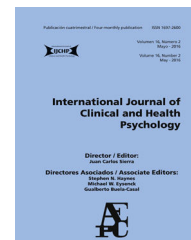




International Journal of Clinical and Health Psychology

www.elsevier.es/ijchp



ORIGINAL ARTICLE

Validation of the Working Alliance Inventory–Observer Short Version with male intimate partner violence offenders



Faraj A. Santirso, Manuel Martín-Fernández, Marisol Lila*, Enrique Gracia, Elena Terreros

University of Valencia, Spain

Received 7 December 2017; accepted 27 February 2018

Available online 27 March 2018

KEYWORDS

Working alliance;
Intimate partner
violence offenders;
Batterer intervention
programs;
Observational scale;
Observational
descriptive study

Abstract The working alliance is a key element to increase intimate partner violence (IPV) offenders' motivation, adherence to treatment, and active participation in batterer intervention programs (BIPs). The objective of the present study is to assess the psychometric properties and factor structure of the Working Alliance Inventory–Observer Short Version (WAI-O-S) with a sample of IPV offenders. The sample was 140 men convicted for IPV and court-mandated to a community-based BIP. Inter-rater agreement and reliability were evaluated by computing the intraclass correlation coefficient. To test the latent structure a Bayesian confirmatory factor analysis approach was used. To test criterion-related validity, the WAI-O-S factorial scores were correlated to protherapeutic behavior, stage of change and motivation to change. The WAI-O-S showed an adequate reliability. Results from Bayesian confirmatory factor analyses showed two first-order factors (*Bond* and *Agreement*), and a second-order factor (*General working alliance*) explaining the relationship between the first-order factors. Results also support the validity of this instrument. The availability of reliable and valid observational measure of the working alliance provides a useful tool to overcome self-report measurement limitations such as social desirability, deception, and denial among IPV offenders.

© 2018 Asociación Española de Psicología Conductual. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

* Corresponding author: Department of Social Psychology, University of Valencia, Avenida Blasco Ibáñez, 21, 46010 Valencia, Spain.
E-mail address: marisol.lila@uv.es (M. Lila).

PALABRAS CLAVE

Alianza terapéutica;
maltratadores;
Programas de
intervención con
maltratadores;
Escala observacional;
Estudio descriptivo
observacional

Validación de la versión breve del *Working Alliance Inventory–Observer* con hombres maltratadores**Resumen**

Antecedentes/Objetivo: La alianza terapéutica es clave para incrementar la motivación, adherencia al tratamiento y participación en los programas de intervención con maltratadores. El objetivo del presente estudio es evaluar las propiedades psicométricas y estructura factorial del *Working Alliance Inventory–Observer Short Version* (WAI-O-S) con una muestra de hombres maltratadores.

Método: Se utilizó una muestra de 140 hombres condenados por violencia de género y derivados a un programa de intervención con maltratadores. El acuerdo inter-jueces y la fiabilidad se evaluaron mediante el coeficiente de correlación intraclase. La estructura latente se obtuvo con un análisis factorial confirmatorio Bayesiano. Las puntuaciones del WAI-O-S se correlacionaron con la conducta proterapéutica, estadio y motivación para el cambio para analizar la validez.

Resultados: El WAI-O-S mostró una fiabilidad adecuada. El análisis factorial confirmatorio dio lugar a dos factores de primer orden (Vínculo y Acuerdo), y un factor de segundo orden (Alianza terapéutica general). Los resultados obtenidos también avalan la validez del instrumento.

Conclusiones: Una medida observacional fiable y válida de la alianza terapéutica, proporciona una herramienta útil para superar las limitaciones de las medidas basadas en auto-informes que se utilizan con población de hombres maltratadores, tales como la deseabilidad social, la negación y el engaño.

© 2018 Asociación Española de Psicología Conductual. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

The working alliance is one of the most extensively studied constructs in psychotherapy research (Del Re, Flückiger, Horvath, Symonds, & Wampold, 2012; Horvath & Symonds, 1991; Martin, Garske, & Davis, 2000). According to Bordin (1979), the working alliance is a collaborative feature of the therapy, composed of three elements: (1) agreement between patient and therapist on the objectives that they aspire to achieve with the treatment, (2) the patient's acceptance and collaboration in the tasks the therapist proposes during therapy to address their problem, and (3) the quality of the patient-therapist bond in terms of mutual trust, appreciation, etc. Several meta-analyses show that the working alliance is one of the best predictors of the results of psychotherapy, indicating that there is a robust although moderate relationship between the working alliance and the indicators of change during treatment (Del Re et al., 2012; Horvath & Symonds, 1991; Martin et al., 2000).

A number of instruments have been developed to measure the working alliance, including the Pennsylvania Helping Alliance Rating Scale (PENN; Luborsky, Crits-Christoph, Alexander, Margolis, & Cohen, 1983), the Vanderbilt Therapeutic Alliance Scale (VTAS; Hartley & Strupp, 1983), and the Working Alliance Inventory (WAI; Horvath & Greenberg, 1989). The latter is one of the most widely used instruments (Andrade-González & Fernández-Liria, 2015) and evaluates the three theoretical dimensions proposed in 1979 by Bordin (i. e., objective, task, and bond). The WAI (Horvath & Greenberg, 1989) presents three different versions to measure the working alliance from different perspectives: the therapist (WAI-T), the patient (WAI-P) and the observer (WAI-O) versions. Tracey and Kokotovic (1989) developed an abbreviated version of the scale composed of 12 items, the Working Alliance Inventory Short Form (WAI-S).

This abbreviated version was adapted for evaluation by an external observer by Tichenor and Hill (1989): The Working Alliance Inventory–Observer Short version (WAI-O-S).

Regarding the WAI factorial structure, research using different versions of this scale (Busseri & Tyler, 2003; Horvath & Greenberg, 1989; Munder, Wilmers, Leonhart, Linster, & Barth, 2010; Tracey & Kokotovic, 1989) tends to support the three-factor structure proposed by Bordin (1979), with some of these studies reporting a high correlation between the 'task' and 'goal' dimensions (Busseri & Tyler, 2003; Horvath & Greenberg, 1989). However, other studies favor a two-factor structure, grouping the 'task' and 'goal' dimensions into a single dimension (Andrusyna, Tang, DeRubeis, & Luborsky, 2001).

