Psychiatry, Psychology and Law, 2024
Vol. 0, No. 0, 1–14, https://dx.doi.org/10.1080/13218719.2024.2441788

Routledge
Taylor & Francis Group

Check for updates

Cross-cultural applicability of the Inventory of Problems–29 (IOP–29): a pre-registered German replication of Akca, Tepedelen, et al. (2023)

Henrike Höpfner[a], Burcu Uysal[b], Mark Stemmler[a], Ali Y. E. Akca[c,d] and Luciano Giromini[d]

[a]*Department of Psychology, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany;* [b]*Department of Psychology, Ibn Haldun University, Istanbul, Türkiye;* [c]*Forensic Psychology Section, Maastricht University, Maastricht, The Netherlands;* [d]*Department of Psychology, University of Turin, Turin, Italy*

Assessing the credibility of psychological problems in forensic evaluations is crucial. The Inventory of Problems–29 (IOP–29) is becoming a valuable symptom validity test (SVT) worldwide and has been validated in over 15 languages. This study evaluates the German IOP–29 and fills a gap by replicating a study of Akca, Tepedelen, et al. We analysed 384 IOP–29 protocols from 128 German-speaking adults (range = 18–87 years) under three conditions: honest, random and feigned responses (post-traumatic stress disorder, PTSD; depression; schizophrenia). Statistical analyses showed that the False Disorder Probability Score (FDS) effectively discriminated between honest and feigned responses (Hedges $g = 3.90$), with a sensitivity of .91 and a specificity of .95 at an FDS cut-off value of $\geq .50$. A new index for detecting careless or random responses also showed promising results. This study confirms the utility of the IOP–29 in the German-speaking population and supports its cross-cultural applicability.

Keywords: depression; feigning; German; IOP–29; malingering; PTSD; schizophrenia; symptom validity test.

Article History: Received 3 July 2024; Accepted 28 November 2024

In recent years, it has become evident that the veracity of psychological problems claimed by clinical and forensic assessees cannot be taken for granted and must be assessed with empirical instruments. Consistent with this trend, professional associations such as the American Academy of Clinical Neuropsychology (AACN) have recently released official statements encouraging the use of symptom validity tests (SVTs) and performance validity tests (PVTs) in nearly all clinical and forensic evaluations (Sweet et al., 2021). More specifically, it is recommended to administer *multiple* SVTs and PVTs, so as to minimise the risks of false-positive and false-negative results (Erdodi, 2023a, 2023b). However, this recommendation can be challenging to implement in real-world practice due to the relatively limited number of validated SVTs currently available (Sherman et al., 2020; Sweet et al., 2021; for a review on available SVTs, see Giromini et al., 2022).

In this context, the Inventory of Problems–29 (IOP–29; Viglione & Giromini, 2020) might be a particularly useful tool. Indeed, at this time, the most widely used and researched SVTs are arguably the validity scales of personality inventories such as the Minnesota

Correspondence: Ali Y. E. Akca, Forensic Psychology Section, Maastricht University, Universiteitssingel 40, 6229 ER, Maastricht, The Netherlands. Email: emre.akca@maastrichtuniversity.nl

© 2024 The Australian and New Zealand Association of Psychiatry, Psychology and Law

Multiphasic Personality Inventory (MMPI–3; Ben-Porath & Tellegen, 2020) and the Personality Assessment Inventory (PAI; Morey, 1991, 2007; for a review on the validity scales of the MMPI, see Burchett & Bagby, 2022; for a review on the validity scales of the PAI, see Kurtz & McCredie, 2022). However, these instruments cannot be administered to everyone due to the cognitive load they require. In addition, they are quite long, making it challenging to incorporate additional time-consuming instruments when one or both of them are administered. Thus, as the IOP–29 is a self-administered instrument composed of only 29 items, taking just 5–10 min to complete, it can be easily integrated into assessment protocols without imposing a significant burden on either the assessor or the assessee.

Although it was introduced relatively recently, in 2017 (Viglione et al., 2017), the IOP–29 has already demonstrated its ability to effectively discriminate credible from non-credible mental health complaints across multiple studies worldwide (for a review, see Giromini & Viglione, 2022; for a meta-analysis, see Puente-López, Pina, López-Nicolás, et al., 2023). Indeed, its effectiveness has been investigated and supported in North America (e.g. Holcomb et al., 2023), South America (e.g. Carvalho et al., 2021), Canada (e.g. Abeare, Razvi, et al., 2021), Australia (e.g. Gegner et al., 2022), Türkiye (Akca, Tepedelen, et al., 2023) and many European countries such as France (Banovic et al., 2022), Italy (Pignolo et al., 2023), Spain (Puente-López, Pina, Rambaud-Quiñones, et al., 2023) and several others. However, to date, there has been no published validation study specifically focusing on the German version of the IOP–29.

Given that approximately 100 million people are estimated to be native German speakers (Eberhard et al., 2020), a German version of the IOP–29 could prove highly beneficial for professionals involved in symptom validity assessment worldwide, particularly in German-speaking regions. Our study thus aimed to fill this gap in the literature by providing initial empirical data on the effectiveness of the German IOP–29 in discriminating feigned from genuine presentations among a sample of adult volunteers. It is worth noting that our research not only served to provide valuable insight into the classification accuracy of the IOP–29 for users dealing with German-speaking respondents, but also to add to the growing body of research on the cross-cultural adaptability of the IOP–29.

