

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Development and Validation of the False Disorder Score: The Focal Scale of the Inventory of Problems

This is the author's manuscript

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1686431> since 2019-01-10T11:31:38Z

Published version:

DOI:10.1080/00223891.2018.1492413

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)



Development and Validation of the False Disorder Score: The Focal Scale of the Inventory of Problems

Journal:	<i>Journal of Personality Assessment</i>
Manuscript ID	JPA-2017-320.R1
Manuscript Type:	General Submission
Keywords:	Inventory of Problems, IOP, feigning, Malingering < Content or Topic, Test Development < Content or Topic, validity

SCHOLARONE™
Manuscripts

Abstract

This article introduces the Inventory of Problems (IOP) – a new, computerized, 181-item tool designed to discriminate bona-fide from feigned mental illness and cognitive impairment – and presents the development and validation of its focal, feigning scale, the False Disorder Score (IOP–FDS). The initial sample included (a) 211 patients and 64 offenders who took the IOP under standard conditions and (b) 210 community volunteers and 64 offenders who feigned mental illness. We split this sample into three subsamples. The first (n = 301) was used to select the variables to generate the IOP–FDS; the second (n = 148) scaled the IOP–FDS into a probability score; and the third (n = 100) tested its validity with an independent dataset. In this third subsample, the IOP–FDS had sensitivity = .90, specificity = .80, and a greater AUC (= .95) than the IOP-29 (= .91). For 40 participants, the PAI was available too. Within this subgroup, the IOP–FDS outperformed the selected PAI validity scales (AUC = .99 vs. AUC ≤ .85).

Keywords: Inventory of Problems, IOP, feigning, malingering, test development, validity.

Development and Validation of the False Disorder Score:

The Focal Scale of the Inventory of Problems

The Inventory of Problems (IOP) is a new, multipurpose, 181-item, computerized test designed to investigate feigning of various psychiatric and cognitive complaints. The aim of the current paper is to report on the development and validation of its focal scale, the False Disorder Score (IOP–FDS) – a measure of the likelihood that the test-taker is presenting a false mental health or cognitive complaint. To accomplish this aim, this article presents three studies. The first study selects from the IOP the items and latencies (i.e., the times to answer the items, or reaction times), to be included in the IOP–FDS. In the second study these data are scaled to create the IOP–FDS. Finally, the IOP–FDS is cross-validated with an independent subsample. Because one might question whether a single measure could detect a wide range of clinical presentations, we also highlight the strategies and techniques we adopted to enable the IOP–FDS to identify a broad array of problems and symptom combinations including neuropsychological impairments, psychosis and schizophrenia, post-traumatic stress disorder (PTSD), and depression.

Background for the Inventory of Problems (IOP)

A leading principle in developing the first version of the IOP (Viglionone & Landis, 1994), which was mentioned by Rogers (1998) in his second edition of his book on malingering and deception, was that an omnibus feigning test should incorporate multiple detection strategies. Based on the literature at the time (Rogers, 1988) and experience with feigning in the military and practice (see for example, Viglionone, Fals-Stewart, & Moxham, 1995) 27 different strategies and 245 corresponding items were formulated for the first version. The reviews of the literature

INTRODUCING THE IOP-FDS

and experience in practice revealed that scales with symptom oriented, keyed true, answered true items were highly redundant with one another and that they dominated most of the feigning measures available at the time. Thus, our item-development strategy focused on potential incremental validity over such items and our pilot research focused on examining whether or not they (a) differentiated bona fide patients and honest probationers from feigners, and (b) whether they added incremental validity beyond keyed true answered true symptom items. This initial research led to many item and detection strategy revisions and refinements and to the development of a second IOP version, comprised of 162 items. Then, based on the research and experience accumulated with the first and second version of the IOP, we again pruned, revised, and added items to create the third and current version of the IOP, which includes 181 items.

Throughout this long developmental period, several doctoral dissertations and other research projects were conducted specifically to test our detection strategies and to refine the IOP item pool. In fact, prior to conducting the three studies described below, about 1,000 participants had already been tested with one of the three developmental versions of the IOP. These research efforts focused on the contrast between patients and feigners, with little concern for discriminating patients from honest non-patients since creating such a scale is easily done but largely irrelevant to the goal of helping examiners in the field to opine whether a given psychiatric presentation is bona fide or feigned. Moreover, to minimize extraneous variance and confounds, in most of these studies, we matched or minimized demographic differences and used community rather than college samples. In some cases, feigners were instructed to fake the symptoms or history of the patient with whom he or she was matched. In the great majority of the cases, feigners were instructed “not to overdo it” (Viglione et al., 2001), so that they would

INTRODUCING THE IOP-FDS

4

more closely resemble successful malingerers in real life situations and to diminish artificially large effect sizes (Rogers & Bender, 2013; Rogers & Gillard, 2011).

The final, 181-item version of the IOP, is administered electronically. In addition to the classic, self-report or symptom validity items (e.g., “I feel terribly sad every day.” [Keyed *True*] “Sometimes, others help me to feel OK.” [Keyed *False*]),¹ the IOP also includes performance validity items in the form of easy cognitive problems, including calculation (e.g., “150 – 50 = ?”), memory (e.g., the test-taker is asked to recall content or simple pictures that were introduced earlier in the test), reasoning (e.g., “A tree is to a forest, as a lightning is to a thunderstorm.”), and other pattern recognition and cognitive items. Additionally, a few Likert-scale symptom ratings, for example, “Rate your problems with depression on a 1 to 7 scale where 1 = No problem7 = Unbearable,” are included. Some of the items were inclusively worded affirmations, such as “It’s killing me,” as we thought that such pronouns and vague language (for example, “it” or “my problem”) might capture a wide variety of false complaints. Of course, only those strategies and item-formats that were supported in the pilot research with the first two IOP versions were retained in the final IOP.

In addition to the prototypical “*True*” versus “*False*” response options, most IOP self-report items offer a third response option, “*Doesn’t Make Sense*.” This is because in our developmental research leading up to the final version of the test, simulators more readily endorsed oddly worded items with pathological content. That is, rather than reading carefully, simulators perhaps ignored odd wording because they were scanning for pathological content to decide whether or not to endorse it. Conversely, patients and control participants did not always understand these items, and thus tended to choose the response option *Doesn’t Make Sense* more

¹ Test items quoted demonstrate the principles described and closely resemble items in the test the structure but are not themselves necessarily included in the test.