Measuring Working Alliance in Batterers' Intervention Programs (BIPs)

The various meta-analyses on the effectiveness of BIPs indicate limited effect sizes of these intervention programs (e.g., Arias, Arce, & Vilariño, 2013; Babcock, Green, & Robie, 2004). A number of new intervention strategies are now being implemented in BIPs with evidence to support their effectiveness among resistant populations (Alexander, Morris, Tracy, & Frye, 2010; Carbajosa, Catalá-Miñana, Lila, & Gracia, 2017; Lila, Gracia, & Catalá-Miñana, 2018; Musser, Semiatin, Taft, & Murphy, 2008; Vargas, Lila, & Catalá-Miñana, 2015). Among these strategies are motivational interviewing, stages of change or strengths-based approaches, and some studies applying them to offenders attending BIPs have consistently showed their ability to promote positive changes, such as acceptance and adherence to the intervention process and lower lev-

els of post-treatment recidivism (Crane, & Eckhardt, 2013; Llor-Esteban, García-Jiménez, Ruiz-Hernández, & Godoy-Fernández, 2016; López-Ossorio, Álvarez, Pascual, García, & Buela-Casal, 2017; López-Ossorio, González-Álvarez, & Andrés-Pueyo, 2016). These new intervention strategies share the idea that the working alliance is a key element to increase 'users' motivation, adherence to treatment, and active participation in BIPs. Improving the working alliance can increase participants' perception that the therapist is concerned about their progress. The working alliance is also associated with a higher motivation to participate and complete the intervention program, reducing dropout and recidivism rates (Babcock et al., 2004; Brown & O'Leary, 2000; Semiatiin, Murphy, & Elliot, 2013; Taft, Murphy, Musser, & Remington, 2004). Several studies have analyzed the relationship between the working alliance and the results of BIPs (Brown & O'Leary, 2000; Semiatiin et al., 2013; Taft et al., 2004). These studies suggest that the working alliance is associated with protherapeutic group behaviors, motivation to change, and stage of change.

Given the importance of this construct for the effectiveness of BIPs, the availability of psychometrically sound observational measures would clearly improve the evaluation of the working alliance in such intervention programs. Men who attend this type of program tend to minimize or deny the violent acts for which they have been convicted (Lila, Oliver, Catalá-Miñana, Galiana, & Gracia, 2014). Assessing the working alliance among BIP participants using self-reported measures is particularly problematic, as social desirability, deception, minimization, and denial are common in this population (Gracia, Rodríguez, & Lila, 2015). Observational measures can overcome these limitations. However, such measures have not so far been used to assess the working alliance in BIPs. In the present study, we aim to address this gap in the literature by validating the WAI-O-S scale (Tichenor & Hill, 1989) with a sample of intimate partner violence (IPV) offenders court-mandated to a community-based BIP. As far as we know, only one study has analyzed the working alliance with observational measures in a group treatment for husband-on-wife spouse abuse (Brown & O'Leary, 2000). It is important to note that in this study the couples attended the treatment on a voluntary basis, while in the present study we analyze the working alliance using an observational measure with men who have been court-mandated to a BIP. The objectives of the present study are: (1) to assess the psychometric properties and factor structure of the WAI-O-S scale with a sample of men convicted of IPV attending a BIP; (2) to analyze the criterion-related validity of this measure by studying the association between the working alliance and a set of variables relevant for BIP intervention processes: protherapeutic group behavior, motivation to change, and stage of change (Brown & O'Leary, 2000; Carbajosa, Catalá-Miñana, Lila, Gracia, & Boira, 2017; Semiatiin et al., 2013; Taft et al., 2004).

Method

Participants

The sample consisted of 140 men who were convicted for IPV and court-mandated to a community-based BIP,

the *Contexto* Program, conducted at the University of Valencia. These offenders had been sentenced to less than two years in prison, had no previous criminal record, and their sentence was suspended on the condition that they attended this community-based intervention program. The criteria for inclusion in this study were: (a) not having a serious mental disorder, (b) not having a serious addiction to alcohol or other substances, and (c) signing an informed consent form. The mean age was 40.26 years ($SD = 11.66$, range 18-76); 50.7% had completed primary or elementary studies, 34.3% had completed high school or vocational training, 10% university studies, and 5% had no schooling; 35.7% were single, 39.3% divorced or separated, and 24.3% married or in a relationship. The majority, 70%, were Spanish, 12.7% were Latin-American, 8.5% from other European countries, 7.8% African, and 0.7% Asian. The median family household income was between 6,000 and 12,000 euros.

Procedure

All participants were assigned to an intervention group. The number of participants per group ranged from 8-12 participants and once started, no more participants were allowed. Two therapists conducted each group. The intervention consisted of 32 weekly group sessions of a cognitive-behavioral intervention including the gender perspective, which is the standard intervention with IPV offenders (Eckhardt et al., 2013; Ferrer-Perez, Ferreiro-Basurto, Navarro-Guzmán, & Bosch-Fiol, 2016). Participants were informed about the nature and purpose of the research, completed a written consent form, and were told that neither participation nor refusal would affect their legal situation. Confidentiality was ensured. The group intervention sessions were recorded on video (there were a total of 14 intervention groups), and one session at the end of each group intervention was evaluated. Four trained research assistants assessed the recorded sessions. Raters previously underwent training during which they assessed the same recorded session separately until they reached an acceptable level of agreement (i. e., not differing by more than one point on each assessed item). Each analyzed video had a recording time of two hours divided into 24 segments. For the observational coding (i. e., working alliance and protherapeutic group behaviors), scores for 5-minute video intervals were used, and averaged for each participant. Stage of change and motivation to change were evaluated at the end of the intervention, as research has found that these measures are positively related to social desirability at the beginning of the intervention, but not at the end (Begun et al., 2003). All participants' data were collected in accordance with the University of Valencia Ethics Committee approved procedures.