This study

Akca, Tepedelen, et al. (2023) recently investigated the effectiveness of the Turkish version of the IOP–29 using a very cost-effective research design. Specifically, they administered the Turkish IOP–29 three times to a sample of adult volunteers: once with instructions to respond honestly (HON condition), once with instructions to respond at random (RND condition) and once with instructions to feign or simulate a mental health disorder (SIM condition). Importantly, while all participants received the same instructions for the HON and RND conditions, in the SIM condition, one third of participants were asked to feign post-traumatic stress disorder (PTSD), one third to feign depression and one third to feign schizophrenia. Modelled after two previously published studies by Giromini et al. (2020) and Winters et al. (2021), a major strength of Akca, Tepedelen, et al.'s (2023) design is that it allows the researcher to address multiple research questions in a very cost-effective manner. Indeed, this design enabled Akca, Tepedelen, et al. (2023) to assess: (a) the extent to which the IOP–29 is similarly versus differently sensitive to the feigning of three different mental disorders; (b) the extent to which the IOP–29 is able to discriminate genuine from feigned presentations; and (c) the extent to which the IOP–29 can identify random responding.

Given the merit of Akca, Tepedelen, et al.'s (2023) research design, the same

approach was recently used in another IOP–29 study aimed at investigating the effectiveness of the Serbian version of the IOP–29 (Volarov et al., 2024). As noted by Volarov et al. (2024), using the same research design helps minimise potential confounding factors. Indeed, by using the same design, any differences in results between the new IOP–29 version (in a language for which there has been no research yet) and the IOP–29 version used in the previous study can be attributed to differences in the effectiveness of the two versions rather than to research design-related factors. Accordingly, for the current study focusing on the German IOP–29, we also employed the same research design as that used by Akca, Tepedelen, et al. (2023) and Volarov et al. (2024).

Method

Participants

Consistent with Akca, Tepedelen, et al. (2023) and Volarov et al. (2024), data collection was conducted online using Qualtrics. In order to participate in the study, participants had to give informed consent. They also had to fulfil the following inclusion criteria: they had to be native German speakers, be at least 18 years old, be fluent in German, not have a diagnosed mental disorder, not self-report poor mental health and not be a psychology student or have studied psychology in the past. Additionally, multiple manipulation checks were interspersed throughout the experiment to ensure adherence to instructions. Participants who failed these checks were excluded from the study.

The final dataset used for statistical analyses included 384 IOP–29 protocols obtained from 128 adult, German-speaking volunteers. The vast majority were from Germany (93.0%), with a small number from Austria (2.3%) and Türkiye (0.8%). A further 3.1% indicated other countries or chose not to specify. The mean age was 35.24 years ($SD = 17.99$, range = 18–87 years); 89 participants (69.5%) identified their gender as female, while 39 (30.5%) identified as male. In terms of education, 7.0% had completed secondary school, 44.5% held a high school diploma, 11.7% had a Bachelor's degree, 31.3% had a Master's degree, and 5.5% specified other types of educational qualifications (4.7% PhD and 0.8% unspecified university degree). In terms of ethnicity, 89.1% identified as White or European, 2.3% as Asian, 3.9% did not identify with any racial or ethnic group, 3.1% preferred not to specify, and 1.6% preferred to use a term that was not provided in the response options. None reported having received a diagnosis of a mental disorder, and when asked to rate their current mental health status, 39 participants (30.5%) responded 'very good'; 59 (46.1%) responded 'fairly good'; and 30 (23.4%) responded 'neither good nor bad'.

Measures

The Inventory of Problems–29 (IOP–29; Viglione & Giromini, 2020)

The IOP–29 is a short self-report measure designed to discriminate between credible and noncredible mental health complaints regardless of the type of symptoms presented. Comprising only 29 items, the IOP–29 stands out from typical SVTs as it integrates multiple feigning detection approaches. Indeed, most SVTs rely on rare, quasi-rare or symptom severity detection strategies, which identify feigned symptoms based on their rarity or extremity. In contrast, the IOP–29 incorporates these strategies but also includes items that assess the respondent's acknowledgment of responsibility for their problems, their ability to cope with these issues, their attitudes towards the testing situation, and more. Additionally, most SVTs rely solely on self-report items presented in either the true–false or Likert scale format. In contrast, the IOP–29 includes 27 self-report items that offer three response options – True, False, or Doesn't Make Sense – and two items that require the respondent to complete open-ended cognitive

or mathematical tasks. This multifaceted approach enables the IOP–29 to evaluate symptom credibility from a different, and arguably broader, perspective compared to the typical SVT.

