



The Cross-Cultural Applicability of the Inventory of Problems – 29 (IOP-29): A Replication of Akca et al. (2023) Using a Serbian Sample

Marija Volarov¹ · Mina Velimirović¹ · Irena Bošković² · Ali Y. E. Akca^{3,4} · Luciano Giromini⁴

Received: 8 May 2024 / Accepted: 16 July 2024 / Published online: 7 August 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

Some widely agreed-upon, official recommendations for professionals conducting psychological assessments suggest employing multiple symptom validity tests (SVTs) to screen the validity of symptom reports. Yet, SVTs are rarely validated in languages other than English, and no free-standing SVT exists in Serbia. To address this gap and stimulate further research on symptom validity within populations from the Balkans, we developed and tested a Serbian version of the Inventory of Problems – 29 (IOP-29). Following the same procedures used in prior IOP-29 validation studies (e.g., Akca et al., 2023), we administered the Serbian IOP-29 to 110 adult volunteers from Serbia. Participants completed the IOP-29 three times under different conditions: responding honestly, randomly, or by feigning a mental disorder (schizophrenia, depression, or post-traumatic stress disorder). We examined the utility of both the False Disorder Probability Score (FDS), which is the chief feigning index of the IOP-29, and of a new index embedded in the IOP-29, which is aimed at detecting random or careless responding. Overall, our results demonstrated that the FDS effectively differentiated between feigned and honest presentations, achieving a sensitivity of 0.86 and a specificity of 0.96 when using the standard cutoff ($FDS \geq 0.50$). In addition, the random responding index also successfully identified random responding, achieving a sensitivity of 0.64 and a specificity greater than 0.90 when using a midrange cutoff of $T \geq 67$. These findings closely align with outcomes of Akca et al. (2023) and support meta-analytic literature reviews on the IOP-29. More broadly, this study advances and encourages further exploration of symptom validity testing in culturally diverse populations.

Keywords Inventory of Problems · IOP-29 · Serbian · Malingering · Random

The American Academy of Clinical Neuropsychology (AACN) recently issued a consensus statement emphasizing the critical role of Performance Validity Tests (PVTs) and Symptom Validity Tests (SVTs) in assessing the credibility of presented clinical complaints (Sweet et al., 2021). Specifically, aligning with the updated multidimensional criteria for cognitive, somatic, and psychiatric malingering by Sherman et al. (2020), Sweet et al. (2021) advocated for

the inclusion of multiple PVTs and multiple SVTs in nearly all clinical and forensic psychological evaluations. One of the main reasons behind such a recommendation is the significant prevalence of invalid (i.e., fabricated or exaggerated) symptom reports, accounting for approximately $15 \pm 15\%$ of all symptom reports. This prevalence is estimated to be even higher in forensic assessments and when psychological complaints are being evaluated (Young, 2015).

Yet, the mandate to use multiple SVTs may be very challenging in certain contexts. Firstly, despite the substantial number of published PVTs, there is a notable scarcity of validated self-report SVTs accessible to professionals conducting validity assessments. Indeed, in the aforementioned consensus statement, Sweet et al. (2021) noted that “there remains a need for the development and validation of new SVTs, including free-standing measures and embedded measures within current self-report symptom measures” (p. 1084). Secondly, SVTs, as highly language-dependent, are arguably more susceptible to cultural

✉ Marija Volarov
marija.volarov@ff.uns.ac.rs

¹ Department of Psychology, Faculty of Philosophy, University of Novi Sad, Novi Sad, Serbia

² Erasmus School of Social and Behavioural Sciences, Erasmus University Rotterdam, Rotterdam, the Netherlands

³ Forensic Psychology Section, Maastricht University, Maastricht, The Netherlands

⁴ Department of Psychology, University of Turin, Turin, Italy

factors compared to nonverbal PVTs, yet the majority of these tests are only available in English. Hence, the administration of SVTs to non-native English speakers is especially challenging.

To address this clinical and forensic demand, the present study aimed to develop and provide an initial validation of a Serbian language version of the Inventory of Problems – 29 (IOP-29; Viglione & Giromini, 2020). The Inventory of Problems – 29 is a brief free-standing SVT that has shown promising cross-cultural adaptability in various studies conducted worldwide (Giromini & Viglione, 2022). The reason we chose Serbian as the target language of our study is that Serbian belongs to the South Slavic branch of the Slavic language family and is widely spoken in several Balkan countries. Thus, due to the linguistic similarities with Croatian, Bosnian, Montenegrin, and North Macedonian, the successful validation of Serbian version could facilitate the development of several similar IOP-29 versions within other Balkan countries. Moreover, due to the conspicuous external migration from the territory of the former Yugoslavia after 1990s, there are now millions of people speaking a Slavic language throughout the European Union and in many other countries (Šrubař & Fňukal, 2010). And yet, quite surprisingly, there are currently no validated free-standing self-report SVTs in Serbian. To our knowledge, Serbian practitioners can only rely on validity scales embedded in other tests, such as in the second edition of the Minnesota Multiphasic Personality Inventory (Biro, 2001; Butcher et al., 2001), lacking any free-standing SVTs options.

Regarding our choice to focus on the IOP-29 as the target SVT of our study, we relied on the review of available SVTs published by Giromini et al. (2022). The authors concluded that the IOP-29 is the shortest validated SVT with a solid research base, making it a viable addition to the multimethod test battery for clinical and forensic assessors. Due to its brevity, administration of IOP-29 does not impose undue burden on either the assessor or the respondent, thus, we concluded that the development and initial validation of a Serbian version of the IOP-29 would be the most efficient and valuable option for Serbian practitioners conducting psychological assessments. In order to compare the psychometric properties of our Serbian IOP-29 version with other available IOP-29 versions, we employed the same research design as previous studies (Akca et al., 2023) to minimize potential confounding factors. More specifically, we adopted the exact procedures used in a study on the Turkish IOP-29 (Akca et al., 2023), which was a replication of studies conducted on English (Winters et al., 2021) and Italian (Giromini et al., 2020a) IOP-29 versions. By doing so, any potential discrepancies in results could then only be attributed to the difference in the effectiveness of the Serbian IOP-29 version compared to other IOP-29 versions.