Instruments

Working Alliance Inventory Shortened Observer-rated version (WAI-O-S; Tichenor & Hill, 1989). This observational scale evaluates the working alliance and is composed of 12 items (e.g., "the participant feels that the therapist appreciates him as a person", "the participant and therapist are working on mutually agreed upon goals"). Raters responded

on a 7-point Likert-type scale ranging from 0 (*conclusive evidence against*) to 7 (*conclusive evidence in favor*). See [Appendix](#).

Observational Coding of Protherapeutic Group Behavior ([Semiatin et al., 2013](#)). This instrument evaluates participants' protherapeutic group behaviors by focusing on their verbalizations. It consists of three items that correspond to the following pro-therapeutic behaviors: (a) Denial/acknowledgment of behavior/responsibility: participants' verbalizations related to recognition or denial of their responsibility for their violent behavior, the consequences of this behavior, and the need for personal change to avoid committing abusive acts in the future; (b) Client role behavior: interpersonal behaviors that occur in the group and are relevant for change. Four types of behavioral roles can be distinguished along two dimensions: confrontation vs. confirmation, and positive progress vs. negative progress; (c) Group value: participants' verbalizations related to the perception of the group, and the treatment program. Raters responded on a 5-point Likert-type scale ranging from 1 (*conclusive evidence against*) to 5 (*conclusive evidence in favor*).

Stage of change. Based on individual interviews, self-reports, and direct observations, therapists rated each participant's stage of change following the classification of [Prochaska, DiClement and Norcross \(1992\)](#). The stages of change are rated as 1 (*precontemplation*), 2 (*contemplation*), 3 (*preparation*), 4 (*action*), and 5 (*maintenance*). For a similar procedure, see [Scott \(2004\)](#) and [Carbajosa et al. \(2017a\)](#).

Motivation to change. The therapists evaluated the motivation to change of each participant on a single item ranging from 1 (*low*) to 5 (*high*).

Data analysis

The following analyses were carried out to assess the psychometric properties of the WAI-O-S. First, inter-rater agreement and reliability were evaluated by computing the intraclass correlation coefficient (ICC). The average scale ICC was estimated using a random two-way ANOVA model, since all raters evaluated each participant ($ICC_{(2,k)}$). This statistic treated raters as a random effect variable, representing a random sample of a larger population of raters (i.e., trained observers). Reliability was also measured in terms of internal consistency with Cronbach's alpha and the McDonald's omega. Values of this statistics equal to or higher than .70 were indicative of good internal consistency. Descriptive statistics and corrected item-total correlations were then obtained for all items.

To test the latent structure underlying the WAI-O-S, a Bayesian confirmatory factor analysis approach was used (CFA). This approach has been shown to perform better in small samples and with skewed observed variables than the classical maximum likelihood estimation in confirmatory factor analysis ([Lee & Song, 2004](#)). Bayesian Structural Equation Modeling is an emergent CFA framework that has proved to be well-suited to skewed distributions of parameter estimates, and it also allows complex latent structures to be tested ([Muthén & Asparouhov, 2012](#)). Given the limited size of the sample in this study, and the potential advantages of

the Bayesian estimation, this methodological approach was followed to test the latent structure of the WAI-O-S.

Five models of different latent complexity were compared in order to obtain the best fit to the data (see [Figure 1](#)). The first (1a) was a one-dimensional model in which all items load onto a single factor. The second (1b), based on the findings of [Andrusyna et al. \(2001\)](#) and [Falkenström, Hatcher, and Holmqvist \(2015\)](#), was an oblique two-dimensional model that set one factor for the original Bond dimension, and a second factor for the Goal and Task items. The third (1c) was an oblique three-dimensional model which posited one factor for each of the original dimensions (i.e., Task, Goal, and Bond; [Munder, Wilmers, Leonharet, Linster, & Barth, 2010](#)). The fourth (1d) and fifth (1e) models were generalizations of the second and third ones, in which a second-order factor, called General working alliance, accounted for the relation between the first order factors of each model. Hence the fourth model (1d) set a second-order model considering the two dimensions of the second model as the first-order factors, whereas the fifth (1e) model considered the three dimensions of the third model as first-order factors.

All models were estimated with the MCMC algorithm, setting 4 chains and 20,000 iterations. The first 10,000 iterations were discarded as a burn-in period. Model convergence was assessed with the potential scale reduction factor (PSR), considering PSR values of 1.05 or lower as indicative of good convergence ([Gelman, Carlin, Stern, & Rubin, 2014](#)). [Asparouhov and Muthén \(2010\)](#) reported, however, that most models usually reach convergence with PSR values between 1.05 and 1.10, hence we considered PSR values of 1.10 or lower as indicative of acceptable convergence. Model parameters were estimated using the expected a posteriori (EAP) method. To delimit the credible intervals of the parameters, posterior SD for each parameter was also obtained. The Deviance Information Criterion (DIC), and the Bayesian Information Criterion (BIC) were obtained to evaluate model fit. The DIC correction proposed by several authors was also computed in this study (DIC_c ; [Ando, 2011](#)). DIC_c penalized more severely than DIC models with a high number of parameters. These criteria are indices of comparative meaning, and, thus, the model with the lowest DIC, DIC_c and BIC values has the best fit.

Finally, to test criterion-related validity, the WAI-O-S factorial scores were correlated to the variables protherapeutic group behavior, stage of change, and motivation to change. Pearson correlations were used when both correlated variables were considered continuous (i.e., with five or more categories). When at least one of the variables was considered ordinal, Spearman correlations were used instead. Descriptive, reliability, and validity tests were carried out with the statistical package R and the psych library for R. Bayesian CFA analyses were performed with Mplus 7.1 ([Muthén & Muthén, 2010](#)).