Despite its brevity, a wide body of research suggests that the validity of the IOP–29 is comparable to, if not superior to, longer and more complex SVTs (e.g. Boskovic et al., 2024; Giromini et al., 2024; Pignolo et al., 2023; Puente-López, Pina, Rambaud-Quiñones, et al., 2023). For instance, in studies where both the IOP–29 and the Structured Inventory of Malingered Symptomatology (SIMS; Smith & Burger, 1997) – a well-established, 75-item long SVT (for a review, see Shura et al., 2022) – were administered, the IOP–29 consistently outperformed the SIMS in detecting noncredible presentations. In fact, Giromini et al. (2018) administered both instruments to 236 experimental feigners and 216 patient controls, finding that the IOP–29 yielded a Cohen's *d* effect size of 1.93, whereas the SIMS had a significantly lower *d* value of 1.39. Similarly, Puente-López, Pina, Rambaud-Quiñones, et al. (2023) administered the IOP–29 and SIMS to a sample of 219 volunteers. When comparing the results from 62 coached feigners versus 37 clinical controls, Cohen's *d* values were 3.74 for the IOP–29 versus 1.96 for the SIMS. Along similar lines, Boskovic et al. (2024) used the same instructions with different groups of nonclinical adult volunteers, asking them to either feign depression (with or without the assistance of internet-based coaching) or respond honestly. Consistent with Giromini et al. (2018) and Puente-López, Pina, Rambaud-Quiñones, et al. (2023), the IOP–29 generated Cohen's *d* effect size values $\geq$3.28, whereas the SIMS generated values $\geq$1.36. A quantitative literature review (Giromini & Viglione, 2022) and a bivariate test accuracy meta-analysis (Puente-López, Pina, López-Nicolás, et al., 2023) both back up these results. They strongly support IOP–29's effectiveness and suggest that it compares favourably to that of other well-established SVTs such as the SIMS. The IOP–29 utilises an online-based scoring system (www.iop-test.com) to generate the False Disorder Probability Score (FDS), an index based on logistic regression that aims to measure negative response bias. Generally speaking, a higher FDS indicates a lower credibility of the clinical presentation. As outlined in the IOP–29 manual (Viglione & Giromini, 2020), the standard cut-off score for the IOP–29 is FDS $\geq$ .50, which is presumed to strike the best balance between sensitivity and specificity. However, in situations requiring a more lenient screening approach, such as initial assessments or large-scale screenings, a cut-off score of FDS $\geq$ .30 may be more suitable. This lower threshold aims to ensure a sensitivity of approximately .90. Conversely, for high-stakes forensic assessments where minimising the risk of false-positive results is crucial, a more conservative cut-off of FDS $\geq$ .65 is recommended to achieve a specificity of at least .90.

In addition to the FDS, another IOP–29 index was recently introduced. This newer index, called the Random Responding Scale (RRS), was developed to assess content-unrelated distortion such as random or careless responding (Giromini et al., 2020). In other words, the main objective of the RRS is to assist the assessor (or researcher) in determining whether the respondent paid sufficient attention to and/or correctly understood the IOP–29 items. Unlike the FDS, which is scaled as a probability score, the RRS uses a *T*-score-based metric, and according to Giromini et al. (2020), a cut-off value of $T \geq$ 61.0 could be helpful in determining whether or not content-unrelated distortion is present in the IOP–29 under consideration. However, in contrast with Giromini et al. (2020), recent research by Akca, Tepedelen, et al. (2023) found that a higher cut-off score, such as $T \geq$ 65.5 or even $T \geq$ 71.0, might be required to achieve an appropriate level of specificity (i.e. $\geq$.90). Thus, as very few studies on the IOP–29 RRS have been published so far (namely,

Akca, Tepedelen, et al., 2023; Giromini et al., 2020; and Winters et al., 2021), the current study also examined the effectiveness of the RRS (in addition to that of the FDS), by inspecting the following cut-off scores: $T \geq$ 61.0 (liberal), $T \geq$ 66.5 (midrange) and $T \geq$ 71.0 (conservative).

Procedure

The current study utilised procedures closely resembling those outlined in Akca, Tepedelen, et al. (2023) and Volarov et al. (2024). Before beginning participant recruitment, formal approval was obtained from the Institutional Review Board of Maastricht University (ERCPN-274_119_11_2023). Aside from the IOP–29 item content, which is copyrighted, all other relevant materials used for the study are available on the Open Science Framework (OSF) platform at: https://osf.io/4nqvc/?view_only=d50860cfa4e745cc94ed57398aa63a77

Participants were recruited through a combination of methods, including online adverts on social media platforms such as Facebook and Instagram, as well as snowballing and billboard advertising. Upon landing on the Qualtrics page where the experiment was conducted, they were provided with a comprehensive overview of the research project and its objectives. Once they provided their informed consent, they were instructed to complete the IOP–29 three times, under three distinct conditions: responding honestly (HON), responding at random (RND) and feigning/simulating a mental disorder (SIM). The order of presentation for these conditions was randomised and counterbalanced across participants to mitigate potential order effects.

In the SIM condition, participants were randomly assigned to simulate one of three psychiatric disorders: PTSD, depression or schizophrenia. To do so, each participant received a detailed vignette presenting a scenario derived from previous studies (e.g. Pignolo et al., 2023) to facilitate their feigning of the assigned disorder, along with a list of characteristic symptoms for that disorder. Additionally, they were explicitly instructed not to 'over-do it' to avoid being easily detected as feigners (Rogers & Gillard, 2011). Furthermore, several pre-test and post-test manipulation checks were incorporated to ensure participants' comprehension and adherence to task instructions.

Each participant had the opportunity to voluntarily enter a prize draw for four e-vouchers, which took place at the end of the entire data collection phase of the study.