INTRODUCING THE IOP-FDS

frequently. With clearly written items, in contrast, endorsing the *Doesn't Make Sense* response option seemed to reflect an attempt to feign cognitive impairment or express an uncooperative intent, behavior that is sometimes adopted by real life malingerers. Indeed, for these items, simulators tended to choose *Doesn't Make Sense* more frequently than did bona fide patients.

Another key feature of the IOP is that some of its items address test-taking behaviors and experiences. Indeed, we wrote items to capture the behavioral dramatization of symptoms seen in clinical interviews (e.g., “My hands are shaking uncontrollably during this test.” [Keyed *True*] “This test does not cover enough of my problems.” [Keyed *True*]). Likewise, we hypothesized that item content which incorporated (a) externalization of responsibility for one’s woes, symptoms, and predicament while minimizing one’s ability to improve them (e.g., “Sometimes, I can think about things that make me feel better.” [Keyed *False*]) and (b) refusal to admit qualified positive attributes (e.g., “I feel attractive sometimes.” [Keyed *False*]) could be particularly effective. Pilot research with the IOP versions one and two confirmed these expectations.

A final distinctive feature of the IOP presented here is that its computerized administration allows recording of item/response latencies, i.e., the reaction time between presentation of an item and the response via computer key. Using this information, we hypothesized that the interaction between the specific content of an item and the response latency to that item also could contribute to discriminating between bona fide and feigned mental problems. Thus, we evaluated latencies and interactions between latencies and the *True*, *False*, and *Doesn't Make Sense* response alternatives.

The Current Studies on the IOP–FDS

INTRODUCING THE IOP-FDS

6

A brief derivative of the IOP, comprised of 29 of the 181 items of the IOP, and named “IOP-29,” was recently introduced in the *Journal of Personality Assessment* (Viglione, Giromini, & Landis, 2017). The current article, in contrast to this previous publication, is the first to report on the full length, 181-item version of the IOP. In the present paper, we describe the development and validation of the 181-item IOP’s focal feigning scale, the IOP-FDS. The validity of this new scale as a measure of feigning is then compared with that of the IOP-29 and, for a small subset of data, with that of the PAI.

Materials and Methods

Sampling Procedures and Participants

The consolidated sample utilized to develop and validate the IOP–FDS is the same one used to scale and cross-validate the IOP-29 (Viglione et al., 2017): It combines data from six dissertation studies (Abramsky, 2005; Connell, 2004; McCullaugh, 2011; O’Brien, 2004; Pizitz, 2001; Woods, 2008).² More specifically, it contains 275 patients or offenders on probation taking the third version of the IOP with standard instructions contrasted to 274 volunteers or offenders on probation taking the IOP with the instruction to feign a psychiatric and or cognitive disorders.

Honest respondents were 38 patients suffering neuropsychological deficits (Pizitz, 2001), 89 patients affected by schizophrenia or psychosis (O’Brien, 2004; Woods, 2008), 40 volunteers suffering from PTSD (Connell, 2004), 44 patients with depression (Abramsky, 2005), and 64 adult offenders on probation being treated for mental health or substance abuse (McCullaugh, 2011). Simulators were 211 non-clinical adult volunteers instructed to either feign (a) neuropsychological deficits ($n = 37$; Pizitz, 2001), (b) schizophrenia or psychosis ($n = 90$; O’Brien, 2004; Wood, 2008), (c) PTSD ($n = 39$; Connell, 2004), or (d) depression ($n = 44$;

² Item selection for the IOP-29 was done with previously collected samples.

Abramsky, 2005), as well as 64 adult offenders on probation instructed to feign a mixture of neuropsychological, depressive, and PTSD symptoms (McCullaugh, 2011). Thus, the data are relevant to the four target diagnostic categories of the IOP – cognitive/neuropsychological impairment, psychosis, PTSD, and depression.

Details on procedures and demographic composition of the sample and sample sources are described in Viglione et al. (2017). Briefly, all participants were adult volunteers, and heterogeneous regarding gender, age, racial characterization, education, and marital status. All had signed an informed consent form prior to being enrolled in the study. To maximize external validity, all simulators were provided with a brief scenario or vignette aimed at improving their feigning abilities, and were instructed not to produce excessively dramatic or severe symptom presentations, or else their performances would easily be detected as fake or feigned (Rogers & Bender, 2013; Rogers & Gillard, 2011, Viglione et al., 2001). Within each of the data sources under consideration, honest respondents (patients and offenders on probation) and simulators did not differ from each other on any important, demographic variables (Viglione et al., 2017).

Approach to Scale Construction & Validation

To apply these data sources to the development and validation of the IOP–FDS, we randomly split this combined sample ($N = 549$) into three³ subsamples: Item Selection Subsample 1 ($n = 301$); Scaling Subsample 2 ($n = 148$); and Cross-validation Subsample 3 ($n = 100$). In parsing the composite sample, we allotted more participants to statistical procedures with less power. Specifically, the Item Selection Subsample 1 required the largest number of participants and the most power as we worked with both the answers and the response latencies for the individual 181 items of the IOP. The Scaling Subsample 2 was used to derive the

³ The reader should endeavor not to confuse the *three versions* of the IOP with the *three subsamples* or research studies presented in this paper. The current paper presents three subsamples with the third and final version of the IOP.

INTRODUCING THE IOP-FDS

8

equations for combining these two indicators (answers and latencies) into our final IOP–FDS and therefore, required fewer participants. The Cross-validation Subsample 3 was the smallest because it was used to cross-validate only one variable (i.e. the final IOP–FDS). Although the assignment of IOP’s to the three subsamples was done randomly, this randomization was stratified by dissertation and equally between honest patient versus simulator. As a result, each subsample contained approximately the same proportion of individuals from each dissertation. (See Table 1).

Study 1 – Item Selection**Research Questions and Analysis Plan**

Study 1 aimed at developing two independent subcomponents, or indicators, to be included in the IOP–FDS. More specifically, we intended to develop a feigning indicator based on the responses given to the IOP, and a feigning indicator using the latencies or reaction times for each item. We thus planned two sets of analyses. First, we tested the Phi association of each item/response combination (e.g., *True* to item 1; *False* to item 1; etc.) to group membership (dummy code, with 0 = honest respondent; 1 = experimental simulator), to select the items that would demonstrate the strongest associations. These items would then be combined to generate our first feigning indicator. Next, we examined all individual patterns of response latencies (e.g., reaction time to endorse *True* on item 1; reaction time to endorse *False* on item 1; etc.) and their point bi-serial correlations to group membership, to select the latency patterns that would demonstrate the strongest associations. These would then be used to generate our second feigning indicator.

