



# Beyond Rare-Symptoms Endorsement: a Clinical Comparison Simulation Study Using the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) with the Inventory of Problems-29 (IOP-29)

Luciano Giromini<sup>1</sup> · Sharon Carfora Lettieri<sup>1</sup> · Salvatore Zizolfi<sup>2</sup> · Daniele Zizolfi<sup>3</sup> · Donald J. Viglione<sup>4</sup> · Emanuela Brusadelli<sup>5,6</sup> · Barbara Perfetti<sup>6</sup> · Daniela Angiola di Carlo<sup>6</sup> · Alessandro Zennaro<sup>1</sup>

Received: 2 March 2019 / Accepted: 22 May 2019 / Published online: 5 June 2019  
© Springer Science+Business Media, LLC, part of Springer Nature 2019

## Abstract

To date, the MMPI-based, rare-symptom detection strategy is considered one of the most effective ones in symptom validity assessment. Because many of the items of the Inventory of Problems-29 (IOP-29) were designed specifically to provide incremental validity over the MMPI F scales, this study tested whether using the IOP-29 in combination with the MMPI-2 would provide higher classification accuracy compared to using either instrument alone. A total of 155 Italian adult individuals contributed to this study. About half ( $n = 93$ ) were experimental malingerers (expMAL) instructed to simulate depression without being detected as feigners. The others were either (a) depressed patients in treatment ( $n = 36$ ) or (b) individuals evaluated for possible malingering associated with work-related stress and considered to be genuinely affected by depression ( $n = 26$ ). All were administered the Italian versions of both the MMPI-2 and the IOP-29. As expected, both instruments were highly effective in discriminating feigned from bona fide depression, with AUC values ranging from .77 to .90. More importantly, when entering the IOP-29 after each of the MMPI-2 scales under consideration (i.e., F, Fb, and Fp), the logistic regression models predicting group membership (0 = patient; 1 = expMAL) improved significantly. Likewise, each of the three MMPI-2 scales under consideration also significantly improved the prediction of group membership, when entered after the IOP-29. These findings thus indicate that using the MMPI-2 together with the IOP-29 could provide incremental validity over using either instrument alone, when testing depression-related complaints.

**Keywords** IOP-29 · MMPI-2 · Malingering · Depression · Validity

Over-reporting of depressive symptoms needs to be assessed carefully in civil forensic settings. Indeed, the duration of absence from work is typically longer for cases of major depression than it is for cases of serious medical problems such as

back pain, hypertension, diabetes mellitus, and heart disease (Druss et al. 2000). Besides, workers' compensation claimants (Repko and Cooper 1983), personal injury claimants (Lees-Haley 1997), and military veterans seeking disability compensation (Frueh et al. 1996; Smith & Frueh, 1996), all typically report significant depressive symptoms in their evaluations. Moreover, possibly because everyone has experienced low mood at some point in his/her life, and information about depressive symptoms is readily accessible to anyone (Lees-Haley and Dunn 1994), major depression symptoms can be feigned easily (Bagby et al. 2000; Nicholson and Martelli 2006; Steffan et al. 2003). In fact, it has been estimated that about 15% of the depressive syndromes diagnosed in litigation or compensation cases are likely feigned (Mittenberg et al. 2002).

To assess the credibility of depression-related presentations, practitioners should always include employ multiple sources of information and multiple tests (Boone 2009; Bush et al. 2005; Heilbronner et al. 2009; Iverson 2006; Larrabee

---

✉ Luciano Giromini  
luciano.giromini@unito.it

<sup>1</sup> Department of Psychology, University of Turin, Turin, Italy

<sup>2</sup> Private Practice, Como, Italy

<sup>3</sup> Faculty of Medicine and Surgery, University of Insubria, Varese, Italy

<sup>4</sup> Alliant International University, San Diego, CA, USA

<sup>5</sup> University of Milano-Bicocca, Milan, Italy

<sup>6</sup> UO di Medicina del Lavoro, ASST-Rhodense, Garbagnate Milanese, Italy

2008). Several stand-alone symptom validity (SVTs) and performance validity (PVTs) tests are available, to that purpose. The Test of Memory Malingering (TOMM; Tombaugh 1996, 1997), Word Memory Test (WMT; Green et al. 1996), and Rey 15-item Memorization Test (RMT; Lezak 1995) are three popular examples of PVTs. The Structured Inventory of Malingered Symptomatology (SIMS; Smith and Burger 1997; Widows and Smith 2005) is a popular example of SVT (Dandachi-FitzGerald et al. 2013; Martin et al. 2015). Additionally, several multiscale personality inventories including one or more validity indicators designed to detect atypical response styles and exaggeration are available as well. Among them, the most investigated one for malingered depression issues (Nicholson and Martelli 2007) is probably the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway and McKinley 1940), in its most updated versions, i.e., the MMPI-2 (Butcher et al. 2001) and MMPI-2-RF (Ben-Porath and Tellegen 2008).

Three MMPI-2 scales are particularly useful to assess symptom validity: F (Infrequency), Fb (Back Infrequency), and Fp (Infrequency-Psychopathology). The F scale was originally designed to measure atypical responding, which occurs in case of random responding or poor understanding of the meaning of the items (Friedman et al. 2015). Because its items address uncommon or deviant behavior, however, elevations of F have been commonly used as an indicator over-reporting or exaggerating. The Fb scale was designed to operate similarly to the F scale, i.e., to detect divergences from normality. Its focus, however, is on the second half of the inventory (Friedman et al. 2015). Thus, what characterizes Fb is that it is sensitive to possible shifts in the respondent's attitude, for example, due to fatigue or poor cooperation during the latter part of the test. Additionally, F and Fb differ in their content: While the former mainly addresses psychosis-related problems, the latter focuses more on acute distress and depression or low self-esteem issues. Lastly, the Fp scale was developed by Arbisi and Ben-Porath (1995a, 1995b) to help practitioners disentangle whether a high score on the F would reflect a "faking bad" response set versus other phenomena such as random responding, poor reading ability, or severe psychopathology (Friedman et al. 2001). Indeed, while F includes items that are endorsed rarely by healthy controls, Fp is comprised of items that are endorsed rarely by both healthy controls and psychiatric patients with known psychopathology. As such, F elevations associated with low Fp scores are deemed to indicate random responding, poor reading ability, or genuine, but severe disturbance, whereas high F scores with high Fp scores might instead suggest an over-reporting or faking bad attitude.

While each of these three validity scales provides useful and unique information, currently, the Fp scale is considered the strongest MMPI-2 scale for discriminating bona fide from feigned psychopathology. Indeed, a meta-analysis of 76 MMPI-2 studies (Rogers et al. 2003) indicates that albeit the

average effect size across studies was slightly higher for F ( $d = 2.21$ ) than for Fp ( $d = 1.90$ ), and the cut-off scores of Fp were much more stable across the different diagnostic targets taken under consideration. Along the same lines, when compared to F and Fb, Fp yielded more similar effect sizes when going from one investigation to another. Conversely, the empirically derived cut-off scores of F were quite variable across the different studies, ranging from  $> 8$  to  $> 30$ , and the average effect size of Fb was remarkably lower ( $d = 1.62$ ) compared to both F and Fp. Based on these findings, Rogers et al. (2003) "recommended the Fp as the primary scale for the assessment of feigning" (p. 173) and "questioned the routine use of Fb" (p. 160).