Results

Reliability and descriptive analyses

The effect of different raters assessing the same participants was first evaluated. The $ICC_{(2,k)}$ for the average of the

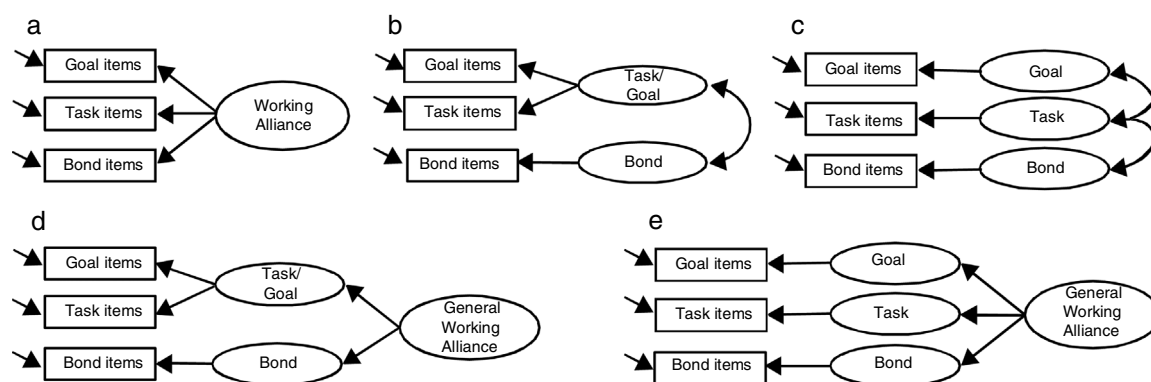


Figure 1 CFA tested models.

Table 1 Item descriptives.

	<i>M</i>	<i>SD</i>	<i>Mdn</i>	<i>Min</i>	<i>Max</i>	<i>Skew</i>	<i>Kurtosis</i>	<i>r</i> _{item-total}
Item 1	4.24	0.23	4.21	3.75	4.91	0.60(.02)	0.32(.02)	0.80
Item 2	4.19	0.21	4.17	3.67	4.88	0.41(.02)	0.14(.02)	0.79
Item 3	4.22	0.24	4.21	3.46	4.91	0.32(.02)	0.64(.02)	0.81
Item 5	4.14	0.21	4.05	3.50	4.88	1.06(.02)	2.35(.02)	0.81
Item 6	4.16	0.22	4.08	3.83	4.88	1.38(.02)	1.63(.02)	0.92
Item 7	4.15	0.23	4.08	3.17	4.88	0.04(.02)	3.04(.02)	0.78
Item 8	4.15	0.24	4.08	3.22	4.91	0.65(.02)	2.85(.02)	0.85
Item 9	4.16	0.23	4.06	3.42	4.88	0.78(.02)	1.30(.02)	0.90
Item 11	4.13	0.21	4.04	3.75	4.88	1.47(.02)	1.98(.02)	0.91
Item 12	4.13	0.21	4.05	3.79	4.88	1.79(.02)	3.14(.02)	0.85

Note. *M* = Mean, *SD* = Standard Deviation, *Mdn* = Median, *Min* = Minimum, *Max* = Maximum, *r*_{item-total} = item-total corrected correlation. In brackets: the standard error for the skew and kurtosis statistics.

four raters' measures was .82, $F(10, 30) = 9.3$, $p < .001$, indicating an excellent level of inter-rater agreement and reliability. The internal consistency of the WAI-O-S was also high (Cronbach's alpha = .96; McDonald's omega = .98).

Item descriptives and item-total corrected correlation are displayed in Table 1. Items 4 and 10 were removed from the scale as they lowered the Cronbach's alpha and the McDonald's omega, they presented low item-total corrected correlations, and, moreover, they were highly correlated. Item scores were centered on 4, the mid-point of the scale, and slightly displaced to the right, reflecting the small differences between the mean and the median and the low standard deviation values. Some items (5, 7, and 12) presented a strong positive skewness and almost all of them showed high kurtosis values, resulting in a leptokurtic distribution of the item scores. Regarding the item-total corrected correlations, all the remaining items showed a very strong relation with the rest of the test.

Confirmatory factor analysis

Fit indices are shown in Table 2. Both DIC progressions are clear: modeling working alliance as a second-order factor improved the model fit to the data as compared to the two and three first-order factor models. Although there were no differences between the one-, two-, and three-factor first-order models, when the second-order factor was taken into account, the model with two first-order factors and

a second-order factor (General working alliance) showed a better fit to the data. BIC values, however, suggest that the two-factor model is the latent structure with the best fit to the data. Given that two out of three indices found that the second-order model with two first-order factors best fitted the data, and since this model had lower DIC, DIC_C, and BIC values than the other second-order model, we decided to keep the second-order with two first-order factors model as the latent structure of the WAI-O-S. All models reached convergence adequately, below the usual cut-off for the PSR factor.

Regarding the selected model (Figure 2), the WAI-O-S Bond items group on a first-order factor, and thus we kept the same name. The remaining Goal and Task items, however, group on a single first-order factor, which we named Agreement, as suggested by Andrusyna et al. (2001). All items loaded positively on the first-order factors, with standardized loadings above .80 in almost all cases. Moreover, both first-order factors contributed equally to the second-order factor, with high standardized loadings. Posterior standard deviations for the standardized loadings were on average low, implying an accurate estimation of these parameters.