Data analysis

Because our study replicates Akca, Tepedelen, et al. (2023) and Volarov et al. (2024), we strictly adhered to the analytic approaches outlined in these articles. Specifically, we first compared the mean IOP–29 FDS values across three feigning conditions using a between-subjects one-way analysis of variance (ANOVA) and used χ^2 analyses to inspect classification accuracy statistics across the three feigning subgroups, considering the three IOP–29 FDS cut-off scores specified in the IOP–29 professional manual (Viglione & Giromini, 2020): $\geq$.30, $\geq$.50, and $\geq$.65. These analyses aimed to identify any differences in the German IOP–29 FDS values among the three feigning presentations – PTSD (SIM PTSD), depression (SIM DEP) and schizophrenia (SIM SCZ) – and to determine whether the three subsamples could be combined for subsequent analysis. Next, we conducted a repeated measures ANOVA to examine how the FDS values varied based on whether participants were instructed to respond honestly, randomly, or simulating a mental disorder.

Following this, we applied analogous procedures to evaluate the effectiveness of the RRS. That is, a between-subjects one-way ANOVA was initially performed with the type of feigning instructions (SIM PTSD, SIM DEP, SIM SCZ) as the independent variable and the RRS values as the dependent variable. This analysis sought to determine whether it would be appropriate to combine the three different subsamples for subsequent analysis.

Then, a repeated measures one-way ANOVA was conducted to assess whether there were significant changes in RRS values across the honest, random, and feigning conditions. Finally, we calculated classification accuracy statistics by examining the RRS cut-offs of $T \geq 61.0$ (liberal), $T \geq 66.5$ (midrange), and $T \geq 71.0$ (conservative).

Results

Effectiveness of the IOP–29 FDS

In line with Akca, Tepedelen, et al. (2023) and Volarov et al. (2024), the mean IOP–29 FDS values, reported in Table 1, did not significantly differ across the three feigning conditions: feigned PTSD (SIM PTSD), depression (SIM DEP), and schizophrenia (SIM SCZ), $F(2,125) = 1.54$, $p = .22$, $\eta^2_p = .02$. The classification accuracy statistics, also presented in Table 1, confirm that the IOP–29 in our German sample was similarly sensitive to the feigning of PTSD, depression and schizophrenia. In fact, the number of cases above versus below the examined cut-offs did not significantly differ across the three feigning conditions for any of the cut-off points, $\chi^2(2) \leq 2.64$, $p \geq .27$. Taken together, these findings provide additional evidence that the IOP–29 is similarly effective in evaluating feigned presentations of these three disorders (for a graphical representation of IOP–29 FDS distributions across the three feigning conditions, see Figure 1). Given this, and in accordance with Akca, Tepedelen, et al. (2023) and Volarov et al. (2024), we combined these three conditions into a single main feigning condition (SIM) for subsequent analyses.

Table 1. IOP–29 FDS mean values and sensitivity by feigning subgroup.

	Feigning subgroup		
	PTSD ($n = 42$)	DEP ($n = 42$)	SCZ ($n = 44$)
IOP–29 FDS			
M	.82	.76	.82
SD	.17	.21	.16
Sensitivity			
FDS ≥ .30	.98	.95	1.00
FDS ≥ .50	.91	.88	.93
FDS ≥ .65	.83	.71	.84

Note: IOP–29 = Inventory of Problems–29; FDS = False Disorder Probability Score; PTSD = post-traumatic stress disorder; DEP = depression; SCZ = schizophrenia.

As expected, there was a statistically significant difference in the IOP–29 FDS scores among the honest (HON), random (RND) and feigning (SIM) conditions, $F(2, 254) = 454.81$, $p < .01$, $\eta^2_p = .78$, with all Bonferroni-corrected pairwise comparisons also being statistically significant. Specifically, the mean IOP–29 FDS values were significantly higher ($p < .01$) in the SIM condition ($M = .80$, $SD = .18$) than in both the RND ($M = .68$, $SD = .19$) and HON conditions ($M = .16$, $SD = .14$), and the mean IOP–29 FDS values in the RND condition were significantly higher ($p < .01$) than those in the HON condition. When comparing the IOP–29 FDS results in the HON versus SIM conditions, a very large Hedges *g* of 3.90 and excellent classification accuracy statistics (Table 2) were observed (for a graphical representation of the distribution of the IOP–29 FDS values for the HON and SIM conditions, see Figure 2).

Effectiveness of the IOP–29 RRS

Similar to the findings for the IOP–29 FDS, the IOP–29 RRS values did not exhibit any significant differences among the three feigning subgroups, $F(2, 125) = 0.02$, $p = .98$, $\eta^2_p = .00$. In fact, the three mean RRS values were nearly identical, with scores of 47.7 ($SD = 8.6$) for SIM PTSD, 48.2 ($SD = 9.0$) for SIM DEP and 47.9 ($SD = 9.9$) for SIM SCZ. Thus, for subsequent analyses, data from these subgroups were pooled and analysed collectively, as done for the IOP–29 FDS.

