VOL 12. NÚMERO 1. ENERO 2021. 44 - 57.

ISSN: 2171-2069
https://doi.org/10.23923/j.rips.2021.01.044

REVISTA IBEROAMERICANA DE PSICOLOGÍA Y SALUD

Revista Oficial de la Federación Iberoamericana de Asociaciones de Psicología (FIAP)
[Official Journal of the Latin-American Federation of Psychological Associations]

Estimating the efficacy of Fptsd Scale to report malingering of PTSD: A meta-analytic review

Yurena Gancedo, Jéssica Sanmarco, Verónica Marcos y Dolores Seijo*.
Psicología Organizacional, Jurídica Forense y Metodología de las Ciencias del Comportamiento.
Facultad de Psicología. Universidad de Santiago de Compostela

• Recibido: 09 - 12 - 2020 • Aceptado: 02 - 02 - 2021 • Avance online: 29 - 01 - 2021

ABSTRACT. The assessment of psychological harm to crime victims has become a regular demand on forensic psychologists. Such evaluation requires not only the diagnosis of psychological harm, but also the differential diagnosis of malingering. For the malingering differential diagnosis of psychological harm, the MMPI-2 has the *Fptsd* scale. However, there is controversy as to the usefulness of this scale. As for this, a random effects meta-analytical review of sizes of experiments corrected by sampling error and criterion unreliability was designed. Eight primary studies were selected from which 12 effect sizes were extracted. The results showed that the *Fptsd* scale was significant and highly sensitive to the malingering of psychological harm in general ($d = 1.51/\delta = 2.08$) and in all the conditions studied (i.e. experimental sample, type of design and contrast group) and of great practical usefulness in the detection and classification of malingering. In addition, it revealed specificity (discriminant validity) before the malingering of psychological damage. The implications for forensic practice of the results, limitations on generalization and increased validity on the *Fp* scale are discussed.

KEYWORDS: MMPI-2, Malingering, *Fptsd* scale, *Fp* scale, Meta-analysis, Posttraumatic Stress Disorder.

Estimación de la eficacia de la escala Fptsd para detectar simulación de TEP: Una revisión meta-analítica

RESUMEN. La evaluación del daño psicológico de las víctimas de delitos se ha convertido en una demanda habitual a los psicólogos/as forenses. Dicha evaluación requiere no sólo del diagnóstico de daño psicológico, sino también del diagnóstico diferencial de simulación. Para el diagnóstico diferencial de simulación de daño psicológico, el MMPI-2 dispone de la escala *Fptsd*. No obstante, existe controversia respecto a la utilidad de esta escala. Por ello diseñamos una revisión meta-analítica de tamaños de efectos aleatorios de experimentos corregidos por el error de muestreo y la falta de fiabilidad del criterio. Se seleccionaron 8 estudios primarios de los que se extrajeron 12 tamaños del efecto. Los resultados mostraron que la escala *Fptsd* era significativa y altamente sensible a la simulación de daño psicológico en general ($d = 1.51/\delta = 2.08$) y en todas las condiciones estudiadas (i.e. muestra experimental, tipo de diseño y grupo de contraste) y de una gran utilidad práctica en la detección y clasificación de la simulación. Además, reveló especificidad ante la simulación de daño psicológico. Se discuten las implicaciones para la práctica forense de los resultados, las limitaciones en la generalización y la validez incrementada sobre la escala *Fp*.

PALABRAS CLAVE: MMPI-2, Simulación, Escala *Fptsd*, Escala *Fp*, Meta-análisis, Trastorno de Estrés Postratumático.

Crimes committed against persons, by their very definition, require a victim, such that without it there is no case (Arce, 2017). In 1985, the General Assembly of the United Nations adopted, by the resolution 40/34, the Declaration of Basic Principles of Justice for Victims of Crime and Abuse of Power defining a victim as a person "who, individually or collectively, have suffered harm, including physical or mental injury, emotional suffering, economic loss or substantial impairment of their fundamental rights, through acts or omissions that are in violation of criminal laws operative within Member States, including

*Correspondencia: Dolores Seijo.
Facultad de Psicología, Campus Vida, s/n
Dirección: 15782. Santiago de Compostela. Spain.
E-mail: mariadolores.seijo@usc.es

©2021 Sociedad Universitaria de Investigación en Psicología y Salud. Publicado por Consejo General de Colegios Oficiales de Psicólogos, España. Este es un artículo Open Access bajo la CC BY-NC-ND licencia (http://creativecommons.org/licencias/by-nc-nd/4.0/).

Citar como/Cite as: Gancedo, Y., Sanmarco, J., Marcos, V. y Seijo, D. (2021). **Estimating the efficacy of Fptsd Scale to report malingering of PTSD: A meta-analytic review.** *Revista Iberoamericana de Psicología y Salud, 12*(1), 44-57. Doi: https://doi.org/10.23923/j.rips.2021.01.044

those laws proscribing criminal abuse of power". Thus, a victim is defined by victimization of physical or psychological damage, economical loss and/or violation of his/her fundamental rights. Given that the burden of proof rests with the prosecution, the assessment of psychological damage becomes on special relevant, being the evidence most demanded from the courts and prosecution to forensic psychologists (Arce, 2018).

The scientific-forensic and clinical literature have determined different disorders as damage. Thus, the disorder that has been identified with the fingerprint of the victimization of a crime is post-traumatic stress disorder (PTSD; Kessler et al., 1995, 2005), both direct and indirect victimization (Gallego et al., 2019; Marcos et al., 2020); o not being a forensic print of damage to other disorders (Brown et al., 2001), nor subsyndromes (O'Donnell et al., 2006). To this disorder, Arce (2018) added Adjustment Disorder when criminal action is not included as Criterion A of PTSD (e.g., psychological violence) and the Acute Stress Disorder when criteria of PTSD are met within 1 month of the traumatic event and resolve within that I-month. On the other hand, in clinical setting other diagnoses should be given instead of, or in addition to PTSD, when criteria were met (American Psychiatric Association [APA], 2000, 2013). In any case, the clinical diagnosis of PTSD that was transferred to the Court required a malingering differential diagnosis (APA, 2000), being also a lawsuit (Arce, 2018). In fact, the clinic diagnosis cannot guarantee either that the cause (Criterion A) is the one referred by the patient, or the reality of the event. However, forensic evidence must comply with the principle of presumption of innocence (i.e., the innocent must *always* be protected from unfounded sentences, while it is sufficient that the guilty be *generally* punished; Supreme Court [Sentence of the Spanish Supreme Court] 213/2002). Therefore, forensic evidence is required to have a zero rate of false positives –that is, to determine that there is psychological harm resulting from the facts under investigation, when this is not the case (e.g., Supreme Court [Sentence of the Spanish Supreme Court] 1029/1997). Consequently, the forensic diagnosis is required in court to ensure that causal link between the act under investigation for each symptom, as well as the reality of the event (Arce, 2018). As for this, no single tool is sufficient, but it is necessary to adopt a multi-method approach that combines interview and psychometric instrumentation (Graham, 2011; Greene, 2011; Rogers, 2008).