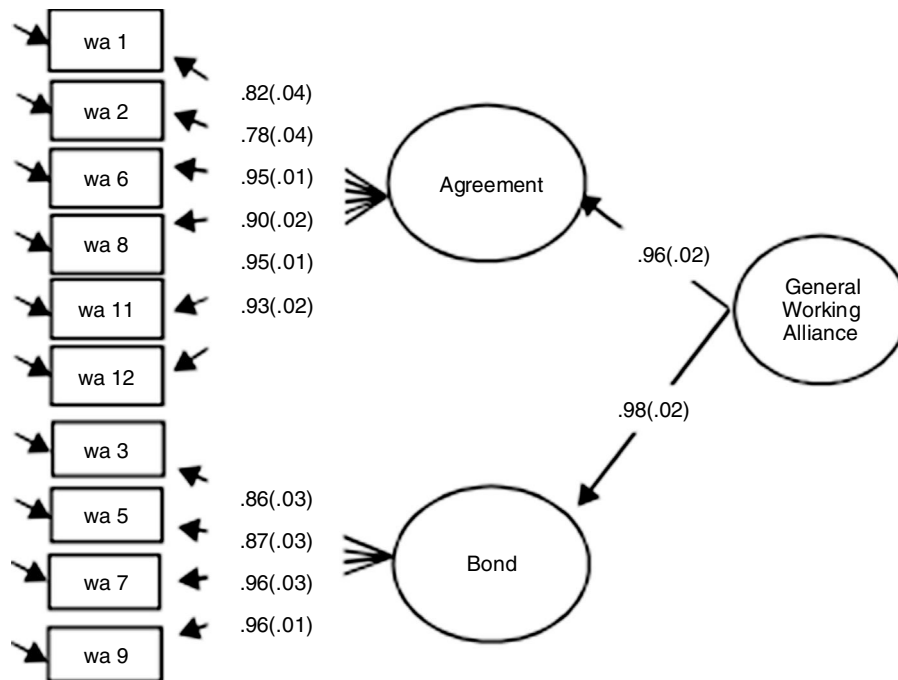
Criterion-related validity

Criterion-related validity was assessed by means of the correlation of various constructs that have been related

Table 2 Bayesian CFA fit indices.

Model	PSR	DIC	DICc	<i>pD</i>	BIC
One Factor	1.00	-1,363.11	-1,333.78	29.33	-1,283.35
Two Factors	1.01	-1,377.33	-1,344.71	32.62	-1,299.95
Three Factors	1.01	-1,377.88	-1,345.09	32.79	-1,292.99
2nd Order – Two Factors	1.08	-1,385.46	-1,358.01	27.45	-1,295.44
2nd Order – Three Factors	1.09	-1,381.87	-1,354.52	27.35	-1,288.22

Note. PSR = potential scale reduction factor, DIC = deviance information criterion, DIC_c = corrected deviance information criterion, *pD* = estimated number of parameters, BIC = Bayesian information criterion.

**Figure 2** CFA second order model with standardized item loadings.**Table 3** Criterion-related validity.

	Agreement(Task/Goal)	Bond	General working alliance
Protherapeutic group Behavior	$r = .69^{**}$	$r = .73^{**}$	$r = .73^{**}$
Stage of change	$r = .29^{**}$	$r = .17^*$	$r = .25^*$
Motivation to change	$r = .33^{**}$	$r = .22^*$	$r = .29^{**}$

Note. * $p < .05$, ** $p < .01$

to the working alliance in multiple studies on IPV. Both two first-order factors (i.e., Bond and Agreement) and the second-order factor (i.e., General working alliance) were significantly and positively associated with the protherapeutic group behavior, stage of change, and motivation to change. Statistics of these associations are shown in Table 3.

Discussion

Results from this study showed that the WAI-O-S could be used to assess the working alliance through external observers in interventional settings. In particular, among the

male batterers attending BIPs, this instrument has showed to be an accurate and consistent approach to evaluate the relationships between the therapists and the offenders. Its factorial structure with this sample supports the grouping of Bordin's (1979) three original factors (i.e., Objective, Task and Bond), into two first-order factors (i.e., Bond and Agreement). We also found a second-order factor (i.e., General working alliance) explaining the relationship between the first-order factors. In addition, our findings showed high correlations between the WAI-O-S scores and a set of relevant variables in the evaluation of BIPs' success with IPV offenders, providing evidence of criterion-related validity.

Results showed an excellent level of inter-rater agreement according to Cicchetti's cut-offs (1994). In our study, we obtained values of agreement superior to those obtained by Andrusyna et al. (2001), Brown and O'Leary (2000) and Strunk, Brotman and DeRubeis (2010), and similar to those obtained by Tichenor and Hill (1989). The discrepancies between our results and other studies could be due to differences in the coding procedure used. The above-mentioned studies scored the working alliance through a single measurement for the entire session. In our study, to evaluate the working alliance we divided the session into 24 segments. Another divergence between our study and that of Andrusyna et al. (2001) is the number of raters. We used four raters in this study, compared to two raters in Andrusyna et al. (2001), which could have given rise to the different results obtained in the two studies.

Regarding the reliability of the WAI-O-S, using both Cronbach's alpha and McDonald's omega, our measure showed high levels of internal consistency. These results are similar to those obtained in previous studies (Tichenor & Hill, 1989). It should be noted that items 4 and 10 were removed from our scale in order to improve the accuracy of the instrument and meet the CFA assumptions of item independence. Andrusyna et al. (2001) also obtained a low item-total correlation with these items. However, in our study, after removing items 4 and 10, we obtained a high item-total correlation.

The factor structure of WAI-O-S indicated that grouping the three original factors into two first-order factors (i.e., Bond and Agreement), and a second-order factor (i.e., General working alliance) improved the model fit to the data. This implies that a second-order factor of General working alliance explains the relationship between the Bond and Agreement factors better than a simple correlation. Thus, the working alliance can be measured as a general second-order factor encompassing the Bond and Agreement items of the WAI-O-S, in which both Bond and Agreement factors have almost the same relevance.

Previous studies using the observational version of the questionnaire with patients in cognitive-behavioral treatment found a latent structure of two independent factors, an Agreement/confidence factor and a Relationship factor (Andrusyna et al., 2001). Other researchers validating the client version of the inventory and using a Bayesian approach have proposed a two-factor structure (Falkenström et al., 2015); whereas other authors have proposed three-factor structures (Munder et al., 2010) using the client version. The study by Andrusyna et al. (2001) is the most comparable to our study as it uses the same version of the questionnaire. However, it should be noted that our study differs from that of Andrusyna et al. (2001) in the study sample and in the methodology, since we used a Bayesian approach, which is more suitable for small sample sizes with skewed and kurtotic observed variables (Lee & Song, 2004). On the other hand, contrary to other studies, we have found that the working alliance could be assessed as a second-order factor model rather than as a factorial model with correlated factors.

