The European Journal of Psychology Applied to Legal Context 9 (2017) 51–56

The European Journal of Psychology Applied to Legal Context

www.elsevier.es/ejpal

Contrasting the efficacy of the MMPI-2-RF overreporting scales in the detection of malingering

Guadalupe Sánchez [a], Amada Ampudia [b], Fernando Jiménez [a,*], Bárbara G. Amado [c]

[a] Universidad de Salamanca, Spain
[b] Universidad Nacional Autónoma de México, UNAM, Mexico
[c] Universidad de Santiago de Compsotela, Spain

ARTICLE INFO

Article history:
Received 12 January 2016
Accepted 24 March 2017
Available online 23 May 2017

Keywords:
MMPI-2-RF
Overreporting
Malingering
Validity scales
Forensic assessment

ABSTRACT

Though it has been the most extensively used instrument for forensic evaluation, the MMPI-2 is being gradually replaced by the MMPI-2-RF version, requiring evidence research to support it. A malingering design was implemented to assess the efficacy of the overreporting validity scales in discriminating between a group of malingerers and the general and clinical populations in a forensic context. Of a total of 878 subjects, 309 were from the general population, 308 from the clinical population, and 261 were instructed to malinger a psychological injury. The results showed that malingerers scored significantly higher than the clinical and general population on the *F-r*, *Fp-r*, *FBS-r*, *Fs* and *RBS* scales. As for the classification of cases, the *F-r*, *Fp-r*, *FBS-r*, *Fs*, and *RBS* scales classified correctly and significantly between malingerers and honest respondents from the general population, and the *F-r* and *Fp-r* scales between malingerers and clinical population. Additionally, the results showed *F-r* incremental validity over *Fp-r*, and vice versa. Thus, *F-r* and *Fp-r* scales are independent and may be accumulated to detect malingering. Forensic practical implications from the results were derived and discussed.

Contraste de la eficacia de las escalas de validez del MMPI-2-RF en la detección de la simulación

Palabras clave:
MMPI-2-RF
Exageración
Simulación
Escalas de validez
Evaluación forense

RESUMEN

Aunque el MMPI-2 ha sido el instrumento psicométrico más usado en la evaluación forense, está siendo reemplazado gradualmente por la versión reestructurada, el MMPI-2-RF precisándose de más evidencia científica para ello. Se utilizó un diseño de investigación de simulación para evaluar la eficacia de las escalas de validez de evaluación de la simulación en la discriminación entre simuladores y las poblaciones general y clínica en el contexto forense. Participaron en el estudio 878 sujetos, 309 de la población general, 308 casos clínicos y 261 instruidos para simular daño psicológico. Los resultados mostraron que los simuladores puntuaban significativamente más alto que los sujetos de las poblaciones general y clínica en las escalas *F-r*, *Fp-r*, *FBS-r*, *Fs* y *RBS*. En la clasificación de casos, las escalas *F-r*, *Fp-r*, *FBS-r*, *Fs* y *RBS* clasificaban correcta y significativamente entre simuladores y respuestas honestas de la población general, y las escalas *F-r* and *Fp-r* entre simuladores y población clínica. Además, los resultados evidenciaron validez incrementada de *F-r* sobre *Fp-r* y viceversa. Se discuten las implicaciones para la práctica forense de los resultados.

* Corresponding author: Facultad de Psicología. Dpto. de Personalidad, Evaluación y Tratamiento Psicológicos. Universidad de Salamanca. Avda. de la Merced, 109. 37005 Salamanca, Spain.
E-mail address: fjimenez@usal.es (F. Jiménez).

http://dx.doi.org/10.1016/j.ejpal.2017.03.002

Malingering is defined by the American Psychiatric Association -APA (2013) as "the intentional production of physical or psychological symptoms disproportionate or false, motivated by external incentives . . ." (p. 726). In the field of mental health, this translates into the reporting of psychiatric symptoms, cognitive disorders, and a combination of both (Pierson & Rosenfeld, 2015). The global prevalence of malingering is estimated to range from 10 to 20%, with a $15 \pm 15\%$ ratio for clinical contexts, and $40 \pm 10\%$ for forensic settings (Young, 2015). This phenomenon conditions psychological practice and must be controlled owing to the high socio-health costs involved (Chafetz & Underhill, 2013), as well as having serious legal implications in forensic evaluations, e.g., a guilty verdict for an innocent defendant (Fariña, Arce, Vilariño, & Novo, 2014). Whereas malingering is undoubtedly of interest for clinical practice, its assessment and control are indispensable mandatory requirements in forensic contexts (Arce, Fariña, & Vilariño, 2015). The conclusions of forensic reports have legal consequences that may prompt to those being evaluated to fake their symptomology. Thus, the APA (2013) asserts in the DSM-5 that malingering should be suspected in evaluations in medical-legal contexts. In contrast, cases of malingering are seldom described in clinical practice for the simple fact that clinicians do not suspect it (Rogers, 2008), that is, priority is given to therapeutic outcomes as opposed to scrutinising the veracity of symptoms.