The first indicator, named “Item-Based Feigning” indicator, would thus be a multi-method scale derived from self-report items, after inspecting all keyed *True* (T) and *False* (F)

items, all cognitive items (Cog), all *Doesn't Make Sense* (D) responses⁴, and all the a-priori determined, item pairs (Pairs), i.e., contradictory responses for two related items, (e.g. answering *True* to “My worst time is when I get up” yet *False* to “My best time is early in the day.”) The second feigning indicator scale, “Latency Feigning,” would instead be derived from response latencies of individual items.

Development of Item-Based Feigning Indicator. Two principles guided the development of the Item-Based Feigning and Latency Feigning indicators. First, we wanted to select (i.e., assign points to) item/response combinations that would perform similarly well from one data source, or diagnostic target, to another. Second, to be selected or given an extra point, an item/response combination should provide incremental validity over the other, already available or selected, item/response combinations in the indicator. According to the first of these principles, we selected the IOP-29 item/response combinations as the core of the IOP-FDS, and assigned them the same item/response weights established for the IOP-29 main scale. These item/response combinations, indeed, have already demonstrated to be applicable to multiple symptom presentations in various data sets both in the U.S. (Viglione et al., 2017) and in Italy (Giromini et al., 2018), with virtually no shrinkage from one sample to another. As such, they represent an already established, fully validated measure of credibility of a wide range of different symptom presentations. Next, we inspected all item/response combinations from all 181 IOP items, and gave one point (or extra point) to those item/response combinations that correlated (*Phi* correlations) with group membership (0 = honest respondent; 1 = simulator) with an effect size of at least $r = .30$ (i.e., *medium*; Cohen, 1992) in each and every one of the five data sources, i.e., head injury (Pizitz, 2001), psychotic (O’Brien, 2004; Woods, 2008), PTSD

⁴ There are six items in which the “D” response alternative corresponds to some version of “Do not have that problem.” Three address denials of hearing voices, one denial of depression, and one denying of being punished physically as a child.

INTRODUCING THE IOP-FDS

10

(Connell, 2004), depressed (Abramsky, 2005), and offenders on probation (McCullaugh, 2011).

This procedure terminated with the creation of our first, “in-progress,” Item-Based Feigning Indicator.

According to our second guiding principle, we then used an iterative procedure, or algorithm, to further refine this in-progress index, and assign extra points to any item/response combinations that would demonstrate incremental validity. More specifically, this in-progress index was then further improved step-by-step, or item-by-item, by using partial correlations, with the final version being the Item-Based Feigning Indicator. That is, at each step, the item/response combination with the highest partial correlation with group membership, after partialing out the in-progress index, was examined. If its average partial correlation across the data sources was (condition one) greater than .10 (*small* effect; Cohen, 1992), and (condition two) all data sources correlated (partial correlations) in the same direction, this item/response combination was added to the index and then applied to the next partial correlation. This repetitive procedure terminated when no more partial correlations met both these conditions.

Development of Latency Feigning Indicator. After developing our Item-Based Feigning Indicator, we next focused on patterns of response latencies. The purpose was to investigate whether, compared to the examinee’s average latency to all the IOP items, taking a longer or shorter time to answer any given individual test item would be related to feigning of mental and/or cognitive disorders. For example, we speculated that when compared to honest patient and offender respondents, simulators might be faster to answer *True* to some keyed *True* items, in that honest respondents might be more resistant than simulators to admit or reveal that they suffer from certain psychopathological problems (e.g., sexual problems).

INTRODUCING THE IOP-FDS

11

To limit the potential impact of outliers (i.e., extremely long item latencies), we truncated the maximum latency interval for each IOP item to the 95th percentile of the combined, patient/forensic group. Next, because latencies are typically positively skewed and non-normal, we applied a square root transformation to all these trimmed, latency values. These values were then standardized, based on z-scores from the combined, patient and offender on probation group, to place all of these latency variables on the same unit of measurement. Lastly, via multiple regression analyses, each item’s latency was predicted based on the particular individual’s average latency for all other IOP items, and the residuals between predicted and observed latencies were used as our target latency parameter. This latency could be understood as the particular item’s characteristic after controlling for the average reaction time for the entire IOP.

Because we anticipated that latencies would produce smaller effect sizes in predicting feigning than would the IOP items, which were specifically designed to achieve that goal, we reduced the cut-off or threshold for item selection to $r = .10$ (*small* effect size; Cohen, 1992). Thus, based on our two guiding principles described above, reaction-time based variables were included in our final Latency Feigning Indicator if they (a) produced a correlation with group membership of at least $|r| = .10$, or (b) produced an average partial correlation with group membership of at least $r = .10$, after partialing out the Item-Based Feigning. If both conditions (a) and (b) were satisfied, the variable was double-weighted. The Latency Feigning Indicator was then calculated as the arithmetic mean of all selected individual item latencies.

Results

Based on item selection and item weight procedures detailed above, the Item-Based Feigning Indicator ultimately included 60 items and 33 item pairs, for a total of 79 different IOP

INTRODUCING THE IOP-FDS

12

items (some of the individual test items included in the indicator were also part of one or more item pairs). The average *Phi* correlation of these 79 items to group membership was .35 (*SD* = .13). Importantly, many of these items received multiple weights. Indeed, the Item-Based Feigning Indicator can potentially range from 0 to 143 points. More in detail, 75 of these points came from IOP-29 item/response combinations and/or from items with *Phi* correlations $\geq .30$ in each of the five data sources mentioned above; 68 came from the iterative partial correlation procedures described in the above sections. Honest responders had an overall mean of 61.6 points (*min* = 55; *max* = 72; *SD* = 3.7), whereas simulators had an overall mean of 76.7 points (*min* = 60; *max* = 88; *SD* = 5.6); this difference is statistically significant, with a very large effect size, $t(259.6) = 27.7, p < .001, d = 3.2$. Point bi-serial correlation of Item-Based Feigning Indicator to group membership was .848, $p < .001$. *AUC* was .971 (*SE* = .010).

