Psychological Injury and Law
https://doi.org/10.1007/s12207-023-09483-3

The Inventory of Problems–29 is a Cross-Culturally Valid Symptom Validity Test: Initial Validation in a Turkish Community Sample

Ali Y. E. Akca[1,2] · **Mehmed S. Tepedelen**[3] · **Burcu Uysal**[3] · **Laszlo A. Erdodi**[4]

Received: 19 May 2023 / Accepted: 24 August 2023
© The Author(s) 2023

Abstract

Because the actuarial evidence base for symptom validity tests (SVTs) is developed in a specific population, it is unclear whether their clinical utility is transferable to a population with different demographic characteristics. To address this, we report here the validation study of a recently developed free-standing SVT, the Inventory of Problems-29 (IOP-29), in a Turkish community sample. We employed a mixed design with a simulation paradigm: The Turkish IOP–29 was presented to the same participants ($N = 125$; 53.6% female; age range: 19–53) three times in an online format, with instructions to respond honestly (HON), randomly (RND), and attempt to feign a psychiatric disorder (SIM) based on different vignettes. In the SIM condition, participants were presented with one of three scripts instructing them to feign either schizophrenia (SIM-SCZ), depression (SIM-DEP), or posttraumatic stress disorder (SIM-PTSD). As predicted, the Turkish IOP–29 is effective in discriminating between credible and noncredible presentations and equally sensitive to feigning of different psychiatric disorders: The standard cutoff ($FDS \geq .50$) is uniformly sensitive (90.2% to 92.9%) and yields a specificity of 88%. Random responding produces FDS scores more similar to those of noncredible presentations, and the random responding score (RRS) has incremental validity in distinguishing random responding from feigned and honest responding. Our findings reveal that the classification accuracy of the IOP–29 is stable across administration languages, feigned clinical constructs, and geographic regions. Validation of the Turkish IOP–29 will be a valuable addition to the limited availability of SVTs in Turkish. We discuss limitations and future directions.

Keywords Symptom validity test · IOP–29 · Cross-cultural validity · Negative response bias · Turkish

Malingering is the intentionally fabrication or exaggeration of physical and/or psychological symptoms motivated by external incentives (American Psychiatric Association, 2022). If incentive kind cannot be determined, such behavior is labeled *feigning* (see Merten & Merckelbach, 2020; Rogers & Bender, 2018). A third concept related to the above is that of *negative response bias* (NRB; or invalid presentation/performance). NRB is the tendency to portray oneself as more mentally disturbed or problematic than one's actual level of functioning, thereby distorting psychometric test results and assessment (Giromini et al., 2022). Estimates of the prevalence of feigning/NRB vary across instruments and settings (Dandachi-FitzGerald et al., 2013; Mittenberg et al., 2002), but converge around 15% (Young, 2015).

Failure to detect NRB can result in misallocation of resources (Chafetz & Underhill, 2013), unnecessary and potentially iatrogenic interventions (van der Heide et al., 2020), denial of necessary care (Knoll & Resnick, 2006), contamination of treatment protocols (Van Egmond et al., 2005), academic research (Abeare et al., 2021; Rienstra et al., 2013), or adjudication (Soliman & Resnick, 2010). Therefore, the credibility of psychological presentations should be objectively assessed using multiple *symptom validity tests* (SVTs; Sherman et al., 2020; Sweet et al., 2021).

SVTs are psychometric instruments designed to assess the credibility of self-reported symptoms and to provide an

✉ Ali Y. E. Akca
aliyunusemre.akca@unito.it

[1] Department of Psychology, University of Turin, Via Verdi 10, 10124 Turin, Italy

[2] Forensic Psychology Section, Maastricht University, Maastricht, The Netherlands

[3] Department of Psychology, Ibn Haldun University, Istanbul, Türkiye

[4] Department of Psychology, University of Windsor, Windsor, Canada

estimate of the likelihood of NRB. The most common detection strategies are pseudosymptoms (deficits that patients with genuine disorders rarely/never endorse), the threshold method (implausibly high symptomatology), or combinations of symptoms that are unlikely to occur together. By design, SVTs can be *free-standing* or *embedded* (i.e., part of a broadband instrument originally developed to assess true pathology; Giromini et al., 2022).

It is recommended to assess NRB using multiple SVTs (Giromini et al., 2022). However, the clinical and practical validity of psychological tests is influenced by linguistic, ethnic, and cultural differences, so they are not automatically applicable to diverse populations (Brantuo et al., 2022; Nijdam-Jones & Rosenfeld, 2017). This is particularly true for SVTs, where idiosyncratic interpretation of item content (Ali et al., 2022) may increase the risk for both false positives (leading to the denial of care or legitimate compensation) and false negatives (misallocation of scarce resources).

SVTs in Türkiye[1]

Determining the classification accuracy of a given instrument in a cross-cultural context is a prerequisite for its use in clinical/forensic settings. Because the actuarial evidence base of SVTs was developed in a specific population (typically native-born citizens who speak the country's official language as their first language), it is unclear whether their clinical utility is transferable to a population with different demographic characteristics (recent immigrants with limited language skills and different cultural norms). Therefore, it should not be assumed that the classification accuracy of SVTs is transferable to examinees with a different ethnic or cultural background.

Little is known about SVTs in Türkiye or Turkish SVTs, although Turkish is spoken by about 200 million people, ranking seventh in the world in terms of number of speakers and area (see Ministry of Culture and Tourism of the Republic of Türkiye, 2023). In addition, Turks make up a significant portion of the population in several Western European countries. Although some instruments have been translated into Turkish (Brockhaus & Peker, 2003; Morel & Marshman, 2008), there is little research on their use. Ardıç et al. (2019) have reported on the validation study of the Turkish version of the Structured Inventory of Malingered Symptomatology (SIMS; Smith & Burger, 1997), a widely used and researched SVT (see Shura et al., 2022). A recent review of psychological tests in Türkiye (Ayhan & Karaman, 2021) identified broadband instruments with embedded validity scales such as the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1943; Savaşır, 1978) and the Miller Forensic Assessment of Symptoms Test (M- FAST; Keyvan et al., 2015; Miller, 2001) as SVTs in use and concluded that additional tests are needed (see also Akca et al., 2023).

The Inventory of Problems

To address the need for additional Turkish instruments such as SVTs, this study was designed to validate a recently developed free-standing SVT, the *Inventory of Problems–29* (IOP–29; Viglione & Giromini, 2020; Viglione et al., 2017), in a Turkish community sample. The IOP–29 can be administered in paper-and-pencil format or online, takes 5–10 min to complete (Giromini et al., 2021), and has a rapidly growing evidence base (Puente-López et al., 2023; Young et al., 2020). Its item content focuses on *how* problems are presented, rather than their mere presence or absence, and assesses the credibility of a wide range of neuropsychiatric presentations (see Giromini et al., 2020b): Depression (Bosi et al., 2022; Ilgunaite et al., 2022), psychosis (Banovic et al., 2022; Winters et al., 2020), PTSD (see Carvalho et al., 2021), neuropsychological deficits (Gegner et al., 2021; Holcomb et al., 2022), and combinations thereof (see Giromini et al., 2020b; Pignolo et al., 2021). In addition to “True” and “False,” “Doesn't make sense” is presented as a third response option. Eliminating the traditional dichotomy of “True” and “False” allows for a more nuanced manifestation of non-credible presentations feigning strategies (rare symptom or amplification; Viglione et al., 2017).

A key difference from other SVTs is that the IOP–29 generates a *False Disorder Probability Score* (FDS) by comparing responses to reference data obtained from real patients and experimental feigners (Giromini et al., 2020b). The FDS is a probability estimate ranging from 0 (unlikely to feign) to 1 (likely to feign) derived from logistic regression (Winters et al., 2020). This scaling method reduces the impact of outliers in research settings, but also facilitates decision making in real-world forensic settings where findings are often expressed as percentages or “on the balance of probabilities” (Giromini et al., 2020b).

The original English version of the IOP–29 has been translated and cross-validated in several countries (Brazil, France, Italy, Lithuania, Portugal, Romania, Slovenia) and is available in several other languages (see official website: http://www.iop-test.com). Emerging data indicate that the validity of the IOP–29 is at least comparable to that of longer and more complex instruments such as the NRB indicators embedded in the MMPI and the Personality Assessment

[1] The country formerly known as "Turkey" has officially changed its name to Türkiye. This change was adopted by the United Nations, the United States and other organizations and countries. See, e.g., United Nations. (2022, June 3).