Hence, forensic evaluation pursues a twofold objective: to measure an individual's clinical status, and to establish a differential diagnosis of malingering (Osuna, López-Martínez, Arce, & Vázquez, 2015). To achieve both objectives, a multi-method and multi-measure technique combining interviews with a psychometric measure, the MMPI being the leading psychometric instrument worldwide (Ben-Porath, 2013; Graham, 2011; Greene, 2011; McDermott, 2012), must be employed. The MMPI integrates sets of personality and clinical factors that have proven to be useful through time, and have been updated on several occasions in order to incorporate the most recent findings in psychopathology. Moreover, the MMPI consists of a series of scales and validity indexes that have shown to be effective in detecting the malingering of symptoms (Ingram & Ternes, 2016; Rogers, Sewell, Martin, & Vitacco, 2003). In short, the MMPI performs the double function of complying with the forensic standard of assessing both clinical status and malingering.

The most recent version of the instrument applied to adults, the MMPI-2-RF (Ben-Porath & Tellegen, 2008/2011), is the restructured form of the MMPI-2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) that was widely used in both forensic contexts (Arce, Fariña, Carballal, & Novo, 2006; Fariña et al., 2014; Nelson, Hoelzle, Sweet, Arbisi, & Demakis, 2010; Rogers et al., 2003; Wolf & Miller, 2014), and clinical practice (Graham, 2011; Jiménez & Sánchez, 2002; Jiménez, Sánchez, & Tobón, 2009; Rogers, 2008). In turn, the MMPI-2 consists of a re-standardization of the original inventory, the MMPI (Hathaway & McKinley, 1940). The new version, the MMPI-2-RF, is shorter (338 vs. 567 items) and contains fewer scales (42 vs. 162). As for the validity of the protocols, the *F-r*, *Fp-r*, *L-r*, *K*-r, *VRIN-r*, *TRIN-r*, and *FBS-r* scales were revised, including scales previously outlined on the MMPI-2, but not available in the commercial version (e.g., *RBS*, *Fs*), whereas other productive scales (i.e., *Fb*, *Ds*, *Wsd*, *Od*, *S*) have been eliminated (Sánchez, Jiménez, Novo, & Silva, 2012). As the scales were modified and eliminated, derived in lack of utility of the composed indexes and configurations (e.g., *F-K*, *L + K*, *F-Fb*). Regarding the clinical scales, the MMPI-2 contains standard clinical scales illustrated in the clinical profile of the commercial version. Moreover, the restructured clinical scales are drawn from the MMPI-2, but are not provided in the commercial version. The MMPI-2-RF, however, only contains the restructured clinical scales. In any case, administering the MMPI-2 is the same as administering the MMPI-2-RF given that the MMPI-2 contains all of the items on the on the latter.

A recent meta-analysis has found that the MMPI-2-RF overreporting validity scales significantly discriminated between honest respondents and malingerers with large effect sizes (Hedges's *g*), ranging from 1.04 for the *FBS-r* Scale to 1.43 for the *Fp-r* Scale (Ingram & Ternes, 2016). Moreover, the evaluation context was found to be a moderator, of which one is the litigant. Nevertheless, these results are subject to considerable variability given that Ns (<400) and/or k (≤ 3) do not guarantee of the stability of sampling estimates (Hunter & Schmidt, 2015). Thus, further studies are required. As for the underreporting validity scales, the MMPI-2-RF is clearly inferior to the MMPI-2 as it has eliminated the *Wsd*, *Od*, *Mp*, *S*, *PMH4* and *Esd* scales, and the *F-K*, *L + K* and *L + K-F* indexes (Arce, Fariña, Seijo, & Novo, 2015; Fariña, Redondo, Seijo, Novo, & Arce, 2017).

Bearing in mind these observations, a malingering design was conducted to compare responses on the MMPI-2-RF under standard instructions (two samples: clinical and general population) and under malingering instructions in a forensic setting (instructions to malinger a psychological injury) and to assess the discriminating capacity (true effect) of the MMPI-2-RF overreporting validity scales for evaluations in forensic setting.

Method

Participants

A total of 878 subjects participated in the study, age range 19 to 69 years ($M = 31.37$, $SD = 11.19$), who were divided into 3 groups: general population consisting of 309 subjects, 163 men (52.75%) and 146 women (47.25%), mean age 32.95 years ($SD = 12.03$); clinical population with 308 participants, 148 men (48.05%) and 160 women (51.95%), mean age 33.74 years ($SD = 11.36$); and malingerers with 261 subjects, 95 men (36.40%) and 166 women (63.60%), mean age 26.70 years ($SD = 8.21$).