The final Latency Feigning Indicator included 63 items, four of which received double weights. Of these 63 items, 25 were included also in the Item-Based Feigning Indicator, whereas 38 were not, so that to calculate both the indicators (i.e., the Item-Based and Latency Feigning Indicators) a total of 117 items (= 79 + 38) would be needed. The average point bi-serial correlation of these 63 reaction time-related variables to group membership was .15 (*SD* = .04). Honest responders had a mean Latency Feigning Indicator value of -.11 (*min* = -.48; *max* = .26; *SD* = .19), simulators had mean value of .12 (*min* = -.47; *max* = .61; *SD* = .14). The difference between the two groups is statistically significant, with a large effect size, $t(278.9) = 12.2, p < .001, d = 1.4$. The point bi-serial correlation of the Latency Feigning Indicator to group membership was $r = .576, p < .001$. Perhaps more importantly, partial correlation with group membership after removing the effects of the Item-Based Feigning Indicator was .209, $p < .001$, which is quite impressive, given the high association between Item-Based Feigning and group

membership. As desired, the Latency Feigning Indicator incremented over Item-Based Feigning Indicator.

Study 2 – Scaling

Research Questions and Analysis Plan

The goal of Study 2 was to test whether the Item-Based Feigning and Latency Feigning indicators would continue to predict group membership when inspecting a new, independent sample, i.e., the Scaling Subsample 2 ($n = 148$). Perhaps more importantly, Study 2 also aimed at testing whether, also within this independent sample, the Latency Feigning Indicator would continue to contribute to predict group membership after controlling for the effects of the Item-Based Feigning Indicator. Indeed, the ultimate goal of Study 3 was to combine the two aforementioned indicators to optimize prediction. As such, we planned to enter our two newly-developed feigning indicators into a logistic regression with group membership as the outcome variable. The regression equation derived from this analysis would then be used to generate the formula to calculate the IOP – FDS score, which would thus be expressed as a probability score ranging between one (certainly a simulator) and zero (certainly not a simulator).

Results

The logistic regression was statistically significant, $\chi^2(2) = 79.73, p < .001$. Nagelkerke R^2 was .555. Both the Item-Based Feigning and Latency Feigning predictors significantly and uniquely contributed to the model. Item-Based Feigning produced an $exp(B)$ greater than 10,000, $p < .001$; the Latency Feigning produced an $exp(B)$ of 63.78, $p = .015$. The IOP–FDS was thus calculated using the regression equation formula obtained from this model. Thus, its score varies from 0.00 to 1.00, reflecting the likelihood of drawing a particular IOP-FDS score from a group of experimental feigners versus the group with bona fide disorders, when the base rate of

feigners is 50%. For example, an IOP-FDS probability score of .75 would suggest a 3 to 1 odds that that IOP-FDS came from a person attempting to feign a mental or cognitive disorder, rather than from a bona fide patient.

Study 3 – Cross-validation

Research Questions and Analysis Plan

The aim of Study 3 was to test the validity (as a measure of feigning) of the IOP-FDS with another, independent subsample ($n = 100$). We thus inspected receiver-operator characteristic curves (AUC 's), diagnostic efficiency statistics, as well as point-biserial correlations with group membership, and Cohen's d effect sizes. Because the IOP-FDS score was designed to be superior to the IOP-29 and to compete with available feigning tools, we also compared the validity of the IOP-FDS to the IOP-29 and PAI. The PAI was selected because a limited number of PAI data were available in Subsample 3.

Results

The IOP-FDS was tested with the independent, Cross-validation Subsample 3, which included 50 honest respondents and 50 simulators. The point bi-serial correlation of IOP-FDS to group membership was $.797, p < .001$; AUC was $.950 (SE = .020)$. The difference between the IOP-FDS of patients vs. simulators corresponded to a Cohen's d of 2.61. Figure 1 shows the distribution of IOP-FDS among honest respondents and simulators.

Diagnostic efficiency results can be found in Table 2. Because positive predictive power (PPP), negative predictive power (NPP), and overall correct classification (OCC) largely depend on the base rates of the conditions under consideration (Meehl & Rosen, 1955), in addition to the .50 base rate of this subsample, we also inspected other common base rates, i.e., .25, .10, and .05, adjusting PPP, NPP, and OCC formulas accordingly (Streiner, 2003). Furthermore, we also

INTRODUCING THE IOP-FDS

inspected various IOP–FDS cut-offs, as they obviously affect diagnostic efficiency statistics, too. When using a cut-off of .50, which matches the base rate of all three subsamples, and the optimal base rate for the use of diagnostic tests, sensitivity is .90, specificity is .80 and the OCC, assuming a base rate of .50, is .87.

Validity for Individual Diagnostic Categories. Because the IOP–FDS was designed to detect feigning across various mental and/or cognitive disorders, we next investigated its validity within each of the five data sources under investigation. As seen in Table 3, the *AUC* values ranged from .88 to 1.00, with a mean of .93 (*SD* = .06). Sensitivity ranged from .86 to 1.00 (*M* = .90, *SD* = .06), and specificity ranged from .57 to 1.00 (*M* = .78, *SD* = .16). Although the size of the individual samples within Subsample 3 was rather small, the IOP–FDS worked fairly well with all data sources. The best performance was observed within the forensic (McCullaugh, 2011) and depression (Abramsky, 2005) data sources; whereas the worst one was within the head injury subgroup (Pizitz, 2001). It is difficult to say whether these differences are associated with substantive, replicable differences connected to type of disorder or are due to variability with the small subsample sizes, which ranged from 14 to 32.

Comparative Validity. For the entire Cross-validation Subsample 3, we examined the IOP-29 score, to inspect whether the IOP–FDS incremented over the IOP-29. Whereas the IOP–FDS produced a point bi-serial correlation with group membership of .797, *p* < .001, the IOP-29 index produced a point bi-serial correlation of .687, *p* < .001. We tested whether the difference between these two correlations was statistically significant using procedures described by Meng, Rosenthal, and Rubin (1992), and found that indeed it was statistically significant, *z* = 3.518, *p* < .001. *AUC* was .950 (*SE* = .020) for the IOP–FDS, and .901 (*SE* = .032) for the IOP-29.

INTRODUCING THE IOP-FDS

16

Although both IOP–FDS and IOP-29 produced very promising results, the IOP–FDS performed significantly better than the IOP-29.

