



Using the Inventory of Problems-29 (IOP-29) with the Inventory of Problems Memory (IOP-M) in Malingering-Related Assessments: a Study with a Slovenian Sample of Experimental Feigners

Maja Maša Šömen¹ · Staša Lesjak¹ · Teja Majaron¹ · Luca Lavopa² · Luciano Giromini² · Donald Viglione³ · Anja Podlesek¹

Received: 1 February 2021 / Accepted: 6 May 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

A recently published article harshly criticized forensic practitioners operating in Slovenia for not including in their assessments any tests specifically designed to assess negative distortion (Areh, 2020). To promote better forensic assessment practice and stimulate future research on symptom and performance validity assessment in Slovenia, the current study translated the Inventory of Problems-29 (IOP-29; Viglione & Giromini, 2020) and its recently developed memory module (IOP-M; Giromini et al., 2020) into Slovene language and tested their validity and effectiveness by conducting a simulation/analogue study. Among 150 volunteers, 50 completed the IOP-29 and IOP-M under standard instructions; 50 were asked to respond as if they suffered from depression; and 50 were asked to respond pretending to suffer from schizophrenia. Statistical analyses showed that (1) the IOP-29 discriminated well between simulators and honest test-takers ($d \geq 3.56$), demonstrating the same effectiveness when inspecting feigned depression (sensitivity = 88%) and feigned schizophrenia (sensitivity = 88%) at an almost perfect specificity (98%); (2) the IOP-M identified 50% of simulators of depression and 80% of simulators of schizophrenia at perfect specificity (100%); and (3) combining the results of the IOP-29 with those of the IOP-M notably improved classification accuracy so as to demonstrate incremental validity. Taken together, these findings provide initial support for using the IOP-29 and IOP-M in applied settings in Slovenia. Limitations related to the design of the study and recommendations for further research are provided.

Keywords Malingering · Psychological assessment · Symptom validity tests · Performance validity tests · Inventory of Problems

Malingered mental illness appears to be as old as mental illness itself (Resnick, 1984). According to the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013), the term “malingering” refers to “the intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by

external incentives, such as avoiding military duty, avoiding work, obtaining financial compensation, evading criminal prosecution, or obtaining drugs” (p. 726). Resnick (1984) also mentions purposes such as wanting to transfer out of prison, avoid civil litigation, or achieve hospital admission, which is most frequent among the homeless. Among prisoners specifically, relocation, medication, compensation, attention, or amusement may all be reasons why inmates may feign mental illness. However, it must be emphasized that, while feigning symptomatology is present in both factitious disorder and malingering, the DSM-5 considers factitious disorder a genuine mental disorder as it is motivated by internal incentives, whereas malingering is considered deliberate behavior and as such not a form of psychopathology (Rogers & Bender, 2018).

Three different models have been proposed (Rogers & Bender, 2018) as explanations for malingering. The

Maja Maša Šömen, Staša Lesjak, and Teja Majaron contributed equally to this paper.

✉ Anja Podlesek
anja.podlesek@ff.uni-lj.si

¹ Department of Psychology, Faculty of Arts, University of Ljubljana, Ljubljana, Slovenia

² Department of Psychology, University of Turin, Turin, Italy

³ California School of Professional Psychology, Alliant International University, San Diego, CA, USA

adaptational model describes malingering as the result of a cost-benefit analysis, where the malingerer predicts the utility of malingering will be greater than any of alternative solutions. The pathogenic model hypothesizes that, at first, malingerers invent their symptoms because of an actual disability that they are experiencing and trying to control and that, only later on, they lose the control over malingering. The criminological model describes malingering as an anti-social act which is more often committed by people with antisocial traits. Regardless of which of these explanations is more suitable, malingering should be recognized and prevented in its early stages as it presents a tremendous cost for society. Indeed, undetected malingering cases are provided with compensations or unnecessary psychiatric treatment creating enormous financial expenses (Chafetz & Underhill, 2013), and more broadly malingering compromises the efficacy of the entire mental health system, as practitioners waste medical resources and time that they should dedicate to provide treatment to true patients (Vigliano et al., 2017).

Malingering is not an uncommon condition. Mittenberg et al. (2003) estimate that 29% of personal injury cases, 30% of disability cases, 19% of criminal cases, and 8% of medical cases probably involve malingering or symptom exaggeration. Larrabee et al. (2009) even suggested that the base rate of malingering in psychological injury cases is 40%, although Young (2015, 2019) has convincingly characterized this estimate as too high. Nevertheless, given that malingering can cause mistaken, life-changing decisions in high stake evaluations and the misuse of mental health and financial resources, forensic and other high-stake evaluations should always evaluate the credibility of symptom presentations and claims of impairment (Bush et al., 2014).

Symptom and Performance Validity Assessment

To evaluate the credibility of presented complaints, forensic assessors rely on various techniques and tests. A widely accepted tool, in this context, is the Structured Interview of Reported Symptoms (SIRS; Rogers, Bagby, et al., 1992; Rogers, Kropp, et al., 1992; for an updated version, see also Rogers et al., 2010, and Rogers et al., 2020). It is a comprehensive, interview measure that is frequently administered to evaluate response styles associated with intentional distortion of self-reported psychiatric symptoms. Another interview that is widely used in the field is the Miller Forensic Assessment of Symptoms Test (M-FAST; Miller, 2001). In contrast to the SIRS, the 25-item M-FAST is typically used for screening purposes only.

In addition to structured interviews, practitioners usually administer both self-report symptom validity tests (SVTs) and performance validity tests (PVTs) too. An SVT is an

instrument designed to evaluate the extent to which the test-takers complain about symptoms or problems that do not really exist in the real clinical world or that occur very rarely (sometimes called “pseudosymptoms”). An example is the Structured Inventory of Malingered Symptomatology (SIMS; Smith & Burger, 1997), a 75-item, true/false questionnaire covering a broad spectrum of improbable symptoms concerning conditions such as psychosis, neurological impairment, and affective disorders. Other examples are the embedded validity scales in multiscale personality inventories such as the Minnesota Multiphasic Personality Inventory (MMPI-3; Ben-Porath & Tellegen, 2020a, b), Personality Assessment Inventory (PAI; Morey, 1991, 2007), and Millon Clinical Multiaxial Inventory (MCMI-IV; Millon et al., 2015).

