Perez, Ferreiro-Basurto, Navarro-Guzmán, & Bosch-Fiol, 2016; Lila, Gracia, & Catalá-Miñana, 2017; Lila, Oliver, Galiana, & Gracia, 2013).

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

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

Conflict of Interest

The authors of this article declare no conflict of interest.
Appendix A

A-IPVAW Scale I think it is acceptable for a man ...

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

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

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

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

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

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

Appendix B

R Code to Obtain the A-IPVAW Scale IRT Scores

# A-IPVAW data.dat
aipvaw_data <- read.table("data path and format")
# aipvaw_data <- read.table("data path and format")
# with respondents on the rows and items on the columns
a_AIPVAW <- c(2.29, 2.25, 2.82, 2.36, 1.96, 2.65, 2.98, 2.41, 2.25, 3.20, 2.44, 2.99, 3.66, 3.15, 2.59, 1.95, 1.71, 2.04, 3.60, 2.11)
b1_AIPVAW <- c(0.52, 0.53, 1.68, 1.49, 1.87, 1.67, 2.25, 1.85, 2.32, 2.44, 2.40, 2.39, 2.57, 2.46, 1.75, 1.55, 1.87, 1.93, 2.44, 1.00)
b2_AIPVAW <- c(1.95, 2.00, 2.86, 2.37, 3.33, 2.88, 3.08, 2.91, 3.45, 3.26, 3.48, 3.42, 3.34, 3.25, 2.84, 2.95, 2.32, 3.10, 3.25, 2.67)
AIPVAW_param <- mirt(aipvaw_data, 1, itemtype = "graded", pars = "values")
AIPVAW_param$est <- FALSE
AIPVAW_param$value$aipv_pars$name == "d1" <- b1_AIPVAW - a_AIPVAW
AIPVAW_param$value$aipv_pars$name == "d2" <- b2_AIPVAW - a_AIPVAW
AIPVAW_IRT <- mirt(AIPVAW_data, 1, itemtype = "graded", pars = AIPVAW_param)
IRTScores <- fscores(AIPVAW_IRT, method = "EAP", full.scores = T)
write.table(IRTScores, "IRTScores.dat", col.names = FALSE, row.names = FALSE)
# return a .dat file with the IRT Scores for each respondent

References

Acebes-Arribas, B. (2016). Estudio de medios de comunicación online. Madrid, España: IAR.
Acceptability of Intimate Partner Violence