With the introduction of the briefer, MMPI-2-RF, the Fb scale was not retained and the F and Fp scales were slightly revised to adjust to the new format of the test (which has decreased from 567 to 338 items) and to the newly collected, normative reference data. Named "F-r" and "Fp-r," these revised counterparts of the MMPI-2 F and Fp scales remained highly consistent with their MMPI-2 predecessors. Indeed, while F-r addresses possible divergences from normality, Fp-r addresses possible divergences from both healthy controls and bona fide psychiatric patients.

It is noteworthy that examination of MMPI-2-RF validity scales' research also leads to similar conclusions to those described above. In fact, a recent meta-analysis of 30 studies by Sharf et al. (2017) suggests that Fp-r may be superior to all other MMPI-2-RF scales for several reasons. First, differently from F-r, which exhibited marked elevations in some bona fide patients affected by mixed diagnoses, major depression, or somatoform disorders (i.e., false positive results), Fp-r was highly specific in all of the studies included in the meta-analysis, with small variations from one tested condition to another. Second, unlike all other MMPI-2-RF validity scales, the Fp-r continued to prove highly useful also in the only one study (Sellbom and Bagby 2010), among those included in the meta-analysis, that compared coached simulators against clinical samples. Third, its effect sizes, receiver operating characteristic (ROC) curves, and empirically derived cut-off scores were particularly stable from one study to another.

All in all, however, the F-r scale also showed some merit in this meta-analytic report. Indeed, its average effect size was  $d = 1.15$ , when comparing all feigners ( $n = 2575$ ) against all genuine patients ( $n = 1836$ ) taken into consideration. Furthermore, Sellbom et al. (2010) suggested that F-r may provide some incremental validity over Fp-r in criminal forensic settings, where malingerers likely present complaints in multiple, rather than one, domains (i.e., psychopathology, cognitive, and somatic).

Both MMPI-2 F and Fp scales, as well as their MMPI-2-RF counterparts F-r and Fp-r, are effective because malingerers likely do not fully know what symptoms are common versus rare for a given, psychopathological condition (Greene 2000).

More specifically, Fp and Fp-r measure the extent to which a test-taker endorses *rare* symptoms, i.e., symptoms that are infrequent among both healthy controls and psychiatric patients, and F and F-r measure endorsement of *quasi rare* symptoms, i.e., symptoms that are infrequent in the normative nonclinical samples but may not be so infrequent among bona fide patients, especially if affected by severe psychopathology. The main idea is that elevation of these scales should raise concerns as to whether a given presentation is credible or not, as it is rather unlikely to find high scores in these scales if the test-takers have answered honestly. Although both the MMPI-2 and MMPI-2-RF use some other detection strategies too (e.g., erroneous stereotypes, obvious-subtle, symptom selectivity), presently, the rare-symptoms detection strategy appears to be the most effective one, in the assessment of feigned mental disorders (Sharf et al. 2017; Rogers et al. 2003). As reviewed above, indeed, both MMPI-2 and MMPI-2-RF meta-analytic studies indicate that the MMPI-2 Fp and its MMPI-2-RF Fp-r produce by far the most stable and satisfactory results across studies. No other MMPI-2 or MMPI-2-RF validity scale reaches similar levels of effectiveness across different studies.

## The Current Study

Nowadays, virtually, all researchers and practitioners would agree that the clinical determination of malingering should never rely on a single measure and should instead use multiple instruments, possibly implementing different feigning strategies (Boone 2009; Bush et al. 2014; Bush et al. 2005; Chafetz et al. 2015; Rogers 2008; Rogers and Bender 2018). To that extent, it might be argued that a tool that could prove particularly useful, when used in combination with the MMPI instruments, is the Inventory of Problems – 29 (IOP-29; Viglione et al. 2017). Comprised of 29 items only, the IOP-29 was indeed designed specifically to provide incremental validity over the classic, MMPI-based, rare-symptom approach scales (Viglione, Giromini et al. 2018).

Rather than focusing on rare-symptoms endorsement, the IOP-29 addresses the subjective experience of the test-taker concerning his or her ability to deal and cope with his or her problems. For example, instead of asking whether or not the respondent has problems falling asleep, it investigates whether s/he feels like there is anything s/he can do about it, whether s/he feels like s/he bears some responsibility for that problem, and so on. Furthermore, in addition to the classic “True” versus “False” response options, the IOP-29 also offers a third possible choice: “Doesn’t make sense.” This is because accumulating experience in the field indicates that feigners may at times present themselves with some confusion, cognitive deficiency, and resistance to the evaluation (Rogers 2008), which may be well captured by this type of response option

(Viglione et al. 2017). Along the same lines, in addition to 26 self-report items, the IOP-29 also presents three cognitive, or PVT items, which also contribute to make the IOP-29 a very different tool, compared to the MMPI instruments. For all these reasons, we hypothesized that using the MMPI together with the IOP-29 would provide some useful incremental validity, over using either instrument alone. The current study tested this hypothesis by administering the MMPI-2 and IOP-29 to a sample of patients with depression-related disorders and to a sample of experimental malingerers (expMAL) instructed to feign depression.

## Method

Three different Italian samples contributed to this research. A first sample included 36 psychiatric patients diagnosed with and in treatment for major depression disorder (MDD) or adjustment disorder with depressive mood (ADDM). A second sample was comprised of 28 adult individuals who met the following three criteria: (1) they had been referred to psychiatric and psychological units of a public hospital for work-related stress issues; (2) they had received a diagnosis of MDD or ADDM; (3) their symptom presentation was deemed to be highly credible. The third sample was comprised of 100 nonclinical adults instructed to feign depressive symptoms elicited by a work-related accident. Thus, a total of 64 patients with depression and 100 expMAL contributed to this study. All signed an informed consent form, and the procedures of this project were reviewed and approved by the applicable ethical committees. Data collection began in March 2018, when the IOP-29 was officially made available to practitioners and ended December 2018.

## Participants and Procedures

All participants were native Italian-speaking adults, who defined themselves as “Italian” or “Caucasian.” As such, all materials were administered in Italian, consistent with standard Italian practice. In addition, because all completed at least Middle School, their reading abilities were considered to be adequate to filling out both the MMPI-2 and IOP-29.

**Depressed Patients in Treatment** All individuals included in this sample ( $n = 36$ ) were consecutive adult patients from a psychiatric ambulatory located in the North of Italy. Two thirds ( $n = 24$ ) were referred for the first time to this ambulatory for psychological assessment and treatment purposes, whereas 12 had been in treatment for months (with SSRI antidepressants and, in some cases, benzodiazepines) and, at the time when the MMPI-2 and IOP-29 were administered, were considered to be in remission. In all cases, the diagnoses of DDM and ADDM had been formulated by the two chief

psychiatrists of the ambulatory via clinical interview, after consulting with each other. For the majority of the sample, the presented depressive symptoms were not considered to be particularly severe.

Twenty-two (i.e., 61.1%) of the patients included in this sample were women, average age was 50.1 ( $SD = 14.0$ ), and average number of years of education was 12.8 ( $SD = 3.5$ ).