Method

Development of the Serbian IOP-29

The development of the Serbian version of the IOP-29 for use in our study followed the classic “back-translation” method (Brislin, 1980; Geisinger, 2003; Van de Vijver & Hambleton, 1996), a widely used approach for producing equivalent versions of a measure across different languages and cultures. First, one of the authors of this article, proficient in both English and Serbian, met with one of the IOP-29 authors to discuss the intended meaning of individual IOP-29 items, which led to the first draft of Serbian IOP-29 version. Next, another author of this article, also proficient in both languages and blind to the item content of the IOP-29, back-translated this newly developed version into English. This allowed the IOP-29 author to assess the degree to which the intended meaning of the IOP-29 item content was preserved in this initial Serbian version. Finally, the two researchers involved in the translation and back-translation process, along with another Serbian psychologist, also an author of this article, proficient in both languages, met with the IOP-29 author to address any possible inconsistencies and make final refinements to the Serbian version. This final Serbian version was then pilot-tested with a few volunteers. This process helped to ensure that the original content was preserved through translation. Following this pilot testing, the Serbian version was deemed ready for empirical testing.

Participants

Prospective participants were recruited online via Qualtrics. Only those meeting specific criteria were eligible to participate: They had to be native Serbian speakers, aged at least 18, fluent in Serbian, and without diagnosed mental disorders or poor self-reported mental health. Informed consent was also required for participation. Additionally, participants who were admitted to the survey but then failed to pass the manipulation or inattention checks dispersed throughout the experiment were also excluded from the study. This procedure led to the exclusion of 31 records.

The final sample comprised valid records only, consisting of 110 adult volunteers from Serbia. The vast majority, i.e., 82.7%, identified themselves as “female.” The remaining participants identified themselves as “male,” except for one person who chose not to disclose their gender identity. Participants’ ages ranged from 19 to 64 ($M = 30.8$; $SD = 8.5$). Approximately a third of the sample were students (36%): 29 undergraduates (70.7%), 6 master’s students (14.6%), and 6 doctoral students (14.6%). Fifteen participants (13.6%)

identified as mental health professionals. All participants reported Serbian as their mother tongue, and 106 (96.4%) identified as Serbian ethnicity. None of the final sample had diagnosed mental disorders, and their self-rated mental health was “extremely good” (28.2%), “good” (58.2%), or “neither good nor bad” (13.6%).

Measures

The Inventory of Problems – 29 (IOP-29; Viglione & Giromini, 2020) The Inventory of Problems – 29 is self-administered free standing SVT designed to discriminate credible from noncredible presentations of cognitive/neuropsychological, depressive/anxious, trauma-related, and/or psychotic problems. The items of the IOP-29 were extracted from a pool of several hundred items whose combined effectiveness had been demonstrated in many independent studies in the United States (Viglione et al., 2017). They assess various aspects of test-takers’ psychological functioning, and the responses are entered on the IOP test scoring platform (www.iop-test.com) which provides the False Disorder Probability Score (FDS). The IOP-29 FDS is a probability value derived from logistic regression, and it ranges from zero to one, with higher values indicating a less credible presentation. According to the IOP-29 manual (Viglione & Giromini, 2020), the standard IOP-29 cutoff score is an FDS value of ≥ 0.50 , which should result in sensitivity and specificity values of around 0.80. However, in cases where higher specificity is required, the IOP-29 authors recommend a cutoff score of $\text{FDS} \geq 0.65$, which should yield a specificity of about 0.90 with a sensitivity of about 0.70. If the assessor is using the IOP-29 as a screening measure and is therefore aiming for higher sensitivity, then the recommended cutoff value would be $\text{FDS} \geq 0.30$, which should result in a sensitivity of about 0.90 with a specificity of about 0.60.

In the last 5 years, numerous research articles have been published addressing the validity of the IOP-29 in different cultural contexts, including the United States (Holcomb et al., 2023), Canada (Abeare et al., 2021), Australia (Gegner et al., 2022), the United Kingdom (Bosi et al., 2022), Romania (Crişan, 2023), Slovenia (Šömen et al., 2021), Brazil (Carvalho et al., 2021), Italy (Giromini et al., 2018; Roma et al., 2020, 2023), Belgium (Blavier et al., 2023), Norway (Grønnerød et al., 2023), Spain (Puente-López et al., 2023a), France (Banovic et al., 2022), the Netherlands (Boskovic et al., 2022), Türkiye (Akca et al., 2023), Lithuania (Ilgu-naite et al., 2022), and Portugal (Giromini et al., 2020b). What is impressive about this extensive body of research is that the validity of the IOP-29 appears to remain very stable when moving from one study in a particular cultural context to another. In fact, a recent quantitative literature review by Giromini and Viglione (2022) found that the weighted

mean sensitivity of the standard IOP-29 cutoff of ≥ 0.50 was 0.86, with a very low weighted standard deviation of 0.07; likewise, the weighted mean specificity was 0.92, with a very low weighted standard deviation of 0.06. Along similar lines, a meta-analytic study by Puente-López et al. (2023b) has recently shown that language is not a significant moderator of the effectiveness of the IOP-29 when using cutoff values ≥ 0.50 and ≥ 0.65 . Given this, it can be assumed that the validity of the IOP-29 is likely to be transferable to other Slavic language versions, too.