Regarding the criterion-related validity of the scale, both the two first-order factors (i.e., Bond and Agreement) and the second-order factor (i.e., General working alliance) were significantly associated with a set of relevant variables

in the evaluation of the effectiveness of BIPs. Thus, we found a positive association between the working alliance and protherapeutic group behaviors (Semiatin et al., 2013). This suggests that when there is agreement between the therapist and the participants on the intervention goals and tasks, and there is an adequate bond between them, the participants are more likely to assume their responsibility, adequately value peer change initiatives, and make positive group verbalizations (Lila, Gracia, & Murgui, 2013). Another variable that was positively associated with the working alliance was motivation to change, in line with other studies with IPV offenders (Carbajosa et al., 2017a; Taft et al., 2004). These results have also been found in other populations with alcohol addiction, where motivation to change was one of the most reliable predictors of the working alliance evaluated by the client and the therapist (Catalá-Miñana, Lila, & Oliver, 2013; Connors et al., 2000; Lila, Gracia, & Catalá-Miñana, 2017). In this regard, our results suggest that a positive bond and agreement on the goals and tasks of the intervention between therapists and participants could facilitate offenders' motivation to change. Previous studies have proposed that the working alliance between the participant and the therapist is not only able to modify their cognitive processing but also their actions and affective state (Brown & O'Leary, 2000; Romero-Martínez, Lila, & Moya-Albiol, 2016). In this regard, Brown and O'Leary (2000) found that the working alliance, observationally assessed, was related to a decrease in psychological and physical aggression. Finally, we found a positive and significant association between stage of change and the therapeutic alliance. Using patients' and therapists' measures, Taft et al. (2004) found a similar association between the stage of change and the working alliance.

The main implication of this study is that the WAI-O-S is a psychometrically sound instrument that can be used to assess the working alliance in intervention programs using external raters. To this end, intervention practitioners could compute the scores of the Agreement and Bond factors (see DiStefano, Zhu, & Mindrila, 2009) and then obtain the General working alliance by summing the participants' scores on the first order factors (i.e., Agreement and Bond). The working alliance could thus be monitored during the different stages of the intervention, allowing researchers and practitioners to assess the impact of the relationship between the therapists and the participants of the program in other relevant intervention outcomes.

This study is not without limitations. First, the sample consisted of men convicted of IPV that were court-mandated to a community-based BIP in Spain. Therefore, results could not be extrapolated to other populations, such as offenders in prison, self-referred men, or other cultural groups (Boira, Carbajosa, & Méndez, 2016; Lila, Gracia, Catalá-Miñana, Santirso, & Romero-Martínez, 2016; Vargas et al., 2015). Second, although the Bayesian approach allowed us to test the internal structure of the scale, further research is needed with larger sample sizes to replicate and test the fit of the proposed model to new data. Third, in order to establish whether the inter-rater agreement and reliability of the WAI-O-S could be generalized, it is necessary to explore the concordance of the raters' evaluations using this instrument in other intervention settings. Finally, future research would benefit from exploring the relationships of the

WAI-O-S with other BIPs' relevant outcomes such as recidivism and dropout rates.

Despite these limitations, given that the working alliance is key in intervention programs with offenders and, in particular, with highly resistant populations such as participants in BIPs, the availability of an observational measure of this construct such as the WAI-O-S can provide a useful tool to overcome self-report measurement limitations such as social desirability, deception, and denial among this population.

Funding and acknowledgements

This research was supported by the Spanish Ministry of Economy and Competitiveness (PSI2011-25434 and PSI2010-21334), and was conducted in collaboration with Home Office Prison Administration (Instituciones Penitenciarias, Ministerio del Interior). Faraj A. Santirso was supported by the FPU Program of the Spanish Ministry of Education, Culture and Sport (grant number FPU15/00864). Manuel Martín-Fernández was supported by the FPI Program of the Spanish Ministry of Economy and Competitiveness (BES15/075576).

Appendix. Spanish version of Working Alliance Inventory Shortened Observer-rated version (WAI-O-S; Tichenor & Hill, 1989).

Item 1	Existe acuerdo sobre las medidas adoptadas para ayudar a mejorar la situación del participante
Item 2	Hay acuerdo sobre la utilidad de la actividad actual en la intervención
Item 3	Hay una simpatía recíproca entre el participante y el terapeuta
Item 5	El participante siente confianza en la habilidad del terapeuta para ayudarle
Item 6	Participante y terapeuta están trabajando metas consensuadas de mutuo acuerdo
Item 7	El participante siente que el terapeuta le valora como persona
Item 8	Hay acuerdo sobre lo que es importante trabajar para el participante
Item 9	Hay confianza mutua entre el participante y el terapeuta
Item 11	El participante y el terapeuta han establecido una buena comprensión de los cambios que podrían ser buenos para el participante
Item 12	El participante cree que la forma en la que están trabajando su problema es correcta