In addition to the IOP, 40 individuals in the Cross-validation Subsample 3 had also previously been administered the Personality Assessment Inventory (PAI; Morey, 1991, 1996, 2007). Sixteen were from the psychosis sample (Woods, 2008) and 24 were from forensic settings (McCullaugh, 2011), with half being feigners. Consistent with previous research (for example, Archer et al., 2006; Blanchard et al., 2003; Edens, Poythress, & Watkins-Clay, 2007; Hawes & Bocaccini, 2009), we selected the Personality Assessment Inventory (PAI) Negative Impression Management (NIM; Morey, 1996), Malingering Index (MAL; Morey, 1996), and Rogers Discriminant Function (RDF; Rogers, Sewell, Morey, & Ustad, 1996) for PAI validity comparisons. The results of these analyses, reported in Table 4, indicate that within this small sample, the IOP–FDS outperformed all three of the selected PAI scales. In fact, using Meng et al. (1992) procedures, we found that the point bi-serial correlation value produced by the IOP–FDS was significantly greater than that produced by the PAI – NIM, $z = 4.835, p < .001$, PAI – MAL, $z = 5.880, p < .001$, and PAI – RDF, $z = 4.941, p < .001$. The mean *AUC* for the three PAI scales was .79, whereas this value was .99, nearly perfect, for the IOP – FDS.

Discussion

This paper presents the development and initial validation of the Inventory of Problems – False Disorder Scale (IOP–FDS), the focal scale of the Inventory of Problems (IOP). After a development period involving approximately 1,000 participants and two previous versions, and the creation of a brief, derivative version (the IOP-29; Viglione et al. 2017), we examined the third and final IOP version. The current studies encompass (a) 275 patients or offenders on probation taking the IOP genuinely, and (b) 274 adult volunteers or offenders on probation

INTRODUCING THE IOP-FDS

taking the IOP with the instruction to feign psychiatric and/or cognitive disorders. With the independent Cross-validation Subsample 3 ($n = 100$), using a probability cutoff core of .50 and considering a base rate of .50, sensitivity was .90, specificity was .80, positive predictive power was .82, negative predictive power was .89, and overall correct classification was .85. The validity of the IOP–FDS (as a measure of feigning) was demonstrated using a variety of data from different conditions and contexts, including head injury, psychotic, PTSD, depressed, and forensic samples, with minimal differences from one data source to another.

The IOP–FDS features a useful and understandable metric, a probability score, to help examiners in the field to opine whether a given presentation is bona fide or feigned. The IOP–FDS score varies from zero to one, representing the probability that the tested individual resembles feigners of psychiatric or cognitive disorders, when the base rate is 50%. For example, a score of .90 suggests that the tested individual is attempting to fake a disorder with approximately 90% probability whereas a score of .50 suggests about a 50% chance. To explain the interpretation of this IOP–FDS probability score, one might imagine two equal stacks of IOP–FDS score outputs, one stack from feigners and one from bona fide patients. Let us say for the case in question, the IOP–FDS score is .80. This score would estimate the probability of such a score coming from the feigning stack to be approximately 80%.

Compared to classically adopted T-scores derived from comparisons to community samples, this IOP–FDS value more directly links the examinee’s score to a theoretical likelihood that he or she is presenting a false condition. To some degree, this approach minimizes the problem produced by the variety of commonly suggested T-score cutoffs across samples, settings, disorders and studies. In practice, it is difficult to be certain of which cut-off to apply to any given, single case (Rogers & Bender, 2013). Only one probability score is produced by the

INTRODUCING THE IOP-FDS

18

IOP–FDS across various disorders. This score can be easily re-calculated to correspond to other base rates (Streiner, 2003) but the initial research suggests that cutoffs do not need to change across disorders, e.g. psychosis/schizophrenia vs. PTSD.

The iterative methodological approach we adopted throughout the studies leading up to the IOP–FDS allowed it to perform similarly across psychiatric and cognitive conditions, as shown in Table 3. This was accomplished by our testing and refining items and strategies across two previous IOP versions, selecting and weighting the IOP–FDS items in the three current, independent, subsamples based on their efficacy and generalizability. Based on research designed to address depression, psychosis, post-traumatic reactions, and cognitive and neuropsychological conditions, as well as combinations thereof, the IOP–FDS should provide excellent coverage for disorders involved with forensic and high stakes determinations. This may be related to a conscious effort in the IOP development to address how complaints are presented rather than the symptoms themselves. Pending further research support, the IOP–FDS has potential to be used across a variety of situations, from homogeneous to diverse symptom presentations. This supports the utility and ecological validity of the IOP in real-life practice, given that malingerers typically present with a multitude of problems rather than with symptoms belonging to a single and specific diagnostic category. Further refinements and developments of the instrument, however, should probably attempt to also develop new diagnosis-specific scales, aimed at providing incremental validity over the current IOP–FDS. For instance, future IOP research could try to develop a depression-specific feigning scale by selecting or double-weighting those items that increment over the IOP–FDS when discriminating between genuine and feigned depression. The 64 IOP items that did not enter in the 117-item, IOP-FDS equation might be particularly suitable to this purpose, as they had received some empirical validation in

past IOP research, but their detection efficacy likely does not generalize very well from one condition to another, given that they were not selected for the IOP-FDS.

Computerization of the IOP allows the recording and analysis of item latency patterns. In the subsample of Study 1, we found that certain latency patterns were associated with the participant belonging to the simulator versus the control groups. Within the independent Subsample 2, the so-derived Latency Feigning indicator demonstrated statistically significant, incremental validity over the Item-Based Feigning indicator. Given these results, and because it is based on within-test patterns rather than individual item latencies, we anticipate that the validity of this measure will generalize across samples and situations, regardless of an individual’s average reading speed, for example. On the other hand, because the Latency Feigning indicator is used to calculate the IOP-FDS, a computer (or tablet) is evidently needed to administer the IOP and record reaction times, together with an administration and scoring software. Readers interested in conducting research with the 181-item IOP are thus encouraged to contact the first or second author of this article in order to obtain it.