Recently, in addition to the FDS, another IOP-29 index has been introduced. Specifically, Giromini et al. (2020c) introduced the Random Responding Scale (RRS), a composite score designed to signal the presence of a careless style or content-unrelated distortion in the test-taker’s IOP-29 responses. Ideally, this scale should help the assessor determine whether the test-taker paid sufficient attention to and/or adequately comprehended the meaning of the IOP-29 items. Scaled as a T-score, RRS values greater than 65T (i.e., 1.5 SDs above the mean) or 70T (i.e., 2 SDs above the mean) might signal the possible careless or random response style. However, to date, only two studies (i.e., Akca et al., 2023; Winters et al., 2021) have independently replicated the original findings of Giromini et al. (2020c), and different cutoff values for the RRS have been proposed. While Giromini et al. (2020c) recommended a cutoff value of $T \geq 61$, Akca et al. (2023) suggested that a cutoff value of $T \geq 67$ was required for the RRS to achieve a specificity of 0.90,¹ and a cutoff value of $T \geq 71$ was required for the RRS to achieve a specificity of 0.95. Accordingly, the calculation of the RRS has not yet been made available on the official IOP tests scoring platform (www.iop-test.com), and our research also aimed to contribute to the further investigation of this newer IOP scale.

Procedure

As noted above, our study is a replication of Akca et al. (2023) within a Serbian sample, so we used the same procedures (e.g., same instructions, same vignettes, etc.) described in Akca et al. (2023). First, before starting participant recruitment, we prepared a document summarizing the entire research project to be submitted for research ethics approval to the institutional review board of the Department of Psychology, University of Novi Sad. Then, after receiving

¹ To be precise, Akca et al. (2023) originally suggested using a cutoff of 66.5 to achieve a specificity of .90. However, because T scores are seldom reported with decimal values and a difference of 0.5 T points corresponds to a negligible Cohen’s d difference of $d = 0.05$, we deemed it preferable to round the scores of the RRS to integer values. Given that, and because our post hoc analyses revealed that using either cutoff ($T \geq 66.5$ versus $T \geq 67$) yielded nearly identical results, in the current article we decided to use $T \geq 67$ as the cutoff for the RRS rather than $T \geq 66.5$, as originally suggested by Akca et al. (2023).

formal approval, the study was advertised on social media, and the Qualtrics survey link was shared on Facebook, Instagram, and LinkedIn. Additionally, recruitment was further promoted through snowballing. Those who agreed to participate were then informed they should complete the same test (i.e., IOP-29) three times under three different conditions: responding honestly (HON condition), responding randomly (RND condition), and feigning or simulating a particular mental health disorder (SIM condition).

The order of administration of HON, RND, and SIM conditions was randomized and counterbalanced across participants. In the HON condition, respondents were instructed to provide honest answers to IOP-29. In the random condition, they were instructed to complete IOP-29 randomly, ideally without reading the items. For the SIM condition, participants were randomly assigned to one of the three groups: schizophrenia, depression, or post-traumatic stress disorder (PTSD), so that each respondent simulated only one of the three disorders. To facilitate feigning, a vignette describing a situation where one might be motivated to feign was provided, along with the list of symptoms of the particular disorder the respondent was asked to simulate (for more details, see Rogers & Gillard, 2011; Viglione et al., 2017). Respondents were also warned to be cautious while feigning, as “overdoing it” would reveal that they were faking the disorder instead of genuinely experiencing its symptoms. Additionally, all respondents were informed that the two “best feigners” (operationalized by us as those with the lowest FDS scores) would be awarded 2,000 RSD (approximately \$18), provided they left their email addresses. Notably, leaving an email address was not mandatory for survey responses to be submitted.

Data Analysis

Consistent with the analytical procedures used by Akca et al. (2023), our first step of data analysis involved comparing the IOP-29 FDS values between the three feigning conditions with a between-subjects comparison (one-way ANOVA). This was done to determine whether there were significant Serbian IOP-29 FDS differences for the different types of feigning presentations (schizophrenia feigning condition [SIM SCZ], depression feigning condition [SIM DEP], and PTSD feigning condition [SIM PTSD]). Next, a repeated-measures ANOVA was conducted to test the extent to which the FDS values varied as a function of instructing the respondent to answer honestly, randomly, or pretend to suffer from a mental disorder. We then calculated the classification accuracy statistics for the three IOP-29 FDS cutoff scores proposed in the IOP-29 professional manual (Viglione & Giromini, 2020, p. 47), i.e., ≥ 0.30 , ≥ 0.50 , and ≥ 0.65 . Finally, we used highly similar procedures for analyzing the effectiveness of the RRS. That is, we first performed

a between-subjects one-way ANOVA using the different feigning instructions as the between-subject factor and the RRS values as the dependent variable. Then, we performed a series of repeated-measures one-way ANOVAs to assess whether the RRS changed across the honest, random, and feigning conditions. Finally, we calculated classification accuracy statistics by inspecting the RRS cutoffs of $T \geq 61$ (liberal), $T \geq 67$ (midrange), and $T \geq 71$ (conservative). Data are publicly available on the following link: https://osf.io/j7d25/?view_only=03fef5cec5ae4dfda894c76a654e58d7.