The results of a repeated measures one-way ANOVA revealed a statistically significant difference in the IOP–29 RRS values based on whether participants were instructed

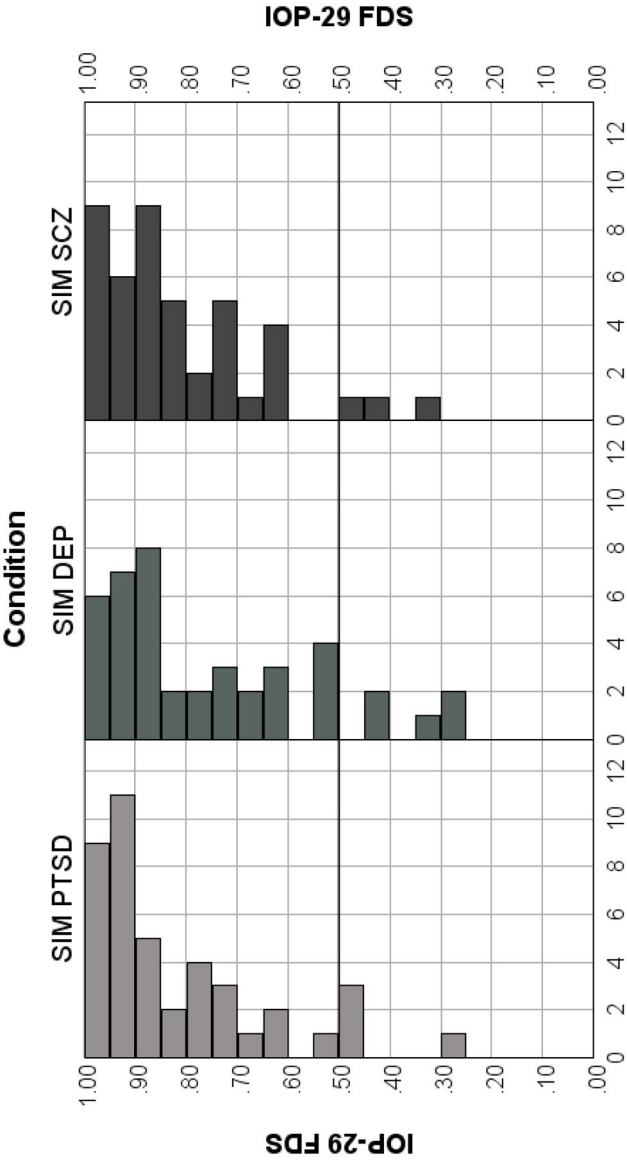

Figure 1. Distribution of Inventory of Problems–29 (IOP–29) False Disorder Probability Score (FDS) values across the three feigning conditions. SIM = feigning/simulating a mental disorder; PTSD = post-traumatic stress disorder; DEP = depression; SCZ = schizophrenia.

Table 2. IOP–29 FDS hit rates within the entire sample (HON vs. SIM conditions).

	Specificity	Sensitivity	OCC
FDS ≥ .30	.88	.98	.93
FDS ≥ .50	.95	.91	.93
FDS ≥ .65	.98	.80	.89

Note: IOP–29 = Inventory of Problems–29; FDS = False Disorder Probability Score; HON = responding honestly; SIM = feigning/simulating a mental disorder; OCC = overall correct classification.

to respond honestly, randomly, or to simulating a mental health disorder, $F(2, 254) = 195.95$, $p < .01$, $\eta^2_p = .61$. Specifically, examination of Bonferroni-corrected pairwise comparisons revealed that the mean IOP–29 RRS values were significantly higher ($p < .01$) in the RND condition than in the HON and SIM conditions, and that the mean IOP–29 RRS values in the HON condition were significantly higher ($p < .01$) than those in the SIM condition. As shown in Table 3, these results suggest that mean IOP–29 RRS values were markedly elevated (approximately between $T = 65$ and $T = 70$) in the RND condition, not elevated at all (approximately $T = 50$) in the HON condition, and very slightly lower than the average (approximately $T = 48$) in the SIM condition. In addition, Table 3 also shows that the IOP–29 RRS demonstrated satisfactory specificity (≥.90) at all three cut-off points, whereas sensitivity values ranged from .36 at $T \geq 71.0$, to .62 at $T \geq 66.5$, and to .83 at $T \geq 61.0$.

Discussion

The Inventory of Problems–29 (IOP–29; Viglione & Giromini, 2020) is a short symptom validity test (SVT) with a rapidly growing research base since its initial publication. Originally developed and validated in the United States (Viglione et al., 2017), more than 30 IOP–29 publications have emerged over the last three years (for a review, see Giromini & Viglione, 2022; for a meta-analysis, see Puente-López, Pina, López-Nicolás, et al., 2023) from more than 15 different countries. These publications include studies conducted in Australia (Gegner et al., 2022), Belgium (Blavier et al., 2024), Brazil (Carvalho et al., 2021), Canada (Abeare, Razvi, et al., 2021), England (Bosi et al., 2024), France (Banovic et al., 2022), Italy (Giromini et al., 2024), Lithuania (Ilgunaite et al., 2022), the Netherlands (Boskovic et al., 2024), Norway (Grønnerød et al., 2023), Portugal (Akca, Martins, et al., 2023), Romania (Crișan, 2023), Serbia (Volarov et al., 2024), Slovenia (Šömen et al., 2021), Spain (Puente-López, Pina, Rambaud-Quiñones, et al., 2023) and Türkiye (Akca, Tepedelen, et al., 2023). Despite this extensive research, no study had yet investigated the German version of the IOP–29, which is why we conducted our study.

Following the procedures of Akca, Tepedelen, et al. (2023) and Volarov et al. (2024), we administered the German IOP–29 to 128 nonclinical native German-speaking adult volunteers under three conditions: responding honestly, responding at random and feigning mental illness. Consistent with the hypothesis that the IOP–29 is a cross-culturally valid and applicable SVT, the German IOP–29 effectively discriminated between honest and feigned presentations with a specificity of .95 and a sensitivity of .91 when using the standard IOP–29 cut-off score. Additionally, a recently introduced IOP–29 score aimed at detecting careless responding showed initial evidence of being useful. Below, we detail how our findings compare to those from previously published studies and offer practical suggestions for using the German IOP–29 in applied settings.