Measurement Instrument

The adapted Spanish version of the MMPI-2 (Hathaway & McKinley, 1999) was applied containing items from the MMPI-2-RF (338 items), with equivalent scores being obtained on either version (Ben-Porath & Tellegen, 2008/2011). The restructured clinical scales, and the overreporting scales: Infrequent Responses (*F-r*), Infrequent Psychopathology Responses (*Fp-r*), Infrequent Somatic Responses (*Fs*), Fake Bad Scale, also known as Symptom Validity (*FBS-r*), and Response Bias Scale (*RBS*) were obtained from the MMPI-2-RF.

The *F-r* Scale is made up of 32 items designed to detect unusual or infrequent responses in the normative population. As a matter of fact, 10% of the normative population responded to these items in the deviant direction. High scores indicate overreporting of a large variety of psychological, cognitive and somatic symptoms. In terms of the reliability of this scale, a Cronbach's α of .818, .863, and .926 was found for the sample of the general population, the clinical population, and malingerers, respectively.

The *Fp-r* Scale analyses infrequent responses by psychiatric inpatient samples throughout 21 items. An elevated score indicates an individual's attempts at self-unfavourable reporting and exaggerated psychopathology. A Cronbach's α of .262, .474, and .651 was obtained for the general population, the clinical population, and the sample of malingerers, respectively.

The *Fs* Scale is composed of 16 somatic content items which are infrequent in medical patient populations. Reporting a wide number of atypical somatic symptoms could be a clue of

malingering. Cronbach's alphas of .697, .709, and .902 were obtained for the sample of the general population, the clinical sample and malingerers, respectively.

The *FBS-r* Scale was designed to be applied more in forensic context than in clinical settings. This scale is made up of 31 items which define somatic and cognitive symptoms that are rarely reported by personal-injury claimants; therefore, a high level of symptoms is associated with over-reporting. The Cronbach's alpha revealed a reliability of .709, .788, and .866 for the general population sample, the clinical sample, and malingerers, respectively.

The *RBS* Scale consists of 28 items that measure over-reporting as an unusual mixed of responses associated with non-credible memory complaints. In short, this scale assesses exaggeration of cognitive dysfunctions. A Cronbach's α value of .609, .767, and .830 was found for the general population sample, the clinical sample, and malingerers, respectively.

Validity scale cut-offs to classify protocols as malingerers were ≥ 7 for *Fp-r* and > 17 for *F-r* (Ben-Porath, 2013), > 16 for *RBS* (Wygant et al., 2010), and ≥ 6 for *Fs* and ≥ 21 for *FBS-r* (Schroeder et al., 2012).

Design and Procedure

A malingering design was implemented to compare responses to the MMPI-2-RF under standard instructions (two samples: clinical and general population), and under malingering instructions. Prior to data analysis, the protocols were screened to detect highly inconsistent responding either due to extreme acquiescence (TRIN raw score ≥18 or T ≥80); random responding (VRIN raw score ≥18 or T ≥ 80); or an extremely high number of non-responding or double response items (i.e., unwillingness to cooperate in the evaluation) equal to or greater than 30, or outliers (*L* raw score >10, and *K* raw score >26), to eliminate these from the study (Arce, Fariña, Seijo, & Novo, 2015; Graham, 2011; Greene, 2011). Under these circumstances, all cases were considered as valid for the study. Participants from the general population were assigned at random to responding under standard instructions or under malingering instructions. The clinical sample was taken randomly from patients of mental health outpatient services. The diagnostic rates were: anxiety disorder (20%), schizophrenia spectrum (16%), substance use disorder (12%), conduct disorder (7%), depressive disorder (12%) somatic symptom disorder (5%), feeding and eating disorders (7%), adjustment disorders (5%), mood disorder (7%), and other mental disorders (9%).

Malingerers were instructed to fake bad a psychological injury, with the aim of getting judicial incentives in order to avoid a criminal prosecution, to obtain financial compensation, and/or to seek revenge. Instructions were written to be easily understandable. No training was provided to malingerers. Nevertheless, malingerers were instructed to prepare consciously the subsequent psychological assessment. A screening to control the engagement with malingering instructions was performed to confront the ability to fake (Fariña et al., 2014); thus, all malingerers simulated at least one clinical diagnosis (T >70) on the MMPI-2-RF.

Data Analysis

Although the correlation between validity scales may be due to the evidence that malingerers were using a combination of malingering strategies, it may also be a consequence of an overlapping of the scales (in fact, share items) and of a measure duplicity. To contrast this, the correlation between scales was computed.

One factor ANOVAs were performed for the comparison of the means between groups (i.e., general population, clinical sample, and malingerers) in the overreporting validity scales of the MMPI-2-RF. Post hoc analysis were performed with the Bonferroni correction (.05/2 = .025).

Table 1
Pearson correlations between the overreporting scales for the total sample.

Scale	*Fs*	*FBS-r*	*F-r*	*Fp-r*	*RBS-r*
Fs	1				
FBS-r	.821***	1			
F-r	.915***	.819***	1		
Fp-r	.835***	.637***	.860***	1	
RBS-r	.878***	.855***	.914***	.755***	1

Note. *Fs*: Infrequent Somatic Responses; *FBS-r*: Symptom Validity; *F-r*: Infrequent Responses; *Fp-r:* Infrequent Psychopathology Responses; *RBS*: Response Bias Scale.
*** $p < .001$.