PVTs, in contrast, are performance-based measures of cognitive ability that are typically aimed at detecting poor cooperation, motivation, or effort. Examples of PVTs are the test of memory malingering (TOMM; Tombaugh, 1996), the Victoria Symptom Validity Test (VSVT; Slick et al., 2005), and the Word Memory Test (WMT; Green et al., 1996). The main reason for their efficacy in discriminating valid versus invalid cognitive symptom presentations is that most feigners do not realize that even brain-injured patients typically perform quite well on simple recognition tasks, and many mistakenly believe that severe memory problems might occur with a variety of different mental health disorders. As such, they tend to exert inadequate or less than optimal effort and often perform more poorly than bona fide patients.

During the past few years, research has shown that combining multiple methods with different symptom validity assessment approaches on a single case can yield substantial incremental validity (Boone, 2013; Erdodi et al., 2017; Giger et al., 2010; Giromini et al., 2019a, b, c; Larrabee, 2008). The underlying assumption is that different tools may tap different feigning strategies, so that using multiple, diverse tests might provide incremental validity compared to using one test alone or two similar measures using the same method or feigning strategies. Statistically, the lower the correlation between any two tests, the greater the potential for incremental validity and better prediction (Bush et al., 2014). In addition, because of the large amount of variance shared by measures using the same method, using tests that differ in the method employed is preferable. Practitioners are therefore encouraged to include multiple SVTs and PVTs in their assessments, so as to provide incremental validity and increased signal detection (Sherman et al., 2020).

The Inventory of Problems-29 and Inventory of Problems—Memory

Because malingerers differ in the preferred strategies of defiance and different situations induce different approaches to

malingering, Viglione and Giromini developed two measures that contain multiple detection strategies, namely, the Inventory of Problems-29 (IOP-29; Viglione et al., 2017; Viglione & Giromini, 2020) and the Inventory of Problems—Memory (IOP-M; Giromini et al., 2020). The IOP-29 is typically conceived of as an SVT, and the IOP-M is a forced-choice PVT; each takes about 10 min to be completed. When used together, the IOP-29 and IOP-M might offer a quick and yet effective and multimethod validity check (Giromini et al., 2020).

The SVT component of the “IOP combo” is the IOP-29, a 29-item, self-administered test focused on the credibility of various symptom presentations (Viglione & Giromini, 2020). Two of its items have an open-ended format, whereas all other 27 offer three response options, i.e., “true,” “false,” and “doesn’t make sense.” Differently from many other measures used in the field, its chief feigning score—the False Disorder probability Score (FDS)—does not compare the responses of the test-taker against a single set of normative reference values obtained from a large sample of non-clinical responders. Instead, it considers two different sets of reference values, one coming from bona fide patients and one coming from experimental simulators. A low score suggests that the IOP-29 under examination closely resembles the IOP-29s included in the bona fide reference sample, and a high score suggests that it closely resembles those included in the simulators reference sample. Derived from logistic regression, the FDS thus is a probability score that ranges from zero to one, with higher scores reflecting less credible symptom presentations. According to the test manual (Viglione & Giromini, 2020), a score of $FDS \geq 0.50$ should offer sensitivity and specificity values of about 80% across different conditions. A higher cutoff, of $FDS \geq 0.65$, would yield a specificity of about 90%, and a lower cutoff, of $FDS \geq 0.30$, would yield a sensitivity of about 90%. Research, so far, has largely supported these claims (e.g., Gegner et al., 2021; Giromini et al., 2018; Giromini et al., 2019a, b, c; Ilgunaitė et al., 2020; Roma et al., 2019; Winters et al., 2020). IOP-29 generates similar validity results to the MMPI and PAI validity scales (Viglione et al., 2017), even outperforming the SIMS (Giromini et al., 2018) and the Rey Fifteen Item Test (Gegner et al., 2021), and providing incremental validity when combined with the TOMM (Giromini et al., 2019a, b, c) and the MMPI-2 (Giromini et al., 2019a, b, c). Thus, in their introductory description and conceptualization of the field of psychological injury and law, the Editor-in-Chief of *Psychological Injury and Law* and his colleagues referred to the IOP-29 as “a newer stand-alone SVT that has the required psychometric properties for use in forensic disability and related assessments. Its research profile is accumulating, a hallmark for use in legal settings” (Young et al., 2020; p. 9).

The PVT component of the “IOP combo” is the IOP-M (Giromini et al., 2020), a performance validity test module designed to be used in combination with the otherwise

free-standing symptom validity test, IOP-29. Its main purpose is to detect feigned memory deficits or, more broadly, cognitive impairment. It is administered immediately after completing the IOP-29. It contains 34 implicit recognition two-alternative-forced-choice test items. The results of the developmental study conducted by Giromini et al. (2020), in which 192 participants were instructed to respond honestly (honest controls) and 168 were instructed to feign mental illness (experimental simulators), suggested that the IOP-M has the potential to yield incremental validity and that it might improve classification accuracy over using the IOP-29 alone. In fact, only 6 of the 168 simulators (i.e., less than 4%) passed both the IOP-29 and IOP-M, and only 3 of the 192 honest responders (i.e., less than 2%) failed both. However, differently from the IOP-29, the IOP-M has not been thoroughly investigated. To our knowledge, only two studies to date — one in Australia by Gegner et al. (2021) focusing on feigned mild traumatic brain injury (mTBI) and one in Brazil by de Francisco Carvalho et al. (2021) focusing on post-traumatic stress disorder (PTSD) — have replicated the initial findings of Giromini et al. (2020). As such, additional research on the effectiveness of the IOP-M would be beneficial.

The Aim of the Present Study

A recent article by Areh (2020) has pointed out that forensic assessors pay little or no attention to possible malingering in Slovenia. In a quite provocative article (its title is *Forensic assessment may be based on common sense assumptions rather than science*), he summarized the psychological tests that had been used more frequently in 166 forensic personality assessments conducted in Slovenia in the period 2003–2018 and argued that “possible malingering of the person being evaluated was not detected” (p. 1). In fact, out of 166 inspected evaluations, 42 on criminal cases and 124 on civil cases, none included any stand-alone SVTs or PVTs, and very few included any broadband personality inventories that incorporate embedded measures of response style. For instance, the MMPI was used in 3 evaluations only, representing less than 2% of the total. As such, he criticized Slovenian forensic practitioners for not including in their assessments “specific psychological instruments used to detect malingering” (p. 7). It should be noted, however, that a brief literature search revealed that commonly used SVTs or PVTs such as the SIMS or TOMM have not been researched or validated in Slovenia. Thus, providing Slovenian practitioners with an empirically sound measure of negative response bias in Slovene language would be beneficial.