In a meta-analysis of MMPI-2 scales in the detection of feigned mental disorders, after inspecting 65 feigning studies and 11 diagnostic studies, Rogers, Sewell, Martin, and Vitacco (2003) reported that F was the most effective scale, producing a Cohen’s *d* effect size of 2.21, whereas Fb produced a *d* of 1.62. A recent meta-analysis of PAI scales conducted by Hawes and Bocaccini (2009) revealed that Cohen’s *ds* were 1.48 and 1.59 for NIM, 1.15 and 1.00 for MAL, and 1.13 and 1.65 for RDF in uncoached and coached malingering studies respectively. In comparison, during our cross-validation research with Subsample 3, which consisted of a series of five clinical comparison simulation studies, the IOP–FDS produced a larger Cohen’s *d* effect size of 2.61. Moreover, in our Study 3, with a small sample of 20 patients and 20 simulators, the

INTRODUCING THE IOP-FDS

20

IOP-FDS outperformed the Personality Assessment Inventory (PAI; Morey, 1991, 1996, 2007) scales: Negative Impression Management (NIM; Morey, 1996), Malinger Index (MAL; Morey, 1996), and Rogers Discriminant Function (RDF; Rogers, Sewell, Morey, & Ustad, 1996) scales. All in all, these findings lend initial support to the use of the IOP-FDS as a valid tool to detect feigned mental or cognitive problems across a variety of conditions and situations.

Although both the IOP-FDS and the IOP-29 yielded encouraging results, the longer IOP version might be preferable over the IOP-29 for a couple of reasons. First, in our independent, Cross-validation Subsample 3, the IOP-FDS significantly outperformed the IOP-29 ($r = .797$ vs. $r = .687$; $z = 3.518$, $p < .001$). Second, future research with the 181-item IOP version might allow for a deeper understanding of the strategies used by each test-taker. Indeed, a number of “descriptive scales” that aim to provide information on how an examinee with a high IOP-FDS score attempts to feign mental health or cognitive problems are currently under development. These include externalizing responsibility for problems, attributing exaggerated symptoms to a trauma, or poor effort on cognitive items. In addition, for those with low IOP-FDS feigning scores who appear to be presenting bona fide complaints, additional diagnostic scales and a measure of cognitive processing are being developed.

Future research might also attempt to address some of the limitations of the current studies. First, in our report, the validity of the IOP-FDS (as a measure of feigning) was tested with clinical comparison, simulation studies only. Additional field research with known-group comparisons of suspected malingers and bootstrapping comparisons are therefore needed to establish ecological validity (Rogers & Cruise, 1998). Moreover, the independent, cross-validation Subsample 3 was comprised of only 100 individuals, which is a modest sample size, raising questions concerning the generalizability of our findings. In this regard, however, it

should be noted that the shrinkage from Subsample 2 to Subsample 3 was virtually nonexistent, which makes it reasonable to assume that the IOP–FDS will likely perform similarly well in future studies. Furthermore, only about a quarter of our participants suffered from or feigned cognitive or neuropsychological disorders, so additional research would be beneficial with this group. Lastly, the generalizability of our findings to non-U.S. contexts also deserves additional research. Accordingly, we are currently researching Italian, Chinese, German, Dutch, and Portuguese versions of the test. Regardless of these limitations, the IOP–FDS demonstrated strong validity in this study as a multipurpose test of feigning, with similar detection efficacy across a variety of clinical conditions.

References

- Abramsky, A. B. (2005). *Assessment of test behaviors as a unique construct in the evaluation of malingered depression on the inventory of problems: Do test behaviors add significant variance beyond problem endorsement strategies?* Unpublished doctoral dissertation, California School of Professional Psychology, San Diego, CA.
- Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment*, 87, 84–94.
- Blanchard D. D., McGrath, R. E., Pogge, D L., Khadivi, A. (2003). A Comparison of the PAI and MMPI-2 as predictors of faking bad in college students. *Journal of Personality Assessment*, 80(2), 197-205
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.
- Connell, K. (2004). *Detecting simulated versus genuine posttraumatic stress disorder*. Unpublished doctoral dissertation, California School of Professional Psychology, San Diego, CA.
- Edens, J. F., Poythress, N. G., Watkins-Clay. M. M., (2007). Detection of Malingering in Psychiatric Unit and General Population Prison Inmates: A Comparison of the PAI, SIMS, and SIRS. *Journal of Personality Assessment*, 88(1):33-42, doi: 10.1080/00223890709336832
- Hawes, S., & Boccaccini, M.T. (2009). Detection of overreporting of psychopathology on the Personality Assessment Inventory: A meta-analytic review. *Psychological Assessment*, 21, 112-124. doi: 10.1037/a0015036

INTRODUCING THE IOP-FDS

23

- 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, Advance Online Publication, March 26*, doi: <https://doi.org/10.1007/s12207-018-9314-1>
- Landis, P. E. (1996). *Detection of simulated posttraumatic stress disorder: A validation study of the Inventory of Problems*. Unpublished doctoral dissertation, California School of Professional Psychology, San Diego, CA.
- McCullaugh, J. M. (2011). *The convergent and ecological validity of the Inventory of Problems with a community-supervised, forensic sample*. Unpublished doctoral dissertation, California School of Professional Psychology, San Diego, CA.
- McDougall, A. (1996). *Rorschach indicators of simulated schizophrenia*. Unpublished doctoral dissertation, California School of Professional Psychology, San Diego, CA.
- Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin*, 52, 194-216.
- Meng, X. L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, 111, 172-175.
- Morey, L. (1991). *Professional manual for the Personality Assessment Inventory*. Odessa, FL: Psychological Assessment Resources
- Morey, L. (1996). *An interpretive guide to the Personality Assessment Inventory (PAI)*. Odessa, FL: Psychological Assessment Resources.
- Morey, L.C. (2007). *Personality Assessment Inventory Professional Manual*, second edition. Odessa, FL: Psychological Assessment Resources.

INTRODUCING THE IOP-FDS

24

- O'Brien, S. M. (2004). *An investigation into the incremental value of test-dependent malingering of schizophrenia*. Unpublished doctoral dissertation, California School of Professional Psychology, San Diego, CA.
- Pizitz, T. D. (2001). *Detection of malingered mild head injury using the tripartite conceptual model of malingering and the inventory of problems*. Unpublished doctoral dissertation, California School of Professional Psychology, San Diego, CA.
- Rogers, R. (1988). *Clinical assessment of malingering and deception*. New York: Guilford Press.
- Rogers, R. (1998). *Clinical assessment of malingering and deception* (2nd ed.). New York: Guilford Press.
- Rogers, R. & Bender, S. D. (2013). Evaluation of malingering and related response styles. In I. B. Weiner (Ed.-in-Chief), J. R. Graham & J. A. Naglieri (Vol. Eds.), *Comprehensive Handbook of Psychology: Assessment Psychology* (2nd Edition Vol. 10, pp. 517-540). Hoboken, NJ: John Wiley & Sons.
- Rogers, R. & Cruise, K. R. (1998). Assessment of malingering with simulation designs: Threats to external validity. *Law and Human Behavior*, 22 (3), 273-285.
- 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). Hoboken, NJ: Wiley.
- Rogers, R., Sewell K. W., Martin M. A., & Vitacco M. J. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI-2 and malingering. *Assessment*, 10(2):160-77.
- Rogers, R., Sewell, K. W., Morey, L. C., & Ustad, K. L. (1996). Detection of feigned mental disorders on the Personality Assessment Inventory: A discriminant analysis. *Journal of Personality Assessment*, 67, 629-640

Streiner, D. L. (2003). Diagnosing Tests: Using and Misusing Diagnostic and Screening Tests. *Journal of Personality Assessment*, 81, 209-219.