Inventory (PAI; Morey, 2007), or popular free-standing SVTs such as SIMS and the Self-Report of Symptom Inventory (SRSI; Boskovic et al., 2022; Giromini et al., 2018, 2019, 2020a; Holcomb et al., 2022; Merten et al., 2016; Pignolo et al., 2021; Roma et al., 2020). The IOP–29 is also robust to coaching (Boskovic et al., 2022; Gegner et al., 2021), a potential threat to the clinical utility of any SVT (Rogers & Bender, 2018).

The developers recommend three levels of cutoffs: a *liberal* FDS ≥ 0.30 for screening-only purposes, a *standard* FDS ≥ 0.50 cutoff, and a *conservative* FDS ≥ 0.65 for high-stakes contexts where specificity is of paramount importance (Viglione & Giromini, 2020). Naturally, these cutoffs offer different tradeoffs in sensitivity and specificity: 0.90 and 0.60 for FDS ≥ 0.30; 0.80 and 0.80 for FDS ≥ 0.50; 0.70 and 0.90 for FDS ≥ 0.65 (Viglione & Giromini, 2020). However, a recent qauntitative review by Giromini and Viglione (2022) reports promising weighted mean sensitivity and specificity values that *exceed* those reported in the test manual: 0.94 and 0.76 for FDS ≥ 0.30, 0.86 and 0.92 for FDS ≥ 0.50, and 0.76 and 0.96 for FDS ≥ 0.65.

Although most of the studies reviewed were based on a simulation design that may inflate (Rogers & Bender, 2018) or obscure (Abeare et al., 2021) the true effect size, a comparison with the criterion group study by Roma et al. (2020) yielded comparable weighted effect size (Cohen's *d* of 3.02 vs. 2.98), sensitivity, and specificity (Giromini & Viglione, 2022). Recently, an independent replication based on psychometrically defined criterion groups and genuine patients found that the standard cutoff (FDS ≥ 0.50) had good classification accuracy, suggesting that it "may be sufficiently specific (0.90–0.91) for routine clinical use" (Holcomb et al., 2022). A second meta-analysis confirmed the IOP–29 as an effective SVT (Puente-López et al., 2023) but cautioned against language of administration (i.e., English vs. non-English) as a potential confounding variable and emphasizing the need for further cross-cultural research.

The present study is a response to this call: it was designed to develop and validate the Turkish version of the IOP–29. In addition, we wanted to test the sensitivity of the Turkish IOP–29 for different types of feigned disorders: schizophrenia, depression, and PTSD. Finally, we wanted to assess the utility of the Random Responding Scale (RRS; Giromini et al., 2020c) in discriminating between random responding and feigning. The RRS was developed as an alternative index to detect careless, uncooperative, or inattentive responding and to distinguish this response pattern from NRB (Giromini et al., 2020b, c; Winters et al., 2020). Indeed, evidence suggests that content-unrelated distortions (CUD; Nichols et al., 1989), such as random responding, can mimic NRB (Burchett et al., 2016; Merckelbach et al., 2019).

The Current Study

The study employed a mixed design using a simulation paradigm (Rogers & Bender, 2018): the Turkish IOP–29 was presented to the same participants three times in an online format, with instructions to respond honestly (HON), at random (RND), and attempt to feign a psychiatric disorder (SIM) based on various vignettes. Individuals who respond with a careless attitude do not respond completely at random, even when prompted (see Giromini et al., 2020c). To increase ecological validity and minimize the risk of artificially inflating effect sizes with computer-generated random responses, we added the condition RND. In the SIM condition, participants were presented with three scripts instructing them to feign schizophrenia (SIM-SCZ), depression (SIM-DEP), or PTSD (SIM-PTSD). We predicted no significant difference between participants' FDS scores in the three feigning sub-conditions, but we hypothesized that FDS scores would be significantly lower in the HON condition compared with SIM (suggesting a valid profile). In addition, we expected that mean FDS values in RND would be between those of HON and SIM but closer to the mean of SIM, as a random response set is unlikely to be credible (see Giromini et al., 2020b).

Method

Participants

Based on *G*Power* analysis (Faul et al., 2009) with three groups (i.e., feigning sub-conditions SIM-SCZ, SIM-DEP, and SIM-PTSD) and with parameters α (alpha) = .05, 1–β (power) = .80, and f (effect size) = 0.25, a minimum sample size of 159 participants with valid responses (i.e., meeting inclusion criteria for analysis) was determined.[2]

Inclusion criteria for participation in the study were age ≥ 18 years old, able to read and understand Turkish, and provide informed consent (N = 189). Exclusion criteria were a self-reported history of psychiatric or neurological disorders (n = 12) and self-reported poor mental health during the study (n = 17). In addition, failure to pass manipulation and inattention checks resulted in exclusion from the analysis (n = 35; see "Procedure" section).

[2] A larger sample size is required to detect a between-subjects factor difference than to detect a within-subjects factor difference. Therefore, our a priori power analysis focused on the between-subjects factor, i.e., testing whether the IOP–29 FDS in the SIM condition varied as a function of the type of disorder being feigned.

Table 1 Demographic Composition of the Sample (N = 125)

	SIM-SCZ ($n=42$)	SIM-DEP ($n=41$)	SIM-PTSD ($n=42$)	Entire Sample ($N=125$)
Gender[a], $\chi^2_{(2)}=0.42$, $p=.81$				
Female	21 (50.0%)	23 (56.1%)	23 (54.8%)	67 (53.6%)
Male	21 (50.0%)	18 (43.9%)	18 (42.9%)	57 (45.6%)
Prefer not to say			1 (2.4%)	1 (0.8%)
Age, $F_{(2,124)}=2.33$, $p=.10$				
M	26.33	29.98	27.10	27.78
SD	6.85	10.30	6.65	8.17
Student?, $\chi^2_{(2)}=2.81$, $p=.25$				
Yes	28 (66.7%)	20 (48.8%)	23 (54.8%)	71 (56.8%)
No	14 (33.3%)	21 (51.2%)	19 (45.2%)	54 (43.2%)
Involved in Relationship, $\chi^2_{(2)}=0.21$, $p=.90$				
Yes (married, cohabiting, dating, …)	24 (57.1)%	23 (56.1%)	22 (52.4%)	69 (55.2%)
No (single, divorced, …)	18 (42.9%)	18 (43.9%)	20 (47.6%)	56 (44.8%)
Education[b], $\chi^2_{(6)}=8.03$, $p=.24$				
High school or less	3 (21.4%)	1 (4.8%)	2 (10.5%)	6 (11.1%)
Associates degree	2 (14.3%)	~	1 (5.3%)	3 (5.6%)
Undergraduate degree	8 (57.1%)	15 (71.4%)	10 (52.6%)	33 (61.1%)
Postgraduate degree	1 (7.1%)	5 (23.8%)	6 (31.6%)	12 (22.2%)

(1) χ^2 = Chi-square value; *F* = Analysis of Variance value; *p* = probability value of type I error; ~ = not applicable (N/A); (2) [a] Chi-square value was calculated after removal of the single case responding "Prefer not to say"; [b] Chi-square value was calculated for non-student participants; (3) HON = honest condition; RND = Random responding condition; SIM-SCZ = feigning schizophrenia condition; SIM-DEP = feigning depression condition; SIM-PTSD = feigning PTSD condition

One hundred and twenty-five Turkish-speaking adults (67 or 53.6% female; see Table 1) completed the study. The majority (97.6%) reported Turkish as their native language and rated their language proficiency highly on a five-point scale (1 = "very poor"; 5 = "very good"). For the entire sample ($N=125$), the mean score for Turkish language proficiency was $M=4.72$ ($SD=.55$; range: 2–5). Mean age was 27.8 ($SD=8.2$; range: 19–53). Participants assigned to between-subject factors (SIM-SCZ, SIM-DEP, and SIM-PTSD) did not differ on gender, age, self-reported Turkish language proficiency, level of education, relationship status, and self-reported mental health (all $ps>.05$; Table 1).

Procedure[3]

Using a translation/back-translation method (e.g., Brislin, 1970; van de Vijver & Hambleton, 1996), the original English version of the IOP–29 was first translated into Turkish by two native speakers who were not involved in the study and then back-translated into English by a third person who was blind to the original English version. After institutional research ethics approval, the study was advertised on social media and recruitment was promoted through snowballing.