To respond to this call, we developed a Slovene adaptation of the IOP-29 and IOP-M and tested their joint validity. The IOP-29 has already been adapted to numerous other languages

— English, German, French, Dutch, Italian, Spanish, Brazilian and European Portuguese, traditional and simplified Chinese, and Lithuanian (www.iop-test.com). Published studies have shown solid support in its initial English language and promising adaptation to other languages (e.g., Ilgunaite et al., 2020). However, the IOP-M has been cross-validated only by Gegner et al. (2021), and no Slovene version of either IOP instrument was available, when we designed our study. The primary purpose of our research project was thus to determine how a healthy Slovene-speaking population would respond to both tests and how many simulators that present themselves as psychologically injured would get detected by the IOP instruments when both test components (i.e., the IOP-29 and IOP-M) are administered. We expected results similar to those with samples speaking other languages, that is that the IOP-29 and IOP-M would both individually discriminate simulators from an honest non-patient sample, and that the IOP-M would identify simulators who were not identified by the IOP-29.

Method

Participants

We decided to test whether the IOP-29 is able to effectively discriminate simulators of depression and schizophrenia from honest responders. Based on previous research (Giromini et al., 2020), no differences were expected between simulators of depression and schizophrenia. Thus, considering an alpha of 0.05, a power of 0.80, and an allocation ratio of 2 to 1 (two simulator groups, one control group), it was determined that 144 participants would be needed (48 in one group and 96 in the other) to detect a Cohen's *d* effect size of 0.50. Accordingly, we aimed to recruit approximately 50 participants for the honest group and approximately 50 participants for each of the two simulator groups.

The overall sample included 150 Slovenian participants, aged 18 to 75 years ($M=30.5$, $SD=13.3$), with 57 (38%) being men. To validate both measures, participants were randomly assigned to three groups with one responding honestly and the other two attempting to convince the examiner that they suffered depression or schizophrenia following a work-related accident causing physical pain. The three groups, each with 50 participants, did not differ significantly with regard to gender, $\chi^2(2)=0.68$, $p=0.71$; age, $F(2, 147)=1.24$, $p=0.29$; and education, $\chi^2(2)=5.03$, $p=0.08$. In the "honest" group, there were 19 (38%) men. This group had an average age of 28.2 years ($SD=11.1$), and 27 (54%) participants had high school education or less, and 23 (46%) participants had a bachelor's degree or more. Simulators of depression had an average age of 32.3 years ($SD=15.5$), with 34 (68%) participants having a high school education or less and 16 (32%) having bachelor's degree or

Table 1 Means (and standard deviation in parentheses) of IOP-29 FDS and IOP-M scores in different groups

	Honest	Simulator	
		Depression	Schizophrenia
IOP-29	0.16 (0.13)	0.75 (0.19)	0.78 (0.21)
IOP-M	33.5 (0.8)	28.6 (4.4)	23.1 (6.7)

more, and 17 (34%) of them were men. The third group, which was simulating schizophrenia, had an average age of 27.5 ($SD=11.1$) and included 21 (42%) men; 23 (46%) participants had a high school education or less, and 27 (54%) had a bachelor degree or more.

Measures

As with other linguistic adaptations of the IOP-29 and the IOP-M, our Slovenian versions were developed by following the classic translation-back translation procedure method (Brislin, 1980; Geisinger, 2003; Van de Vijver & Hambleton, 1996). All participants were then administered both the IOP-29 and the IOP-M in Slovene language.

Inventory of Problems-29 (Viglione & Giromini, 2020).

The IOP-29 is a self-administered test designed to assist practitioners to evaluate the credibility of symptom presentations related to various psychiatric or cognitive disorders. Its main purpose is to discriminate bona fide patients from feigners. It is composed of 29 items and administered via classic paper-and-pencil format or online, using a tablet or a PC. Items address diverse mental health symptoms, attitudes towards one's own condition, test-related behaviors, claims of impairment, and problem-solving abilities. As noted above, the chief feigning measure of the IOP-29 is the FDS, a probability value derived from logistic regression, which compares the responses of the test-taker against those provided by a group of bona fide patients and those provided by a group of experimental feigners (Viglione & Giromini, 2020). The higher the score, the lower the credibility of the

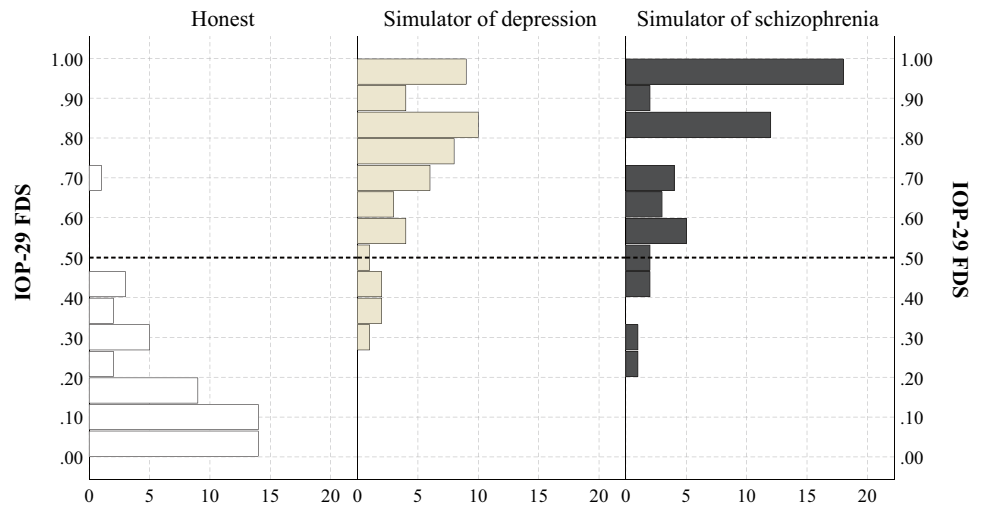
Table 2 Classification accuracy of IOP-29 FDS and IOP-M

	Honest	Simulator	
		Depression	Schizophrenia
IOP-29			
FDS $\geq .50$	49 (98%) ^a	6 (12%)	6 (12%)
FDS $< .50$	1 (2%)	44 (88%) ^b	44 (88%) ^b
IOP-M			
# of correct ≥ 30	50 (100%) ^a	25 (50%)	10 (20%)
# of correct < 30	0 (0%)	25 (50%) ^b	40 (80%) ^b

^aSpecificity

^bSensitivity

Fig. 1 Distribution of IOP-29 FDS scores by group



presentation. According to the test authors, the cutoff score of $FDS \geq 0.50$ ensures the best balance between sensitivity and specificity (Giromini et al., 2018; Viglione & Giromini, 2020; Viglione et al., 2017).