With regard to the interview, it has been proven that the standard clinical interview (i.e., SCID) is not valid, as it facilitates the malingering of symptomatology by asking directly about the presence of symptoms, so that the interviewee only responds whether if they are present or not; and, moreover, it has not measures for the control of the malingering (Arce, 2018). For this reason, specific tools for these functions have been developed within the forensic field. Between them, the Structured Inventory of Reported Symptoms (SIRS; Rogers et al., 1992), as well as the Forensic-Clinical Interview (Arce & Fariña, 2001) stand out. However, the first one presents a problem: it does not allow to establish a causal link between the investigated facts (crime) and each symptom. Besides, it seems specific to criminal insanity evaluation, but not as much harm. As for the second, it has proven effective both in the diagnosis of psychological harm and in establishing the causal link between investigated facts and the symptoms and in controlling malingering, but it does not correctly classify all malingered protocols (Arce et al., 2002, 2006, 2009; Fariña et al., 2014; Vilariño et al., 2013).

On the other hand, the most commonly used psychometric instrument within the forensic field is the MMPI-2 and its restructured form, the MMPI-2-RF (Archer et al., 2006; Graham, 2011; Greene, 2011; Rogers et al., 2003; Sharf et al., 2017). The MMPI-2 allows an overall evaluation of the forensic task: the psychological harm assessment (PTSD), comorbid clinical disorders with PTSD and malingering differential diagnosis. However, no psychometric instrument establishes the causal link between facts and symptoms and therefore they are insufficient forensic evidence on their own (Arce, 2018). On the other hand, the MMPI-2-RF evaluates comorbid clinical disorders and malingering differential diagnosis, but not psychological damage. This better adjustment to the forensic task, the greatest persistence in time, the availability of more scales (*F, K, Fb, F-K, Fp,*

Fs, RBS, O-S, Ds, Ds-r, Obvious, Subtle, FBS, LW, Fptsd), of scientific evidence validating a decision criterion admissible in court and malingering measurement indexes (Arce et al., 2002, 2006, 2009, 2015; Fariña et al., 2014); together with the fact that the revised MMPI-2-RF scales (*F-r, K-r, Fp-r, Fs, FBS-r, RBS*) have not been shown more sensitive to malingering (validity scales are shorter versions of their MMPI-2 namesakes; Arce, 2018; Greene, 2011), have led to the continued use of MMPI-2 than the RF version.

For the detection of psychological damage malingering (PTSD), Elhai et al. (2002) developed the *Fptsd* scale (Infrequency-Posttraumatic Stress Disorder). This scale consists of 32 items, which war veterans with this disorder score with low frequency (< 20%). This scale was highly correlated with *F* ($r = .59$) and *Fb* ($r = .51$), and especially with *Fp* ($r = .81$), sharing 65.6% of the variance. Although in the original seminal study, the *Fptsd* scale showed incremental validity over *F, Fb* and *Fp* scales (Elhai et al., 2002), subsequent studies of mean differences (not case studies as it is the target of forensic practice) found that effect sizes for the discriminatory between honest and malingering *Fptsd* responses did not exceed those of *F, Fb* and/or *Fp*; that Fptsd did not systematically increase validity over *F, Fb* or *Fp* (Efendov, 2006; Marshall & Bagby, 2006); or that it did so with very little difference (Arbisi et al., 2006). However, these studies did not consider whether the *F, Fb* or *Fp* scales were being sensitive to PTSD malingering or whether malingers had adopted it as an indiscriminate symptom endorsement malingering strategy. In these cases, *F, Fb* and *Fb* scales increase validity as they have as target the measurement of general malingering, not specific of PTSD. Thus, this systematic source of error has not been controlled, so the variance may be attributable to the method than the construct (Podsakoff et al., 2003).

In this context, a meta-analytical review to know the sensitivity of the *Fptsd* scale to malinger was performed. As for this, the following hypotheses were tested (Arce, 2018):

1) The *Fptsd* Scale is sensitive to malingering of psychological harm (PTSD) in any condition.

2) The *Fptsd* Scale is more sensitive to simulation research studies so in differential prevalence and known-groups designs.

3) The *Fptsd* Scale is more sensitive to malinger in mentally healthy population (students, community sample) than in population where harm is known (veterans).

4) The *Fptsd* Scale is more sensitive to malinger when compared to normative population than to a known-harm population (veterans or patients/litigants).

5) The *Fptsd* Scale is less sensitive (specificity) to the simulation detection of other disorders than PTSD (discriminant validity).

METHOD

• SEARCH FOR STUDIES

The strategy used in the literature search was to find all studies in which the effectiveness of the MMPI-2 *Fptsd* scale in detecting malingering was measured. To this end, at first, those previous meta-analytical reviews were identified that studied the validity scales, not finding that included the scale subject to this study. Subsequently, the descriptors were chosen from which to identify the studies using that scale, selecting *Fptsd scale, Fp scale, Fp-r scale, malingering, faking bad, simulation, MMPI-2, MMPI-2-Rf*. These descriptors were combined with search algorithms performing scans on the research databases the reference in psychology, Web of Science, Scopus and PsycInfo; of doctoral theses, communications and conference proceedings, and ephemeral works, Proquest; as well as the Google Scholar meta-search engine. The following selection criteria were applied to the study bank obtained: 1) to measure the effectiveness of the MMPI-2 *Fptsd* scale in the detection of malingering; and 2) to provide the effect size or the data needed to calculate it. Following the application of these criteria, 8 primary studies were selected, obtaining from these a total of 12 effect sizes (see Annex I). The study search flowchart is shown in Figure 1.