**Depressed Patients with Work-Related Stress** Individuals included in this sample were depressed patients evaluated for possible exaggeration and considered highly unlikely to be malingers. Because all had external incentives to look depressed (e.g., lawsuits in progress), they were first evaluated through an extensive clinical interview by a medical doctor on the Occupational Health Unit of a hospital located in the North of Italy. Then, if this doctor believed that their complaints were bona fide, they were sent to a different unit of the same hospital for a second clinical interview, this time performed by a psychiatrist. Diagnoses of MDD and ADDM were formulated in this occasion. Then, all of these patients returned to the Occupational Health Unit, where two experienced psychologists conducted another extensive clinical interview and reviewed, together with the doctor from the first interview, all relevant information concerning the cases, including clinical histories and any potentially useful materials such as email and photos. This three-step, thorough, examination terminated with the identification of 28 patients deemed to be genuinely affected by MDD or ADDM. All individuals who did not receive one of these psychiatric diagnoses or whose symptom presentation was not considered fully credible were excluded from the current study.

The administration of both MMPI-2 and IOP-29 occurred at the end of this three-step examination. Slightly more than the half of this genuinely depressed sample (i.e., 15, or 53.6%) were women, average age was 48.9 ( $SD = 8.3$ ), and average number of years of education was 14.8 ( $SD = 3.2$ ).

**Experimental Malingers** A nonclinical sample comprised of 100 adult participants instructed to feign depression also contributed to this research. These were recruited via convenience and snowball sampling procedures in various Italian cities (mainly located in the North of Italy). Inclusion criteria required being 18 or more, not having been diagnosed with any major psychiatric disorders, and being able (and willing) to read and sign an informed consent form. In line with standard guidelines on how to conduct a simulation study (Rogers and Bender 2018), all were given a vignette depicting a situation in which a person might decide to fake depression, a brief list of symptoms characterizing this psychopathological condition, a cautionary statement “not to over-do it” or else their performance would not be believable, and a small economic incentive to do their best to successfully feign depression without looking like feigners (see Appendix 1). Lastly, at

the end of the experiment, they were inquired about their feigning strategies, so to ascertain that everyone followed the instructions. In terms of demographic variables, 62 (i.e., 62.0%) were women, average age was 51.0 ( $SD = 17.0$ ), and average number of years of education was 14.0 ( $SD = 3.7$ ).

## Measures

**The Minnesota Multiphasic Personality Inventory-2 (Butcher et al. 2001)** The Minnesota Multiphasic Personality Inventory-2 (MMPI-2) is the probably most popular measure of general psychopathology for forensic and psychiatric assessment. It is comprised of 567 “True” or “False” items, and offers several validity and clinical scales, as well as content components and supplementary scales. In this study, the official Italian version of the MMPI-2 was used (Pancheri and Sirigatti 1995).

As reviewed above, among all MMPI-2 validity scales addressing negative response bias, Fp is probably the most supported one, from an empirical standpoint, but F and—to a lesser extent—Fb have some merit too. According to Rogers et al.’s (2003) meta-analysis, optimal cut scores for these F scales may be  $F_{raw} > 20$  or  $F_{raw} > 24$ ;  $Fb_{raw} > 18$  or  $Fb_{raw} > 20$ ; and  $Fp_{raw} > 7$ . It should be noted that the MMPI-2 is generally considered too long and time consuming to be used as a screening test for malingering. It follows then that all these cut scores favor specificity (less than 5% or 2% of patients should be classified as feigning) over sensitivity, which would be favored in screening tests.

**The Inventory of Problems-29 (Viglione et al. 2017)** The Inventory of Problems-29 (IOP-29) is a relatively new, self-administered test, comprised of 27 “True,” “False,” or “Doesn’t make sense” items and two open-ended cognitive items. Its chief feigning scale, the False Disorder Probability Score (FDS), was derived from logistic regression, and therefore, it consists of a probability score. More specifically, the IOP-29 FDS provides the likelihood that a given IOP-29 comes from a sample of experimental feigners versus a sample of bona fide patients, when the a-priori expectations are 50% and 50%. The higher the score, the more likely the score represents noncredible complaints. In this study, the cross-culturally adapted version of the IOP-29 for use with Italian populations has been used (Giromini et al. 2018).

According to the results of a clinical comparison simulation study conducted by Giromini et al. (2018), despite it having only 29 items, the IOP-29 offered a better classification accuracy compared to the 75-item SIMS. Furthermore, two recent studies from Portugal (Giromini et al. 2019a) and Italy (Giromini et al. 2019b) have shown that the IOP-29 FDS may be similarly sensitive to different types of mental health complaints, such as those related to depression, PTSD, psychosis, or mild traumatic brain injury. As a general principle, an  $FDS \geq .50$  should offer the best balance between sensitivity

and specificity, offering an average overall correct classification percentage of about 80%. Because the IOP-29 is so short, however, one could also use it as a screening instrument. If that were the case, cut scores of  $FDS \geq .30$  or  $FDS \geq .15$  may be preferable, as they should produce higher sensitivity rates, of 90% and 95% respectively. Conversely, in forensic contexts where specificity is likely more important than sensitivity, cut scores of  $FDS \geq .65$  or  $FDS \geq .70$  may be more appropriate as they should offer higher specificity rates, of 90% and 95% respectively (for details on these cut scores, please see Giromini et al. 2018).

### Protocol Screening and Statistical Analyses

Prior to analyzing the data, all 164 available MMPI-2 and IOP-29 records were screened for content-unrelated distortions, such as inconsistencies, and inadequate item endorsement. Thus, records with MMPI-2 Cannot Say (CNS)  $\geq 30$ , True Response Inconsistency (TRIN)  $T \geq 80$ , Variable Response Inconsistency (VRIN)  $T \geq 80$ , or more than 3 missing responses on the IOP-29 were excluded. This approach reduced the sample size to 155 valid cases, as 8 people had an invalid MMPI-2 and one person had an invalid IOP-29. Of these 155 valid cases, 62 were depressed patients (36 depressed patients in treatment and 26 depressed patients assessed for work-related stress) and 92 were expMAL. Next, we compared the patient and expMAL groups on gender, age, and years of education, to evaluate whether the two groups were sufficiently balanced on these demographic variables. None of these analyses produced statistically significant results, all  $p \geq .41$ .

Subsequently, we focused on Cohen's  $d$  effect sizes, ROC curves, and classification accuracy statistics by contrasting the patients' data against those of expMAL. To evaluate incremental validity, we then performed a series of hierarchical logistic regressions, with group (0 = patient; 1 = expMAL) as criterion variable and the MMPI-2 and IOP-29 scores as predictors. Lastly, we inspected MMPI-2 clinical scales to evaluate whether the expMAL did elevate the depression-related scales, as one would expect.

### Results

Table 1 reports on average MMPI-2 and IOP-29 scores produced by the depressed patients and the expMAL included in this study. As shown in Table 2, the MMPI-2 scale that produced the highest effect size and AUC was F: When considering the entire sample ( $N = 155$ ), it produced a Cohen's  $d$  of 1.48 and an AUC of .89. With that same sample (i.e., when considering the entire group), the IOP-29 FDS produced relatively similar, perhaps slightly superior results, with Cohen's  $d = 1.80$  and  $AUC = .89$ . According to Rogers et al.'s (2003)

characterization of Cohen's  $d$  values from experimental malingering studies, the IOP-29 FDS showed "very large" effect sizes (i.e.,  $\geq 1.75$ ), MMPI-2 scales F and Fb showed "large" effect sizes (i.e.,  $\geq 1.25$ ), and MMPI-2 Fp showed "moderate" effect sizes (i.e.,  $\geq .75$ ).