Participants were compensated for their time with the opportunity to enter a raffle to win an Amazon gift card worth 250 Turkish Liras (TL; about $13 U.S. Dollar [USD]). After provided their demographic information, participants were asked to complete the IOP-29 three times under different instructions: HON, RND, and SIM – the within-subjects factor. For the between-subjects factor, participants were randomly assigned to one of the three vignettes in the SIM condition (SIM-SCZ, SIM-DEP, and SIM-PTSD). The Turkish versions of the scenarios are available on the *open science framework* (OSF) platform (https://osf.io/6xksf/). To facilitate feigning, participants were presented with a vignette containing characteristic symptoms for each of the three psychiatric disorders. Participants were also warned not to "overdo it" to avoid being detected as feigners.[4] They were informed that two participants would be rewarded with a 100 TL (approximately 5 USD)

[3] We began collecting data on January 26, 2023, and early in the morning of February 6, 2023, a magnitude 7.7 earthquake struck Türkiye, followed by a second magnitude 7.6 earthquake on the same day that directly affected approximately 13 million people and resulted in devastating consequences. Because the scenario of one of the sub-feigning conditions (i.e., SIM-PTSD) required participants to act as if they were survivors of an earthquake, we immediately stopped data collection. It would not have been ethical under these circumstances to ask participants to pretend to be in this condition.

[4] This warning has been implemented in the vignette, i.e. it has been scripted.

Amazon gift card if they managed to successfully feign the assigned disorder (operationalized as IOP–29 FDS < .50).

The design and vignettes have been used in previous publications (e.g., Giromini et al., 2020b; Pignolo et al., 2021). To neutralize potential order effects, the sequence of the main conditions (i.e., HON, RND, and SIM) was counterbalanced, in addition to random assignment to the feigning sub-condition (SIM-SCZ, SIM-DEP, and SIM-PTSD). For each condition, pretest (understanding of the instructions) and posttest manipulation checks (execution of the task) were employed. Following methodological recommendations (e.g., Meade & Craig, 2012; Ziegler, 2015), an item was added to the Turkish IOP-29 items to check for inattentive responding (HON condition only).

Data Analysis

First, we compared FDS scores between the three feigning conditions with a between-subjects comparison (one-way ANOVA) to determine whether there were significant Turkish IOP–29 false disorder probability score (FDS) differences for different types of feigning presentations (schizophrenia feigning condition [SIM-SCZ], depression feigning condition [SIM-DEP], and PTSD feigning condition [SIM-PTSD]), followed by a Tukey-corrected post hoc contrast. The repeated measures ANOVA was followed by Bonferroni-corrected post hoc contrasts. Effect size estimates were expressed in partial eta squared (η^2_p) and Hedges' *g*. Next, we computed classification accuracy [area under the curve (AUC), sensitivity, specificity, positive predictive power (PPP), negative predictive power (NPP), and overall correct classification (OCC)] with honest responders (HON) versus feigners of psychiatric disorders (SIM) as criterion groups across commonly used FDS cutoffs ($\geq .30$, $\geq .50$, and $\geq .65$) and alternatives. We repeated the process for the random responding scale (RRS). The above analyzes were performed using IBM SPSS 25. Data can be retrieved from the OSF platform (https://osf.io/6xksf/).

Results

Fig. 1 shows the distribution of Turkish IOP–29 FDS scores for the three feigning sub-conditions (SIM-SCZ, SIM-DEP, and SIM-PTSD). One-way ANOVA revealed no difference in FDS score: $F(2, 122) = 0.052$, $p = .949$. This was further tested by a Tukey-corrected post hoc comparison, which also showed no independent significant difference (*p*s > .05; *g*s between .03 and .07). Therefore, the three sub-conditions were combined into one main feigning condition (SIM). The conditions HON, RND, and SIM formed the within-subjects factor. Table 2 shows the mean FDS scores for each main

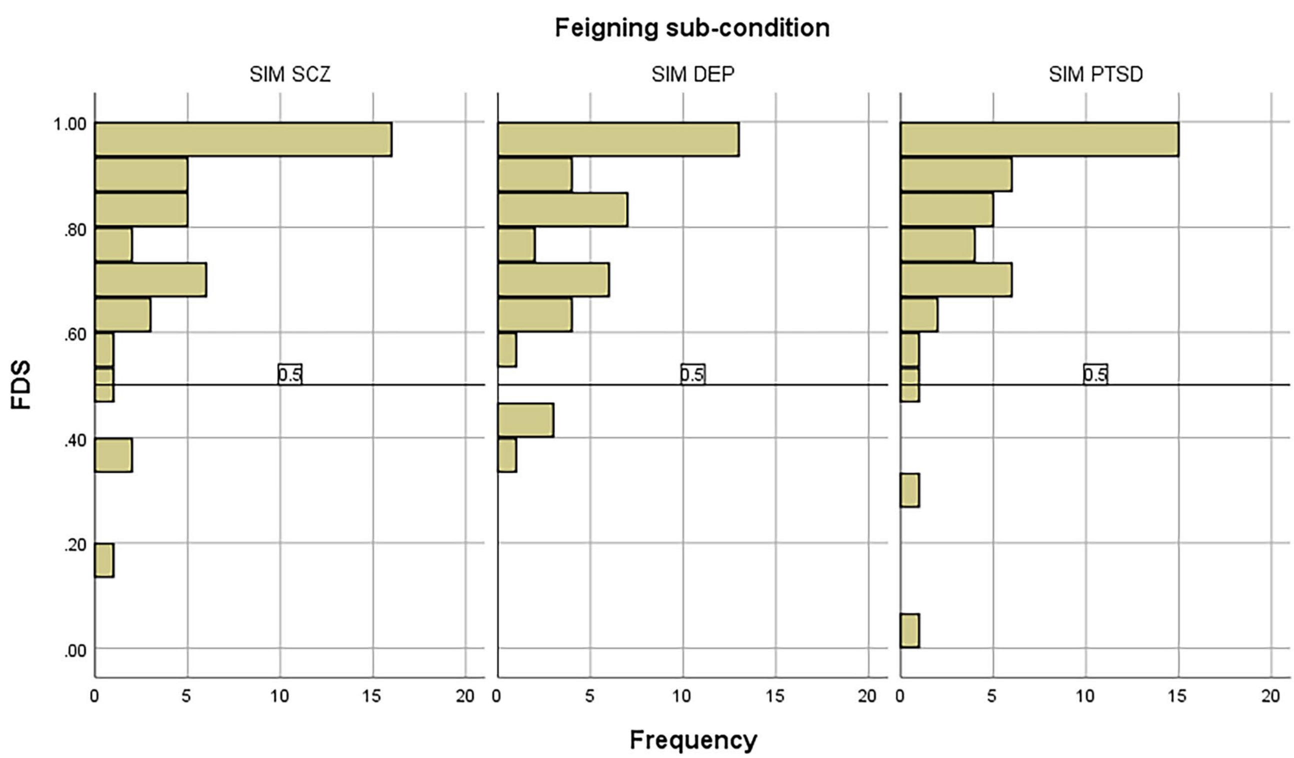

Fig. 1 Graphical representation of Turkish Inventroy of Problems–29 (IOP–29) false disorder probability scores (FDS) across the three feigning sub-conditions (between-subjects factor)

Table 2 Descriptive statistics for Turkish IOP-29 FDS across all conditions, compared to Giromini et al. (2020b) and Winters et al. (2020)

	Turkish IOP–29 FDS			Giromini et al. (2020b)			Winters et al. (2020)		
	N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
Main conditions									
vHON (Standard instructions condition)	125	.25	.22	400	.22	.17	151	.14	.14
RND (Random responding condition)	125	.66	.21	400	.65	.27	~	~	~
SIM (Experimental feigning condition)	125	.80	.19	400	.82	.20	151	.82	.18
Feigning sub-conditions									
SIM-SCZ (Schizophrenia)	42	.80	.20	100	.85	.19	~	~	~
SIM-DEP (Depression)	41	.79	.18	100	.86	.16	~	~	~
SIM-PTSD (PTSD)	42	.80	.20	100	.76	.24	~	~	~

FDS False Disorder Probability score, *PTSD* posttraumatic stress disorder, ~ not applicable (N/A)

condition and the feigning sub-conditions along with the results of two previous studies that used the same design.