Inventory of Problems–Memory (Giromini et al., 2020). The IOP-M is administered immediately after completing the IOP-29. It consists of 34 two-alternative-forced-choice items. Each item presents two words or brief statements — one that was part of the IOP-29-item content (target) and one that was not (foil). To preserve the standard IOP-29 administration procedure, incidental memory is tested: there is no mention of a subsequent memory test or an expectation to remember the IOP-29 items. Based on Giromini et al. (2020) findings, at least 30 of the 34 IOP-M items should be answered correctly by individuals who do not suffer from relatively severe cognitive problems. As such, if the total number of IOP-M items answered correctly is lower than 30, the performance is considered to be non-credible. Conversely, a total score of ≥ 30 is interpreted as a credible result.

Procedure

The study was approved by the Ethical Committee of the Faculty of Arts, University of Ljubljana. Participants were recruited from the general population via convenience snowball sampling. That is, we first distributed flyers in the faculty and invited our family members, friends, and acquaintances to participate in the study. We also asked our participants to help us spread the word and invite their acquaintances to participate as well, if there was interest. Although participation was completely voluntary, all participants were informed that, upon the completion of data analysis, three of them would receive a 20€ Amazon voucher. Individuals who have met the inclusion criteria (having Slovenian nationality, not having any psychiatric or cognitive disorders, no familiarity with the IOP-29) were then asked to sign an informed consent form and were divided into three groups of 50. The first group was asked to answer as honestly as possible, the second group was instructed to

Fig. 2 Distribution of IOP-M scores by group

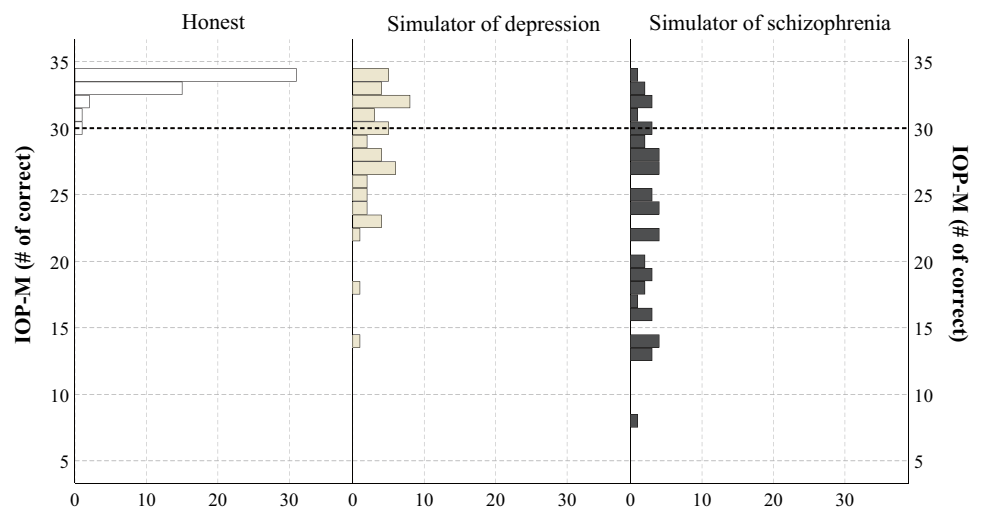
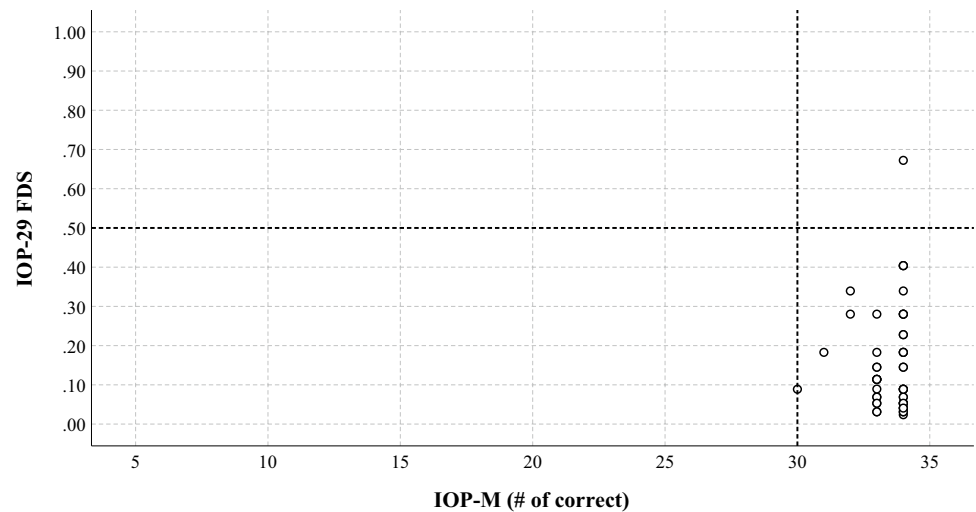


Fig. 3 Scatterplot IOP-29 versus IOP-M scores within the “honest” group



simulate schizophrenia, and the third group was instructed to simulate depression. Specifically, participants assigned to the schizophrenia and depression group were presented with a short vignette on the situation in which being diagnosed with mental illness would lead to an economic advantage and were instructed to take the tests as if they wanted to convince the examiner that they were experiencing symptoms associated with schizophrenia or depression, respectively. A list of symptoms of the disorder to be feigned was presented. Additionally, both groups were cautioned not to overdo the expression of the disorder in order to not be detected as feigners. All participants were administered a short sociodemographic questionnaire in addition to the IOP-29 and IOP-M. For each participant, FDS values were calculated

using the official IOP-29 scoring program, which can be found at www.iop-test.com, and IOP-M errors were counted using a scoring sheet created by the test authors (Giromini et al., 2020).