References

- Alexander, P. C., Morris, E., Tracy, A., & Frye, A. (2010). Stages of change and the group treatment of batterers: A randomized clinical trial. *Violence and Victims*, 25, 571–587. <http://dx.doi.org/10.1891/0886-6708.25.5.571>
- Ando, T. (2011). Predictive Bayesian model selection. *American Journal of Mathematical and Management Sciences*, 31, 13–38. <http://dx.doi.org/10.1080/01966324.2011.10737798>
- Andrade-González, N., & Fernández-Liria, A. (2015). Spanish adaptation of the Working Alliance Inventory (WAI). Psychometric properties of the patient and therapist forms (WAI-P and WAI-T). *Anales de Psicología*, 31, 524–533. <http://dx.doi.org/10.6018/analesps.31.2.177961>
- Andrusyna, T. P., Tang, T. Z., DeRubeis, R. J., & Luborsky, L. (2001). The factor structure of the Working Alliance Inventory in cognitive-behavioral therapy. *The Journal of Psychotherapy Practice and Research*, 10, 173–178.
- Asparouhov, T., & Muthén, B. (2010). *Bayesian analysis using Mplus: Technical implementation*. Retrieved from: <http://statmodel2.com/download/Bayes3.pdf>.
- Arias, E., Arce, R., & Vilariño, M. (2013). Batterer intervention programmes: A meta-analytic review of effectiveness. *Psychosocial Intervention*, 22, 153–160. <http://dx.doi.org/10.5093/in2013a18>
- Babcock, J. C., Green, C. E., & Robie, C. (2004). Does batterers' treatment work? A metaanalytic review of domestic violence treatment. *Clinical Psychology Review*, 23, 1023–1053. <http://dx.doi.org/10.1016/j.cpr.2002.07.001>
- Begun, A. L., Murphy, C., Bolt, D., Weinstein, B., Strodthoff, T., Short, L., & Shelley, G. (2003). Characteristics of the Safe at Home instrument for assessing readiness to change intimate partner violence. *Research on Social Work Practice*, 13, 80–107. <http://dx.doi.org/10.1177/1049731502238758>
- Boira, S., Carbajosa, P., & Méndez, R. (2016). Fear, conformity and silence. Intimate partner violence in rural areas of Ecuador. *Psychosocial Intervention*, 25, 9–17. <http://dx.doi.org/10.1016/j.psi.2015.07.008>
- Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research and Practice*, 16, 252–260.
- Brown, P. D., & O'Leary, K. D. (2000). Therapeutic alliance: Predicting continuance and success in group treatment for spouse abuse. *Journal of Consulting and Clinical Psychology*, 68, 340–345. <http://dx.doi.org/10.1037/0022-006X.68.2.340>
- Busseri, M. A., & Tyler, J. D. (2003). Interchangeability of the Working Alliance Inventory and Working Alliance Inventory. *Short Form. Psychological Assessment*, 15, 193–197. <http://dx.doi.org/10.1037/1040-3590.15.2.193>
- Carbajosa, P., Catalá-Miñana, A., Lila, M., & Gracia, E. (2017). Differences in treatment adherence, program completion, and recidivism among batterer subtypes. *The European Journal of Psychology Applied to Legal Context*, 9, 93–101. <http://dx.doi.org/10.1016/j.ejpal.2017.04.001>
- Carbajosa, P., Catalá-Miñana, A., Lila, M., Gracia, E., & Boira, S. (2017). Responsive versus treatment-resistant perpetrators in batterer intervention programs: Personal characteristics and stages of change. *Psychiatry, Psychology and Law*, 24, 936–950. <http://dx.doi.org/10.1080/13218719.2017.1347933>
- Catalá-Miñana, A., Lila, M., & Oliver, A. (2013). Consumo de alcohol en hombres penados por violencia contra la pareja: factores individuales y contextuales. *Adicciones*, 25, 19–28.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284–290. <http://dx.doi.org/10.1037/1040-3590.6.4.284>
- Connors, G. J., DiClemente, C. C., Dermen, K. H., Kadden, R., Carroll, K. M., & Frone, M. R. (2000). Predicting the therapeutic alliance in alcoholism treatment. *Journal of Studies on Alcohol*, 61, 139–149. <http://dx.doi.org/10.15288/jsa.2000.61.139>
- Crane, C. A., & Eckhardt, C. I. (2013). Evaluation of a single-session brief motivational enhancement intervention for partner abusive men. *Journal of Counseling Psychology*, 60, 180–187. <http://dx.doi.org/10.1037/a0032178>
- Del Re, A. C., Flückiger, C., Horvath, A. O., Symonds, D., & Wampold, B. E. (2012). Therapist effects in the therapeutic alliance-outcome relationship: A restricted-maximum likelihood meta-analysis. *Clinical Psychology*, 32, 642–649. <http://dx.doi.org/10.1016/j.cpr.2012.07.002>