Comparison of FDS values with repeated measures ANOVA between HON, RND, and SIM conditions was significant: $F(2, 248) = 265.0$, $p < .001$, $\eta^2_p = .68$ (extremely large). FDS scores were higher in SIM ($M = .80$, $SD = .19$) compared to HON ($M = .25$, $SD = .22$; $g = 2.68$, very large effect) and RND ($M = .66$, $SD = .21$, $g = 0.70$, medium-large effect). The mean FDS score was higher for RND than for HON ($g = 1.91$, very large effect; Table 3).

The FDS achieved an excellent AUC of .95 ($SE = .01$) in discriminating between HON and SIM. The standard cutoff (FDS ≥ .50) correctly classified 88.0% of HON (specificity) and 91.2% of SIM (sensitivity). For the feigning sub-conditions (i.e., SIM-SCZ, SIM-DEP, and SIM-PTSD), sensitivity ranged from 90.2% to 92.9%. Use of the liberal cutoff (FDS ≥ .30) resulted in the predictable tradeoff between decreased specificity (71.2%) and improved sensitivity (97.6%). Conversely, the more conservative cutoff (FDS ≥ .65) achieved high specificity (93.6%) at the expense of sensitivity (81.6%; Table 4). The sequence of the experimental conditions was not related to the FDS. Table 5 shows the positive predictive power (PPP), negative predictive power (NPP), and overall correct classification (OCC) at different cutoffs and base rates for the Turkish IOP–29 FDS.

Classification Accuracy of the RRS

RRS scores across feigning sub-conditions did not differ significantly: $F(2, 122) = 2.33$, $p = .101$ (Fig. 2). This was further tested by a Tukey-corrected post hoc comparison, which also showed no independent significant difference (*ps* > .05; *gs* between 0.23 and 0.43). However, a very large main effect was found for the main conditions (HON, RND, and SIM): $F(1.854, 229.875) = 74.7$, $p < .001$, $\eta^2_p = .38$.

The highest RRS scores were obtained in the RND condition ($M = 67.2$, $SD = 9.7$), followed by SIM ($M = 55.3$, $SD = 11.2$) and HON ($M = 54.0$, $SD = 8.0$). As expected, random responding (RND condition) was associated with higher scores relative to honest responders (HON condition; $g = 1.48$, very large) and SIM ($g = 1.13$ large). The contrast between HON and SIM conditions was *not* significant ($p = .837$).

The recommended RRS cutoff (T ≥ 61; Giromini et al., 2020c) correctly identified 79.2% of HON (specificity) and 71.2% of SIM (specificity). For the RND condition,

Table 3 One-way repeated measures ANOVA outcomes for the main within-subject factor conditions HON, RND, and SIM (N = 125)

Condition (I)		Condition (J)	Mean Difference	*SE*	*p*	95% CI for Mean Difference: Lower	95% CI for Mean Difference: Upper	Hedges' *g*
$HON_{IOP–29\ FDS}$	-	$RND_{IOP–29\ FDS}$	-0.412	.026	< .001	-0.474	-0.350	1.91
$HON_{IOP–29\ FDS}$	-	$SIM_{IOP–29\ FDS}$	-0.551	.025	< .001	-0.613	-0.489	2.68
$RND_{IOP–29\ FDS}$	-	$SIM_{IOP–29\ FDS}$	-0.139	.024	< .001	-0.196	-0.081	0.70

(1) P-value and confidence intervals corrected using Bonferroni method; (2) *SE* = Standard error of the mean difference between conditions I and J; *p* = probability value of type I error; (3) HON = honest condition; RND = Random responding condition; SIM = experimental feigning main condition; (4) IOP–29 = Inventory of Problems–29; FDS = false disorder probability score

Table 4 Classification accuracy of the Turkish IOP–29 for the main conditions HON and SIM, and the feigning sub-conditions (N = 125)

	Specificity	Sensitivity			
	HON (*n* = 125)	*SIM* (*n* = 125)	*SIM-SCZ* (*n* = 42)	*SIM-DEP* (*n* = 41)	*SIM-PTSD* (*n* = 42)
IOP–29 FDS cutoff					
FDS ≥ .70	117 (93.6%)	94 (75.2%)	31 (73.8%)	30 (73.2%)	33 (78.6%)
Conservative FDS ≥ .65*	117 (93.6%)	102 (81.6%)	34 (81.0%)	32 (78.0%)	36 (85.7%)
FDS ≥ .60	115 (92.0%)	111 (88.8%)	37 (88.1%)	36 (87.8%)	38 (90.5%)
FDS ≥ .55	115 (92.0%)	111 (88.8%)	37 (88.1%)	36 (87.8%)	38 (90.5%)
Standard FDS ≥ .50*	110 (88.0%)	114 (91.2%)	38 (90.5%)	37 (90.2%)	39 (92.9%)
FDS ≥ .45	103 (82.4%)	116 (92.8%)	39 (92.9%)	37 (90.2%)	40 (95.2%)
FDS ≥ .40	94 (75.2%)	119 (95.2%)	39 (92.9%)	40 (97.6%)	40 (95.2%)
FDS ≥ .35	94 (75.2%)	119 (95.2%)	39 (92.9%)	40 (97.6%)	40 (95.2%)
Liberal FDS ≥ .30*	89 (71.2%)	122 (97.6%)	41 (97.6%)	41 (100%)	40 (95.2%)
FDS ≥ .25	87 (69.6%)	123 (98.4%)	41 (97.6%)	41 (100%)	41 (97.6%)
FDS ≥ .15	58 (46.4%)	123 (98.4%)	41 (97.6%)	41 (100%)	41 (97.6%)

HON honest condition, *SIM* experimental feigning main condition, *SIM-SCZ* feigning schizophrenia condition, *SIM-DEP* feigning depression condition, *SIM-PTSD* feigning PTSD condition, *IOP–29* Inventory of Problems–29, *FDS* false disorder probability score

[*]Recommended cutoffs

the sensitivity was 80.8%. Because there was no difference between HON and SIM in RRS scores, we combined the data from HON and SIM and applied the recommended cutoff T ≥ 61. This resulted in a specificity of 75.2%. Comparison of RND with HON, SIM, and the combination of these two conditions (i.e., non-random responding) resulted in an AUC of .85 (*SE* = .03), .80 (*SE* = .03), and .83 (*SE* = .02), respectively.

The cutoff value T ≥ 61 was recommended to balance false positives and false negatives. However, we explored alternative cutoff values based on the commonly used specificity standards of 90% and 95%. As with the recommended cutoff value (T ≥ 61), we combined the data to examine these alternative cutoff values. Examination of the AUC results showed that T ≥ 66.5 and T ≥ 71.0 for

Table 5 Positive predictive power (PPP) and negative predictive power (NPP) at different cutoffs and base rates for the Turkish IOP-29 FDS

	Base rate = 50%			Base rate = 40%[a]			Base rate = 30%			Base rate = 15%[b]		
Cutoffs	*PPP*	*NPP*	*OCC*	*PPP*	*NPP*	*OCC*	*PPP*	*NPP*	*OCC*	*PPP*	*NPP*	*OCC*
FDS ≥ .70	.922	.791	.844	.887	.850	.862	.834	.898	.881	.675	.955	.908
FDS ≥ .65*	.927	.836	.876	.895	.884	.888	.845	.922	.900	.692	.967	.918
FDS ≥ .60	.917	.892	.904	.881	.925	.907	.826	.950	.910	.662	.979	.915
FDS ≥ .55	.917	.892	.904	.881	.925	.907	.826	.950	.910	.662	.979	.915
FDS ≥ .50*	.884	.909	.896	.835	.938	.893	.765	.959	.890	.573	.983	.885
FDS ≥ .45	.841	.920	.876	.779	.945	.866	.693	.964	.855	.482	.985	.840
FDS ≥ .40	.793	.940	.852	.719	.959	.832	.622	.973	.812	.404	.989	.782
FDS ≥ .35	.793	.940	.852	.719	.959	.832	.622	.973	.812	.404	.989	.782
FDS ≥ .30*	.772	.967	.844	.693	.978	.818	.592	.986	.791	.374	.994	.752
FDS ≥ .25	.764	.978	.840	.683	.985	.811	.581	.990	.782	.364	.996	.739
FDS ≥ .15	.647	.967	.724	.550	.978	.672	.440	.985	.620	.245	.994	.542