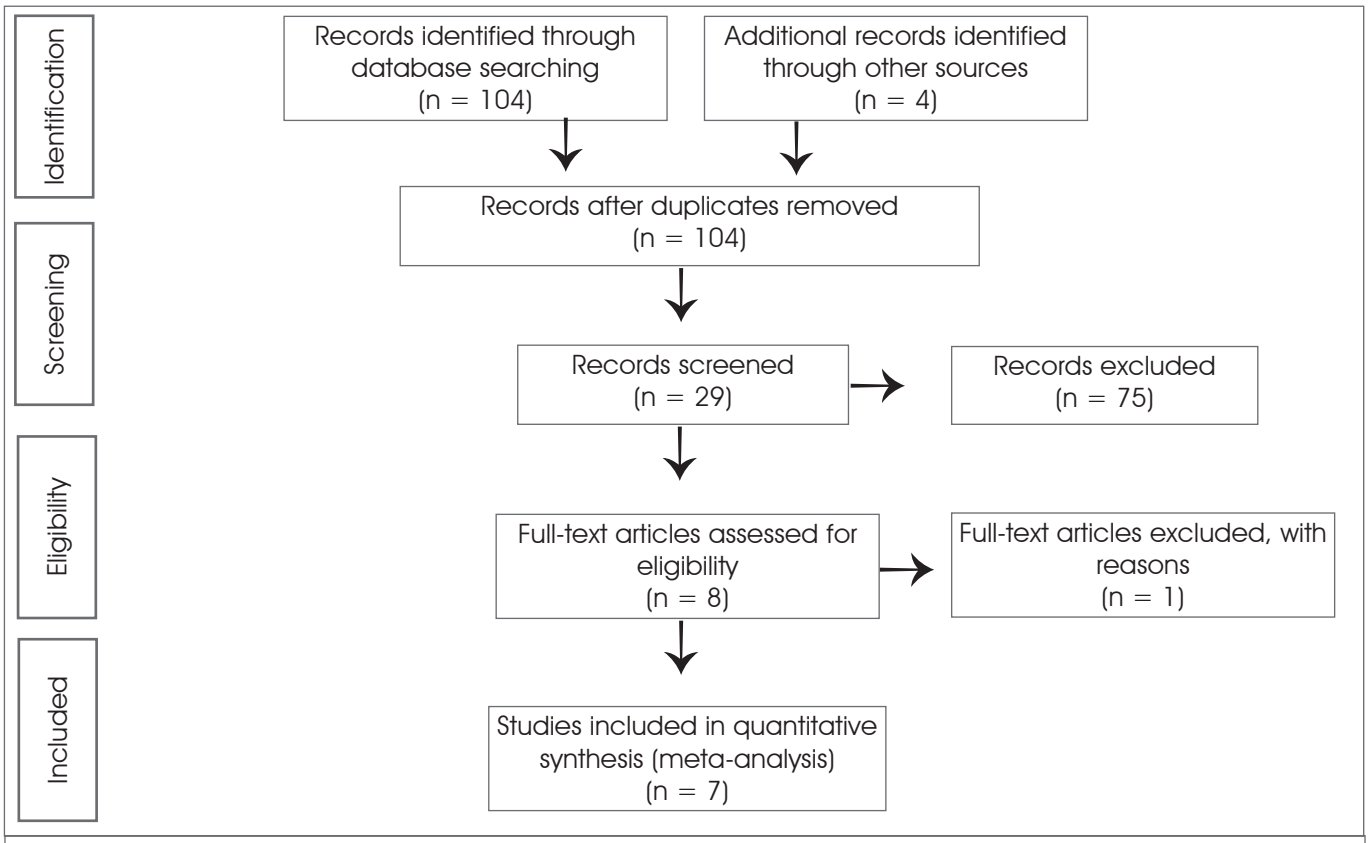

Figure 1: Flow diagram of the meta-analysis.

•CODING OF PRIMARY STUDIES

The studies were codified in: a) article reference; b) design characteristics (i.e., design type, simple type used, malingering instructions); c) characteristics of the sample (i.e., sample size, sex, age, area of origin); d) mean and standard deviation of the groups compared or, failing that, the data needed to calculate the effect size; and e) effect size. Two trained and experienced raters separately evaluated the primary studies in the referred variables. As for the study of the agreement for nominal scales, kappa coefficient is usually adopted, which corrects the random effect in concordance. Nevertheless, this index is subject to a source of error: it does not verify the exact correspondence of the codings, such that two errors (the coding of one category by the two raters in different places) are coded as one hit (Arce et al., 2000). This Kappa setting, which verifies the exact correspondence in the codings, named it true kappa ($\bar{k}$). The agreement observed in this study between-raters was total, $\bar{k} = 1$. In addition, these raters were consistent in other contexts (i.e., studies) and with other raters (i.e., contexts; Arce et al., 2020). Thus, in contrast the concordance between- and within-raters, as well as with other raters (studies and contexts), raters are stable in time (test-retest) and between raters, that is, the same and other raters would had coded the studies equally in the analyzed variables, which indicates that the coding is consistent (reliable) and valid (evaluated the coded variables).

•DATA ANALYSIS

The effect size was taken directly from the primary studies when they provided it in *d* or was transformed to this when it was with another index, and, if not provided, d was calculated with Cohen's (when $N1 = N2$ and for the comparison with a test value), Hedges' (when $N1 \neq N2$) or Glass' (when the standard deviations differ) formula with the means and standard deviations of the malingering and honest responding (control group) groups. With this information, excel spreadsheets were created so that the calculations were accurate (the correct operation was verified by comparing it with manual execution).

Meta-analysis of random effects correcting the effect size by sampling error and the unreliability criterion were performed (Schmidt & Hunter, 2015). Succinctly, the following statistics were calculated: the effect size weighted for the sampling error (*dw*); the standard deviation of *d* (SD_d); the standard deviation of *d* predicted by artifactual errors (SD_{pre}); the standard deviation of *d*, after removal of variance due to artifactual errors (SD_{res}); the mean true effect size, corrected for criterion unreliability (δ); the standard deviation of δ (SD_δ); the variance accounted by artifactual errors (%Var); the 95% confidence interval for *d* (95% CI_d); and the 80% credibility interval for δ (80% CI_δ). The magnitude of the effect size was interpreted in terms of the probability superiority of the effect size (PS_{ES}; Monteiro et al., 2018).

Though effect sizes and their magnitude are valuable for deriving the implications for forensic practice, it was complemented with the q statistic for the comparison of the effect size between two conditions (Amado et al., 2015); and with the study of cases as the forensic task has as target *N* of 1 trial i. e., the U1 statistic (Cohen, 1988), Effect Incremental Index (EII; Redondo et al., 2019), and the probability of an inferiority score (PIS; Arias et al., 2020).

•CRITERION RELIABILITY

The reliability of the *Fptsd* scale was taken from Elhai et al. (2002), $\alpha = .53$.

RESULTS

•STUDY OF OUTLIERS

No extreme [±3*IQR], nor outliers [±1.5*IQR] values were observed. Additionally, all the primary studies were inside the Chauvenet's criterion (± 2*SD*).