Accuracy classification of the MMPI-2-RF overreporting scales was estimated with sensitivity, specificity, diagnostic odds ratio (DOR), and the Area Under the Curve (AUC). DORs were better estimators than negative and positive predictive power as these vary according to the base ratio, that is unknown for the MMPI-2-RF validity scales (Fariña et al., 2014), while the DORs do not (Glas, Lijmer, Prins, Bonsel, & Bossuyt, 2003).

Undoubtedly, a comparative analysis of measures provides valuable data for drawing evidence-based conclusions that have practical implications ($N = 1$ designs), which should be complemented with an analysis of case studies (APA, 2013), particularly in forensic contexts (Amado, Arce, Fariña, & Vilariño, 2016). As for the analysis of the ability to classify overreporting cases between populations, and the incremental validity, binary logistic and multinomial regression were performed. Finally, the cumulative classification of the scales was estimated to derive forensic judgement criteria (Arce et al., 2006).

Results

Correlations between Overreporting Validity Scales

The results of the correlation analysis (see Table 1) between scales revealed a high significant and very high correlations, generally above .80. Thus, the scales were either measuring the same malingering strategy or subjects were combining strategies.

Comparison of Means between Populations

The comparison of means showed statistically significant differences between populations (see Table 2) on all the validity scales analysed. Post hoc analysis revealed malingerers scored significantly higher on all of the overreporting scales, i.e., *F-r*, *Fp-r*, *Fs*, *FBS-r*, and *RBS*, than the general and clinical population, in all of the comparisons a large effect size ($g > .80$) was observed, and the ability to discriminate malingering from honest responding was significantly (see CI overlapping at Table 2) higher in the general than in the clinical population.

Classification Accuracy

The classification rate of the malingering validity scales (see Table 3) was sensitive to discriminating between malingerers and the general population with values ranging from 43.68% on the *FBS-r* Scale to 96.17% on the *Fp-r* scale. Comparatively (see CIs overlapping at Table 3), the *Fp-r* Scale was significantly more sensitive to malingering than the other scales, but it was also significantly less specific. The *FBS-r* and *RBS* scales were significantly less sensitive to malingering than the *F-r*, *Fp-r*, and *Fs* scales. However, the *FBS-r*, *RBS* and *F-r* scales were significantly more specific (> 90%) than the *Fp-r* and *Fs* scales. The between contexts comparisons (general population vs. malingerers, and clinical population vs. malingerers), showed both were sensitive to detecting malingering, whilst the *Fs* and *RBS* scales were more specific in discriminating between

Table 2
One-factor ANOVA for Mean Contrast of General, Clinical, and Malingering Populations.

Population	General		Clinical		Malingerers				
Scale	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F* (2, 875)	g_1 [95% CI]	g_2 [95% CI]
F-r	5.25	4.39	10.47	6.21	24.77	7.12	79.931***	3.36 [3.10, 3.61]	2.15 [1.94, 2.35]
Fp-r	5.42	1.58	6.16	2.27	13.92	2.96	1146.89***	3.65 [3.38, 3.92]	2.96 [2.72, 3.20]
Fs	1.81	2.12	3.87	2.86	11.33	4.57	684.40***	2.74 [2.51, 2.97]	1.99 [1.79, 2.19]
FBS-r	7.77	4.10	12.64	5.31	17.62	5.43	279.96***	2.07 [1.87, 2.27]	0.93 [0.76, 1.10]
RBS	7.21	3.33	11.13	4.84	18.28	4.92	454.59***	2.67 [2.39, 2.95]	1.46 [1.27, 1.65]

Note. g_1: Hedges' effect size for malinger vs. general comparison; g_2: Hedges' effect size for malinger vs. clinical comparison; 95% CI: 95% credibility interval.
*** $p < .001$.

Table 3
Classification Accuracy.

Population	Scale	AUC [95%CI]	SE_{AUC}	*Se* [95%CI]	*Sp* [95%CI]	DORs [95%CI]
General/Malingerers	*F-r*	.945 [.928, .962]	.009	84.29 [79.17, 88.37]	91.57 [89.02, 93.59]	58.30 [37.62, 90.34]
	Fp-r	.963 [.946, .979]	.008	96.17 [92.85, 98.04]	72.93 [69.21, 76.37]	67.63 [35.08, 130.40]
	Fs	.917 [.895, .940]	.012	84.67 [79.59, 88.71]	82.01 [78.70, 84.91]	25.19 [16.98, 37.37]
	FBS-r	.828 [.796, .859]	.016	43.68 [37.61, 49.94]	95.30 [93.24, 96.77]	15.72 [10.07, 24.56]
	RBS	.896 [.872, .920]	.012	69.35 [63.31, 74.81]	91.57 [89.02, 93.59]	24.58 [16.69, 36.21]
Clinical/Malingerers	*F-r*	.917 [.894, .941]	.012	84.29 [79.17, 88.37]	85.39 [80.83, 89.04]	31.36 [19.81, 49.65]
	Fp-r	.954 [.936, .973]	.009	96.17 [92.85, 98.04]	65.26 [59.62, 70.51]	47.15 [24.03, 92.53]
	Fs	.887 [.858, .916]	.015	84.67 [79.59, 88.71]	70.78 [65.30, 75.73]	13.38 [8.82, 20.30]
	FBS-r	.752 [.711, .794]	.021	43.68 [37.61, 49.94]	91.56 [87.73, 94.31]	8.41 [5.26, 13.46]
	RBS	.844 [.812, .877]	.017	69.35 [63.31, 74.81]	84.74 [80.12, 88.47]	12.56 [8.36, 18.88]