PPP Positive predictive power, *NPP* Negative predictive power, *OCC* Overall correct classification, *IOP–29* Inventory of Problems–29, *FDS* false disorder probability score; PPP and NPP for base rates were calculated using Streiner's (2003) formulas

*Recommended cutoffs

[a]Larrabee et al. (2009)

[b]Young (2015)

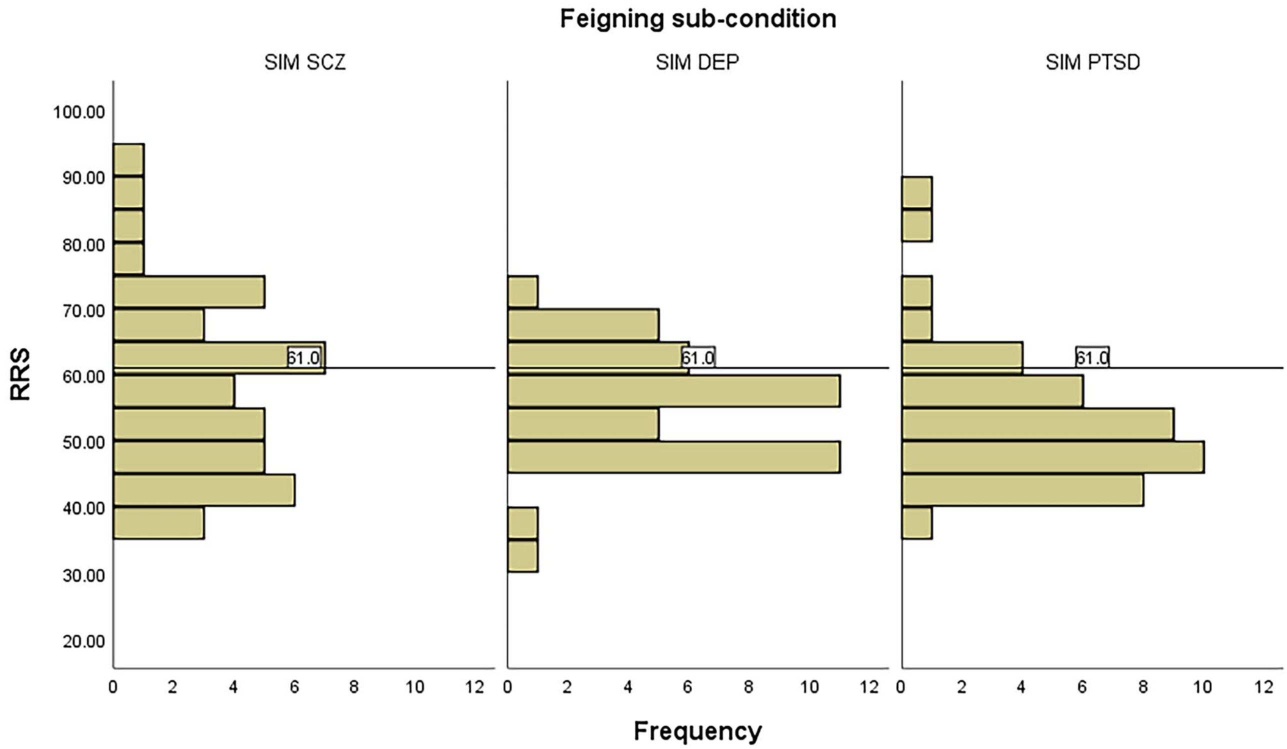

Fig. 2 Graphical representation of Turkish Inventroy of Problems–29 (IOP–29) random responding scores (RRS) across the three feigning sub-conditions (between-subjects factor)

the HON and SIM conditions yielded a specificity of 90% and 95%, respectively. In the present sample, the specificity was 90.4% at T ≥ 66.5 and 96.0% at T ≥ 71.0. Table 6 shows the PPP, NPP, and OCC at different cutoffs and base rates for the Turkish IOP–29 RRS.

FDS and RRS were positively correlated in the RND condition [$r(123) = .43$, $p < .001$], but not the HON and SIM conditions ($r = .16$ and $r = -.10$, respectively; $p > .05$).

Discussion

There is a growing consensus that NRB can undermine the validity of psychological assessments and therefore should be carefully monitored in both forensic and clinical contexts (e.g., Sherman et al., 2020; Sweet et al., 2021). The credibility of clinical presentation should be assessed by combining multiple sources of information, which include

Table 6 Positive predictive power (PPP) and negative predictive power (NPP) at different cutoffs and base rates for the Turkish IOP-29 RRS

			Base rate = 30%			Base rate = 10%			Base rate = 5%			Base rate = 2%		
Cutoffs	Se	Sp	*PPP*	*NPP*	*OCC*	*PPP*	*NPP*	*OCC*	*PPP*	*NPP*	*OCC*	*PPP*	*NPP*	*OCC*
T ≥ 71.0	.304	.960	.765	.763	.763	.458	.925	.894	.286	.963	.927	.134	.985	.947
T ≥ 66.5	.616	.904	.733	.846	.818	.416	.955	.875	.252	.978	.890	.116	.991	.898
T ≥ 61*	.808	.752	.583	.901	.769	.266	.972	.758	.146	.987	.755	.062	.995	.753

(1) * = Recommended cutoff; (2) Se = sensitivity; Sp = specificity; PPP = Positive predictive power; NPP = Negative predictive power; OCC = Overall correct classification; IOP–29 = Inventory of Problems–29; RRS = random responding scale score; (3) PPP and NPP for base rates were calculated using Streiner's (2003) formulas; (4) Sensitivity were calculated based on the random responding (RND) condition; Specificity were calculated based on the collapsed data from honest (HON) and feigning (SIM) condition; (5) The MMPI literature suggests that the prevalence of inconsistent or careless responses in forensic settings is likely between 5 and 20% (Sellbom et al., 2010; Wise, 2009; Wygant et al., 2010), with some studies reporting an even higher rate of about 30% (Gu et al., 2017). A short test such as the IOP-29 is probably less likely to observe careless or random responding than a long test such as the MMPI, which is why we chose to report rates between 2 and 30%

psychometric testing (Bush et al., 2014; Schutte et al., 2015; Sweet et al., 2021). In addition, it is important that SVTs are cross-validated in the target population in which they are used (e.g., Nijdam-Jones & Rosenfeld, 2017). The availability of SVTs in non-English speaking populations in general (Crişan, 2023) and in Turkish in particular (Ayhan & Karaman, 2021) is remarkably limited.

To fill this knowledge gap, this study reports the results of the initial validation of the Turkish version of the IOP–29, a free-standing SVT. The main objective was to investigate the cross-cultural validation of the IOP–29 (e.g., Viglione & Giromini, 2020) for Turkish speakers. We compared the response sets of 125 participants who were asked to complete the IOP–29 three times according to different instructions: Standard (HON), random responding (RND), and feigning a psychiatric disorder (SIM). In the SIM condition, participants were randomly assigned to SIM-SCZ, SIM-DEP, or SIM-PTSD. This design allowed us to compare the sensitivity of the Turkish IOP–29 to feigning different psychiatric disorders. Notably, the IOP–29. was designed to be applicable to a wide range of symptom presentations rather than symptoms belonging to a single, specific diagnosis (e.g., schizophrenia), which promotes ecological validity and relevance in a real-world setting (forensic and/or clinical) where engagement in NRB is typically associated with a variety of symptom presentations or combinations thereof (Giromini et al., 2020b; Viglione et al., 2019). This design therefore also allowed us to test the cross-cultural validity of this specific feature of IOP–29. Finally, we tested the classification accuracy of the RRS – an index specifically designed to distinguish random responding from feigned or honest presentations.

The results converge on a number of conclusions. As predicted, the Turkish IOP–29 was equally sensitive to the feigning of different psychiatric disorders: The standard cutoff (FDS ≥ .50) was uniformly sensitive (90.2% to 92.9%). This result is consistent with previous studies using the same design (e.g., Giromini et al., 2020b), suggesting that the classification accuracy of the IOP–29 is stable across different languages of administration, feigned clinical constructs, and geographic regions.