Table 3 reports on the classification accuracy of selected MMPI-2 F, Fb, and Fp cut scores, as well as IOP-29 FDS cut scores. As expected, using MMPI-2 cut scores from Rogers et al.'s (2003) meta-analysis ensured very high specificity values, ranging from .94 to 1.00, depending on the sample under consideration. Sensitivity, for those same cut scores, ranged from .33 to .52.

The classification accuracy of the IOP-29 also was in line with previous research and expectations. Consistent with Giromini et al. (2018), using  $FDS \geq .70$  and  $FDS \geq .65$  yielded specificity values of about .95 and .90 (.92 and .89 respectively, considering the entire sample), whereas using  $FDS \geq .15$  and  $FDS \geq .30$  generated sensitivity values of about .95 and .90 (.97 and .89 respectively, considering the entire sample). Also in line with Giromini et al.'s (2018) findings,  $FDS \geq .50$  provided the best balance between sensitivity and specificity (.75 and .87 respectively, considering the entire, combined sample), with an approximate overall correct classification rate of 80%.

Tables 4 and 5 present the results of our incremental validity analyses, which focused on the entire sample so to maximize statistical power. Table 4 demonstrates that entering the IOP-29 after each of the three MMPI-2 validity scales under investigation significantly improved the prediction of group membership (0 = patient; 1 = expMAL). Likewise, but in the opposite direction, each of the three selected MMPI-2 scales also significantly improved our logistic regression models, when entered after the IOP-29 FDS (Table 5). Interestingly, the model with the highest  $\chi^2$  was the one that included MMPI-2 F together with IOP-29 FDS,  $\chi^2(2) = 105.06$ ,  $p < .001$ . Also noteworthy is that neither MMPI-2 Fb,  $\chi^2(1) = 2.10$ ,  $p = .15$ , nor MMPI-2 Fp,  $\chi^2(1) = .02$ ,  $p = .90$ , significantly improved the prediction of group membership when entered after MMPI-2 F. That is, the only scale that yielded some incremental validity over MMPI-2 F, in this study, was the IOP-29 FDS.

Because entering MMPI-2 F together with IOP-29 FDS produced the best model, we created a composite score, calculated as the  $Z$  average of the MMPI-2 F and IOP-29 FDS scores (for details on how to calculate this variable, see Appendix 2). As expected, when considering the entire sample ( $N = 155$ ), this  $Z$  average index produced slightly higher Cohen's  $d$  ( $d = 1.85$ ) and  $AUC$  ( $= .93$ ) values compared to all of the MMPI-2 and IOP-29 scales under investigation (Fig. 1). To further investigate whether combining the IOP-29 FDS with the MMPI-2 F scale would improve classification accuracy compared to the MMPI-2 F scale alone, we performed ROC analyses. Given that our a-priori selected cut scores for F

**Table 1** MMPI-2 and IOP-29 scores across groups

	Depressed patients in treatment ( $n = 36$ )		Depressed patients with work-related stress ( $n = 26$ )		All depressed patients ( $n = 62$ )		ExpMAL ( $n = 93$ )	
	<i>M</i>	SD	<i>M</i>	SD	<i>M</i>	SD	<i>M</i>	SD
MMPI-2 F Raw	8.6	5.7	9.7	4.2	9.0	5.1	23.1	11.6
MMPI-2 Fb Raw	5.5	5.2	5.9	4.4	5.7	4.8	16.9	9.1
MMPI-2 Fp Raw	2.6	1.9	2.5	1.4	2.5	1.7	6.6	4.9
IOP-29 FDS	.25	.23	.30	.21	.27	.22	.71	.26

(see Table 3) yielded specificity values of .95 ( $F > 20$ ) and .98 ( $F > 24$ ), we selected cut scores for our  $Z$  average index with the same specificity values. We then examined whether  $Z$  average cut scores would yield increased sensitivity. The  $Z$ -average cut score of  $Z \geq 1.5$  produced a specificity of .95 and a sensitivity of .66, and the  $Z \geq 1.8$  produced a specificity of .98 and a sensitivity of .60. With the same specificity values, the MMPI-2 F cut scores produced notably lower sensitivity values of .52 and .38. This pattern thus demonstrates that adding the IOP-29 to the most valid MMPI-2 F validity scale remarkably improved the prediction of group membership.

Lastly, we inspected MMPI-2 clinical scales across the three groups to evaluate the extent to which our expMAL could reproduce adequate elevations in the depression-related indicators. As depicted in Fig. 2, the expMAL group showed elevations in several scales, including—but not limited to—Scale 2 (D, Depression). Conversely, the group of depressed patients with work-related stress showed notable elevations on scales 1 (Hs, Hypochondriasis), 2 (D, Depression), and 3 (Hy, Hysteria), but lower scores on all other scales, as one might expect in the case of depression-related conditions. The group of depressed patients in treatment instead showed markedly lower scores compared to both other groups.

## Discussion

This study was designed to test whether using the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher et al. 2001) together with the recently developed, Inventory of

Problems-29 (IOP-29; Viglione et al. 2017) would provide incremental validity in evaluating the credibility of presented depressive symptoms, compared to using either test alone. Examination of MMPI-2 and IOP-29 data from 93 experimental malingers (expMAL), 36 patients in treatment for depression, and 26 depressed patients assessed for work-related stress confirmed this hypothesis. In fact, a series of hierarchical logistic regressions with group membership as criterion variable (0 = depressed patient; 1 = expMAL) and the selected MMPI-2 and IOP-29 scales as predictors demonstrated that using both instruments together yielded a statistically significantly better prediction than using either instrument alone. Importantly, both the IOP-29 scale FDS and the MMPI-2 scales F, Fb, and Fp also demonstrated effectiveness in differentiating bona fide from feigned depression when considered alone, with relatively large Cohen's  $d$  ( $\geq 1.28$  for MMPI-2 F and  $\geq 1.64$  for IOP-29 FDS) and excellent AUC ( $\geq .88$  for both MMPI-2 F and IOP-29 FDS) values (for thresholds for characterizing AUC values, please see Hosmer and Lemeshow 2000). Taken together, these findings thus suggest that including both the MMPI-2 and IOP-29 in multimethod forensic assessments might be a particularly suitable choice, when evaluating depression-related complaints.