Note. *AUC*: Area Under the Curve; 95% CI: 95% confidence intervals; *Se*: Sensitivity; *Sp*: Specificity; SE_{AUC}: Area Under the Curve Standard Error; DORs: Diagnostic Odds Ratio.

Table 4
Contrast of the Capacity of the Overreporting Scales to Classify Malingering Cases between Populations.

Population	*B*	*SE*	*Wald*	*df*	*p*	*OR*	95% CI	
Scale							*LL*	*UL*
General Population								
RBS-Cases = 0	1.648	0.571	8.325	1	0.004	5.198	1.697	15.923
Fs-Cases = 0	1.135	0.382	8.821	1	0.003	3.111	1.471	6.581
FBS-r-Cases = 0	1.431	0.728	3.859	1	0.049	4.181	1.003	17.426
Fp-r-Cases = 0	2.758	0.397	48.200	1	0.000	15.763	7.237	34.336
F-r-Cases = 0	2.449	0.496	24.405	1	0.000	11.577	4.381	30.589
Clinical Sample								
RBS-Cases = 0	0.314	0.346	0.827	1	0.363	1.369	0.696	2.695
Fs-Cases = 0	0.004	0.331	0.000	1	0.991	1.004	0.524	1.922
FBS-r-Cases = 0	0.561	0.346	2.626	1	0.105	1.752	0.889	3.453
Fp-r-Cases = 0	2.551	0.384	44.070	1	0.000	12.814	6.035	27.210
F-r-Cases = 0	2.007	0.355	31.931	1	0.000	7.439	3.709	14.921

Note. Reference category: Population of malingerers; *df*: degrees of freedom, *LL*: lower limit; *UL*: upper limit.

malingerers and the general population (82.01% and 91.57% for the *Fs* and *RBS* scales, respectively) than between malingerers and the clinical population (70.78% and 84.74%).

In terms of diagnostic accuracy, DORs (the ratio between of the probability of a correct classification of malingering and the probability of an incorrect classification of malingering) ranged extensively from 8.41 to 67.63. Comparatively (see CIs overlapping at Table 3), the diagnostic accuracy of *F-r* was higher between malingerers and the general population than for the *Fs*, *FBS-r*, and *RBS*; and for the *Fp-r* Scale than for the *FBS-r* Scale. In the discrimination between malingerers and the clinical population *Fp-r* achieved significantly higher diagnostic accuracy than *Fs*, *FBS-r*, and *RBS*; and *F-r* than *FBS-r* or *RBS*. The scales maintained their diagnostic accuracy in the between contexts comparison (see CIs overlapping at Table 3).

The superiority probability (AUC), that is, higher scores on the scale for the malingering population than for honest responding, ranged from .752 on the *FBS-r* Scale for the clinical population to .963 on the *Fp-r* Scale for the general population. Comparatively (see CIs overlapping at Table 3), the probability of obtaining higher score for malingerers was significantly less on the *FBS-r* Scale than on the other scales in comparison to the general and clinical populations. The performance of the *FBS-r* Scale was significantly better in the general population than in the clinical population. Nevertheless, interpreting the AUC as an effect size, the magnitude in all of the scales was more than large (> .75).

Case Classification Analysis of the Overreporting Scales

As the explanatory hypothesis resulting from case classification of the overreporting scales may be malingering or severity distress (Ben-Porath & Tellegen, 2008/2011; Graham, 2011; Greene, 2011), a multinomial logistic regression was performed to discriminate case classification of the overreporting scales among populations, i.e., to inform of real malingering (general population vs. malingerers) and between malingering and clinical severity (clinical population vs. malingerers). The results (see Table 4) revealed that all of the scales significantly and correctly discriminated (the greater probability of classifying a population of malingerers as such) with a large effect size (OR > 4.25) for the *F-r*, *Fp-r*, and *RBS* scales; and a moderate effect size ($2.47 < OR < 4.25$) for the *Fs* and *FBS-r* scales between malingerers and the general population. Notwithstanding, the ability to classify cases between malingerers and the clinical population was significant and with large effect

Table 5
Incremental Validity from *F-r* to *Fp-r*, and from *Fp-r* to *F-r*.