The False Disorder Probability Score (FDS) of the IOP–29 was designed and scaled to discriminate genuine from feigned mental health complaints regardless of the type of symptoms presented (Viglione & Giromini, 2020). This approach ensures that the same test score does not need to be interpreted

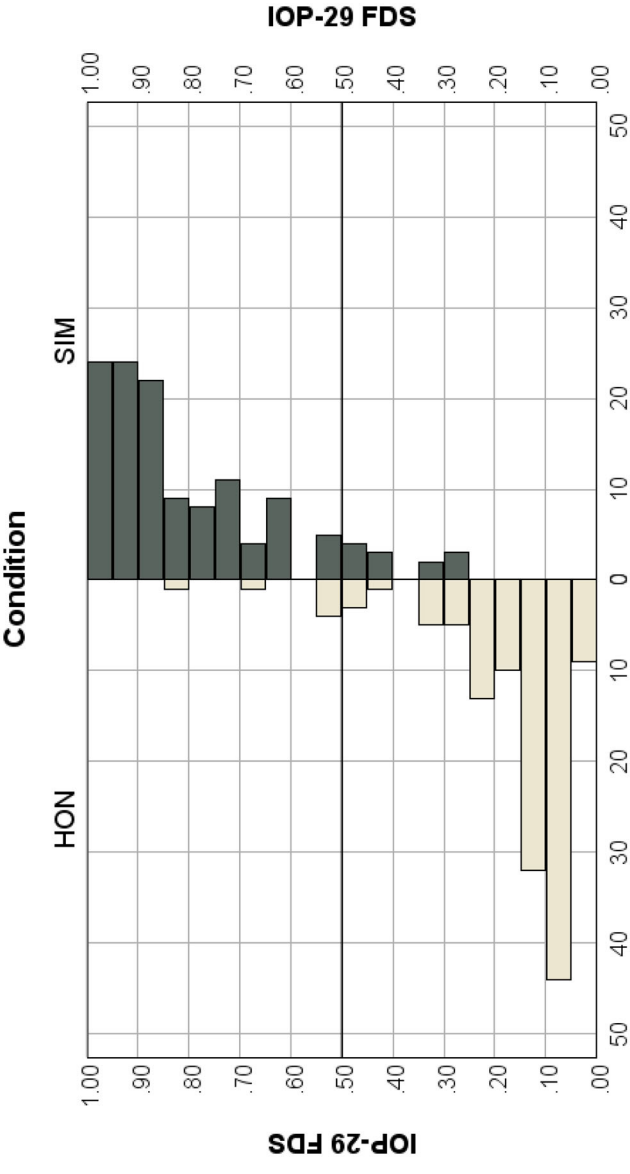

Figure 2. Distribution of Inventory of Problems–29 (IOP–29) False Disorder Probability Score (FDS) values in the responding honestly (HON) and feigning/simulating a mental disorder (SIM) conditions.

Table 3. Mean IOP–29 RRS values and hit rates by condition.

	Condition		
	HON	SIM	RND
IOP–29 RRS			
M	50.5	47.9	67.3
SD	7.6	9.2	9.2
Hit rates			
RRS ≥ 61.0			
Specificity	.92	.91	—
Sensitivity	—	—	.83
RRS ≥ 66.5			
Specificity	.98	.98	—
Sensitivity	—	—	.62
RRS ≥ 71.0			
Specificity	.98	.98	—
Sensitivity	—	—	.36

Note: IOP–29 = Inventory of Problems–29; RRS = Random Responding Scale; HON = responding honestly; SIM = feigning/simulating a mental disorder; RND = responding randomly.

differently for evaluees with different mental health problems, which is particularly important in forensic contexts where standardised tests should provide fully objective and standardised results. In our study, the German IOP–29 FDS performed equally well in the assessment of feigned PTSD, depression and schizophrenia, which is consistent with previously published IOP–29 research. In particular, the studies by Akca, Tepedelen, et al. (2023) and Volarov et al. (2024), which used the same research design as that in our study, also found that the type of feigned symptoms had no significant effect on the mean score or sensitivity of the IOP–29 FDS. Furthermore, and perhaps more importantly, the quantitative literature review by Giromini and Viglione (2022) reported that the overall effectiveness of the IOP–29 FDS remained consistent across different target conditions – depression/anxiety, PTSD, neuropsychological issues and mixed/other – with weighted mean area under the curve (AUC) values ranging from a minimum of .93 to a maximum of .96. Based on these data, we believe it is reasonable to expect that the German IOP–29 FDS will likely perform similarly well regardless of the specific symptoms presented by the evaluee, whether related to PTSD, depression or schizophrenia. From an applied perspective, this consistency, subject to further replication, would simplify the interpretation procedures and make the IOP–29-based determinations more homogeneous across different evaluees and forensic cases.

In the quantitative literature review by Giromini and Viglione (2022), when experimental feigners were compared against non-clinical controls, as was done in our study, the weighted mean Cohen's *d* yielded by the IOP–29 FDS was 3.59 (weighted $SD = 0.98$), and the standard cut-off score of FDS ≥ .50 yielded a weighted mean sensitivity of .89 (weighted $SD = .07$) at a weighted mean specificity of .93 (weighted $SD = .03$). The results observed in our study with the German IOP–29 perfectly align with these findings: Hedges *g* was 3.90, and using the standard IOP–29 FDS cut-off score (FDS ≥ .50) yielded a sensitivity of .91 at a specificity of .95. Once again, thus, the IOP–29 FDS seems to perform similarly well across languages and geographical contexts, which is excellent news for professionals working with evaluees from diverse cultural settings.