•GENERAL META-ANALYSIS OF THE EFFICACY OF THE *FPSTD* SCALE TO DISCRIMINATE BETWEEN MALINGERING AND HONEST RESPONDING

The results of the general meta-analysis for the *Ftpsd* Scale (see Table 1) revealed a significant (when the confidence interval has no zero, indicating the effect size was significant), positive (higher scores among malingering protocols in comparison with honest responding), generalizable (the credibility interval has no zero), and more than large ($\delta > 1.20$) mean true effect size ($\delta = 1.83$). This confirms hypothesis 1: the *Fptsd* scale is highly sensitive to detection of malingering of PTSD. In practical terms, the magnitude of the observed effect is greater than 90.15% of all possible effect sizes ($PS_{ES} = .9015$), with the exact capacity of discrimination between honest and malingering responding populations being 78.03% (U1 = .7803), the probability that the malingering responding population will get a score lower (error) than the population mean of honest responding is 3.36% (PIS = .0336), and the increase in effect size due to the correct classification of the malingering is of 95.69% (EII = .9569). In addition, the *Fptsd* scale has been shown to be significantly sensitive to malinger under all conditions studied (see Table 1), thus confirming the first hypothesis. Nevertheless, the percentage of variance explained by the artifactual errors is lower than 75% (75% rule: if artifactual variance explains less than 75%, moderators mediated the results i.e., heterogeneous data; Hunter et al., 1982), advertising of heterogeneity in primary studies. Thus, the results are mediated by moderators. As moderators that could be extracted from primary studies we identified the type of experimental design, the control group and the type of population instructed to malinger.

•TYPE OF STUDY DESIGN

The results of the meta-analysis for the *Fptsd* Scale for studies that use a simulation research design (see Table 1) displayed a significant, positive, generalizable, and more than large mean true effect size ($\delta = 2.08$). In practical terms, the magnitude of the effect size was greater than 92.92% of all possibilities ($PS_{ES} = .9292$), with the exact capacity for discrimination between honest and malingering responding populations being 82.47% (U1 = .8247), the probability that the population of malingering responding get a score lower (error) to the population mean of honest responding is 1.87% (PIS = .0187) and the increase in the effect size due to the correct classification of malingering is of 97.73% (EII = .9773).

Table 1
Meta-Analyses for the *Fptsd* Scale

K	*N*	d_w	SD_d	SD_{pre}	SD_{res}	δ	SD_δ	%Var	5% CI_d	80% CI_δ
TOTAL										
10	2314	1.33	0.5291	0.1457	0.5087	1.83	0.6987	7.58	1.24, 1.42	0.94, 2.72
SIMULATION RESEARCH DESIGN										
9	1937	1.51	0.3575	0.1551	0.3221	2.08	0.4424	18.81	1.31, 1.51	1.51, 2.65
MALINGERING SAMPLE: VETERANS (Oversimulation)										
2	325	1.29	0.8054	0.1731	0.7866	1.77	1.0805	4.62	1.05, 1.53	0.39, 3.15
MALINGERING SAMPLE: COMMUNITY SAMPLE/STUDENTS (Healthy people)										
6	207	3.20	0.7826	0.5193	0.5854	4.40	0.8041	44.04	2.79, 3.61	3.37, 5.43
COMPARISON WITH VETERANS										
3	1468	1.33	0.5532	0.1000	0.5441	1.83	0.7473	3.27	1.22, 1.44	0.87, 2.79
COMPARISON WITH LITIGANTS/PATIENTS										
5	672	1.34	0.4533	0.1914	0.4109	1.85	0.5645	17.83	1.17, 1.51	1.12, 2.58
COMPARISON WITH THE NORMATIVE SAMPLE										
8	532	2.03	1.2288	0.3044	1.1905	2.79	1.6353	6.14	1.82, 2.24	0.70, 4.89

Note. *k* = number of effect sizes; *N* = total sample size; d_w = sample size weighted mean effect size; SD_d = standard deviation of d; SD_{pre} = standard deviation predicted for sampling error alone; SD_{res} = standard deviation of *d* after removing sampling error variance; δ = mean true effect size; SD_δ = the standard deviation of δ; %Var = percent of observed variance accounted by artifactual errors; 95% CI_d = 95% confidence interval for *d*; 80% CI_δ = 80% credibility interval for δ.

With respect to the designs of known groups (*k* = 0) and differential prevalence (*k* = 1), meta-analyses could not be performed due to the scarcity of studies with both types of designs. However, with differential prevalence designs it was found a study with a positive and significant effect size, *d* = 1.01[0.79, 1.22], which corrected by criterion unreliability resulted in δ = 1.39[1.16, 1.62], being the effect size significantly lower than that obtained with simulation designs (*d* = 1.51), q_s = 0.261, *p* < .01. Consequently, hypothesis 2 is confirmed in relation to the increased sensitivity to malinger of the *Fptsd* scale in simulation designs than in differential prevalence designs, as long as it could not be compared in known group designs.

• MALINGERING POPULATION TYPE

The results of the meta-analysis for veterans as malingerers (oversimulation i.e., exaggeration of genuine harm) showed (see Table 1) a significant, positive, generalizable, and more than large mean true effect size (δ = 1.29). In practical terms, the magnitude of the observed effect is greater than 81.86% of all possible effect sizes (PS_{ES} = .8186), with the exact capacity for discrimination between honest

and malingering responding populations being 64.96% (U1 = .6496), the probability that the malingering responding population will get a lower score (error) than the population mean of honest responding is 9.85% (PIS = .0985), and the increase in effect size due to the correct classification of malingering is of 84.83% (EII = .8483).

The results of the meta-analysis for samples from community and students (healthy samples) instructed to malinger (see Table 1) revealed a significant, positive, generalizable, and more than large mean true effect size (δ = 4.40). In practical terms, the magnitude of the observed effect is greater than 100% of all possible effect sizes (PS_{ES} = 1.00), with the exact ability of discrimination between honest and malingering responding populations being 98.59% (U1 = .9859), the probability that the malingering responding population will get a lower score (error) than the average of the population of honest responding is < .000005 (PIS = .000005) and the increase in the size of the effect due to the correct classification of the malingering is of 99.99% (EII = .9999).

The sensitivity of the *Fptsd* scale is significantly higher in healthy samples (students, community sample; δ = 4.40) than in known harm samples (veterans, oversimulation; δ = 1.29), q_s =0.922, $p < .001$, supporting the third hypothesis.

•CONTROL GROUP

The results of the meta-analysis for studies that use veterans as comparison group (see Table 1) showed a significant, positive, generalizable, and more than large mean true effect size (δ = 1.83). In practical terms, the magnitude of the observed effect is greater than 90.15% of all possible effect sizes (PS_{ES} = .9015), with the exact ability for discrimination between honest and malingering responding populations being 78.03% (U1 = .7803), the probability that the malingering responding population will get a lower score (error) than the population mean of honest responding is 3.36% (PIS = .0336), and the increase in effect size due to the correct classification of malingering is of 95.69% (EII = .9569).