Modelo	$\chi^2(df)$	*w*	$\chi^2(df)$	Δw
F-r/Fp-r				
Step 1: *F-r*	302.97(1)***	.59		
Step 2: *Fp-r*	366.57(2)***		63.60(1)***	.27
Fp-r/F-r				
Step 1: *Fp-r*	267.71(1)***	.55		
Step 2: *F-r*	366.57(2)***		98.86(1)***	.34

*** $p < .001$

Table 6
Accumulative Study.

	f	%	Incremental classification
General Population			
0	249	80.6	80.6 (true negative)
1	53	17.1	19.4 (false positive)
2	7	2.3	2.3 (false positive)
Clinical Sample			
0	191	62.0	62.0 (true negative)
1	82	26.6	38.0 (false positive)
2	35	11.4	11.4 (false positive)
Malingerers			
0	10	3.8	3.8 (false negative)
1	31	11.9	96.2 (true positive)
2	220	84.3	84.3 (true positive)

Note. Population = Malingerers; *f*: frequency.

sizes for the *F-r* and *Fp-r* scales, but not for the *Fs*, *FBS-r*, and *RBS* scales. As for the ability to discriminate on the *F-r* and *Fp-r* scales between malingerers and the general and clinical populations, the results were comparable. In short, only the *F-r* and *Fp-r* scales were valid (i.e., significant discrimination) in the classification of inter-population malingering.

Incremental Validity of the F-r *and* Fp-r *Scales*

As the *F-r and Fp-r* scales correctly and significantly classified malingerers from the general and clinical populations, and a high correlation was observed between them ($r = .860$), incremental validity was analysed to determine if there was concurrent validity of one over the other. The results (see Table 5) revealed the *F-r* scale significantly increased the classification of malingering as compared to the *Fp-r* Scale and vice versa; in other words, both added validity to the other. In consequence, both scales were independent.

Incremental Malingering Classification of the Valid Overreporting Scales

The accumulative analysis of the classification of malingering on the *F-r* and *Fp-r* scales (see Table 6) shows two indicators of malingering correctly classified 84.3% of malingerers (true positives), and 80.6% and 62.0% (true negatives) in the general population and clinical population, respectively, but erroneously classified 38.0% of clinical cases (at least on one scale malingering was erroneously classified as such) as false positives (it failed to discriminate between clinical severity and malingering), and in 19.4% of the general population; and 3.8% as false negatives (classification of malingering as honest responding).

Discussion

The generalization of the results of the present study is subject to several limitations that should be borne in mind. First, though care was taken to control implication in the task, subjects under malingering instructions do not perform the same task as malingerers in real-life forensic evaluation (Fariña, Arce, & Real, 1994; Konecni & Ebbesen, 1992). Second, the design was based on the assumption of general malingering, with an expected malingering context effect; in other words, it was conjectured that subjects would perform differently in malingering psychological injury to malingering mental insanity. Thus, the results are not directly generalizable to specific malingering contexts. Third, overreporting scales do not provide a differential diagnosis of malingering, but diagnostic impressions that require a multi-method approach. Fourth, the responses of subjects may be due to systematic bias owing to the tendency to report inexistent symptoms in the belief that it is important to do so for a specific reason (Greene, 2011). Fifth, the malingering hypothesis derived from the overreporting scales is compatible with other alternative hypothesis, severity distress being the most prominent in forensic assessment. These scales are insufficient for discriminating between both hypotheses and fail to meet the requirements of a forensic task.

Bearing in mind the above limitations, the following conclusions may be drawn. In line with the model, the results showed malingerers scored higher in all overreporting scales in comparison to the control groups, i.e., general and clinical population. Moreover, overreporting scales performed significantly better at discriminating malingerers from the general population than from the clinical population. The best discriminative capacity was for the *Fp-r* scale and for the *F-r* scale. The former discriminated significantly better than the *RBS*, *FBS-r*, *Fs* and *F-r* scales with the clinical population, and the *RBS*, *FBS-r* and *Fs* scales with the general population, whereas the latter performed significantly higher than the *RBS* and *FBS-r* scales with the clinical population, and the *RBS*, *FBS-r* and *Fs* scales with the general population. Furthermore, *F-r* and *Fp-r* scales diagnosed significantly better malingerers in comparison to the general and clinical population, respectively. However, the *Fp-r* Scale was less specific (i.e., correct classification of non-malingerers) than the other overreporting scales. These results are in agreement with previous literature on the original scales of the MMPI-2 (Rogers et al., 2003), and support the underlying models for these scales: *F-r* was created with items infrequently endorsed by the general population and *Fp-r* with items rarely endorsed by psychiatric patients (clinical sample). In short, the F-Family scales discriminated better between malingerers and honest responding (clinical and general population).