The fact that the IOP-29 FDS provided some incremental validity over the use of MMPI-2 F scales is not too surprising. As briefly reviewed in the introduction, many of the detection strategies used by the IOP-29 items were aimed exactly at offering some incremental validity over the classic rare-symptom approach implemented by the MMPI-2 F scales (Viglione et al. 2018). Indeed, while the MMPI-2 F scales primarily focus on symptom endorsement, the emphasis of

**Table 2** Comparison between ExpMAL and depressed patients: Cohen's  $d$  and area under the curve values (AUC)

	ExpMAL vs. depressed patients in treatment		ExpMAL vs. depressed patients with work-related stress		ExpMAL vs. all depressed patients	
	<i>d</i>	AUC	<i>d</i>	AUC	<i>d</i>	AUC
MMPI-2 F Raw	1.41	.89	1.28	.88	1.48	.89
MMPI-2 Fb Raw	1.39	.87	1.32	.87	1.46	.87
MMPI-2 Fp Raw	.93	.77	.93	.79	1.04	.78
IOP-29 FDS	1.82	.90	1.64	.88	1.80	.89

**Table 3** Classification accuracy of selected MMPI-2 and IOP-29 cut scores

	ExpMAL vs. depressed patients in treatment		ExpMAL vs. depressed patients with work-related stress		ExpMAL vs. all depressed patients	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
MMPI-2 F Raw						
>20	.52	.94	.52	.96	.52	.95
>24	.38	.97	.38	1.00	.38	.98
MMPI-2 Fb Raw						
>18	.41	1.00	.41	1.00	.41	1.00
>20	.33	1.00	.33	1.00	.33	1.00
MMPI-2 Fp Raw						
>7	.33	.97	.33	1.00	.33	.98
IOP-29 FDS						
≥.70	.62	.94	.62	.88	.62	.92
≥.65	.67	.89	.67	.88	.67	.89
≥.50	.75	.86	.75	.88	.75	.87
≥.30	.89	.78	.89	.62	.89	.71
≥.15	.97	.50	.97	.31	.97	.42

the IOP-29 FDS is more on how a person manages life despite symptoms, and on this person’s beliefs surrounding the possibility of influencing the severity and expression of problems. Combining the results of the MMPI instruments together with those of the IOP-29 might thus prove particularly useful because it potentially allows to understand not only *what*

symptom(s) the person is experiencing, but also *how* s/he is managing to cope with them. In this article, to appreciate how one might integrate the results of the MMPI-2 with those of the IOP-29, we have calculated the Z average of the MMPI-2 F and IOP-29 FDS scores. As shown in Fig. 1, this composite score showed superior effectiveness compared to either one

**Table 4** Incremental validity of IOP-29 FDS over MMPI-2 Validity Scales: logistic regression models

	OR step 1 [95% CI]	<i>p</i>	OR step 2 [95% CI]	<i>p</i>	χ <sup>2</sup> model	<i>p</i>	Δ χ <sup>2</sup>	<i>p</i>
MMPI-2 F and IOP-29								
Step 1					81.06	< .001	–	–
MMPI-2 F Raw	1.28 [1.18, 1.39]	< .001	1.23 [1.11, 1.35]	< .001				
Step 2					105.06	< .001	24.00	< .001
IOP-29 FDS	–	–	81.24 [11.41, 578.34]	< .001				
MMPI-2 Fb and IOP-29								
Step 1					72.16	< .001	–	–
MMPI-2 Fb Raw	1.26 [1.17, 1.36]	< .001	1.17 [1.07, 1.27]	.001				
Step 2					93.10	< .001	20.93	< .001
IOP-29 FDS	–	–	53.85 [8.42, 344.31]	< .001				
MMPI-2 Fp and IOP-29								
Step 1					45.40	< .001	–	–
MMPI-2 Fp Raw	1.54 [1.29, 1.83]	< .001	1.40 [1.14, 1.73]	.002				
Step 2					94.53	< .001	49.13	< .001
IOP-29 FDS	–	–	196.54 [32.74, 1179.93]	< .001				

OR, odds ratio; CI, confidence interval

**Table 5** Incremental validity of MMPI-2 Validity Scales over IOP-29 FDS: logistic regression models

	OR step 1 [95% CI]	<i>p</i>	OR step 2 [95% CI]	<i>p</i>	$\chi^2$ model	<i>p</i>	$\Delta \chi^2$	<i>p</i>
IOP-29 and MMPI-2 F								
Step 1					78.95	< .001	–	–
IOP-29 FDS	333.04 [62.52, 1773.96]	< .001	81.24 [11.41, 578.34]	< .001				
Step 2					105.06	< .001	26.11	< .001
MMPI-2 F Raw	–	–	1.23 [1.11, 1.35]	< .001				
IOP-29 and MMPI-2 Fb								
Step 1					78.95	< .001	–	–
IOP-29 FDS	333.04 [62.52, 1773.96]	< .001	53.85 [8.42, 344.31]	< .001				
Step 2					93.10	< .001	14.14	< .001
MMPI-2 Fb Raw	–	–	1.17 [1.07, 1.27]	.001				
IOP-29 and MMPI-2 Fp								
Step 1					78.95	< .001	–	–
IOP-29 FDS	333.04 [62.52, 1773.96]	< .001	196.54 [32.74, 1179.93]	< .001				
Step 2					94.53	< .001	15.58	< .001
MMPI-2 Fp Raw	–	–	1.40 [1.14, 1.73]	.002				

OR, odds ratio; CI, confidence interval

instrument used alone. Future studies might thus further investigate this approach and perhaps provide additional information on what cut scores one might want to use, if s/he intended to adopt this *Z* average score in his/her practice.

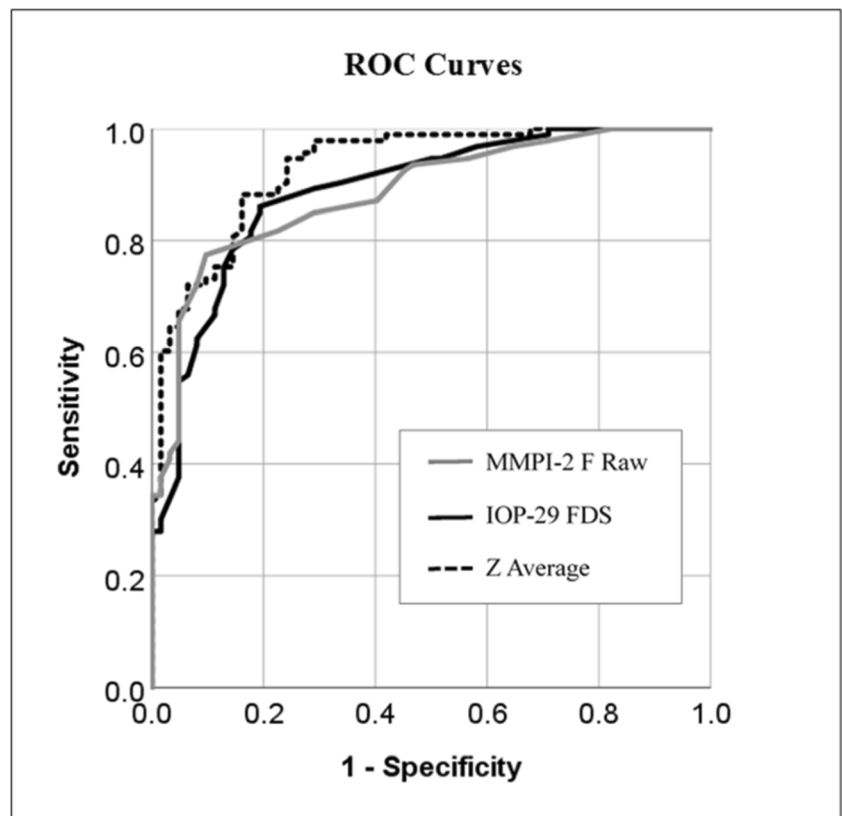
Differently from recent MMPI meta-analytic research, our study found that the F and Fb provided superior effectiveness in detecting experimental feigning compared to the Fp. This unexpected finding is quite difficult to explain. On the one hand, one might say that the fact that many of the patients included in the first of the two patient samples, i.e., the one comprised of depressed patients in treatment (many of which were in remission) suffered from very mild depressive symptoms (see Fig. 2) may have favored F and Fb over Fp. Indeed, while the F and Fb are elevated by endorsement of symptoms that are infrequent in the MMPI-2 normative nonclinical samples, the Fp is elevated by endorsement of symptoms that are infrequent among psychiatric patients. As such, the fact that our patients were not suffering from severe psychopathology may have boosted the specificity of F and Fb, without influencing the overall effectiveness of Fp (Rogers et al. 2003). This explanation, however, does not fit well with the fact that this same pattern of finding, with F and Fb offering better classification accuracy than Fp, was observed also when comparing our expMAL group against the sample of patients assessed for work-related stress and presumably affected by genuine depression (Table 2). These patients, indeed, reported remarkably more severe psychological problems compared to

the depressed patients in treatment sample, as shown in Fig. 2. Future studies might thus attempt to clarify whether this unexpected finding is specific to our sample or perhaps depends on other variables such as the type of vignette we used, the specific instructions we gave to our expMAL, and so on.