Consistent with previous reports (Giromini et al., 2020b; Šömen et al., 2021) and our a priori prediction, the mean FDS was lower for HON than for SIM ($g = 2.68$, extremely large effect). Compared with Winters et al. (2020), the FDS for the HON condition was higher (.14 versus .25; $g = .60$, medium effect) and more variable (*SD* of .14 versus .22), whereas the FDS for SIM was similar (*M* of .80 versus .82; *SD* of .18 versus .19), suggesting that the presentation of feigned psychiatric disorders may be culturally invariant despite significant differences in honest responding (HON). Classification accuracy (88.0% specificity at 91.2% sensitivity) of the standard FDS cutoff (≥ .50) was similar to values reported in previous studies with similar designs (Banovic et al., 2022; Carvalho et al., 2021; Giromini & Viglione, 2022; Ilgunaite et al., 2022).

As hypothesized, the average FDS in RND was intermediate between the values of the HON and SIM conditions but closer to the latter (Table 2), consistent with the reports of Giromini et al. (2020b), despite the differences in experimental conditions. We instructed participants to respond randomly without considering item content, whereas Giromini et al. (2020b) asked their participants to feign a psychiatric/neuropsychological disorder and to respond randomly (i.e., randomly/uncooperatively) and compared them with an additional sample instructed to respond completely at random (similar to the present study). The latter group yielded a mean FDS of .50 ($g = .63$, medium effect), which was lower and less variable than our results. It has long been known that people have difficulty producing truly random responses (Towse & Neil, 1998; Wagenaar, 1972); the present results suggest additional cross-cultural variability in the construct of *randomness*. From a practical perspective, the IOP–29 is likely to flag random response sets as noncredible, even though such profiles contain weaker psychometric evidence of invalid responding than individuals who intentionally present themselves as having a serious mental disorder.

The RRS, a novel alternative outcome measure to distinguish between random responders, feigners and honest responders, performed well in this sample (g: 1.13–1.48, large effects). However, the proposed cutoff of T ≥ 61 to balance false positives and false negatives (Giromini et al., 2020c) yielded lower specificity than Winters et al. (2020): 79.2% vs. 96.7% for HON and 71.2% vs. 84.1% for SIM. This discrepancy further highlights possible cultural differences in the conceptualization of random responding. Raising the cutoff to T ≥ 66.5 resulted in a specificity of 90.4% for the combined sample of HON and SIM, with a sensitivity of 61.6%. This classification accuracy resonates with the *Larrabee limit* – the seemingly inescapable tradeoff in validity testing that states that if specificity is set at 90%, sensitivity tends to hover around 50% (Crişan et al., 2021; Lichtenstein et al., 2019). Setting an even more conservative cutoff value (T ≥ 71) resulted in a further improvement in specificity (96%), but at the expense of sensitivity (30.4%). The fact that FDS and RRS were positively correlated in the RND condition, but not HON and SIM, provides preliminary evidence of the convergent and divergent validity of RSS. Although the target construct of RSS (distortions unrelated to item content) is clinically relevant, further research is needed to clarify its differential sensitivity to specific confounding factors (e.g., difficulty in comprehension, distraction, boredom, fatigue, etc.).

Limitations

Inevitably, the study has a number of limitations. First, the results of simulation designs have low external validity (Rai et al., 2019; Rogers & Bender, 2018). Examinees in the real world have much stronger incentives to successfully feign and may use different strategies (Abeare et al., 2022; An et al., 2017). Nevertheless, our results are consistent with previous research in Italy and the United Kingdom based on the same research design (Giromini et al., 2020b; Winters et al., 2020), suggesting that the Turkish, Italian, and English versions of the IOP–29 have similar classification accuracy when tested under the same conditions. Second, exclusion of participants with psychiatric/neurological disorders was based on self-report that could not be independently verified. Additional screening measures for specific disorders would have improved the diagnostic purity of the sample (but also increased the time required of participants). Third, although we employed manipulation checks to verify compliance with the instructions, the true fidelity with which instructions were executed could not be determined with certainty (Abeare et al., 2021). Finally, the vignettes used in this study likely did not elicit the motivation that is typical of real malingerers, which weakens the ecological validity of the experimental design.

Conclusion and Future Directions

The results suggest that the Turkish version of the IOP-29 is effective in discriminating between credible and non-credible presentations, that it is comparatively sensitive to a variety of feigned psychiatric disorders, and that the RRS has incremental validity in discriminating between random responses and feigned as well as honest responses. However, replication in psychometrically defined criterion groups based on genuine patients with a wider range of diagnoses (including neuropsychological disorders) is needed to establish its ecological validity. Incorporating the newly developed memory module (IOP–M; Giromini et al., 2020d; see also Erdodi et al., 2023) in future studies could further expand the clinical utility of the instrument among Turkish-speaking examinees (e.g., Brockhaus & Pecker, 2003), as there is ample evidence of the efficient utility of combining SVTs with performance validity tests (PVTs; see Holcomb et al., 2022).

As mentioned above, one of the main limitations of our study is the lack of a clinical sample and the use of a simulation design. Regarding the lack of a clinical sample, the use of nonclinical controls as opposed to real patients has the effect of likely overestimating the specificity results. The use of a simulation design may have the effect of maximizing internal validity, but external validity is questionable. In other words, there is no guarantee that subjects would actually respond similarly to our experimental participants in real assessments. Because of the impact of simulation designs on classification accuracy, it is of utmost importance for Turkish IOP research to switch to criterion group sampling (or known group sampling) rather than simulation designs in the future.

Overall, the combination of its brevity, ease of administration/scoring, and the stability of its classification accuracy across different countries and language communities makes the IOP–29 a cost-effective and empirically supported choice of SVTs.

Acknowledgements The authors would like to thank Luciano Giromini for his unconditional support and the use of the Inventory of Problems instrument. Irena Boskovic is thanked for her support in providing the survey platform. The authors would also like to thank Sinan Canan, Veysi Çeri, Ahmet Demirden, Enis Doko, and Ali Zafer Sağıroğlu for their interest in advertising the study through their social network.

Author Contributions *Conceptualization*: Ali Y. E. Akca, Mehmed S. Tepedelen, & Burcu Uysal; *Data curation*: Ali Y. E. Akca; *Formal analysis*: Ali Y. E. Akca; *Investigation*: Ali Y. E. Akca, Mehmed S. Tepedelen; *Methodology*: Ali Y. E. Akca, Mehmed S. Tepedelen, & Burcu Uysal; *Project administration*: Ali Y. E. Akca; *Resources*: Ali Y. E. Akca; *Supervision*: Ali Y. E. Akca & Burcu Uysal; *Validation*: Ali Y. E. Akca; *Writing – original draft*: Ali Y. E. Akca, Mehmed S. Tepedelen, Burcu Uysal, & Laszlo A. Erdodi; *Writing – review & editing*: Ali Y. E. Akca, Mehmed S. Tepedelen, Burcu Uysal, & Laszlo A. Erdodi.

Funding Open access funding provided by Università degli Studi di Torino within the CRUI-CARE Agreement. The authors did not receive support from any organization for the submitted work.

Availability of Data and Material Data and supplemental material are available on the *Open Science Framework* (OSF; https://osf.io/6xksf/).

Declarations

Ethics Approval This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of Ibn Haldun University (Istanbul, Türkiye; Date 28 Nov. 2022/No. 2022/07–7).

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

Conflict of Interest All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Abeare, C. A., Hurtubise, J., Cutler, L., Sirianni, C., Brantuo, M., Makhzoun, N., & Erdodi, L. (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, C. A., An, K., Tyson, B., Holcomb, M., Cutler, L., May, N., & Erdodi, L. A. (2022). The Emotion Word Fluency Test as an embedded validity indicator, alone and in multivariate verbal fluency validity composite. *Applied Neuropsychology: Child, 11*(4), 713–724. https://doi.org/10.1080/21622965.2021.1939027

Akca, A. Y. E., Çiller, A., Karataş, E., & Tepedelen, M. S. (2023). *Psikolojide Olumsuz Yanıt Yanlılığı, Temaruz ve Belirti Geçerliliği: Türkiye Odaklı Bir Gözden Geçirme [Negative Response Bias, Malingering and Symptom Validity in Psychology: A Türkiye-Focused Review]*. Manuscript submitted for publication.

Ali, S., Crisan, I., Abeare, C. A., & Erdodi, L. A. (2022). Cross-cultural performance validity testing: Managing false positives in examinees with limited English proficiency. *Developmental Neuropsychology, 47*(6), 273–294. https://doi.org/10.1080/87565641.2022.2105847

American Psychiatric Association. (2022). *Diagnostic and Statistical Manual of Mental Disorders* (5th, Text Revision ed.). American Psychiatric Association.