As for case studies ($N = 1$), the overreporting validity scales correctly and significantly classified malingering in contrast to honest responding in the general population, but only the *Fp-r* and *F-r* scales in contrast to the clinical cases. Succinctly, the *F-r* and *Fp-r* scales correctly and significantly classified malingering from honest responding, i.e., in contrast to the general and clinical population, otherwise, the accuracy of the *RBS*, *Fs*, and *FBS-r* in the classification of cases was not significant in inter-contexts. Namely, these scales are not generally valid to inform on overreporting as they do not perform adequately in the identification of malingering among clinical cases. Although *F-r* and *Fp-r* scales are strongly correlated, they were independent in malingering classification and their efficacy may be added as significant incremental validity of one over the other. Thus, the combination of both malingering scales improved the classification significantly. Nevertheless, the resulting combination was insufficient for forensic practice as false positives and negatives occurred, which is inadmissible in evaluations in forensic contexts. This underscores the differential diagnosis of malingering cannot rest on the *Fp-r* and *F-r* scales alone. Hence, malingering differential diagnosis requires a multimethod approach consisting of a clinical interview and a psychometric measure. Specifically, the psychometric measure with vast empirical support for this task is the MMPI-2, and within the clinical interviews – as the standard clinical interview is not valid as it facilitates

malingering (recognition task for malingerers) and has no techniques to detect malingering – the forensic clinical interview with empirical support is the SIRS (Rogers, Bagby, & Dickens, 1992), for criminal insanity assessment, and the Forensic-Clinical Interview (Vilariño, Arce, & Fariña, 2013), for psychological injury assessment.

Future research should focus on the incremental validity of the standard and revised validity scales, and the incremental validity of the overrreporting measures of the MMPI-2 (e.g., the MMPI-2 includes additional scales such as the Dissimulation Scale *Fptsd*, and indexes such as *F-K*), the MMPI-2-RF, and the combination of the scales and indexes of both versions of the MMPI.

References

Amado, B. G., Arce, R., Fariña, F., & Vilariño, M. (2016). CBCA reality criteria in adults: A meta-analytic review. *International Journal of Clinical and Health Psychology*, *15*, 29–36. http://dx.doi.org/10.1016/j.ijchp.2016.01.002

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

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 simulación [Evaluating psychological injury in motor vehicle accidents (MVA): Development and validation of a protocol for detecting simulation]. *Psicothema*, *18*, 278–284. Retrieved from http://www.psicothema.com/pdf/3210.pdf

Arce, R., Fariña, F., Seijo, D., & Novo, M. (2015). Assessing impression management with the MMPI-2 in child custody litigation. *Assessment*, *22*, 769–777. http://dx.doi.org/10.1177/1073191114558111

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*, 72–80. http://dx.doi.org/10.1016/j.rips.2015.04.002

Ben-Porath, Y. S. (2013). Forensic applications of the Minnesota Multiphasic Personality Inventory-2 Restructured Form. In R. P. Archer, & E. M. A. Wheeler (Eds.), *Forensic use of clinical assessment instruments* (pp. 63–107). New York: Rutledge.

Ben-Porath, Y. S., & Tellegen, A. (2008/2011). *MMPI-2-RF (Minnesota Multiphasic Personality Inventory-2 Restructured Form): Manual for administration, scoring, and interpretation*. Minneapolis, MN: University of Minnesota Press.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Manual for the restandardized Minnesota Multiphasic Personality Inventory: MMPI-2*. Minneapolis, MN: University of Minnesota Press.

Chafetz, M., & Underhill, J. (2013). Estimated costs of malingered disability. *Archives of Clinical Neuropsychology*, *28*, 633–639. http://dx.doi.org/10.1093/arclin/act038

Fariña, F., Arce, R., & Real, S. (1994). Ruedas de identificación: De la simulación y la realidad [Lineups: A comparison of high fidelity research and research in a real context]. *Psicothema*, *6*, 395–402. Retrieved from https://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. *The Spanish Journal of Psychology*, *17*, e32. http://dx.doi.org/10.1017/sjp.2014.30, 1-10

Fariña, F., Redondo, L., Seijo, D., Novo, M., & Arce, R. (2017). A meta-analytic review of the MMPI validity scales and indexes to detect defensiveness in custody evaluations. *International Journal of Clinical and Health Psychology.*, http://dx.doi.org/10.1016/j.ijchp.2017.02.002. Advance online publication.

Glas, A. S., Lijmer, J. G., Prins, M. H., Bonsel, G. J., & Bossuyt, P. M. (2003). The diagnostic odds ratio: A single indicator of test performance. *Journal of Clinical Epidemiology*, *56*, 1129–1135. http://dx.doi.org/10.1016/S0895-4356(03)00177-X

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* (3th ed.). Boston, MA: Allyn & Bacon.

Hathaway, S. R., & McKinley, J. C. (1940). A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *Journal of Psychology*, *10*, 249–254. http://dx.doi.org/10.1080/00223980.1940.9917000

Hathaway S. R., & McKinley J. C. (1999). *Inventario Multifásico de Personalidad de Minnesota-2. Manual*. [The Minnesota Multiphasic Personality Inventory-2. Manual]. Madrid, Spain: TEA Ediciones.