Results

Table 1 shows the scores obtained on the two instruments in different groups of participants. As expected, the IOP-29 FDS values statistically significantly differed among the three groups, $F(2, 147) = 193.52, p < 0.001$. More specifically, Bonferroni corrected post hoc tests revealed that the “honest” group scored notably lower than both simulators

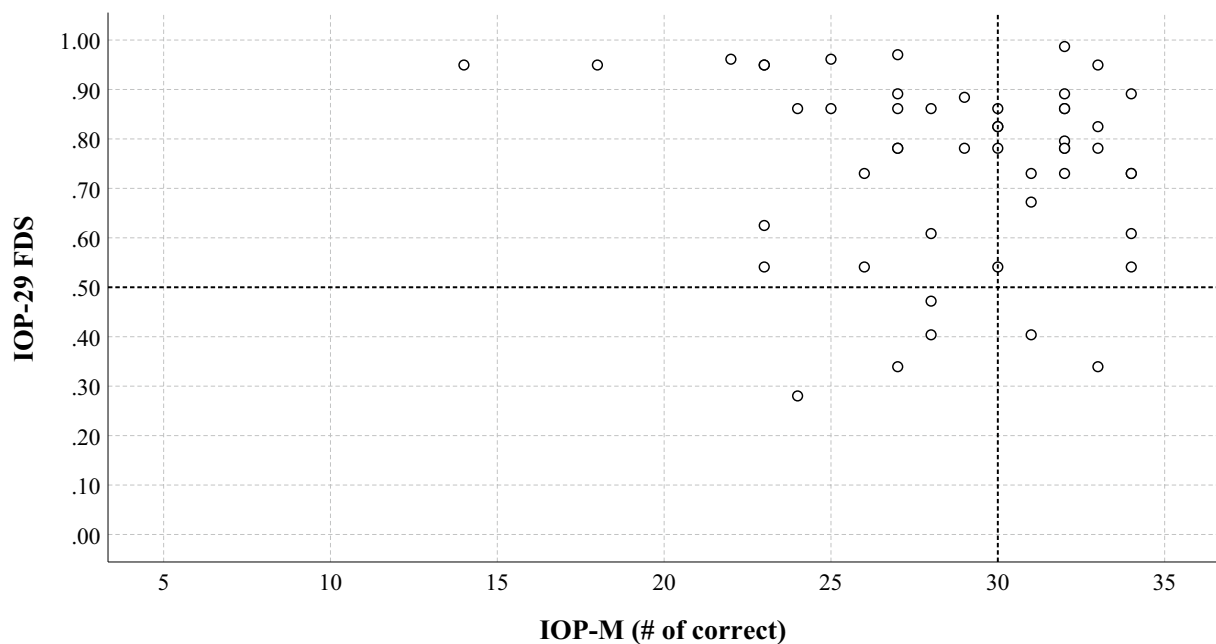
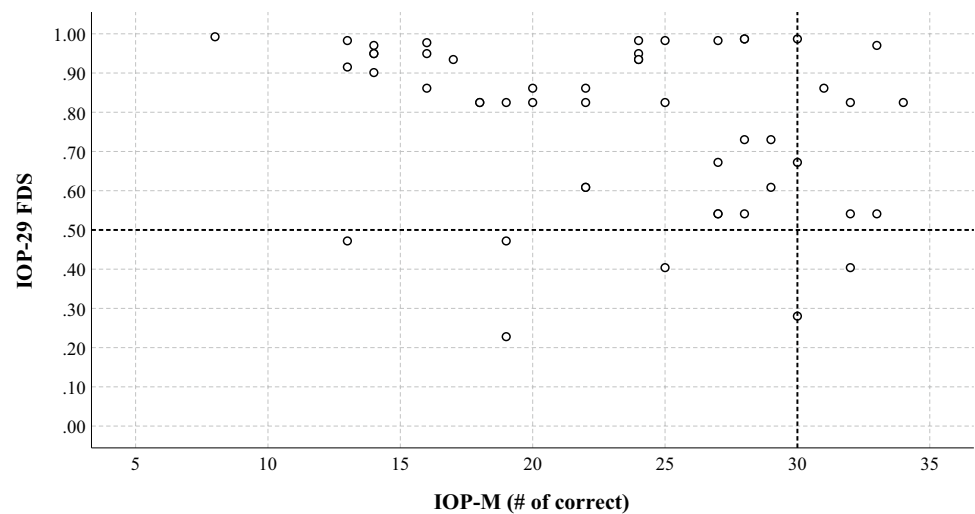


Fig. 4 Scatterplot IOP-29 versus IOP-M scores within the simulators of depression

Fig. 5 Scatterplot IOP-29 versus IOP-M scores within the simulators of schizophrenia



of depression ($d = 3.69$, $p < 0.001$) and simulators of schizophrenia ($d = 3.56$, $p < 0.001$), whereas the two simulator groups did not statistically significantly differ from each other ($d = 0.14$, $p \approx 1.00$). The IOP-M yielded statistically significant group differences as well, $F(2, 147) = 62.93$, $p < 0.001$. In this case, however, all pairwise comparisons were statistically significant: the “honest” group had a notably higher IOP-M score than both the simulators of depression ($d = 1.55$, $p < 0.001$) and the simulators of schizophrenia ($d = 2.18$, $p < 0.001$), and the simulators of depression scored higher than the simulators of schizophrenia ($d = 0.97$, $p < 0.001$).

In terms of classification accuracy, the standard IOP-29 cutoff score of $FDS \geq 0.50$ yielded a specificity of 0.98 and a sensitivity of 0.88 (Table 2). Noteworthy, the same exact sensitivity results emerged when inspecting feigned depression and feigned schizophrenia. The IOP-M showed a perfect specificity, but its sensitivity was notably higher in the schizophrenia simulators’ group (0.80) than it was for the simulators of depression (0.50). To offer a better appreciation of the performance of both the IOP-29 and IOP-M, Figs. 1 and 2 show full frequency distributions of FDS scores across the three groups.