The results of the meta-analysis for studies that use litigants/patients as comparison group (see Table 1) exhibited a significant, positive, generalizable, and more than large mean true effect size (δ = 1.85). In practical terms, the magnitude of the observed effect is greater than 90.49% of all possible effect sizes (PS_{ES} = .9049), with the exact ability for discrimination between honest and malingering responding populations being 78.42% (U1 = .7842), the probability that the malingering responding population will get a lower score (error) than the population mean of honest responding of 3.22% (PIS = .0322), and the increase in effect size due to the correct classification of malingering is of 95.89% (EII = .9589).

Finally, the results of the meta-analysis for studies that use the normative sample as comparison group (see Table 1) revealed a significant, positive, generalizable, and more than large mean true effect size (δ = 2.79). In practical terms, the magnitude of the observed effect is greater than 97.56% of all possible effect sizes (PS_{ES} = .9756), with the exact ability for discrimination between honest and malingering responding populations being 91.13% (U1 = .9113), the probability that the malingering responding population will get a lower score(error) than the population mean of honest responding of 0.26% (PIS = .0026), and the increase in effect size due to the correct classification of malingering is of 99.71% (EII = .9971).

The comparison of the observed effects about the sensitivity to malingering of the *Fptsd* scale taking the normative sample as contrastive group was significantly higher than when a sample of veterans, q_s = 0.315, $p < .01$, or a sample of patients/litigants, q_s = 0.308, $p < .05$, were the contrastive group, confirming the fourth hypothesis.

•STUDY OF THE SPECIFICITY

The *Fptsd* scale has been shown to be totally insensitive to the malingering of depression, d = -0.22[δ = 0.30] (Lange et al., 2010), and, although with a positive, significant and medium effect size (d = 0.51[δ =0.70]; Whitney et al., 2008), an effect size larger than 64.06%

of all possible effect sizes (PS_{ES} = .6406), is significantly less sensitive, q_s = 0.362, p < .01, in malingering detection of neurocognitive dysfunctions than in PTSD malingering (d = 1.31). Thus, the results ratify the fifth hypothesis giving support to a discriminant validity to the *Fptsd* scale of PTSD from other disorders.

DISCUSSION

The results of meta-analytic studies of the Fptsd scale confirm sensitivity to malingering of psychological harm in any condition: samples of veterans, students or patients; simulation and differential prevalence designs; and contrast with normative population or population with known-PTSD (veterans or patients/litigants). In this way, it correctly detects classifies malingering (whole simulation) and oversimulation (exaggeration of genuine harm). The magnitude of sensitivity is extraordinarily high in any condition: the overall effect is larger than 90% of all possible and in oversimulation greater than 80%. In addition, it is sensitive to the specificity of PTSD malingering (discriminant validity).

However, these results are subject to limitations in their generalization. First, the effects are due almost exclusively (9 out of 10 sizes) to simulation designs, which have a high internal validity, but with weak external validity (Rogers, 2018). It has been found that simulation designs in this research setting to produce significantly different results than field studies (i.e., known and differential prevalence groups) and that participants follow different response strategies in simulation studies than in field studies (Fariña et al., 1994). In this particular case, malingering participants in simulation studies would follow a strategy of success maximization (i.e., harm regardless of the consequences of being detected as a malingerer), while malingering participants in field study conditions would adopt the strategy of combining harm malingering with detection minimization (Fariña et al., 1994). Consequently, the sensitivity of the scale in simulation designs is very oversized (ceiling effect; thus, the observed effect sizes are so high that they are impossible), meanwhile with genuine malingerers would decrease very significantly.

In fact, while studies with strong external validity (i.e., known group comparisons) were not developed, in the only study (Tolin et al., 2010) with differential prevalence design (moderate external validity) the effect size drops to d = 0.39, supporting this prediction. As the power of the study (1-β) is high, .94, and counterbalancing type I and II errors ($\alpha/\beta \approx 1$), certain stability of this result is expected. Paradoxically, Rogers (2018) concludes that such designs should not be employed, despite its empirical usefulness (i.e., in groups where malingering is suspected, the scale detects it even though not the entire population simulates). Therefore, the observed effect is lessened (floor effect) compared to the true one. This result gives it a double sensitivity to the scale. On the one hand, it detects malingering and, on the other hand, it discriminates against non-malingering in contexts of suspected malingering. In turn, and in line with expectations, the observed effect is 3.87 times smaller, OR = 3.87, and significantly lower, q_s = 0.503, p < .05, than that registered in studies with simulation design. Only the differential prevalence design allows testing this double sensitivity, i.e., the ability to discriminate. If not significant effect of the scale was found with differential prevalence designs and significantly less than studies with simulation design, the scale would be invalid. Only with this type of designs there is the statistical certainty (the registered prevalence is significantly higher than in the population where malingering is not suspected) of the classification of genuine malingerers. Therefore, the differential prevalence contrast is more valid when the population of genuine cases is contrasted (in this case, PTSD). However, in the differential prevalence groups of malingering there are also genuine cases classified as malingerers (error). Second, only with known-PSTD groups designs can be quantified the rate of false positives, that is, genuine cases of PTSD classified by the scale as malingerers. In fact, the family of MMPI-2 *F* scales, including the *Fptsd* scale, were created on the basis of selecting items infrequently endorsed by the normative sample or clinical cases (in *Fptsd* scale, PTSD combat veteran sample). But genuine cases also endorse these items. For example, the rate of inaccurate (Fp > 3) of MMPI-2 profiles (two highest elevated