To our knowledge, prior to our study, only three other studies (Akca, Tepedelen, et al., 2023; Volarov et al., 2024; Winters et al., 2021) had independently corroborated the findings initially presented by Giromini et al. (2020) on the potential utility of the IOP–29 Random Responding Scale (RRS) as a measure aimed at detecting content-unrelated distortion. Although our study adds further evidence to these previous studies, several factors warrant careful consideration before endorsing the use of the IOP–29 RRS in clinical and forensic settings. A primary concern is the current ambiguity surrounding the optimal cut-off score for the IOP–29 RRS in real-world evaluations. Our study found that an RRS cut-off of ≥61.0 yielded promising results, demonstrating a sensitivity of .83 and

an adequate specificity of .91–.92. However, previous research (Akca, Tepedelen, et al., 2023) reported less satisfactory hit rates using this same cut-off (i.e. specificity = .75, sensitivity = .81). Therefore, there is a need for further investigation and replication. Moreover, all existing studies have involved participants responding randomly to the IOP–29 items, a scenario that may not accurately represent the behaviour likely to be exhibited in high-stakes forensic contexts. Before pushing for the use of the IOP–29 RRS in real-life forensic settings, it is important to do more ecologically valid and in vivo research with people who speak different languages or are easily distracted. Simply put, in light of these considerations, we believe it is premature to implement the IOP–29 RRS in clinical and forensic contexts. For the time being, its use should be confined to research settings to exclude invalid records based on highly conservative cut-off scores.

Like any other empirical investigation, our study has important limitations that warrant acknowledgment. First, our sample lacked clinical or forensic participants, which prevented us from adequately assessing specificity. Second, using a simulation design raises concerns about generalisability to real-world contexts, as the motivations and risks involved in high-stakes assessments differ considerably from those in our experimental study (Abeare, Hurtubise, et al., 2021; Rai et al., 2023). Third, participants completed the IOP–29 multiple times, whereas in real-world settings, there is usually only a single administration. Fourth, our study relied on Qualtrics for data collection, which differs from typical administration procedures and does not allow the direct interaction between experimenter and participant common in real-world assessments. Fifth, due to the exclusive use of the IOP–29, we were unable to assess comparative validity or examine score elevations on the clinical scales of other instruments. Finally, although we provided a vignette with characteristic symptoms, real-life malingerers are likely to prepare more exhaustively when they are assessed in high-stakes contexts.

On the other hand, most of the limitations listed above are common to all simulation studies, and we did take several precautions to maximise external validity, such as including multiple manipulation checks, informing participants that if they 'over-did it' while feigning a disorder, their presentation would not look credible, and so on. All in all, thus, our study still holds the merit of being the first to offer initial evidence based on 384 IOP–29 protocols from 128 German-speaking participants, suggesting that the German IOP–29 is likely to perform comparably to all other versions.

Ethical standards

Declaration of conflicts of interest

Henrike Höpfner has declared no conflicts of interest.

Burcu Uysal has declared no conflicts of interest.

Mark Stemmler has declared no conflicts of interest.

Ali Y. E. Akca has declared no conflicts of interest.

Luciano Giromini declares that he owns a share in the corporate (LLC) that possesses the rights to Inventory of Problems.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee of Maastricht University (ethics approval reference: ERCPN-274_119_11_2023) and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

Informed consent

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

ORCID

Henrike Höpfner http://orcid.org/0009-0008-3892-1039
Burcu Uysal http://orcid.org/0000-0002-0202-2655
Mark Stemmler http://orcid.org/0000-0003-4321-1302
Ali Y. E. Akca http://orcid.org/0000-0002-3450-7059
Luciano Giromini http://orcid.org/0000-0002-9540-4803

Data availability statement

Data and supplemental material are available on the Open Science Framework (OSF): https://osf.io/4nqvc/?view_only=d50860cfa4e745cc94ed57398aa63a77.

References

Abeare, C. A., Hurtubise, J. L., Cutler, L., Sirianni, C., Brantuo, M., Makhzoum, N., & Erdodi, L. A. (2021). Introducing a forced choice recognition trial to the Hopkins Verbal Learning Test – Revised. *The Clinical Neuropsychologist*, *35*(8), 1442–1470. https://doi.org/10.1080/13854046.2020.1779348

Abeare, K., Razvi, P., Sirianni, C. D., Giromini, L., Holcomb, M., Cutler, L., Kuzmenka, P., & 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*(1), 2–16. https://doi.org/10.1007/s12207-021-09402-4

Akca, A. Y. E., Martins, A. T., Brazzoni, L., Jiménez-Ros, A. M., Di Girolamo, M., Barbosa, F., & Zennaro, A. (2023). The association of the dark tetrad with the ability to feign schizophrenia. *Psychology & Neuroscience*, *16*(3), 302–315. https://doi.org/10.1037/pne0000323

Akca, A. Y. E., 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., Zappalà, 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

Ben-Porath, Y. S., & Tellegen, A. (2020). *Minnesota multiphasic personality inventory-3; manual for administration, scoring, and interpretation*. University of Minnesota Press.