Tombaugh, T.N. (1996). *Test of Memory Malingering*. Toronto, Canada: Multi-Health.

Viglione, D.J., Fals Stewart, W. & Moxham, E. (1995). Maximizing internal and external validity in MMPI malingering research: A study in a military population. *Journal of Personality Assessment*, 65, 502-513.

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*, 5, 534-544. doi.org/10.1080/00223891.2016.1233882

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*, 5, 534-544. doi.org/10.1080/00223891.2016.1233882

Viglione, D. J. & Landis, P. (1994). The Development of an Objective Test for Malingering. Paper presented at the biennial meeting American Psychology-Law Society, Santa Fe, New Mexico.

Viglione, D. J., Wright, D., Dizon, N. T., Moynihan, J. E., DuPuis, S., & Pizitz, T. D. (2001). Evading detection on the MMPI-2: Does caution produce more realistic patterns of responding? *Assessment*, 8(3), 237-250.

Wood, S. M. (2008). *Unique contributions of performance and self-report methods in the detection of malingered psychotic symptoms*. Unpublished doctoral dissertation, California School of Professional Psychology, San Diego, CA.

INTRODUCING THE IOP-FDS

26

For Peer Review Only

Table 1. Composition of the Study Subsamples and Data Sources.

	Pizitz (2001) Head Injury	O’Brien (2004) Psychosis	Woods (2008) Psychosis	Connell (2004) PTSD	Abramsky (2005) Depression	McCullaugh (2011) Forensic Sample
Item Selection Subsample 1 (<i>n</i> = 341)						
Honest responders	21	24	25	22	24	35
Simulators	20	25	25	21	24	35
Total	41	49	50	43	48	70
Scaling Subsample 2 (<i>n</i> = 148)						
Honest responders	10	12	12	11	12	17
Simulators	10	12	12	11	12	17
Total	20	24	24	22	24	34
Cross-Validation Subsample 3 (<i>n</i> = 100)						
Honest responders	7	8	8	7	8	12
Simulators	7	8	8	7	8	12
Total	14	16	16	14	16	24
Composite Sample, all data (<i>N</i> = 549)						
Honest responders	38	44	45	40	44	64
Simulators	37	45	45	39	44	64
Total	75	89	90	79	88	128

INTRODUCING THE IOP-FDS

28

Table 2. Diagnostic Efficiency Statistics of the IOP-FDS within Validation Subsample 3 ($n = 100$)

Cut-off	Sensitivity	Specificity	<i>Base Rate = .50^a</i>			<i>Base Rate = .25</i>			<i>Base Rate = .10</i>			<i>Base Rate = .05</i>		
			PPP	NPP	OCC	PPP	NPP	OCC	PPP	NPP	OCC	PPP	NPP	OCC
0.05	1.00	0.16	0.54	1.00	0.58	0.28	1.00	0.37	0.12	1.00	0.24	0.06	1.00	0.20
0.10	1.00	0.34	0.60	1.00	0.67	0.34	1.00	0.51	0.14	1.00	0.41	0.07	1.00	0.37
0.20	0.98	0.56	0.69	0.97	0.77	0.43	0.99	0.67	0.20	1.00	0.60	0.11	1.00	0.58
0.50	0.90	0.80	0.82	0.89	0.85	0.60	0.96	0.83	0.33	0.99	0.81	0.19	0.99	0.81
0.80	0.78	0.96	0.95	0.81	0.87	0.87	0.93	0.92	0.68	0.98	0.94	0.51	0.99	0.95
0.90	0.58	0.98	0.97	0.70	0.78	0.91	0.88	0.88	0.76	0.96	0.94	0.60	0.98	0.96
0.95	0.44	1.00	1.00	0.64	0.72	1.00	0.84	0.86	1.00	0.94	0.94	1.00	0.97	0.97

^a Current sample base rate.

INTRODUCING THE IOP-FDS

Table 3. Validity of the IOP–FDS by Data Source (Cross-Validation Subsample 3, $n = 100$).

	Pizitz (2001) Head Injury ($n = 14$)	O’Brien (2004) & Woods (2008) Psychosis ($n = 32$)	Connell (2004) PTSD ($n = 14$)	Abramsky (2005) Depression ($n = 16$)	McCullaugh (2011) Forensic Sample ($n = 24$)
IOP – FDS					
Sensitivity	.86	.88	.86	1.00	.92
Specificity	.57	.75	.71	.88	1.00
PPP (base rate = .50)	.67	.78	.75	.89	1.00
NPP (base rate = .50)	.80	.86	.83	1.00	.92
OCC (base rate = .50)	.71	.81	.79	.94	.96
Correlation with group membership ^a	.69**	.70**	.73**	.93**	.96**
AUC	.88*	.91**	.88*	1.00**	1.00**

^a Point bi-serial correlation with group membership (dummy code: 0 = honest respondent; 1 = simulator); * $p < .05$; ** $p < .01$.