One more consideration deserves mentioning. When compared to other similar IOP-29 experimental malingering studies, ours has produced slightly lower sensitivity results, when using the standard cut score of  $FDS \geq .50$ . In fact, when investigating feigning of depression-related symptoms via malingering experimental paradigm, using that same cut score previous studies showed sensitivity rates ranging from .79 to .96 (Giromini et al. 2018; Giromini et al. 2019a, b; Viglione et al. 2017). In our study, with that cut score sensitivity was .75. Because the exact same instructions used in our study were used also in Giromini et al. (2019a) and Giromini et al. (2019b), we speculate that our reduced sensitivity has possibly to do with the fact that the administration of the MMPI-2 may somehow negatively impact the IOP-29's ability to detect feigned depression. Indeed, it is possible that our participants felt like they had already convinced the examiner about their depressive symptoms with the 567 MMPI-2 items, so that they did not have to continue over-reporting depression-related problems also when responding to the IOP-29. Alternatively, it is also possible that, given the length of the MMPI-2, some fatigue had occurred while responding to the two tests, so that the IOP-29 was attended to by our

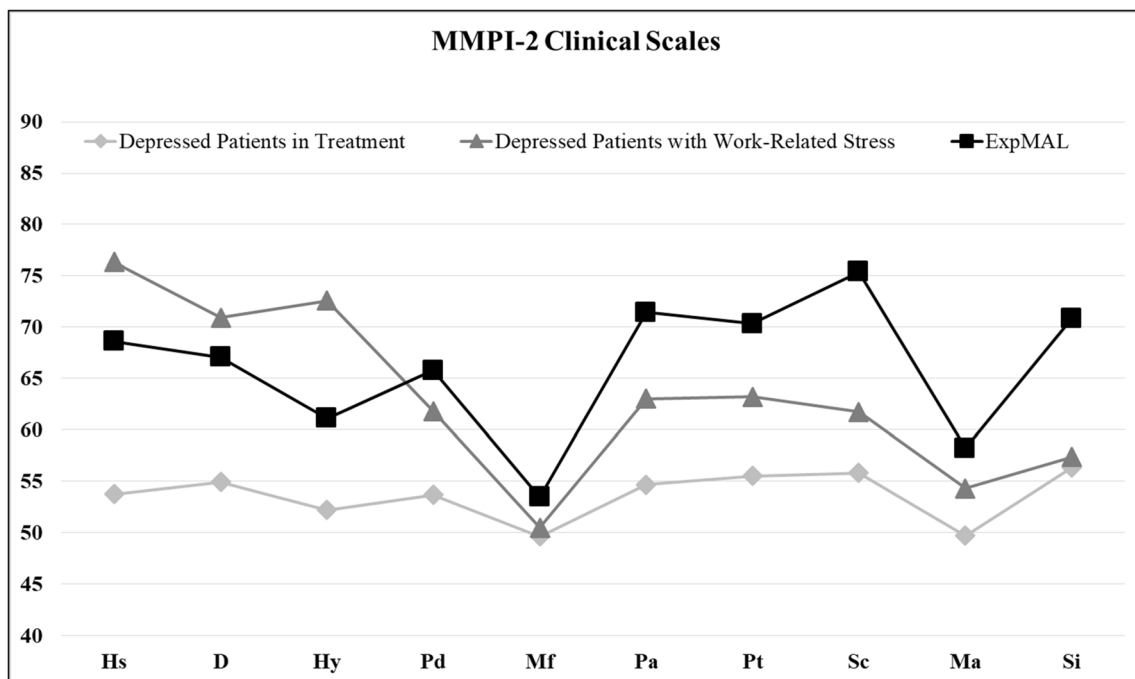


**Fig. 1** ROC curve of MMPI-2 F, IOP-29 FDS, and their Z average



participants with relatively less attention, compared to Giromini et al.'s (2019a) and Giromini et al.'s (2019b) studies. Indeed, in those previous studies, participants only had to fill out the IOP-29 and be examined with the TOMM (Giromini

et al. 2019a) or fill out the IOP-29 alone (Giromini et al. 2019b), which obviously required notably less cognitive effort compared to filling out a long and complex personality inventory such as the MMPI-2. Additional research using both the



**Fig. 2** MMPI-2 Clinical Scales: average scores by group

IOP-29 and MMPI-2 would therefore be highly beneficial, to better understand the possible influence of MMPI-2 administration on IOP-29 sensitivity results.

Lastly, it should be noted that like all malingering-related studies, ours also have some limitations that need to be considered. First, external validity may be questioned, given that our expMAL were instructed to feign depression using an experimental paradigm, so that it is unknown whether real-life malingerers in high stakes contexts would really behave like our experimental participants did in our study. Second, although our Table 2 reveals that there were no notable differences between the results of the MMPI-2 F scales and IOP-29 FDS across the two different patient samples, our inclusion of a patient sample characterized by very mild depressive symptoms may have boosted the effect sizes of our study, to some extent. Third, the patients included in the group of individuals assessed for work-related stress were considered highly unlikely to be malingerers. Although all of them had been thoroughly screened by a series of interviews performed by experienced psychiatrists and psychologists, we cannot rule out that some of them may have in fact over-reported their symptoms. Indeed, the limitation of clinical judgment in determining the credibility of a response set has long been known (Heaton et al. 1978). Fourth, using the MMPI-2 and IOP-29 only may have limited ecological validity, given that real-life symptom validity assessment typically is performed by using a multitude of instruments. Fifth, our inclusion criteria required our expMAL to report that they had not been diagnosed with any major psychiatric disorders. However, given that depression is a high-prevalence mental disease and self-report has its limitations, we cannot rule out that some of our expMAL participants did in fact suffer from depression. If that was the case, our results could be inaccurate regarding the actual effectiveness of the MMPI-2 and IOP-29 to detect feigned depression. Sixth, our study could not evaluate the possible impact of administration order, which in previous studies has shown to have the potential to significantly influence test scores (Erdodi and Lajiness-O'Neill 2014; Ryan et al. 2010; Zuccato et al. 2018). Future research randomizing administration sequence and examining its potential impact on MMPI-2 and IOP-29 scores would therefore be beneficial.