Results

Effectiveness of the IOP-29 FDS: Condition Differences

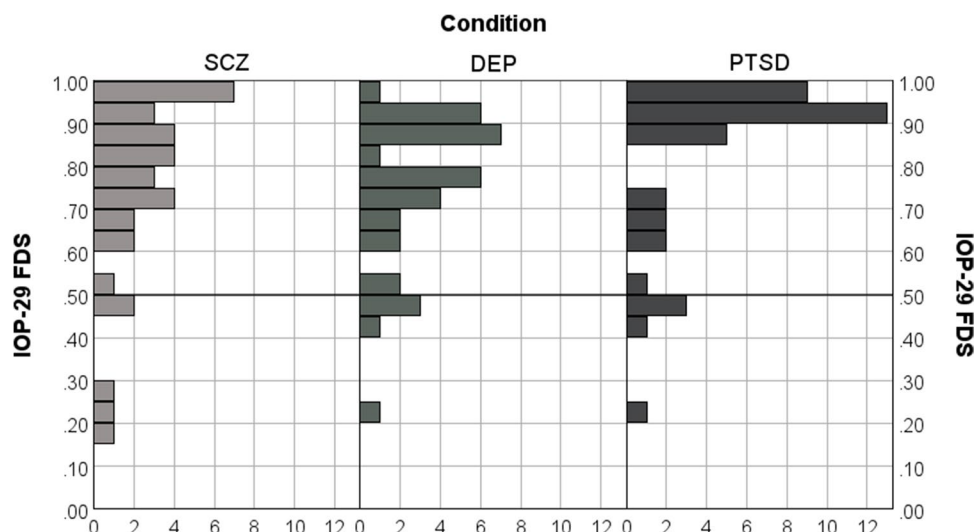
The average IOP-29 FDS values did not significantly differ across the three different feigning conditions, $F_{(2,107)} = 1.60$, $p = .21$, $\eta^2_p = .03$, indicating that the IOP-29 performed equally well in detecting feigned schizophrenia (SIM SCZ; $n = 35$), depression (SIM DEP; $n = 36$), and PTSD (SIM PTSD; $n = 39$). Visual examination of Fig. 1 further supports this observation, revealing a strikingly similar distribution of IOP-29 FDS values across the three feigning conditions. Consequently, these feigning conditions were combined into one main feigning condition (SIM) for subsequent analyses, consistent with Akca et al. (2023).

As expected, the IOP-29 FDS values significantly differed across the HON, RND, and SIM conditions, $F_{(2,218)} = 450.19$, $p < .001$, $\eta^2_p = .81$, with the SIM condition ($M = .77$; $SD = .20$) yielding the highest values, followed by the RND condition ($M = .71$; $SD = .17$), and in turn followed by the HON condition ($M = .15$; $SD = .15$). Notably, all Bonferroni-corrected pairwise comparisons were statistically significant, indicating that SIM yielded significantly higher values than both RND (Hedge's $g = 0.32$) and HON (Hedge's $g = 3.51$), and RND also yielded significantly higher values than HON (Hedge's $g = 3.49$). The distribution of IOP-29 FDS values across the HON and SIM conditions is graphically represented in Fig. 2.

Effectiveness of the IOP-29 FDS: Classification Accuracy

Examination of classification accuracy statistics further confirmed that the Serbian IOP-29 was similarly effective in detecting feigning of schizophrenia, depression, and PTSD. In fact, when considering the standard IOP-29 cutoff (i.e., $FDS \geq .50$), sensitivity values were .86, .86, and .87 for SIM SCZ, SIM DEP, and SIM PTSD, respectively. With the liberal IOP-29 cutoff (i.e., $FDS \geq .30$), sensitivity was 0.91 for SIM SCZ, .97 for SIM DEP, and .97 for SIM

Fig. 1 Graphical representation of the distribution of the IOP-29 FDS across the three feigning conditions



PTSD, while with the conservative IOP-29 cutoff (i.e., $FDS \geq .65$), sensitivity was .77 for SIM SCZ, .75 for SIM DEP, and .80 for SIM PTSD. Notably, the percentage of cases above versus below the cutoff did not significantly differ across the three feigning conditions for any of these cutoffs: for the liberal cut-off score, $\chi^2_{(2)} = 1.92$, $p = .38$, for the standard cut-off score, $\chi^2_{(2)} = 0.37$, $p = .98$, and for the conservative cut-off score, $\chi^2_{(2)} = .22$, $p = .90$. Classification accuracy statistics for the combined sample are presented in Table 1.

Effectiveness of the IOP-29 RRS: Random Responding Across Conditions

The IOP-29 RRS, unlike the IOP-29 FDS, yielded statistically significant differences across different simulating conditions, $F_{(2,107)} = 9.32$, $p < .001$, $\eta^2_p = .15$. Specifically,

Bonferroni-corrected pairwise comparisons indicated that SIM SCZ ($M = 53.8$; $SD = 9.7$) produced significantly higher IOP-29 RRS values compared to both SIM DEP ($M = 48.0$; $SD = 8.5$; $p = .02$) and SIM PTSD ($M = 45.3$; $SD = 7.6$; $p < .001$). In contrast, SIM DEP and SIM PTSD conditions obtained similar scores ($p = .53$), so, for the subsequent analyses, the data from these subgroups were combined ($n = 75$), while the data from the SIM SCZ subgroup were analyzed separately ($n = 35$).

When excluding from the analyses the participants who, in the feigning condition, were instructed to feign schizophrenia, the IOP-29 RRS values varied significantly as a function of instructing the respondent to answer honestly, randomly, or pretend to suffer from PTSD or depression, $F_{(2,148)} = 187.01$, $p < .001$, $\eta^2_p = .72$. More in detail, in this subgroup ($n = 75$), the mean IOP-29 RRS value was significantly higher in the RND ($M = 69.9$; $SD = 8.7$) than in

Fig. 2 Graphical representation of the distribution of the IOP-29 FDS across the HON and SIM conditions