clinical scales at a T score of 65 or higher) ranges from 15.61% (1-9/9-1) to 64.25 (6-8/8-6) (Greene, 2011). PTSD comorbidity or multi-comorbidity ranges from 80 to 98.8% (APA, 2013; Brady et al., 2000; Kessler et al., 1995), being strongly related to severity (Kessler et al., 2005; Vilariño et al., 2018). Thus, the expected probability of genuine PTSD cases classified by the scale as malingerers is high and should be known to correct this source of error. Therefore, the exact sensitivity cannot be estimated with a single type of design, but all three are required, since the discrimination index between true positives (malingerers classified as such) and false positives (genuine patients classified as malingerers) requires all three results (Monteiro et al., 2018). Third, the results of the meta-analysis in certain conditions may be subject to a degree of variability given that *Ns* < 400 or *k* ≤ 3 is no guarantee of the stability of sampling estimates (Schmidt & Hunter, 2015). Thus, more studies are needed to guarantee the stability of sampling estimates. Fourth, it has been taken as a reference for the creation of the scale, veteran patients diagnosed with PTSD, when the judicial demand for malingering differential diagnosis for this population is negligible (in clinical setting malingering is not diagnosed, is suspected; the structured clinical interview [SCID] has not a measure of malingering; and the clinical criteria of suspicion of malingering are ineffective; APA, 2000, 2013; Arce, 2018; Rogers, 2008; Rogers & Vitacco, 2002), while the main judicial claim is as evidence of harm to support victimization in criminal cases (Arce et al., 2009; United Nations, 1985) or damage in civil cases (e.g., compensations for motor vehicle accidents) (Arce et al., 2006; Blanchard & Hickling, 2004). That is why the validity of the scale must be verified in samples of crime and accident victims. Fifth, the samples of participants in simulation conditions are almost exclusively male (only 13% were women, mainly students, and with only male studies), as were those of veterans who served to create the scale. However, in the judicial context, women are evaluated more than men as they are more likely to be victimized and gender differences have been found in PTSD (Lehavot et al., 2018; Street & Dardis, 2018). Therefore, studies with a female population are needed because there are expected to be differences between men and women in the development of PTSD and in the malingering differential diagnosis (APA, 2000, 2013). Given that a high correlation between the *Fptsd* scale and the *Fp* scale and, by extension, its revised *Fp-r* version, has been observed and that it has been hypothesized that the first one does not increase validity compared to the second ones, the effect sizes of the meta-analytical reviews of the MMPI-2 and MMPI-2-RF in PTSD malingering with those obtained in the present study were compared. For the *Fp* scale Rogers et al. (2003) reported an effect size for the comparison between PTSD malingerers and patients with PTSD, $d = 1.22[0.73, 1.15]$), equal, $q_s = 0.050$, *ns*, to that obtained with the *Fptsd* scale ($d = 1.34$ [1.17, 1.51]); while Sharf et al. (2017) for the *Fp-r* scale of an effect size, $d = 0.94$ [0.66, 1.21]), significantly lower, $q_s = 0.174$, $p < .05$, than that found in this study. However, the results of the meta-analyses by Rogers et al. are miscalculated. Thus, for the *Fp* scale the reported effect size is 1.22 ($N = 392$), but there are not only two studies with contrast of malingers of PTSD with genuine patients with PTSD ($N = 352$). For these two, the arithmetic mean of the effect sizes would be 1.215, which is what they calculated, while the corrected meta-analysis for sampling error would be 1.172 (random effects) and for inverse variance method (fixed effects) 1.169. Likewise, there is no effect size on the Fp-r scale in the primary studies lower than 0.94, so the result of the meta-analysis by Sharf et al. (2017) cannot be $d = 0.94$.

In any case, this controversy is of no interest because, simply as a statistical issue, the *Fptsd* scale increases validity over *Fp* and *Fp-r*. The scale construction mode is the same: infrequently endorsed items (< 20%) by genuine patients. The *Fp* and *Fptsd* scales share 20 items to which *Fptsd* adds 12 items from PTSD populations, that is, it adds more measure (validity), while *Fp* (27 items) adds 4 items of defensive measure (noise), which relates to genuine cases, but not to malinger (Fariña et al., 2014; Garrido-Macías et al., 2020), and 3 of populations with other disorders (not among those selected with PTSD patients), and *Fp-r* (21 items endorsed infrequently by psychiatric inpatients of any disorder, eliminates items related to defensiveness) with 17items

shared with *Fp* i.e., *Fp-r* is a short version of *Fp*. In view of this, it is worth asking: can a measure with fewer items be more valid since this implies less validity because with each item a part of the measure is lost? Is a measure adjusted to the measurement object (genuine patients of PTSD) more valid than a generalist measure (genuine patients of any disorder)? In any case, this controversy is irrelevant to forensic practice because the differential diagnosis of malingering requires the combination of multiple criteria (Arce, 2018). So that because the principle of presumption of innocence that carries the burden of proof requires the probability of a false positive (classification of malingered PTSD as genuine) be 0. Thus, all the classification criteria of malingering must be combined resulting in a decision rule that makes the false positive rate was zero (e.g., Arce et al., 2015; Fariña et al., 2014).

• **Conflict of interest.**

The authors declare no conflict of interest.

• **Funding.**

Funding: This research has been sponsored by a grant of the Spanish Ministry of Economy, Industry and Competitiveness (PSI2017-87278-R), and by a grant of Xunta de Galicia (GPC), in the context of the project "Revisión meta-analítica de los instrumentos forenses de evaluación del daño psicológico [Meta-analytic review of forensic psychometric instrument used to evaluate psychological damage]".

REFERENCES

References marked with an asterisk indicate studies included in the meta-analysis.

Amado, B. G., Arce, R., & Herraiz, A. (2015). Psychological injury in victims of child sexual abuse: A meta-analytic review. *Psychosocial Intervention, 24,* 49-62. http://dx.doi.org/10.1016/j.psi.2015.03.002

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th ed. text. rev.) Washington, DC: American Psychiatric Association.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: American Psychiatric Association.

*Arbisi, P. A., Ben-Porath, Y. S., & McNulty, J. (2006). The ability of the MMPI-2 to detect feigned PTSD within the context of compensation seeking. *Psychological Services, 3*(4), 249-261. https://doi.org/10.1037/1541-1559.3.4.249

Arce, R. (2017). Análisis de contenido de las declaraciones de testigos: Evaluación de la validez científica y judicial de la hipótesis y la prueba forense [Content analysis of the witness statements: Evaluation of the scientific and judicial validity of the hypothesis and the forensic proof]. *Acción Psicológica, 14*(2), 171-190. https://doi.org/10.5944/ap.14.2.21347

Arce, R. (2018). Evaluación del daño psicológico: Psicometría, entrevista y técnica forense [Assessment of psychological harm: Psychometry, interview and forensic technique]. In E. Carbonell, D. Pineda & M. Novo (Eds.), *Psicología jurídica: Ciencia y profesión* (pp. 235-244). Santiago de Compostela, Spain: Sociedad Española de Psicología Jurídica y Forense.

Arce, R., Arias, E., Novo, M., & Fariña, F. (2020). Are interventions with batterers effective? A meta-analytical review. *Psychosocial Intervention, 29*(3), 153-164. https://doi.org/10.5093/pi2020a11

Arce, R., & Fariña, F. (2001). *Construcción y validación de un procedimiento basado en una tarea de conocimiento para la medida de la huella psíquica en víctimas de delitos: La entrevista clínico-forense* [Construction and validation of an interview based-knowledge task for the measure of the psychological harm in victims of crimes: The forensic-clinical interview]. Santiago de Compostela, Spain: Universidad de Santiago de Compostela.