An, K. Y., Kaploun, K., Erdodi, L. A., & Abeare, C. A. (2017). Performance validity in undergraduate research participants: A comparison of failure rates across tests and cutoffs. *The Clinical Neuropsychologist, 31*(1), 193–206. https://doi.org/10.1080/13854046.2016.1217046

Ardıç, F. C., Köse, S., Solmaz, M., Kulacaoğlu, F., & Balcıoğlu, Y. H. (2019). Reliability, validity, and factorial structure of the Turkish version of the Structured Inventory of Malingered Symptomatology (Turkish SIMS). *Psychiatry and Clinical Psychopharmacology, 29*(2), 182–188. https://doi.org/10.1080/24750573.2019.1599237

Ayhan, H., & Karaman, H. (2021). Adli Psikolojik Değerlendirmenin Temel İlkeleri: Destekleyici Bir Unsur Olarak Psikolojik Testler [Basic Principles of Forensic Psychological Assessment: Psychological Tests as a Supporting Component]. *Uluslararası Türk Kültür Coğrafyasında Sosyal Bilimler Dergisi*, *6*(1), 36–47. https://dergipark.org.tr/en/pub/turksosbilder/issue/64427/937080

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

Bosi, J., Minassian, L., Ales, F., Akca, A. Y. E., Winters, C., Viglione, D. J., Zennaro, A., & 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*, 1–13. 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*, 1–13. https://doi.org/10.1080/23279095.2022.2057856

Brantuo, M., An, K., Biss, R., Ali, S., & Erdodi, L. A. (2022). Neurocognitive profiles associated with limited English proficiency in cognitively intact adults. *Archives of Clinical Neuropsychology, 37*(7), 1579–1600. https://doi.org/10.1093/arclin/acac019

Brislin, R. W. (1970). Back-Translation for Cross-Cultural Research. *Journal of Cross-Cultural Psychology, 1*(3), 185–216. https://doi.org/10.1177/135910457000100301

Brockhaus, R., & Peker, Ö. (2003, July 16–20). *Testing effort in Turkish-speaking subjects: Validation of a translation of the Word Memory Test (WMT)* Twenty-Sixth Annual International Neuropsychological Society Mid-Year Conference, Berlin, Germany. https://www.researchgate.net/publication/318101365_Testing_effort_in_Turkish-speaking_subjects_Validation_of_a_translation_of_the_Word_Memory_Test_WMT

Burchett, D., Dragon, W. R., Smith Holbert, A. M., Tarescavage, A. M., Mattson, C. A., Handel, R. W., & Ben-Porath, Y. S. (2016). "False feigners": Examining the impact of non-content-based invalid responding on the Minnesota Multiphasic Personality Inventory-2 Restructured Form content-based invalid responding indicators. *Psychological Assessment, 28*(5), 458–470. https://doi.org/10.1037/pas0000205

Bush, S. S., Heilbronner, R. L., & Ruff, R. M. (2014). Psychological assessment of symptom and performance validity, response bias, and malingering: Official position of the Association for Scientific Advancement in Psychological Injury and Law. *Psychological Injury and Law, 7*(3), 197–205. https://doi.org/10.1007/s12207-014-9198-7

Carvalho, L. D. F., 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

Chafetz, M., & Underhill, J. (2013). Estimated Costs of Malingered Disability. *Archives of Clinical Neuropsychology, 28*(7), 633–639. https://doi.org/10.1093/arclin/act038

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. Advance Online Publication*. https://doi.org/10.1037/pne0000316

Crişan, I., Sava, F. A., Maricuţoiu, L. P., Ciumăgeanu, M. D., Axinia, O., Gîrniceanu, L., & Ciotlăuş, L. (2021). Evaluation of various detection strategies in the assessment of noncredible memory performance: Results of two experimental studies. *Assessment, 29*(8), 1973–1984. https://doi.org/10.1177/10731911211040105

Dandachi-FitzGerald, B., Ponds, R. W., & Merten, T. (2013). Symptom validity and neuropsychological assessment: A survey of practices and beliefs of neuropsychologists in six European countries. *Archives of Clinical Neuropsychology, 28*(8), 771–783. https://doi.org/10.1093/arclin/act073

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. Advance online publication. *Journal of Forensic Psychology Research and Practice*.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Gegner, J., Erdodi, L. A., Giromini, L., Viglione, D. J., Bosi, J., & Brusadelli, E. (2021). An Australian study on feigned mTBI using the Inventory of Problems – 29 (IOP-29), its Memory Module (IOP-M), and the Rey Fifteen Item Test (FIT). *Applied*

Neuropsychology: Adult, 1–10. https://doi.org/10.1080/23279095.2020.1864375

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*(6), 504–516. https://doi.org/10.1080/23279095.2019.1570929

Giromini, L., Lettieri, S. C., Zizolfi, S., Zizolfi, D., Viglione, D. J., Brusadelli, E., Perfetti, B., di Carlo, D. A., & Zennaro, A. (2019). Beyond Rare-Symptoms Endorsement: A Clinical Comparison Simulation Study Using the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) with the Inventory of Problems-29 (IOP-29). *Psychological Injury and Law, 12*(3), 212–224. https://doi.org/10.1007/s12207-019-09357-7

Giromini, L., Pignolo, C., Young, G., Drogin, E. Y., Zennaro, A., & Viglione, D. J. (2021). Comparability and Validity of the Online and In-Person Administrations of the Inventory of Problems-29. *Psychological Injury and Law, 14*(2), 77–88. https://doi.org/10.1007/s12207-021-09406-0

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*. https://doi.org/10.1007/s12207-021-09437-7

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. (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*(3), 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: New Articles. *New Issues. Psychological Injury and Law, 15*(1), 1–21. https://doi.org/10.1007/s12207-022-09444-2

Gu, W., Reddy, H. B., Green, D., Belfi, B., & Einzig, S. (2017). Inconsistent responding in a criminal forensic setting: An evaluation of the VRIN-r and TRIN-r scales of the MMPI–2–RF. *Journal of Personality Assessment, 99*(3), 286–296. https://doi.org/10.1080/00223891.2016.1149483

Hathaway, S. R., & McKinley, J. C. (1943). *The Minnesota Multiphasic Personality Inventory* (Rev. ed.). University of Minnesota Press.

Holcomb, M., Pyne, S., Cutler, L., Oikle, D. A., & Erdodi, L. A. (2022). Take Their Word for It: The Inventory of Problems Provides Valuable Information on Both Symptom and Performance Validity. *Journal of Personality Assessment*, 1–11. 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

Keyvan, A., Ger, M. C., Ertürk, S. G., & Türkcan, A. (2015). The Validity and Reliability of the Turkish Version of Miller Forensic Assessment of Symptoms Test (M-FAST). *Noro Psikiyatr Ars, 52*(3), 296–302. https://doi.org/10.5152/npa.2015.7587

Knoll, J., & Resnick, P. J. (2006). The detection of malingered post-traumatic stress disorder. *Psychiatric Clinics of North America, 29*(3), 629–647. https://doi.org/10.1016/j.psc.2006.04.001

Larrabee, G. J., Millis, S. R., & Meyers, J. E. (2009). 40 Plus or Minus 10, a New Magical Number: Reply to Russell. *The Clinical Neuropsychologist, 23*(5), 841–849. https://doi.org/10.1080/13854040902796735

Lichtenstein, J. D., Flaro, L., Baldwin, F., Rai, J. K., & Erdodi, L. A. (2019). Further evidence for embedded validity tests in children within the Conners' Continuous Performance Test – Second Edition. *Developmental Neuropsychology, 44*(2), 159–171. https://doi.org/10.1080/87565641.2019.1565536

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437–455. https://doi.org/10.1037/a0028085

Merckelbach, H., Dandachi-FitzGerald, B., van Helvoort, D., Jelicic, M., & Otgaar, H. (2019). When Patients Overreport Symptoms: More Than Just Malingering. *Current Directions in Psychological Science, 28*(3), 321–326. https://doi.org/10.1177/0963721419837681

Merten, T., & Merckelbach, H. (2020). Factitious disorder and malingering. In J. R. Geddes, N. C. Andreasen, & G. M. Goodwin (Eds.), *New Oxford Textbook of Psychiatry* (Third ed., pp. 1342–1349). Oxford University Press.