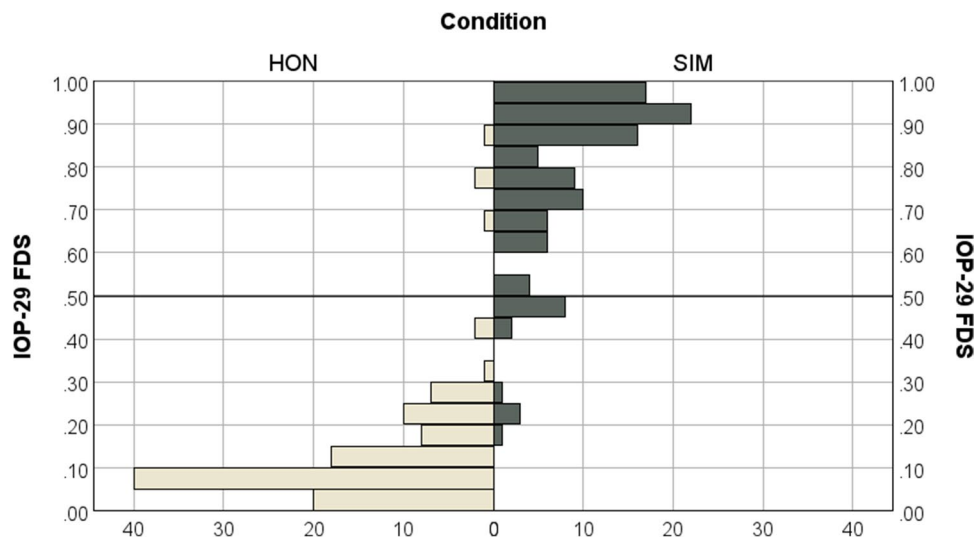


Table 1 IOP-29 FDS classification accuracy statistics for the combined sample

		HON		SIM	
		<i>n</i>	%	<i>n</i>	%
<i>Liberal cut-off</i>					
	IOP-29 FDS ≥ .30	7	6.4	105	<u>95.5</u> ^b
	IOP-29 FDS < .30	103	<u>93.6</u> ^a	5	4.5
<i>Standard cut-off</i>					
	IOP-29 FDS ≥ .50	4	3.6	95	<u>86.4</u> ^b
	IOP-29 FDS < .50	106	<u>96.4</u> ^a	15	13.6
<i>Conservative cut-off</i>					
	IOP-29 FDS ≥ .65	4	3.6	85	<u>77.3</u> ^b
	IOP-29 FDS < .65	106	<u>96.4</u> ^a	25	22.7

Classification accuracy statistics are highlighted using underlined font

^a Specificity

^b Sensitivity

the SIM ($M = 46.6$; $SD = 8.1$; $p < .001$) and HON ($M = 50.1$; $SD = 7.5$; $p < .001$) conditions. Additionally, the mean IOP-29 RRS value in the HON condition was significantly higher than in the SIM condition ($p = .01$). Conversely, when considering exclusively the participants who, in the feigning condition, were instructed to feign schizophrenia ($n = 35$), the main effect of condition remained statistically significant, $F_{(2,68)} = 35.9$, $p < .001$, $\eta^2_p = .51$. In this case, however, the difference between the SIM and HON conditions did not reach statistical significance ($p \approx 1.00$), despite the mean IOP-29 RRS value still being significantly higher in the RND condition ($M = 69.8$; $SD = 10.5$) compared to both the SIM ($M = 53.8$; $SD = 9.7$; $p < .001$) and HON ($M = 51.5$; $SD = 7.2$; $p < .001$) conditions. In summary, the mean IOP-29 RRS values were very high (i.e., approximating $T = 70$) in the RND condition, at the upper end of the normal range (i.e., approximating $T = 54$) in the SIM SCZ condition, around the center of the normal range (i.e., approximating $T = 50$) in the HON condition, and at the lower end of the normal range (i.e., approximating $T = 47$) in the SIM DEP and SIM PTSD conditions (see Table 2).

Examination of classification accuracy statistics, as reported in Table 3, reveals that the liberal IOP-29 RRS cutoff of $T \geq 61$ yields suboptimal specificity (i.e., specificity $< .90$) in both the HON and SIM SCZ groups. Conversely, the midrange ($T \geq 67$) and conservative ($T \geq 71$) IOP-29 RRS cutoffs yield adequate ($\geq .90$) and excellent ($\geq .97$) specificity values across all non-RND groups, with sensitivity values of .64 and .49, respectively.

Additional Analyses

As a post hoc analysis to explore the potential confounding effect of the order in which conditions were presented to participants, we conducted two mixed ANOVAs. In these analyses, we included condition (HON, RND, SIM) as a

within-subject factor, administration order (the six possible administration order combinations) as a between-subject factor, and the FDS and RRS as the dependent variables. The results of these post hoc analyses indicated that neither the interaction effects (condition by administration order) nor the main effects of administration order were statistically significant for either the FDS or RRS. Thus, the order in which the conditions were presented did not influence the FDS and RRS values or their effectiveness.

Discussion

Given the recommendation for professionals to use multiple symptom validity tests (SVTs) in nearly all of their assessments (Sweet et al., 2021), our study aimed to address the limited availability of validated SVTs in languages other than English. Specifically, we developed a Serbian version of the Inventory of Problems – 29 (IOP-29; Viglione & Giromini, 2020), and tested its validity by administering it repeatedly under three different conditions to 110 adult volunteers from Serbia. Volunteers were instructed to answer honestly in one condition, randomly in another, and to feign a specific mental illness in yet another condition. In the simulating condition, about a third were asked to feign schizophrenia, about a third were asked to feign depression, and about a third were asked to feign PTSD. This approach closely followed the research design and procedures of Akca et al. (2023), allowing us to attempt replication in a Serbian context.