Arce, R., Fariña, F., Carballal, A., & Novo, M. (2006). Evaluación del daño moral en accidentes de tráfico: Desarrollo y validación de un protocolo para la detección de la simulación [Evaluating psychological injury in motor vehicle accidents (MVA): Development and validation of a protocol for detecting simulation]. *Psicothema, 18*(2), 278-283. Retrieved from http://www.psicothema.com/psicothema.asp?id=3210

Arce, R., Fariña, F., Carballal, A., & Novo, M. (2009). Creación y validación de un protocolo de evaluación forense de las secuelas psicológicas de la violencia de género [Creation and validation of a forensic protocol to assess psychological harm in battered women]. *Psicothema, 21*(2), 241-247. Retrieved from http://www.psicothema.com/pdf/3621.pdf

Arce, R., Fariña, F., & Fraga, A. (2000). Género y formación de juicios en un caso de violación [Gender and juror judgment making in a case of rape]. *Psicothema, 12*(4), 623-628. Retrieved from http://www.psicothema.com/pdf/381.pdf

Arce, R., Fariña, F., & Vilariño, M. (2015). Daño psicológico en casos de víctimas de violencia de género: estudio comparativo de las evaluaciones forenses [Psychological injury in intimate partner violence cases: A contrastive analysis of forensic measures]. *Revista Iberoamericana de Psicología y Salud, 6*(2), 72-80. https://doi.org/10.1016/j.rips.2015.04.002

Arce, R., Pampillón, M. C., & Fariña, F. (2002). Desarrollo y evaluación de un procedimiento empírico para la detección de la simulación de enajenación mental en el contexto legal [Development and validation of an empirical procedure to detect malingering of mental insanity in legal setting]. *Anuario de Psicología, 33*(3), 385-408. Retrieved from https://revistes.ub.edu/index.php/Anuario-psicologia/article/view/8769/10980b

Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment, 87*, 84-94. http://dx.doi.org/10.1207/s15327752jpa870107

Arias, E., Arce, R., Vázquez, M. J., & Marcos, V. (2020). Treatment efficacy on the cognitive competence of convicted intimate partner violence offenders. *Anales de Psicología/Annals of Psychology, 36*(3), 427-435. https://doi.org/10.6018/analesps.428771

Blanchard, E. B., & Hickling, H. J. (2004). What are the psychosocial effects of MVAs on survivors? In E. B. Blanchard & E. J. Hickling (Eds.), *After the crash: Psychological assessment and treatment of survivors of motor vehicle accidents* (2nd ed., pp. 57-97). Washington, DC: American Psychological Association.

Brady, K. T., Killeen, T. K., Brewerton, T. & Lucerini, S. (2000). Comorbidity of psychiatric disorders and post traumatic stress disorder. *Journal of Clinical Psychiatry, 61*(suppl 7), 22-32.

Brown, T. A., Campbell, L. A., Lehman, C. L., Grisham, J. R., & Mancill, R. B. (2001). Current and lifetime comorbidity of the DSM-IV anxiety and mood disorders in a large clinical sample. *Journal of Abnormal Psychology, 110,* 585-599. http://dx.doi.org/10.1037//0021-843X.110.4.585

Cohen, J. (1988). *Statistical power analysis for behavioral sciences* (2nd ed.). Hillsdale, NJ: LEA.

*Efendov, A. A. (2007). *Incremental predictive validity of the TSI, MMPI-2, and SIMS in identifying coached and uncoached feigning of PTSD in workplace accident victims* (Doctoral thesis. University of Toronto, Canada). Retrieved from https://search.proquest.com/psycinfo/docview/622026284/

*Elhai, J. D., Naifeh, J. A., Zucker, I. S., Gold, S. N., Deitsch, S. E., & Frueh, B. C. (2004). Discriminating malingered from genuine civilian posttraumatic stress disorder: A validation of three MMPI-2 infrequency scales (F, Fp, and Fptsd). *Assessment, 11*(2), 139-144. https://doi.org/10.1177/1073191104264965

*Elhai, J. D., Ruggiero, K. J., Frueh, B. C., Beckham, J. C., Gold, P. B., & Feldman, M. E. (2002). The Infrequency-Posttraumatic Stress Disorder Scale (Fptsd) for the MMPI-2: Development and initial validation with veterans presenting with combat-related PTSD. *Journal of Personality Assessment, 79*(3), 531-549. https://doi.org/10.1207/S15327752JPA7903_08

Fariña, F., Arce, R., & Real, S. (1994). Ruedas de identificación: De la simulación y la realidad

[Lineups: A comparision of high fidelity research and research in a real context]. *Psicothema, 6*(3), 395-402. Retrieved from http://www.psicothema.com/pdf/935.pdf

Fariña, F., Arce, R., Vilariño, M., & Novo, M. (2014). Assessment of the standard forensic procedure for the evaluation of psychological injury in intimate-partner violence. *Spanish Journal of Psychology, 17,* e32. http://dx.doi.org/10.1017/sjp.2014.30

Gallego, R., Novo, M., Fariña, F., & Arce, R. (2019). Child-to-parent violence and parent-to-child violence: A meta-analytic review. *European Journal of Psychology Applied to Legal Context, 11*(2), 51-59. http://dx.doi.org/10.5093/ejpalc2019a4

Garrido-Macías, M., Valor-Segura, I., & Expósito, F. (2020). Which tactics of sexual violence predict leaving the relationship? The role of dependence towards partner. *European Journal of Psychology Applied to Legal Context, 12*(2), 53-60. https://doi.org/10.5093/ejpalc2020a6

Graham, J. R. (2011). *MMPI-2: Assessing personality and psychopathology* (5th ed.). New York, NY: Oxford University Press.

Greene, R. L. (2011). *The MMPI-2-MMPI-2-RF: An interpretive manual* (3rd ed.). Boston, MA: Allyn & Bacon.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies*. Beverly Hills, CA: Sage.