Hunter, J. E., & Schmidt, F. L. (2015). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

Ingram, P. B., & Ternes, M. S. (2016). The detection of content-based invalid responding: A meta-analysis of the MMPI-2-Reestrutured Form's (MMPI-2-RF) over-reporting validity scales. *The Clinical Neuropsychologist*, *30*, 473–496. http://dx.doi.org/10.1080/13854046.2016.1187769

Jiménez, F., & Sánchez, G. (2002). Sensibilidad al fingimiento de la Escala Psiquiátrica *Fp* de Arbisi y Ben-Porath (1995, 1998) en la adaptación española del MMPI-2 [Sensitivity to the fake-bad of the Psychiatric Scale Fp of Arbisi and Ben-Porath, 1995, 1998, in the Spanish adaptation of the MMPI-2]. *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica*, *14*, 117–134. Retrieved from http://gredos.usal.es/jspui/bitstream/10366/55768/1/DPETP_SensibilidadfingimientoFp_MMPI-2.pdf

Jiménez, F., Sánchez, G., & Tobón, C. (2009). A social desirability scale for the MMPI-2. Which of the two: Wiggins (WSD-R) or Edwards (ESD)? *European Journal of Psychology Applied to Legal Context*, *1*, 147–163.

Konecni, V. J., & Ebbesen, E. B. (1992). Methodological issues on legal decision-making, with special reference to experimental simulations. In F. Lösel, D. Bender, & T. Bliesener (Eds.), *Psychology and law. International perspectives* (pp. 413–423). Berlin, Germany: Walter de Gruyter.

McDermott, B. E. (2012). Psychological testing and the assessment of malingering. *Psychiatric Clinics of North America*, *35*, 855–876. http://dx.doi.org/10.1016/j.psc.2012.08.006

Nelson, N. W., Hoelzle, J. B., Sweet, J. J., Arbisi, P. A., & Demakis, G. J. (2010). Updated meta-analysis of the MMPI-2 symptom validity scale (FBS): Verified utility in forensic practice. *The Clinical Neuropsychologist*, *24*, 701–724. http://dx.doi.org/10.1080/13854040903482863

Osuna, E., López-Martínez, M., Arce, R., & Vázquez, M. J. (2015). Analysis of response patterns on the MMPI-2 in psychiatric prison inmates. *International Journal of Clinical and Health Psychology*, *15*, 29–36. http://dx.doi.org/10.1016/j.ijchp.2014.09.002

Pierson, A., & Rosenfeld, B. (2015). Malingering. In L. R. Cautin, & S. O. Lilienfeld (Eds.), *The encyclopedia of clinical psychology* (pp. 1–6). Hoboken, NY: Wiley and Sons. http://dx.doi.org/10.1002/9781118625392

Rogers, R. (2008). *Clinical assessment of malingering and deception* (3rd. ed.). New York, NY: The Guilford Press.

Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *Structured Interview of Reported Symptoms (SIRS) and professional manual*. Odessa, 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*, 160–177. http://dx.doi.org/10.1177/1073191103252349

Sánchez, G., Jiménez, F., Novo, R., & Silva, D. (2007). Contribución de la escala revisada de Disimulación de Gough (Ds-r) a la validez del MMPI-2 [Contribution of Gough Dissimulation Scale revised – Ds-r – to validity of the MMPI-2]. *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica*, *23*, 101–120. Retrieved from https://gredos.usal.es/jspui/bitstream/10366/55793/1/DPETP_ContribucionDs-r_MMPI-2.pdf

Schroeder, R. W., Baade, L. E., Peck, C. P., VonDran, E. J., Brockman, C. J., Webster, B. K., & Heinrichs, R. J. (2012). Validation of MMPI-2-RF validity scales in criterion group neuropsychological samples. *The Clinical Neuropsychologist*, *26*, 129–146. http://dx.doi.org/10.1080/13854046.2011.639314

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*, 1–21.

Wolf, E. J., & Miller, M. W. (2014). The Minnesota Multiphasic Personality Inventory-2 Restructured Form and posttraumatic stress disorder: Forensic applications and considerations. *Psychological Injury and Law*, *7*, 143–152. http://dx.doi.org/10.1007/s12207-014-9193-z

Wygant, D. B., Selbon, M., Gervais, R. O., Ben-Porath, Y. S., Stafford, K. P., & Freeman, D. B. (2010). Further validation of the MMPI-2-RF Response Bias Scale: Findings from disability and criminal forensic settings. *Psychological Assessment*, *22*, 745–756. http://dx.doi.org/10.1037/a0020042

Young, G. (2015). Malingering in forensic disability-related assessments: Prevalence 15 ± 15%. *Psychological Injury and Law*, *8*, 188–199. http://dx.doi.org/10.1007/s12207-015-9232-4