Consistent with Akca et al. (2023), the IOP-29 was able to detect invalid responses with similar effectiveness across three different feigning conditions – schizophrenia, depression, and PTSD – suggesting that it may be used with clinical examinees exhibiting symptoms of any of these psychiatric conditions. Perhaps more importantly, the effect size of the difference between the average IOP-29 FDS value in

Table 2 IOP-29 RRS: Descriptive statistics

		IOP-29 RRS		
		<i>n</i>	<i>M</i>	<i>SD</i>
HON	Entire sample	110	50.6	7.5
SIM	SIM DEP & SIM PTSD	75	46.6	8.1
	SIM SCZ	35	53.8	9.7
RND	Entire sample	110	69.9	9.3

the honest versus simulating conditions was a very large Hedge's *g* value of 3.51, which is similar to, but slightly larger than, both the Hedge's *g* value of 2.68 reported in Akca et al. (2023) and the weighted mean Cohen's *d* value of 3.02 reported in the quantitative literature review by Giromini and Viglione (2022). In addition, in our sample, the standard IOP-29 FDS cutoff score of $\geq .50$ yielded a specificity of .96 and a combined sensitivity of .86. These findings also are similar to both Akca et al. (2023), who reported a specificity of 88% and a combined sensitivity of .91, and Giromini and Viglione (2022), who reported a weighted mean specificity of .92 and a weighted mean sensitivity of .86. Taken together, these results thus suggest that, pending additional research on clinical and forensic populations, the Serbian version of the IOP-29 is likely to perform similarly well as all other available versions of the IOP-29.

While not yet available through the official IOP test's scoring platform (www.iop-test.com), Giromini et al. (2020c) recently introduced a new index that is also based on the IOP-29 responses, that is aimed at identifying possible careless or random responding. This scale, called the Random Responding Scale (RRS) and scaled as a T-score, was developed to help professionals distinguish between intentional feigning (i.e., content-related distortion) and factors such as poor attention, misunderstanding of the IOP-29 item content, or lack of cooperation with the testing condition (i.e., content-unrelated distortion). Prior to our study, only three research articles – Akca et al. (2023), Giromini et al. (2020c), and Winters et al. (2021) – had reported on the effectiveness of this scale, prompting us to contribute additional data on its potential utility.

Overall, our findings support the notion that the IOP-29 RRS may enhance our understanding of test takers' approaches to responding to IOP-29 items. Indeed, our study demonstrates that the RRS effectively discriminated content-unrelated distortion from other response styles such as honest responding or content-related distortion. In addition, it is noteworthy that the RRS was slightly more elevated when participants were instructed to feign schizophrenia compared to when they were instructed to feign another disorder (depression or PTSD) or respond honestly. One possible explanation for this finding is that participants may have believed that schizophrenia involves disorganization, leading them to intentionally endorse items to appear more confused or disorganized. Alternatively, the higher inconsistency observed in the schizophrenia condition might be due to participants' limited knowledge of the clinical manifestations of schizophrenia, which could have led to an unintentionally less consistent presentation compared to disorders they might be more familiar with, such as depression or PTSD. In the absence of sufficient data to determine which explanation is more likely, further research on the IOP-29 scale (i.e., RRS)

Table 3 IOP-29 RRS:
Classification accuracy

		HON		SIM DEP & SIM PTSD		SIM SCZ		RND	
		<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Liberal cutoff									
	IOP-29 RRS T ≥ 61	13	11.8	6	8.0	10	28.6	95	86.4 ^b
	IOP-29 RRS T < 61	97	88.2 ^a	69	92.0 ^a	25	71.4 ^a	15	13.6
Midrange cutoff									
	IOP-29 RRS T ≥ 67	2	1.8	1	1.3	2	5.7	70	63.6 ^b
	IOP-29 RRS T < 67	108	98.2 ^a	74	98.7 ^a	33	94.3 ^a	40	36.4
Conservative cutoff									
	IOP-29 RRS T ≥ 71	1	0.9	0	0.0	1	2.9	54	49.1 ^b
	IOP-29 RRS T < 71	109	99.1 ^a	75	100.0 ^a	35	97.1 ^a	56	50.9

Classification accuracy statistics are highlighted using underlined font

^aSpecificity

^bSensitivity

is warranted. Indeed, looking ahead, one might speculate that the RRS could even prove useful in contributing to differentiating between credible and non-credible presentations of schizophrenia, as Giromini et al. (2020c) found that, unlike participants who feigned schizophrenia in our study, real-world patients genuinely affected by schizophrenia did not typically elevate the RRS.

In Akca et al. (2023), the average IOP-29 RRS values were approximately 67T in the random responding condition, approximately 55T in the simulating condition, and approximately 54T among honest responders. In our study, the random responding condition yielded a similarly high value (about 70T), while the simulating and honest conditions produced relatively lower values (ranging from about 47T to about 54T). Aligned with Akca et al. (2023), our results indicated that the threshold of $T \geq 67$ resulted in a specificity of $\geq .90$, while a threshold of $T \geq 71$ achieved a specificity of $\geq .95$ across all attentive responders. Therefore, pending future replications, IOP-29 RRS values exceeding these cutoffs could potentially indicate the presence of careless or random responding in the collected IOP-29 data, particularly in research settings.

When considering the implications of our results for clinical and forensic settings, several limitations should be noted. First and most importantly, our sample, as those in the studies we replicated, lacked clinical or forensic participants, which limited our ability to accurately test the specificity of the Serbian IOP-29 FDS. In other words, our study primarily informs about its sensitivity, and further research involving patients is necessary to gauge how clinical and forensic respondents perform on the IOP-29 when presented with the Serbian version. Although our results are in good agreement with previous studies (Akca et al., 2023; see also Giromini & Viglione, 2022), and despite recent meta-analysis indicating that language does not significantly impact the effectiveness

of the IOP-29 when the standard and conservative cutoffs are used (Puente-López et al., 2023b), caution is still required when interpreting the results of our Serbian version of the IOP-29.

The fact that we used a simulation design also raises questions about the generalizability of our findings to real-world contexts. On one hand, the use of a simulation design is beneficial for preserving the internal validity of the study and ensuring a sufficient number of feigners are included in the sample (Giromini et al., 2022). However, the external validity of simulation studies is inherently uncertain, as there is no guarantee that individuals assessed in real-world settings will respond in the same way and use the same feigning strategies as our experimental participants. After all, the costs of being detected as a feigner and the benefits of remaining undetected when assessed in real-world settings are drastically different from those associated with a simulation study such as ours.