Blavier, A., Palma, A., Viglione, D. J., Zennaro, A., & Giromini, L. (2024). 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*, *24*(5), 595–614. https://doi.org/10.1080/24732850.2023.2203130

Bosi, J., Minassian, L., Ales, F., Akca, A. Y. E., Winters, C., Viglione, D. J., Zennaro, A., & Giromini, L. (2024). 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*, *31*(6), 1234–1246. https://doi.org/10.1080/23279095.2022.2115910

Boskovic, I., Akca, A. Y. E., & Giromini, L. (2024). 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*, *31*(4), 626–638. https://doi.org/10.1080/23279095.2022.2057856

Burchett, D., & Bagby, R. M. (2022). Assessing negative response bias: A review of the non-credible overreporting scales of the MMPI-2-RF and MMPI-3. *Psychological Injury and Law*, *15*(1), 22–36. https://doi.org/10.1007/s12207-021-09435-9

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*(1), 58–70. https://doi.org/10.1007/s12207-021-09403-3

Crișan, 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

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2020). *Ethnologue: Languages of the world* (23rd ed.). SIL International.

Erdodi, L. A. (2023a). Cutoff elasticity in multivariate models of performance validity assessment as a function of the number of components and aggregation method. *Psychological Injury and Law*, *16*(4), 328–350. https://doi.org/10.1007/s12207-023-09490-4

Erdodi, L. A. (2023b). Multivariate models of performance validity: The Erdodi Index captures the dual nature of non-credible responding (continuous and categorical). *Assessment*, *30*(5), 1467–1485. https://doi.org/10.1177/10731911221101910

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

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., Pignolo, C., Zennaro, A., & Sellbom, M. (2024). Using the MMPI-2-RF, IOP-29, IOP-M, and FIT in the in-person and remote administration formats: A simulation study on feigned mTBI. *Assessment*, *31*(8), 1626–1642. https://doi.org/10.1177/10731911241235465

Giromini, L., Viglione, D. J., Pignolo, C., & Zennaro, A. (2018). A clinical comparison, simulation study testing the validity of SIMS and IOP-29 with an Italian sample. *Psychological Injury and Law*, *11*(4), 340–350. https://doi.org/10.1007/s12207-018-9314-1

Giromini, L., Viglione, D. J., Pignolo, C., & Zennaro, A. (2020). 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., Young, G., & Sellbom, M. (2022). Assessing negative response bias using self-report measures: New articles, new issues. *Psychological Injury and Law*, *15*(1), 1–21. https://doi.org/10.1007/s12207-022-09444-2

Grønnerød, C., Rekkedal Rolfsnes, 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

Kurtz, J. E., & McCredie, M. N. (2022). Exaggeration or fabrication? Assessment of negative response distortion and malingering with the personality assessment inventory. *Psychological Injury and Law*, *15*(1), 37–47. https://doi.org/10.1007/s12207-021-09433-x

Morey, L. C. (1991). *Personality assessment inventory (PAI). Professional manual*. Psychological Assessment Resources.

Morey, L. C. (2007). *Personality Assessment Inventory (PAI). Professional manual* (2nd ed.). Psychological Assessment Resources.

Pignolo, C., Giromini, L., Ales, F., & Zennaro, A. (2023). Detection of feigning of different symptom presentations with the PAI and IOP-29. *Assessment*, *30*(3), 565–579. https://doi.org/10.1177/10731911211061282

Puente-López, E., Pina, D., López-Nicolás, R., Iguacel, I., & Arce, R. (2023). 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-Quiñones, P., Ruiz-Hernández, J. A., Nieto-Cañaveras, M. D., Shura, R. D., Alcazar-Crevillén, A., & Martinez-Jarreta, B. (2023). 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*, *38*(3), 738–762. https://doi.org/10.1080/13854046.2023.2249171

Rai, J. K., Gervais, R. O., & Erdodi, L. A. (2023). A large-scale investigation of the classification accuracy of various performance validity tests in a medical-legal setting. *Psychology & Neuroscience*, *16*(3), 225–243. https://doi.org/10.1037/pne0000320

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.

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: The Official Journal of the National Academy of Neuropsychologists*, *35*(6), 735–764. https://doi.org/10.1093/arclin/acaa019

Shura, R. D., Ord, A. S., & Worthen, M. D. (2022). Structured Inventory of Malingered Symptomatology: A psychometric review. *Psychological Injury and Law*, *15*(1), 64–78. https://doi.org/10.1007/s12207-021-09432-y

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. http://jaapl.org/content/jaapl/25/2/183.full.pdf

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*(2), 104–113. https://doi.org/10.1007/s12207-021-09412-2

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. (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

Viglione, D. J., & Giromini, L. (2020). *Inventory of Problems-29: Professional manual*. IOP-Test, LLC.

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

Volarov, M., Velimirović, M., Bošković, I., Akca, A. Y. E., & Giromini, L. (2024). The cross-cultural applicability of the Inventory of Problems – 29 (IOP-29): A Replication of Akca et al. (2023) using a Serbian sample. *Manuscript Under Review; Conference Paper Presented:* Velimirović, M., Volarov, M., Bošković, I., & Giromini, L. (2023, October 26–28). Assessing the sensitivity of the Serbian translation of Inventory of Problems-29. [Conference presentation abstract]. Current Trends in Psychology, Novi Sad, Serbia.

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