Kessler, R. C., Chiu, W. T., Demler, O., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry, 62*(6), 617-627. https://doi.org/10.1001/archpsyc.62.6.617

Kessler, R. C., Sonnega, A., Hughes, M. & Nelson, C. B. (1995). Posttraumatic stress disorder in the national comorbidity survey. *Archives of General Psychiatry, 52,* 1048-1060. https://doi.org/10.1001/archpsyc.1995.03950240066012

*Lange, R. T., Sullivan, K. A., & Scott, C. (2010). Comparison of MMPI-2 and PAI validity indicators to detect feigned depression and PTSD symptom reporting. *Psychiatry Research, 176*(2-3), 229-235. https://doi.org/10.1016/j.psychres.2009.03.004

Lehavot, K., Katon, J. G., Chen, J. A., Fortney, J. C., & Simpson, T. L. (2018). Post-traumatic stress disorder by gender and veteran status. *American Journal of Preventive Medicine, 54*(1), e1-e9. https://doi.org/10.1016/j.amepre.2017.09.008

Marcos, V., Gancedo, Y., Castro, B., & Selaya, A. (2020). Victimización de violencia entre novios, gravedad percibida en las conductas de violencia entre novios, mitos del amor romántico, y dependencia emocional entre chicos y chicas adolescentes [Dating violence victimization, perceived gravity in dating violence behaviors, sexism, romantic love myths and emotional dependence between female and male adolescents]. *Revista Iberoamericana de Psicología y Salud, 11*(2), 132-145. https://doi.org/10.23923/j.rips.2020.02.040

*Marshal, M. B., & Bagby, R. M. (2006). The incremental validity and clinical utility of the MMPI-2 Infrequency Posttraumatic Stress Disorder Scale. *Assessment, 13*(4), 417-429. https://doi.org/10.1177/1073191106290842

Monteiro, A., Vázquez, M. J., Seijo, D., & Arce, R. (2018). ¿Son los criterios de realidad válidos para clasificar y discernir entre memorias de hechos auto-experimentados y de eventos vistos en vídeo? [Are the reality criteria valid to classify and to discriminate between memories of self-experienced events and memories of video-observed events?]. *Revista Iberoamericana de Psicología y Salud, 9*(2), 149-160. https://doi.org/10.23923/j.rips.2018.02.020

O'Donnell, M. L., Creamer, M. Bryant, R. A., Schnyder, U., & Shalev, A. (2006). Posttraumatic disorders following injury: Assessment and other methodological considerations. In G. Young, A. W. Kane & K. Nicholson (Eds.), *Psychological knowledge in courts: PTSD, pain and TBI* (pp. 70-84). New York, NY: Springer.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method bias in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879-903. http://dx.doi.org/10.1037/0021-9010.88.5.879

Redondo, L., Fariña, F., Seijo, D., Novo, M., & Arce, R. (2019). A meta-analytical review of the responses in the MMPI-2/MMPI-2-RF clinical and restructured scales of parents in child custody dispute. *Anales de Psicología, 35*(1) 156-165. http://dx.doi.org/10.6018/analesps.35.1.338381

Rogers, R. (2008). Structured interview and dissimulation. *In R. Rogers (Ed.), Clinical assessment of malingering and deception* (3rd ed., pp. 301-322). New York, NY: The Guilford Press.

Rogers, R. (2018). Researching response styles. In R. Rogers & S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (4th ed., pp. 592-614). New York; NY: Guilford Press.

Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *The SIRS test manual*. Tampa, FL: Psychological Assessment Resources.

Rogers, R., Sewell, K. W., Martin, M. A., & Vitacco, M. J. (2003). Detection of feigned mental disorders. A meta-analysis of the MMPI-2 and malingering. *Assessment, 10*(2), 160-177. http://dx.doi.org/10.1177/1073191103252349

Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting errors and bias in research findings* (3rd ed.). Thousand Oaks, CA: Sage.

Sharf, A. J., Rogers, R., Williams, M. M., & Henry, S. A. (2017). The effectiveness of the MMPI-2-RF in detecting feigned mental disorders and cognitive deficits: A meta-analysis. *Journal of Psychopathology and Behavioral Assessment, 39*(3), 441-455. https://doi.org/10.1007/s10862-017-9590-1

Street, A. E., & Dardis, C. M. (2018). Using a social construction of gender lens to understand gender differences in posttraumatic stress disorder. *Clinical Psychology Review, 66*, 97-105. https://doi.org/10.1016/j.cpr.2018.03.001

*Tolin, D. F., Steenkamp, M. M., Marx, B. P., & Litz, B. T. (2010). Detecting symptom exaggeration in combat veterans using the MMPI–2 symptom validity scales: A mixed group validation. *Psychological Assessment, 22*(4), 729-736. https://doi.org/10.1037/a0020973

Tribunal Supremo (Sala de lo Penal, Sección 2ª). *Sentencia Nº 1029/1997 de 29 de septiembre, 1997*.

Tribunal Supremo (Sala de lo Penal, Sección 2ª). *Sentencia Nº 213/2002 de 14 de febrero, 2002*.

United Nations (1985). *Declaration of basic principles of justice for victims of crime and abuse of power.* New York, NY: United Nations. Retrieved from https://www.un.org/en/genocideprevention/documents/atrocity-crimes/Doc.29_declaration%20victims%20crime%20and%20abuse%20of%20power.pdf

Vilariño, M., Amado, B. G., Vázquez, M. J., & Arce, R. (2018). Psychological harm in women victims of intimate partner violence: Epidemiology and quantification of injury in mental health markers. *Psychosocial Intervention, 27*(3), 145-152. https://doi.org/10.5093/pi2018a23

Vilariño, M., Arce, R., & Fariña, F. (2013). Forensic-clinical interview: Reliability and validity for the evaluation of psychological injury. *European Journal of Psychology Applied to Legal Context, 5*(1), 1-21. Retrieved from https://journals.copmadrid.org/ejpalc/archivos/articulo20180219092235.pdf

*Whitney, K. A., Davis, J. J., Shepard, P. H., & Herman, S. M. (2008). Utility of the Response Bias Scale (RBS) and other MMPI-2 validity scales in predicting TOMM performance. *Archives of Clinical Neuropsychology, 23*(7-8), 777-786. https://doi.org/10.1016/j.acn.2008.09.001

Appendix 1. Primary Studies

Author(s) and publication year	*n*	Type of design	Sample	Source	Effect size: *d*
Arbisi et al. (2006)	SG: 35 CG: 55	SR	Veterans	Paper	1.63
Efendov (2006)	SG1: 29 SG2: 27 SG3: 31 CG: 84	SR	SG: general population CG: PTSD patients-litigants	Doctoral thesis	SG1: 2,07 SG2: 1,73 SG3: 0,73
Elhai et al. (2002)	SG: 61 CG: 940	SR	SG: students CG: veterans	Paper	1.66
Elhai et al. (2004)	SG: 39 CG: 41	SR	SG: students CG: PTSD patients	Paper	0.82
Lange et al. (2010)	SG: 14 CG: 20	SR	Students	Paper	0.11
Marshall & Babgy (2006)	SG: 67 CG1: 73 CG2: 186	SR	SG: students CG1: students CG2: PTSD patients	Paper	CG1: 1.60 CG2: 1.32
Tolin et al. (2010)	SG: 290 CG: 87	DP	Veterans	Paper	0.39

Note. SG: Simulation group; CG: Comparison group; SR: Simulation research; DP: Differential Prevalence.