To test the incremental validity, three scatterplots were examined. As shown in Fig. 3, the only false-positive generated by the IOP-29 FDS was correctly classified as a (true) negative outcome by the IOP-M. Figure 4 shows that, when inspecting the simulator depression group, four of the six false-negative classifications generated by the IOP-29 FDS were correctly classified as (true) positive outcomes by the IOP-M. Likewise, Fig. 5 shows that, when inspecting the simulator schizophrenia group, four of the six false-negative classifications generated by the IOP-29 FDS were correctly classified as (true) positive outcomes by the IOP-M. Also noteworthy, in the “honest” group, no one failed both the IOP-29 and IOP-M, and in each of the simulator subgroups,

only two cases out of 50 passed both tests. The overall frequency with which both the IOP-29 and IOP-M misclassified the same case was 4 out of 150, i.e., 2.7%.

Discussion

The aim of our study was to test the validity of our Slovenian adaptation of the Inventory of Problems-29 (IOP-29; Viglione & Giromini, 2020) and Inventory of Problems–Memory (IOP-M; Giromini et al., 2020). Our statistical analyses showed that, when comparing the mean values of the IOP-29 FDS in the honest versus the two simulating groups, the IOP-29 discriminated significantly between the groups. However, no differences between the two simulating groups emerged. The average IOP-29 FDS values found in the two simulating groups (0.75 for the depression simulants and 0.78 for the schizophrenia simulants) are similar to those observed in experimental simulator samples in other studies (e.g., 0.77 in Ilgunaite et al., 2020; 0.82 in Giromini et al., 2019a, c). Honest versus both simulator groups also differed for the IOP-M, but the pairwise comparison between the two simulating groups was significant as well. More importantly, combining the results of the IOP-29 with those of the IOP-M remarkably increased classification accuracy, both when inspecting feigned depression and when inspecting feigned schizophrenia. All in all, thus, our study provides additional support for the growing research base for using the IOP-29 and IOP-M in applied settings. Besides, this article fills an important gap within the Slovenian forensic context (Areh, 2020), given that (to our knowledge) no stand-alone SVTs or PVTs with research support were available to Slovenian practitioners, prior to this publication.

The effect sizes generated by the IOP-29 when comparing the honest group against the two simulator groups were very

large, $d \geq 3.56$ (for characterization of d values in malingering research, please see Rogers et al., 2003). Besides, the Slovenian IOP-29 yielded excellent classification accuracy both in terms of specificity (98%) and sensitivity (88% in both simulating groups). Given that, the Slovenian IOP-29 proved to be very accurate in our study. Also, noteworthy, there were no significant differences between simulators who faked depression versus schizophrenia. This finding is consistent with a previous study conducted by Giromini et al. (2019c), in which IOP-29 scores produced by feigners of depression, schizophrenia, mTBI, and PTSD did not significantly differ from each other. As such, our study seems to confirm Viglione and Giromini's (2020) claim that the IOP-29 likely performs similarly well when used in remarkably different contexts and with different symptom presentations. Such generalizability of a single cutoff score across different disorders, cultures, and languages has rarely if ever demonstrated for other SVTs.

Another fact that deserves mentioning is that, in our investigation, the IOP-29 performed similarly well as it previously did in other simulation studies conducted in Italy (Giromini et al., 2019b), Lithuania (Ilgunaite et al., 2020), Portugal (Giromini et al., 2019a), UK (Winters et al., 2020), and Australia (Gegner et al., 2021). To some extent, thus, our investigation also contributes to the growing empirical research suggesting that the IOP-29 can be applied cross-culturally with no need to make any significant adjustments to its FDS formula and cutoffs.

When compared to the IOP-29, the IOP-M showed higher specificity (100%), but lower sensitivity ($\approx 90\%$ for the IOP-29; $\leq 80\%$ for the IOP-M). Furthermore, the sensitivity of the IOP-M was notably higher when considering simulators of schizophrenia (80%) rather than when considering simulators of depression (50%). This finding could be possibly attributed to the nature of the IOP-M and the difference between the two conditions. More specifically, one might speculate that, while people simulating schizophrenia likely linked symptoms of abnormal reality interpretation with memory deficits, depression simulators perhaps did not recognize memory deficits as the most common symptom of depression — and this may be the reason why the difference between the two simulating groups was found on IOP-M. Nevertheless, the examination of scatterplots revealed that the combination of both test components (IOP-29 and IOP-M) yielded promising results regarding the classification of the participants both when testing feigned depression and when testing feigned schizophrenia. Indeed, the information that, with only two exceptions, at least one of the test components was able to identify the feigners' testing results as invalid is very encouraging. Similarly, importantly, the only one false-positive result generated by the IOP-29 was correctly classified as a credible performance by the IOP-M. Thus, it does seem that adding the IOP-M contributes

to both the sensitivity and specificity of the original IOP-29, consistent with Giromini et al. (2020) and Gegner et al. (2021).

Several limitations, however, should be underscored. First, since we did not include in our study any other SVTs or PVTs, comparative validity could not be investigated. On the other hand, the lack of such instruments in Slovene language is exactly one of the primary reasons why this study was initiated. Second, our findings are limited by the fact that we did not have access to any clinical samples, so that our study essentially is a sensitivity study. Actual overall classification accuracy is likely to be lower because of reduced specificity if clinical samples were employed. Thus, research with genuinely impaired individuals and feigners is needed to better appreciate the specificity of the Slovenian IOP-29 and IOP-M. Third, this study only investigated feigned depression and feigned schizophrenia, so that additional research is needed to appreciate the extent to which the Slovenian IOP-29 and IOP-M could be used in applied settings in which other problems (e.g., pain, mTBI symptoms, etc.) may be feigned. Fourth, as is the case for all simulation/analogue studies, the ecological validity of our investigation may be questioned, as there is no way to assess whether real-life malingerers would adopt the same strategies utilized by our experimental simulators, when pretending to be mentally ill. Additionally, group assignment was a quasi-independent variable, and we did not manipulate the valid versus invalid response condition but randomly assigned participants to the given instruction. The internal validity of such a research paradigm depends on the fidelity with which participants perform the given instructions (Rai et al., 2019). In a study using the experimental malingering paradigm, An et al. (2019) found that the group that feigned cognitive decline performed well despite being asked to try to feign cognitive deficits, possibly because they made an effort to achieve credible feigning as required by the instructions given, and that some participants in the control group performed worse than expected according to their abilities, possibly due to lack of interest and low effort, leading to an underestimation of the difference between the control and simulating groups. Unfortunately, we could not rely on established SVTs or PVTs as criteria for monitoring participants' compliance with the given instructions, as such instruments do not exist in Slovene. In the absence of a gold standard, we could recommend using tests such as the TOMM or the Rey Fifteen Item Test, which is not language-based, to determine the extent to which simulators follow the feigning instructions. Another option would be to include bilingual participants who speak both Slovene and English (or some other languages, e.g., Italian) and give them the IOP in Slovene and another consolidated SVT (e.g., the MMPI, SIMS, etc.) in another language to check their compliance with the feigning instructions.