Other aspects to consider regarding the potential generalizability of our findings are as follows: First, in contrast to the typical symptom validity assessment battery used in real-world evaluations, which generally includes multiple tests administered in multiple formats, our study only incorporated the IOP-29 and consistently administered it online. Secondly, even though the order in which the different conditions were presented did not significantly influence the results of the two IOP-29 scales under investigation, our participants completed the IOP-29 multiple times, which contrasts with real-world practice, where the IOP-29 is typically completed only once. Future studies should therefore employ a between-subject design, requiring participants to complete the IOP-29 only once, as would be the case in a real-world setting. Thirdly, the vast majority of participants in our study described themselves as “female,” with the average age being around 30 years, and approximately one-third of the sample consisted

of students. These and some other similar limitations lead us to emphasize that our study should be regarded as an “initial” validation of the Serbian version of the IOP-29. Further research, for example using a criterion-groups design with an ecologically valid sample and considering the Serbian Minnesota Multiphasic Personality Inventory as a criterion variable, is essential before we can confidently conclude that our Serbian IOP-29 works as well as other, more widely used IOP-29 versions. In addition, this further research might also investigate, either via simulation or a criterion-groups design, the memory module of the IOP-29 (i.e., the IOP-M; Giromini et al., 2020d), which has already demonstrated promising results in several other countries and languages (e.g., Crişan, 2023; Erdodi et al., 2023; Puente-López et al., 2023b), and is therefore likely to maintain its validity in Serbia as well.

With all these considerations in mind, our study is the first to evaluate the effectiveness of a Serbian version of the IOP-29 in an experimental sample from Serbia. Not only does our study show that the performance of the Serbian IOP-29 FDS is likely to be in line with a priori expectations and the results of other international studies, but it also highlights the potential utility of the newly developed IOP-29 RRS as an indicator of content-unrelated distortion. Furthermore, from a broader perspective, our study also underscores the importance of conducting symptom validity studies in culturally diverse settings. Indeed, improving our understanding of how SVTs such as the IOP-29 perform in diverse populations, such as those from the Balkans, lays the foundation for tailored and effective assessment strategies in clinical and forensic settings worldwide.

Data Availability Data are publicly available on the following link: https://osf.io/j7d25/?view_only=03fef5cec5ae4dfda894c76a654e58d7.

Declarations

Conflict of Interest Last author declares that he owns a share in the corporate (LLC) that possesses the rights to Inventory of Problems. The rest of the authors declare no conflict of interest.