Merten, T., Merckelbach, H., Giger, P., & Stevens, A. (2016). The Self-Report Symptom Inventory (SRSI): A New Instrument for the Assessment of Distorted Symptom Endorsement. *Psychological Injury and Law, 9*(2), 102–111. https://doi.org/10.1007/s12207-016-9257-3

Miller, H. A. (2001). *M-FAST: Miller Forensic Assessment of Symptoms Test professional manual*. Psychological Assessment Resources.

Ministry of Culture and Tourism of The Republic of Türkiye. (2023). *Language*. Retrieved January 25th from https://www.ktb.gov.tr/EN-117846/language.html

Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology, 24*(8), 1094–1102. https://doi.org/10.1076/jcen.24.8.1094.8379

Morel, K. R., & Marshman, K. C. (2008). Critiquing symptom validity tests for posttraumatic stress disorder: A modification of Hartman's criteria. *Journal of Anxiety Disorders, 22*(8), 1542–1550. https://doi.org/10.1016/j.janxdis.2008.03.008

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

Nichols, D. S., Greene, R. L., & Schmolck, P. (1989). Criteria for assessing inconsistent patterns of item endorsement on the MMPI: Rationale, development, and empirical trials. *Journal of Clinical Psychology*, *45*(2), 239–250. https://doi.org/10.1002/1097-4679(198903)45:2<239::AID-JCLP2270450210>3.0.CO;2-1

Nijdam-Jones, A., & Rosenfeld, B. (2017). Cross-cultural feigning assessment: A systematic review of feigning instruments used with linguistically, ethnically, and culturally diverse samples. *Psychological Assessment, 29*, 1321–1336. https://doi.org/10.1037/pas0000438

Pignolo, C., Giromini, L., Ales, F., & Zennaro, A. (2021). Detection of Feigning of Different Symptom Presentations With the PAI and IOP-29 [Epub ahead of print]. *Assessment*. 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, Advance Online Publication. https://doi.org/10.1037/pas0001209

Rai, J. K., An, K. Y., Charles, J., Ali, S., & Erdodi, L. A. (2019). Introducing a forced choice recognition trial to the Rey Complex Figure Test. *Psychology & Neuroscience, 12*(4), 451–472. https://doi.org/10.1037/pne0000175

Rienstra, A., Groot, P. F. C., Spaan, P. E. J., Majoie, C. B. L. M., Nederveen, A. J., Walstra, G. J. M., de Jonghe, J. F. M., van Gool, W. A., Olabarriaga, S. D., Korkhov, V. V., & Schmand, B. (2013). Symptom validity testing in memory clinics: Hippocampal-memory associations and relevance for diagnosing mild cognitive impairment. *Journal of Clinical and Experimental Neuropsychology, 35*(1), 59–70. https://doi.org/10.1080/13803395.2012.751361

Rogers, R., & Bender, S. D. (Eds.). (2018). *Clinical assessment of malingering and deception* (4 ed.). The Guilford Press.

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*(1), 57–65. https://doi.org/10.1007/s12207-019-09368-4

Savaşır, I. (1978). Minnesota Çok Yönlü Kişilik Envanterinin Türkçeye uyarlanışı ve standardizasyon projesi. *Türk Psikoloji Dergisi, 1*(1), 18–24.

Schutte, C., Axelrod, B. N., & Montoya, E. (2015). Making sure neuropsychological data are meaningful: Use of performance validity testing in medicolegal and clinical contexts. *Psychological Injury and Law, 8*(2), 100–105.

Sellbom, M., Toomey, J. A., Wygant, D. B., Kucharski, L. T., & Duncan, S. (2010). Utility of the MMPI–2–RF (Restructured Form) validity scales in detecting malingering in a criminal forensic setting: A known-groups design. *Psychological Assessment, 22*, 22–31. https://doi.org/10.1037/a0018222

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/acaa019

Shura, R. D., Ord, A. S., & Worthen, M. D. (2022). Structured inventory of malingered symptomatology: A psychometric review. *Psychological Injury and Law, 15*, 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). *The Journal of the American Academy of Psychiatry and the Law, 25*(2), 183–189.

Soliman, S., & Resnick, P. J. (2010). Feigning in adjudicative competence evaluations. *Behavioral Sciences and the Law, 28*(5), 614–629. https://doi.org/10.1002/bsl.950

Šömen, M. M., Lesjak, S., Majaron, T., Lavopa, L., Giromini, L., Viglione, D., & 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

Streiner, D. L. (2003). Diagnosing Tests: Using and Misusing Diagnostic and Screening Tests. *Journal of Personality Assessment, 81*(3), 209–219. https://doi.org/10.1207/S15327752JPA8103_03

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. *Clinical Neuropsychologist, 35*(6), 1053–1106. https://doi.org/10.1080/13854046.2021.1896036

Towse, J. N., & Neil, D. (1998). Analyzing human random generation behavior: A review of methods used and a computer program for describing performance. *Behavior Research Methods, Instruments & Computers, 30*, 583–591.

United Nations. (2022, June 3). *The country name "Türkiye" is replacing "Turkey" at the UN*. United Nations. Retrieved February 9, 2023 from https://turkiye.un.org/en/184798-turkeys-name-changed-turkiye

van de Vijver, F., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist, 1*, 89–99. https://doi.org/10.1027/1016-9040.1.2.89

van der Heide, D., Boskovic, I., van Harten, P., & Merckelbach, H. (2020). Overlooking Feigning Behavior May Result in Potential Harmful Treatment Interventions: Two Case Reports of Undetected Malingering. *Journal of Forensic Sciences, 65*(4), 1371–1375. https://doi.org/10.1111/1556-4029.14320

Van Egmond, J., Kummeling, I., Balkom, T., & a. (2005). Secondary gain as hidden motive for getting psychiatric treatment. *European Psychiatry, 20*(5–6), 416–421. https://doi.org/10.1016/j.eurpsy.2004.11.012

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

Viglione, D. J., Giromini, L., Landis, P., McCullaugh, J. M., Pizitz, T. D., O'Brien, S., Wood, S., Connell, K., & Abramsky, A. (2019). Development and Validation of the False Disorder Score: The Focal Scale of the Inventory of Problems. *Journal of Personality Assessment, 101*(6), 653–661. https://doi.org/10.1080/00223891.2018.1492413

Wagenaar, W. A. (1972). Generation of random sequences by human subjects: A critical survey of literature. *Psychological Bulletin, 77*(1), 65–72.

Winters, C. L., Giromini, L., Crawford, T. J., Ales, F., Viglione, D. J., & Warmelink, L. (2020). An Inventory of Problems-29 (IOP-29) study investigating feigned schizophrenia and random responding in a British community sample. *Psychiatry, Psychology and Law, 28*(2), 235–254. https://doi.org/10.1080/13218719.2020.1767720

Wise, E. A. (2009). Selected MMPI–2 scores of forensic offenders in a community setting. *Journal of Forensic Psychology Practice, 9*, 299–309. https://doi.org/10.1080/15228930902936048

Wygant, D. B., Sellbom, M., Gervais, R. O., Ben-Porath, Y. S., Stafford, K. P., Freeman, D. B., & Heilbronner, R. L. (2010). Further validation ofthe MMPI–2 and MMPI–2–RF Response Bias scale: Findings from dis-ability and criminal forensic samples. *Psychological Assessment, 22*, 745–756. https://doi.org/10.1037/a0020042

Young, G. (2015). Malingering in Forensic Disability-Related Assessments: Prevalence 15 ± 15 %. *Psychological Injury and Law, 8*(3), 188–199. https://doi.org/10.1007/s12207-015-9232-4

Young, G., Foote, W. E., Kerig, P. K., Mailis, A., Brovko, J., Kohutis, E. A., McCall, S., Hapidou, E. G., Fokas, K. F., & Goodman-Delahunty, J. (2020). Introducing Psychological Injury and Law. *Psychological Injury and Law, 13*(4), 452–463. https://doi.org/10.1007/s12207-020-09396-5

Ziegler, M. (2015). "F*** You, I Won't Do What You Told Me!" – Response biases as threats to psychological assessment. *European Journal of Psychological Assessment, 31*(3), 153–158. https://doi.org/10.1027/1015-5759/a000292

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