Despite all these limitations, our study still has the merit to be the first to report on the utility of using the MMPI-2 together with the IOP-29 when assessing the credibility of depression-related complaints. All in all, our findings indicate that the IOP-29 may provide useful incremental validity over the classic rare-symptoms endorsement detection strategy scales of the MMPI-2. Given that, researchers are encouraged to continue to investigate the utility of using the IOP-29 in combination with other popular instruments such as the Personality Assessment Inventory (PAI; Morey 1991, 2007) or the recently developed and very promising Self-Report Symptom Inventory (SRSI; Merten et al. 2016).

**Acknowledgments** We thank Lucrezia Frinco for her help in the data collection.

## Compliance with Ethical Standards

**Conflict of Interest** Luciano Giromini and Donald J. Viglione declare that they own a share in the corporate (LLC) that possesses the rights to Inventory of Problems. The other authors declare that they have no conflict of interest.

**Ethical Approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

**Informed Consent** Informed consent was obtained from all individual participants included in the study.

## Appendix 1. Instructions for Experimental Malingerers (ExpMAL)

### Instructions and Vignette to Feign Depression

In this study, you will be asked to complete a series of tests that may measure a variety of changes that people experience following a diagnosis of Major Depression Disorder. When you are answering the questions on the tests, I would like you to put yourself in the shoes of a person who has had an accident at work and is now suffering from mental health problems—namely, depression—related to that accident and for which he has requested to be put on disability. That is, try to answer the questions of these tests like you think that a person that really suffers from depression might do. To help you provide a credible presentation, please read the following text, and try to pretend that you are the person depicted in this scenario.

“You are an administrator at a small, well-established firm. Your boss has been trying to cut expenses by having the cleaning crew work before regular work hours are over, thus getting the job done at a cut rate. You have repeatedly informed him that this is not a safe working condition for the employees, but he has not changed the procedure. One day, near the end of the day, you are leaving to do a special errand for your boss. As you cross a freshly mopped floor, you slip and fall, landing hard on your tailbone. As a result, you have been out of work for 2 weeks on disability and continue to experience a fair amount of pain, particularly when you sit for any length of time. The workers compensation physician insists that he can find nothing to explain the pain and refuses to authorize any more time off or disability payments, stating that you are able to return to work, a job that requires long periods of time sitting at your computer. You are angry with your boss for the injury you have and frustrated at the physician’s apparent collusion with your boss to unreasonably limit your recovery time (thereby cutting off his disability

payments). Before terminating your case, the physician refers you to the staff psychologist for a routine evaluation. You correctly realize that this evaluation is your only opportunity to remain on disability under your employer's obligation. You have no additional coverage and need an income until you are fully recovered. You also feel that your boss is responsible, and that money should come from the company through workers compensation. You know well that workers compensation will continue providing benefits to patients who are psychologically disturbed as a result of a work-related accident. This would not be too unusual because you have tried to take measures to avoid the problem, and now are suffering as a result of your boss's negligence. So, your only choice is to present yourself as having significant depression on the tests that the psychologist is going to give you. You therefore decide to attempt to present yourself as having a major depression as the result of your accident, to remain on disability."

### Description of Symptoms of Depression and Cautionary Statement

Now, please take a look at the symptoms that characterize a Major Depression Disorder. Keep in mind that depressed patients typically have 5 or more of the following symptoms, but most likely not all of them.

1. Depressed mood most of the day, nearly every day (e.g., feeling sad, empty, hopeless)
2. Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day
3. Significant weight loss when not dieting or weight gain, or decrease or increase in appetite nearly every day
4. Insomnia or hypersomnia nearly every day
5. Psychomotor agitation or retardation nearly every day
6. Fatigue or loss of energy nearly every day
7. Feelings of worthlessness or excessive or inappropriate guilt nearly every day
8. Diminished ability to think or concentrate, or indecisiveness, nearly every day
9. Recurrent thoughts of death, recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide.

When you take the tests and try to pretend you suffer from a Major Depressive Disorder, please keep in mind that if you present your condition in an extremely dramatic way, your performance may not be believable, and the examiner might understand that you do not suffer from depression but are only faking it. So, try to not "over-do it".

If you will be able to produce test results that are consistent with those produced by people who really suffer from Major Depression Disorder and you will not look like a feigner, you may win a small prize consisting of a 20€ gift card!

## Appendix 2. Formula Used to Calculate the Z Average of MMPI-2 F and IOP-29 FDS

$$Z \text{ average} = \frac{\left( \frac{\text{MMPI2 F raw} - 9.0}{5.1} \right) + \left( \frac{\text{IOP29 FDS} - .27}{.22} \right)}{2}$$

Note: For each scale, Z values were calculated using the patients' data only, so to avoid possible outliers or extreme variability.

## References

- Arbisi, P. A., & Ben-Porath, Y. S. (1995a). *Identifying changes in infrequent responding to the MMPI-2*. Paper presented at the 30th Annual Symposium on Recent Developments in the Use of the MMPI-2, St. Petersburg, FL.
- Arbisi, P. A., & Ben-Porath, Y. S. (1995b). An MMPI-2 infrequent response scale for use with psychopathological populations. The Infrequency-Psychopathology scale, F(p). *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 7, 424–431.
- Bagby, R. M., Nicholson, R. A., Buis, T., & Bacchiochi, J. R. (2000). Can the MMPI-2 Validity scales detect depression feigned by experts? *Assessment*, 7, 55–62.
- Ben-Porath, Y. S., & Tellegen, A. (2008). *The Minnesota Multiphasic Personality Inventory-2 Restructured Form: Manual for administration, scoring, and interpretation*. Minneapolis, MN: University of Minnesota.
- Boone, K. B. (2009). The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *The Clinical Neuropsychologist*, 23(4), 729–741.
- Bush, S., Ruff, R., Troster, A., Barth, J., Koffler, S., Pliskin, N., et al. (2005). Symptom validity assessment: Practice issues and medical necessity. *Archives of Clinical Neuropsychology*, 20(4), 419–426.
- Bush, S. S., Heilbronner, R. L., & Ruff, R. M. (2014). Psychological assessment of symptom and performance validity, response bias, and malingering: Official position of the Association for Scientific Advancement in Psychological Injury and Law. *Psychological Injury and Law*, 7, 197–205.
- Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, B. (2001). *Minnesota Multiphasic Personality Inventory—2: Manual for administration, scoring and interpretation* (rev ed.). Minneapolis, MN: University of Minnesota.
- Chafetz, M. D., Williams, M. A., Ben-Porath, Y. S., Bianchini, K. J., Boone, K. B., Kirkwood, M. W., Larrabee, G. J., & Ord, J. S. (2015). Official position of the American Academy of clinical neuropsychology Social Security Administration policy on validity testing: Guidance and recommendations for change. *The Clinical Neuropsychologist*, 29(6), 723–740.
- Dandachi-FitzGerald, B., Ponds, R. W. H. M., & Merten, T. (2013). Symptom validity and neuropsychological assessment: A survey of practices and beliefs of neuropsychologists in six European countries. *Archives of Clinical Neuropsychology*, 28(8), 771–783.
- Druss, B. G., Rosenheck, R. A., & Sledge, W. H. (2000). Health and disability costs of depression in a major U.S. Corporation. *American Journal of Psychiatry*, 157, 1274–1278.
- Erdodi, L. A., & Lajiness-O'Neill, R. (2014). Time-related changes in Conners' CPT-II Scores: A replication study. *Applied Neuropsychology: Adult*, 21(1), 43–50.