- DiStefano, C., Zhu, M., & Mindrila, D. (2009). *Understanding and using factor scores: Considerations for the applied researcher. Practical Assessment, 14*, 1–11.
- Eckhardt, C. I., Murphy, C. M., Whitaker, D. J., Sprunger, J., Dykstra, R., & Woodard, K. (2013). The effectiveness of intervention programs for perpetrators and victims of intimate partner violence. *Partner Abuse, 4*, 196–231. <http://dx.doi.org/10.1891/1946-6560.4.2.196>
- Falkenström, F., Hatcher, R. L., & Holmqvist, R. (2015). Confirmatory factor analysis of the patient version of the Working Alliance Inventory-Short Form Revised. *Assessment, 22*, 581–593. <http://dx.doi.org/10.1177/1073191114552472>
- Ferrer-Perez, V. A., Ferreiro-Basurto, V., Navarro-Guzmán, C., & Bosch-Fiol, E. (2016). Batterer intervention programs in Spain: The professionals' perspective. *Psychosocial Intervention, 25*, 159–168. <http://dx.doi.org/10.1016/j.psi.2016.06.001>
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). *Bayesian data analysis (2)*. Boca Raton, FL, USA: Chapman & Hall/CRC.
- Gracia, E., Rodriguez, C. M., & Lila, M. (2015). Preliminary evaluation of an analog procedure to assess acceptability of intimate partner violence against women: The Partner Violence Acceptability Movie Task. *Frontiers in Psychology, 6*, 1567. <http://dx.doi.org/10.3389/fpsyg.2015.01567>
- Hartley, D. E., & Strupp, H. H. (1983). The therapeutic alliance: Its relationship to outcome in brief psychotherapy. In J. Masling (Ed.), *Empirical studies of psychoanalytical theories (1)* (pp. 1–38). Hillsdale, NJ: Analytical Press.
- Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology, 36*, 223–233. <http://dx.doi.org/10.1037/0022-0167.36.2.223>
- Horvath, A. O., & Symonds, B. D. (1991). Relation between working alliance and outcome in psychotherapy: A meta-analysis. *Journal of Counseling Psychology, 38*, 139–149. <http://dx.doi.org/10.1037/0022-0167.38.2.139>
- Lee, S. Y., & Song, X. Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research, 39*, 653–686. <http://dx.doi.org/10.1207/s15327906mbr3904.4>
- Lila, M., Gracia, E., & Catalá-Miñana, A. (2017). More likely to dropout, but what if they don't? Partner violence offenders with alcohol abuse problems completing batterer intervention programs. *Journal of Interpersonal Violence, 32*, 165–185. <http://dx.doi.org/10.1177/0886260517699952>
- Lila, M., Gracia, E., & Catalá-Miñana, A. (2018). Individualized Motivational Plans in batterer intervention programs: A randomized clinical trial. *Journal of Consulting and Clinical Psychology, 86*, 1–11. <http://dx.doi.org/10.1037/ccp0000291>
- Lila, M., Gracia, E., Catalá-Miñana, A., Santirso, F. A., & Romero-Martínez, Á. (2016). Alcohol abuse in Latin American immigrants participating in intervention programs for abusers: The importance of adherence to treatment. *Universitas Psychologica, 15*, 1–16. <http://dx.doi.org/10.11144/Javeriana.upsy15-4.caa1>
- Lila, M., Gracia, E., & Murgui, S. (2013). Psychological adjustment and victim-blaming among intimate partner violence offenders: The role of social support and stressful life events. *The European Journal of Psychology Applied to Legal Context, 5*, 147–153. <http://dx.doi.org/10.5093/ejpalc2013a4>
- Lila, M., Oliver, A., Catalá-Miñana, A., Galiana, L., & Gracia, E. (2014). The intimate partner violence responsibility attribution scale (IPVRAS). *The European Journal of Psychology Applied to Legal Context, 6*, 29–36. <http://dx.doi.org/10.5093/ejpalc2014a4>
- Llor-Esteban, B., García-Jiménez, J. J., Ruiz-Hernández, J. A., & Godoy-Fernández, C. (2016). Profile of partner aggressors as a function of risk of recidivism. *International Journal of Clinical and Health Psychology, 16*, 39–46. <http://dx.doi.org/10.1016/j.ijchp.2015.05.004>
- López-Ossorio, J. J., Álvarez, J. L. G., Pascual, S. B., García, L. F., & Buela-Casal, G. (2017). Risk factors related to intimate partner violence police recidivism in Spain. *International Journal of Clinical and Health Psychology, 17*, 107–119. <http://dx.doi.org/10.1016/j.ijchp.2016.12.001>
- López-Ossorio, J. J., González-Álvarez, J. L., & Andrés-Pueyo, A. (2016). Predictive effectiveness of the police risk assessment in intimate partner violence. *Psychosocial Intervention, 25*, 1–7. <http://dx.doi.org/10.1016/j.psi.2015.10.002>
- Luborsky, L., Crits-Christoph, P., Alexander, L., Margolis, M., & Cohen, M. (1983). Two helping alliance methods for predicting outcomes of psychotherapy. A counting signs vs. a global rating method. *The Journal of Nervous and Mental Disorders, 171*, 480–491.
- Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 68*, 438–450. <http://dx.doi.org/10.1037/0022-006X.68.3.438>
- Munder, T., Wilmers, F., Leonhart, R., Linster, H. W., & Barth, J. (2010). Working Alliance Inventory-Short Revised (WAI-SR): Psychometric properties in outpatients and inpatients. *Clinical Psychology and Psychotherapy, 17*, 231–239. <https://doi.org/10.1002/cpp.658>
- Musser, P. H., Semiatin, J. N., Taft, C. T., & Murphy, C. M. (2008). Motivational interviewing as a pregroup intervention for partner-violent men. *Violence and Victims, 23*, 539–557. <http://dx.doi.org/10.1891/0886-6708.23.5.539>
- Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods, 17*, 313–335. <https://doi.org/10.1037/a0026802>
- Muthén, L. K., & Muthén, B. O. (2010). *Mplus User's Guide*. USA, Los Angeles: Muthén & Muthén.
- Prochaska, J. O., DiClemente, C. C., & Norcross, J. C. (1992). In search of how people change: Applications to addictive behaviors. *American Psychologist, 47*, 102–114. <http://dx.doi.org/10.1037/0003-066X.47.9.1102>
- Romero-Martínez, Á., Lila, M., & Moya-Albiol, L. (2016). Testosterone and attention deficits as possible mechanisms underlying impaired emotion recognition in intimate partner violence perpetrators. *The European Journal of Psychology Applied to Legal Context, 8*, 57–62. <http://dx.doi.org/10.1016/j.ejpalc.2016.01.001>
- Scott, K. L. (2004). Stage of change as a predictor of attrition among men in a batterer treatment program. *Journal of Family Violence, 19*, 37–47. <http://dx.doi.org/10.1023/B:JOFV.0000011581.01231.1e>
- Semiatin, J., Murphy, C., & Elliot, J. (2013). Observed behavior during group treatment for partner-violent men: Acceptance of responsibility and promotion of change. *Psychology of Violence, 3*, 126–139. <http://dx.doi.org/10.1037/a0029846>
- Strunk, D. R., Brotman, M. A., & DeRubeis, R. J. (2010). The process of change in cognitive therapy for depression: Predictors of early inter-session symptom gains. *Behaviour Research and Therapy, 48*, 599–606. <http://dx.doi.org/10.1016/j.brat.2010.03.011>
- Taft, C. T., Murphy, C. M., Musser, P. H., & Remington, N. A. (2004). Personality, interpersonal, and motivational predictors of the working alliance in group cognitive-behavioral therapy for partner violent men. *Journal of Consulting and Clinical Psychology, 72*, 349–354. <http://dx.doi.org/10.1037/0022-006X.72.2.349>
- Tichenor, V., & Hill, C. E. (1989). A comparison of six measures of working alliance. *Psychotherapy, 26*, 195–199. <https://doi.org/10.1037/h0085419>
- Tracey, T., & Kokotovic, A. M. (1989). Factor structure of the Working Alliance Inventory. *Psychological Assessment: A*

- Journal of Consulting and Clinical Psychology*, 1, 207–210. <http://dx.doi.org/10.1037/1040-3590.1.3.207>
- Vargas, V., Lila, M., & Catalá-Miñana, A. (2015). Do cultural differences influence batterer intervention program outcomes? A study with Spanish and Latin American offenders. *Psychosocial Intervention*, 24, 41–47. <http://dx.doi.org/10.1016/j.psi.2015.03.001>