Nevertheless, this study is the first to independently replicate Giromini et al. (2020) encouraging findings concerning the potential utility of the IOP-M when investigating feigned depression and feigned schizophrenia, and the first to contribute to the study of the IOP-29 and IOP-M within a Slovenian sample. Given the encouraging results, we invite Slovenian researchers and practitioners to contact the corresponding author, in case they were interested in using or further researching the IOP instruments.

Funding The authors acknowledge the financial support from the Slovenian Research Agency (research core funding No. P5-01110).

Data Availability For test security reasons, our data will not be placed in an open access repository. However, we will be willing to share it with interested readers upon reasonable request.

Declarations

Ethics Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The study was approved by the Ethical Committee of Faculty of Arts, University of Ljubljana, Approval No. 172-2019.

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

Conflict of Interest The fifth and sixth authors declare that they own a share in the corporate (LLC) that possesses the rights to Inventory of Problems. The other five authors declare that they have no conflict of interest to report.

References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders*. (5th ed.). American Psychiatric Association.
- An, K. Y., Charles, J., Ali, S., Enache, A., Dhuga, J., & Erdodi, L. A. (2019). Reexamining performance validity cutoffs within the Complex Ideational Material and the Boston Naming Test-Short Form using an experimental malingering paradigm. *Journal of Clinical and Experimental Neuropsychology*, *41*(1), 15–25. <https://doi.org/10.1080/13803395.2018.1483488>
- Areh, I. (2020). Forensic assessment may be based on common sense assumptions rather than science. *International Journal of Law and Psychiatry*, *71*, 101607. <https://doi.org/10.1016/j.ijlp.2020.101607>
- Ben-Porath, Y. S., & Tellegen, A. (2020a). *MMPI-3 Manual for administration, scoring, and interpretation*. University of Minnesota Press.
- Ben-Porath, Y. S., & Tellegen, A. (2020b). *MMPI-3 Technical manual*. University of Minnesota Press.
- Boone, K. B. (2013). *Clinical practice of forensic neuropsychology*. New York, NY: Guilford.
- Brislin, R. W. (1980). Translation and content analysis of oral and written material. In H. C. Triandis & J. W. Berry (Eds.), *Handbook of cross-cultural psychology* (Vol.1., pp. 389–444). Boston, MA: Allyn & Bacon.
- 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*(3), 197–205. <https://doi.org/10.1007/s12207-014-9198-7>
- de Francisco Carvalho, L., Reis, A., Colombarolli, M. S., Pasian, S. R., Miguel, F. K., Erdodi, L. A., Viglione, D. J., & Giromini, L. (2021). *Discriminating feigned from credible PTSD symptoms: A validation of a Brazilian version of the Inventory of Problems-29 (IOP-29)*. Advanced online publication. <https://doi.org/10.1007/s12207-021-09403-3>
- Chafetz, M., & Underhill, J. (2013). Estimated costs of malingered disability. *Archives of Clinical Neuropsychology*, *28*(7), 633–639. <https://doi.org/10.1093/arclin/act038>
- Erdodi, L. A., Abeare, C. A., Lichtenstein, J. D., Tyson, B., & T., Kucharski, B., Zuccato, B. G., & Roth, R. M. (2017). Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) processing speed scores as measures of non-credible responding – The third generation of embedded performance validity indicators. *Psychological Assessment*, *29*(2), 148–157. <https://doi.org/10.1037/pas0000319>
- Gegner, J., Erdodi, L. A., Giromini, L., Viglione, D. J., Bosi, J., & Brusadelli, E. (2021). An Australian study on feigned mTBI using the Inventory of Problems–29 (IOP-29), its Memory Module (IOP-M), and the Rey Fifteen Item Test (FIT). *Applied Neuropsychology: Adult*. Advance online publication. <https://doi.org/10.1080/23279095.2020.1864375>
- Geisinger, K. F. (2003). Testing and assessment in cross-cultural psychology. In J. R. Graham, J. A. Naglieri, & I. B. Weiner (Eds.), *Handbook of psychology (Vol. 10). Assessment Psychology* (pp. 95–118). New Jersey, NJ: John Wiley & Sons. <https://doi.org/10.1002/0471264385.wei1005>
- Giger, P., Merten, T., Merckelbach, H., & Oswald, M. (2010). Detection of feigned crime-related amnesia: A multi-method approach. *Journal of Forensic Psychology Practice*, *10*, 440–463. <https://doi.org/10.1080/15228932.2010.489875>
- 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 Malingering (TOMM) in symptom validity assessment: A study with a Portuguese sample of experimental feigners. *Applied Neuropsychology: Adult*, *27*(6), 504–516. <https://doi.org/10.1080/23279095.2019.1570929>
- Giromini, L., Carfora Lettieri, S., Zizolfi, S., Zizolfi, D., Viglione, D., Brusadelli, E., Perfetti, B., Angiola di Carlo, D., & Zennaro, A. (2019b). 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). *Psychological Injury and Law*, *12*, 212–224. <https://doi.org/10.1007/s12207-019-09357-7>
- Giromini, L., Viglione, D., 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*, 340–350. <https://doi.org/10.1007/s12207-018-9314-1>
- Giromini, L., Viglione, D. J., Pignolo, C., & Zennaro, A. (2019c). An Inventory of Problems–29 sensitivity study investigating feigning of four different symptom presentations via malingering experimental paradigm. *Journal of Personality Assessment*, *102*, 563–572. <https://doi.org/10.1080/00223891.2019.1566914>
- Giromini, L., Viglione, D., Zennaro, A., Maffei, A., & Erdodi, L. A. (2020). SVT meets PVT: Development and initial validation of the Inventory of Problems-Memory (IOP-M). *Psychological Injury and Law*, *13*, 261–274. <https://doi.org/10.1007/s12207-020-09385-8>