- Friedman, A. F., Bolinsky, P. K., Levak, R. W., & Nichols, D. S. (2015). *Psychological assessment with the MMPI-2/MMPI-2-RF* (3rd ed.). New York: Routledge/Taylor & Francis Group.
- Frueh, B. C., Smith, D. W., & Barker, S. E. (1996). Compensation seeking status and psychometric assessment of combat veterans seeking treatment for PTSD. *Journal of Traumatic Stress, 9*(3), 427–439.
- Giromini, L., Viglione, D. J., Pignolo, C., & Zennaro, A. (2018). A clinical comparison, simulation study testing the validity of SIMS and IOP-29 with an Italian sample. *Psychological Injury and Law, 11*(4), 340–350. <https://doi.org/10.1007/s12207-018-9314-1>.
- Giromini, L., Barbosa, F., Coga, G., Azeredo, A., Viglione, D. J., & Zennaro, A. (2019a). Using the inventory of problems – 29 (IOP-29) with the Test of Memory Malingered (TOMM) in symptom validity assessment: A study with a Portuguese sample of experimental feigners. *Applied Neuropsychology: Adult*, [Epub ahead of print, 1–13]. <https://doi.org/10.1080/23279095.2019.1570929>.
- Giromini, L., Viglione, D. J., Pignolo, C., & Zennaro, A. (2019b). An inventory of problems – 29 (IOP-29) sensitivity study investigating feigning of four different symptom presentations via malingered experimental paradigm. *Journal of Personality Assessment*, [Epub ahead of print, 1–10]. <https://doi.org/10.1080/00223891.2019.1566914>.
- Green, P., Allen, L. M., & Astner, K. (1996). *The Word Memory Test: A user's guide to the oral and computer administered forms, US version 1.1*. Durham, NC: CogniSyst.
- Greene, R. L. (2000). *The MMPI-2: An interpretive manual*. Boston: 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.
- Heaton, R. K., Smith, H. H., Lehman, R. A. W., & Vogt, A. T. (1978). Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology, 46*(5), 892–900.
- Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & conference, p. (2009). American Academy of Clinical Neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist, 23*, 1093–1129.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: Wiley.
- Iverson, G. L. (2006). Ethical issues associated with the assessment of exaggeration, poor effort, and malingering. *Applied Neuropsychology, 13*(2), 77–90.
- Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist, 22*(4), 666–679.
- Lees-Haley, P. (1997). MMPI-2 base rates for 492 personal injury plaintiffs: Implications and challenges for forensic assessment. *Journal of Clinical Psychology, 53*(7), 745–755.
- Lees-Haley, P. R., & Dunn, J. T. (1994). The ability of naive subjects to report symptoms of mild brain injury, post-traumatic stress disorder, major depression and generalized anxiety disorder. *Journal of Clinical Psychology, 50*(2), 252–256.
- Lezak, M. D. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.
- Martin, P. K., Schroeder, R. W., & Odland, A. P. (2015). Neuropsychologists' validity testing beliefs and practices: A survey on North American professionals. *The Clinical Neuropsychologist, 29*(6), 741–776.
- Merten, T., Merckelbach, H., Giger, P., & Stevens, A. (2016). The Self-Report Symptom Inventory (SRSI): A new instrument for the assessment of distorted symptom endorsement. *Psychological Injury and Law, 9*, 102–111.
- Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology, 24*, 1094–1102.
- Morey, L. C. (1991). *Personality assessment inventory. Professional manual*. Odessa, FL: Psychological Assessment Resources.
- Morey, L. C. (2007). *Personality Assessment Inventory (PAI). Professional manual* (2nd ed.). Odessa, FL: Psychological Assessment Resources.
- Nicholson, K., & Martelli, M. (2006). The confounding effects of pain, psychoemotional problems or psychiatric disorder, premorbid ability structure, and motivational or other factors on neuropsychological test performance. In G. Young, A. Kane, & K. Nicholson (Eds.), *Psychological Knowledge for Court: PTSD, Chronic Pain and TBI* (pp. 335–351). New York: Springer Science+Business Media.
- Nicholson, K., & Martelli, M. F. (2007). Malingering: Traumatic brain injury. In G. Young, A. W. Kane, & K. Nicholson (Eds.), *Causality of psychological injury* (pp. 427–475). New York: Springer.
- Pancheri, P., & Sirigatti, S. (1995). *MMPI-2: Adattamento italiano – Manuale*. Firenze: O.S. Organizzazioni Speciali.
- Repko, G. R., & Cooper, R. (1983). A study of the average workers' compensation case. *Journal of Clinical Psychology, 39*, 287–295.
- Rogers, R. (2008). Detection strategies for malingering and defensiveness. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 14–35). New York, NY: Guilford Press.
- Rogers, R., & Bender, D. (2018). *Clinical assessment of malingering and deception*. New York, NY: Guilford Press.
- 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.
- Ryan, J. J., Glass, L. A., Hinds, R. M., & Brown, C. N. (2010). Administration order effects on the test of memory malingered. *Applied Neuropsychology, 17*, 246–250.
- Sellbom, M., & Bagby, R. M. (2010). Detection of overreported psychopathology with the MMPI-2 RF form validity scales. *Psychological Assessment, 22*(4), 757–767. <https://doi.org/10.1037/a0020825>.
- Sellbom, M., Toomey, A., Wygant, D., Kucharski, L. T., & Duncan, S. (2010). Utility of the MMPI-2-RF (Restructured Form) Validity Scales in detecting malingering in a criminal forensic setting: A known groups design. *Psychological Assessment, 22*, 22–31.
- 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.
- Smith, G. P., & Burger, G. K. (1997). Detection of malingering: Validation of the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of the American Academy on Psychiatry and Law, 25*, 180–183.
- Smith, D. W., & Frueh, B. C. (1996). Compensation seeking, comorbidity, and apparent exaggeration of PTSD symptoms among Vietnam combat veterans. *Psychological Assessment, 8*, 3–6.
- Steffan, J. S., Clopton, J. R., & Morgan, R. D. (2003). An MMPI-2 scale to detect malingered depression (Md Scale). *Assessment, 10*(4), 382–392.
- Tombaugh, T. N. (1996). *Test of Memory Malingered (TOMM)*. New York: Multi Health Systems.
- Tombaugh, T. N. (1997). The Test of Memory Malingered (TOMM): Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment, 9*(3), 260–268.
- Viglione, D. J., Giromini, L., & Landis, P. (2017). The development of the Inventory of Problems–29: A brief self-administered measure for discriminating bona fide from feigned psychiatric and cognitive complaints. *Journal of Personality Assessment, 99*(5), 534–544.
- Viglione, D. J., Giromini, L., Landis, P., McCullough, J. M., Pizitz, T. D., O'Brien, S., et al. (2018). Development and validation of the false disorder score: The focal scale of the inventory of problems. *Journal*

- of Personality Assessment*, [Epub ahead of print], 1–9. <https://doi.org/10.1080/00223891.2018.1492413>.
- Widows, M. R., & Smith, G. P. (2005). *SIMS-Structured Inventory of Malingered Symptomatology. Professional manual*. Lutz, FL: Psychological Assessment Resources.
- Zuccato, B. G., Tyson, T. T., & Erdodi, L. A. (2018). Early bird fails the PVT? The effects of timing artifacts on performance validity tests. *Psychological Assessment*, *30*(11), 1491–1498.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.