References

- Abeare, K., Razvi, P., Sirianni, C. D., Giromini, L., Holcomb, M., Cutler, L., & Erdodi, L. A. (2021). Introducing alternative validity cutoffs to improve the detection of non-credible symptom report on the BRIEF. *Psychological Injury and Law*, 14, 2–16. <https://doi.org/10.1007/s12207-021-09402-4>
- Akca, A. Y., Tepedelen, M. S., Uysal, B., & Erdodi, L. A. (2023). The Inventory of Problems–29 is a cross-culturally valid symptom validity test: Initial validation in a Turkish community sample. *Psychological Injury and Law*, 16(3), 289–301. <https://doi.org/10.1007/s12207-023-09483-3>
- Banovic, I., Filippi, F., Viglione, D. J., Scrima, F., Zennaro, A., Zapalà, A., & Giromini, L. (2022). Detecting coached feigning of schizophrenia with the Inventory of Problems – 29 (IOP-29) and its Memory Module (IOP-M): A simulation study on a french community sample. *International Journal of Forensic Mental Health*, 21(1), 37–53. <https://doi.org/10.1080/14999013.2021.1906798>
- Biro, M. (2001). *Dijagnostička procena ličnosti MMPI-202 (2nd Ed.)* [Clinical Personality Assessment – MMPI 202]. Društvo psihologa Srbije.
- Blavier, A., Palma, A., Viglione, D. J., Zennaro, A., & Giromini, L. (2023). A natural experiment design testing the effectiveness of the IOP-29 and IOP-M in assessing the credibility of reported PTSD symptoms in Belgium. *Journal of Forensic Psychology Research and Practice*, [Epub ahead of print]. <https://doi.org/10.1080/24732850.2023.2203130>
- Bosi, J., Minassian, L., Ales, F., Akca, A. Y. E., Winters, C., Viglione, D. J., ... & Giromini, L. (2022). The sensitivity of the IOP-29 and IOP-M to coached feigning of depression and mTBI: An online simulation study in a community sample from the United Kingdom. *Applied Neuropsychology: Adult*, [Epub ahead of print]. <https://doi.org/10.1080/23279095.2022.2115910>
- Boskovic, I., Akca, A. Y. E., & Giromini, L. (2022). Symptom coaching and symptom validity tests: An analog study using the structured inventory of malingered symptomatology, self-report symptom inventory, and inventory of problems-29. *Applied Neuropsychology: Adult*. [Epub ahead of print]. <https://doi.org/10.1080/23279095.2022.2057856>
- 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). Allyn & Bacon.
- Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, B. (2001). *MMPI-2 (Minnesota Multiphasic Minnesota Multiphasic Personality Inventory-2 (MMPI-2))*. Minneapolis University of Minnesota Press.
- 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). *Psychological Injury and Law*, 14, 58–70. <https://doi.org/10.1007/s12207-021-09403-3>
- Crişan, I. (2023). English versus native language administration of the IOP-29-M produces similar results in a sample of Romanian bilinguals: A brief report. *Psychology & Neuroscience*, 16(3), 254–260. <https://doi.org/10.1037/pne0000316>
- Erdodi, L., Calamia, M., Holcomb, M., Robinson, A., Rasmussen, L., & Bianchini, K. (2023). M is For Performance Validity: The IOP-M Provides a Cost-Effective Measure of the Credibility of Memory Deficits during Neuropsychological Evaluations. *Journal of Forensic Psychology Research and Practice*, 24(4), 434–450. <https://doi.org/10.1080/24732850.2023.2168581>
- Gegner, J., Erdodi, L. A., Giromini, L., Viglione, D. J., Bosi, J., & Brusadelli, E. (2022). 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*, 29(5), 1221–1230. <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). John Wiley & Sons.
- Giromini, L., & Viglione, D. J. (2022). Assessing negative response bias with the Inventory of Problems-29 (IOP-29): A quantitative literature review. *Psychological Injury and Law*, 15(1), 79–93. <https://doi.org/10.1007/s12207-021-09437-7>
- Giromini, L., Barbosa, F., Coga, G., Azeredo, A., Viglione, D. J., & Zennaro, A. (2020a). 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, 504–516. <https://doi.org/10.1080/23279095.2019.1570929>
- Giromini, L., Pignolo, C., Zennaro, A., & Viglione, D. J. (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., Viglione, D. J., Pignolo, C., & Zennaro, A. (2020b). An Inventory of Problems–29 sensitivity study investigating feigning of four different symptom presentations via malingering experimental paradigm. *Journal of Personality Assessment*, 102(4), 563–572. <https://doi.org/10.1080/00223891.2019.1566914>
- Giromini, L., Viglione, D. J., Pignolo, C., & Zennaro, A. (2020c). An Inventory of Problems-29 study on random responding using experimental feigners, honest controls, and computer-generated data. *Journal of Personality Assessment*, 102(6), 731–742. <https://doi.org/10.1080/00223891.2019.1639188>
- Giromini, L., Viglione, D. J., Zennaro, A., Maffei, A., & Erdodi, L. A. (2020d). 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>
- Giromini, L., Young, G., & Sellbom, M. (2022). Assessing Negative Response Bias Using Self-Report Measures: Introducing the Special Issue. *Psychological Injury and Law*, 15, 1–21. <https://doi.org/10.1007/s12207-022-09444-2>
- Grønnerød, C., Rekkedal Rølfesnes, S., & Gustavson, K. (2023). Validity of the Norwegian Version of Inventory of Problems–29 (IOP-29): A simulation study with experimental feigning of depression and a nonclinical control group. *Psychology & Neuroscience*, 16(2), 147–154. <https://doi.org/10.1037/pne0000309>
- Holcomb, M., Pyne, S., Cutler, L., Oikle, D. A., & Erdodi, L. A. (2023). Take Their Word for It: The Inventory of Problems Provides Valuable Information on Both Symptom and Performance Validity. *Journal of Personality Assessment*, 105(4), 520–530. <https://doi.org/10.1080/00223891.2022.2114358>
- Ilgunaite, G., Giromini, L., Bosi, J., Viglione, D. J., & Zennaro, A. (2022). 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*, 29(2), 155–162. <https://doi.org/10.1080/23279095.2020.1725518>
- Puente-López, E., Pina, D., López-Nicolás, R., Iguacel, I., & Arce, R. (2023a). The Inventory of Problems-29 (IOP-29): A systematic review and bivariate diagnostic test accuracy meta-analysis. *Psychological Assessment*, 35(4), 339–352. <https://doi.org/10.1037/pas0001209>
- Puente-López, E., Pina, D., Rambaud-Quñones, P., Ruiz-Hernández, J. A., Nieto-Cañaveras, M. D., Shura, R. D., Alcazar-Crevillén, A., & Martínez-Jarreta, B. (2023b). Classification accuracy and resistance to coaching of the Spanish version of the Inventory of Problems -29 and the Inventory of Problems - Memory: a simulation study with mTBI patients. *The Clinical Neuropsychologist*, [Epub ahead of print]. <https://doi.org/10.1080/13854046.2023.2249171>
- Rogers, R., & Gillard, N. D. (2011). Research methods for the assessment of malingering. In B. Rosenfeld & S. Penrod (Eds.), *Research Methods in Forensic Psychology* (pp. 174–188). Wiley.
- Roma, P., Giromini, L., Burla, F., Ferracuti, S., Viglione, D. J., & Mazza, C. (2020). 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>
- Roma, P., Giromini, L., Sellbom, M., Cardinale, A., Ferracuti, S., & Mazza, C. (2023). The ecological validity of the IOP-29: A follow-up study using the MMPI-2-RF and the SIMS as criterion variables. *Psychological Assessment*, 35(10), 868–879. <https://doi.org/10.1037/pas0001273>
- 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(6), 735–764. <https://doi.org/10.1093/arclin/acia019>
- Šömen, M. M., Lesjak, S., Majaron, T., Lavopa, L., Giromini, L., Viglione, D. J., & Podlesek, A. (2021). 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. *Psychological Injury and Law*, 14, 104–113. <https://doi.org/10.1007/s12207-021-09412-2>
- Šrubař, M., & Fňukal, M. (2010). Migration from the countries of former Yugoslavia and its impact on international relations. *Acta Universitatis Palackianae Olomucensis-Geographica*, 41(2), 21–36.
- Sweet, J. J., Heilbronner, R. L., Morgan, J. E., Larrabee, G. J., Rohling, M. L., Boone, K. B., Kirkwood, M. W., Schroeder, R. W., Suhr, J. A., & Participants, C. (2021). American Academy of Clinical Neuropsychology (AACN) 2021 consensus statement on validity assessment: Update of the 2009 AACN consensus conference statement on neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, 35(6), 1053–1106. <https://doi.org/10.1080/13854046.2021.1896036>
- Van de Vijver, F., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist*, 1, 89–99.
- 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*. IOP-Test, LLC.
- Winters, C. L., Giromini, L., Crawford, T. J., Ales, F., Viglione, D. J., & Warmelink, L. (2021). An Inventory of Problems–29 (IOP–29) study investigating feigned schizophrenia and random responding in a British community sample. *Psychiatry, Psychology and Law*, 28(2), 235–254. <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>

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.