- Green, P., Allen, L. M., & Astner, K. (1996). *The Word Memory Test: A user's guide to the oral and computer administered forms*. CogniSyst Inc.
- Ilgunaite, G., Giromini, L., Bosi, J., Viglione, D. J., & Zennaro, A. (2020). A clinical comparison simulation study using the Inventory of Problems-29 (IOP-29) with the Center for Epidemiologic Studies Depression Scale (CES-D) in Lithuania. *Applied Neuropsychology: Adult*. Advance online publication. <https://doi.org/10.1080/23279095.2020.1725518>
- Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist*, 22, 666–679. <https://doi.org/10.1080/13854040701494987>
- Larrabee, G. J., Millis, S. R., & Meyers, J. E. (2009). 40 plus or minus 10, a new magical number: Reply to Russell. *The Clinical Neuropsychologist*, 23, 841–849. <https://doi.org/10.1080/13854040902796735>
- Miller, H. A. (2001). *M-FAST: Miller forensic assessment of symptoms test professional manual*. Psychological Assessment Resources Inc.
- Millon, T., Grossman, S., & Millon, C. (2015). *Millon Clinical Multiaxial Inventory-IV (MCMI-IV) manual*. Bloomington, MN: NCS Pearson.
- Mittenberg, W., Patton, C., Morgan, E., & Condit, D. (2003). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, 24, 1094–1102. <https://doi.org/10.1076/jcen.24.8.1094.8379>
- 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.). Psychological Assessment Resources.
- Rai, J. K., An, K. Y., Charles, J., Ali, S., & Erdodi, L. A. (2019). Introducing a forced choice recognition trial to the Rey Complex Figure Test. *Psychology & Neuroscience*, 12(4), 451–472. <https://doi.org/10.1037/pne0000175>
- Resnick, P. J. (1984). The detection of malingered mental illness. *Behavioral Sciences & the Law*, 2(1), 21–38. <https://doi.org/10.1002/bsl.2370020104>
- 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., & Bender, D. (2018). *Clinical assessment of malingering and deception*. (3rd ed.). Guilford Press.
- Rogers, R., Kropp, P., Bagby, M., & Dickens, S. (1992). Faking specific disorders: A study of the structured interview of reported symptoms (SIRS). *Journal of Clinical Psychology*, 48(5), 643–648. [https://doi.org/10.1002/1097-4679\(199209\)48:5%3c643::AID-JCLP2270480511%3e3.0.CO;2-2](https://doi.org/10.1002/1097-4679(199209)48:5%3c643::AID-JCLP2270480511%3e3.0.CO;2-2)
- Rogers, R., Sewell, K. W., & Gillard, N. D. (2010). *Structured interview of reported symptoms, second edition: professional test manual*. (2nd ed.). Psychological Assessment Resources.
- Rogers, R., Sewell, K. W., Martin, M. A., & Vitacco, M. J. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI-2 and malingering. *Assessment*, 10(2), 160–177. <https://doi.org/10.1177/1073191103010002007>
- Rogers, R., Velsor, S. F., & Williams, M. M. (2020). *A brief commentary on SIRS versus SIRS-2 critiques*. Advance online publication. <https://doi.org/10.1007/s12207-020-09379-6>
- Roma, P., Giromini, L., Burla, F., Ferracuti, S., Viglione, D. J., & Mazza, C. (2019). Ecological validity of the inventory of problems-29 (IOP-29): An Italian study of court-ordered, psychological injury evaluations using the Structured Inventory of Malingered Symptomatology (SIMS) as criterion variable. *Psychological Injury and Law*, 13, 57–65. <https://doi.org/10.1007/s12207-019-09368-4>
- Sherman, E. M. S., Slick, D. J., & Iverson, G. L. (2020). Multidimensional malingering criteria for neuropsychological assessment: A 20-year update of the malingered neuropsychological dysfunction criteria. *Archives of Clinical Neuropsychology*, 35, 735–764. <https://doi.org/10.1093/arclin/aca019>
- Slick, D., Hopp, G., Strauss, E., & Thompson, G. B. (2005). *VSVT Victoria Symptom Validity Test*. Odessa, FL: Psychological Assessment Resources.
- 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.
- Tombaugh, T. N. (1996). *Test of memory malingering (TOMM)*. New York, USA: Multi Health Systems.
- Van de Vijver, F., & Hambleton, R. K. (1996). Translating tests. *European Psychologist*, 1(2), 89–99. <https://doi.org/10.1027/1016-9040.1.2.89>
- 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. <https://doi.org/10.1080/00223891.2016.1233882>
- Viglione, D. J., & Giromini, L. (2020). *Inventory of Problems–29: Professional manual*. Columbus, OH: IOP-Test, LLC.
- Winters, C. L., Giromini, L., Crawford, T. J., Ales, F., Viglione, D. J., & Warmelink, L. (2020). An Inventory of Problems–29 (IOP–29) study investigating feigned schizophrenia and random responding in a British community sample. *Psychiatry, Psychology and Law*. Advance online publication. <https://doi.org/10.1080/13218719.2020.1767720>.
- Young, G. (2015). Malingering in forensic disability-related assessments: Prevalence 15±15%. *Psychological Injury and Law*, 8(3), 188–199. <https://doi.org/10.1007/s12207-015-9232-4>
- Young, G. (2019). The cry for help in psychological injury and law: Concepts and review. *Psychological Injury and Law*, 12(3–4), 225–237. <https://doi.org/10.1007/s12207-019-09360-y>
- Young, G., Foote, W. E., Kerig, P. K., Mailis, A., Brovko, J., Kohutis, E. A., McCall, S., Hapidou, E. G., Fokas, K. F., & Goodman-Delahunty, J. (2020). Introducing psychological injury and law. *Psychological Injury and Law*, 13, 452–463. <https://doi.org/10.1007/s12207-020-09396-5>

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