INTRODUCING THE IOP-FDS

30

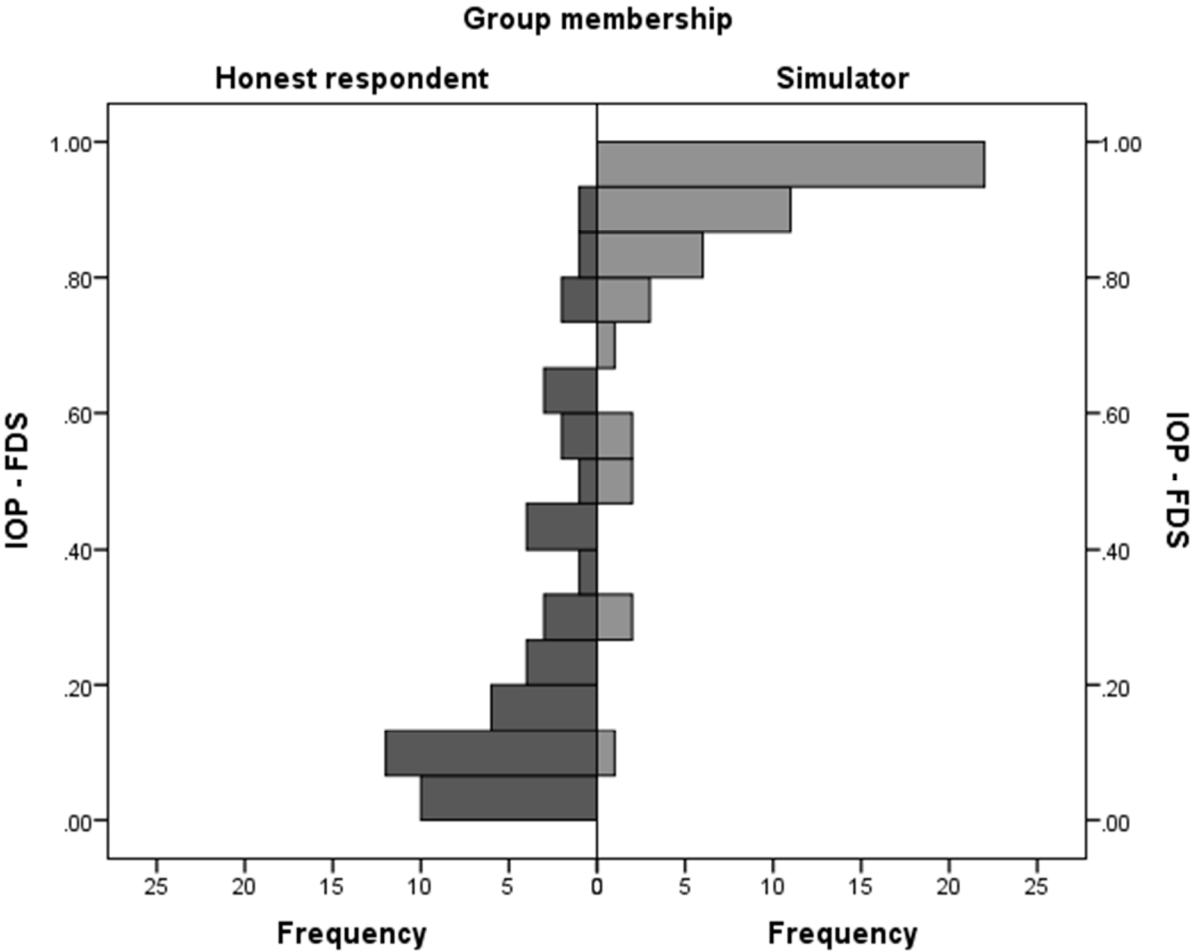
Table 4. Comparative Validity of the IOP–FDS against the PAI ($n = 40$).

	Point bi-serial correlation with group membership	<i>AUC</i>
IOP – FDS	.92**	.99**
PAI – NIM	.63**	.85**
PAI – MAL	.35*	.70*
PAI – RDF	.54**	.83**

Note: This is a subgroup from Cross-Validation Subsample.

* $p < .05$; ** $p < .01$.

Figure 1. Distribution of IOP–FDS among Honest Respondents ($n = 50$) and Simulators ($n = 50$) of Cross-Validation Subsample 3



Abstract

~~Recently, we introduced the Inventory of Problems–29 (IOP–29), a brief, paper and pencil instrument designed to discriminate bona fide from feigned mental illness, which was derived from a more comprehensive, computerized, 181-item tool, the Inventory of Problems (IOP). To provide background, the current article begins with an overview of both these instruments. Then, it turns to its main purpose~~ This article introduces the Inventory of Problems (IOP) – a new, computerized, 181-item tool designed to discriminate bona-fide from feigned mental illness and cognitive impairment – and presents the development and validation of its focal, feigning scale, the False Disorder Score (IOP–FDS). The initial sample included (a) 211 patients and 64 offenders who took the IOP under standard conditions and (b) 210 community volunteers and 64 offenders who feigned mental illness. We split this sample into three subsamples. The first ($n = 301$) was used to select the variables to generate the IOP–FDS; the second ($n = 148$) scaled the IOP–FDS into a probability score; and the third ($n = 100$) tested its validity with an independent dataset. In this third subsample, the IOP–FDS had sensitivity = .90, specificity = .80, and a greater AUC (= .95) than the IOP-29 (= .91). For 40 participants, the PAI was available too. Within this subgroup, the IOP–FDS outperformed the selected PAI validity scales ($AUC = .99$ vs. $AUC \leq .85$).

Keywords: Inventory of Problems, IOP, feigning, malingering, test development, validity.

Comment [LG1]: R1, Issue 1: “the opening sentence of the Abstract should focus on the aim to develop the FDS rather than on the more tangential fact that the IOP-29 exists. So much attention to the IOP-29 at the outset maintains one very confusing aspect of the original paper, i.e., why are we describing a short version of an instrument when scoring of the long form has not yet been derived?”

Comment [LG2]: REPLY: Thank you for this suggestion, we have revised the abstract accordingly.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

INTRODUCING THE IOP-FDS

2

**Development and Validation of the False Disorder Score:
The Focal Scale of the Inventory of Problems**

The Inventory of Problems (IOP) is a new, multipurpose, 181-item, computerized test designed to investigate feigning of various psychiatric and cognitive complaints. The aim of the current paper is to report on the development and validation of its focal scale, the False Disorder Score (IOP-FDS) – a measure of the likelihood that the test-taker is presenting a false mental health or cognitive complaint. To accomplish this aim, this article presents three studies. The first study selects from the IOP the items and latencies (i.e., the times to answer the items, or reaction times), to be included in the IOP-FDS. In the second study these data are scaled to create the IOP-FDS. Finally, the IOP-FDS is cross-validated with an independent subsample. Because one might question whether a single measure could detect a wide range of clinical presentations, we also highlight the strategies and techniques we adopted to enable the IOP-FDS to identify a broad array of problems and symptom combinations including neuropsychological impairments, psychosis and schizophrenia, post-traumatic stress disorder (PTSD), and depression.

~~A brief derivative of the IOP, comprised of 29 items and named “IOP-29,” was recently introduced in the *Journal of Personality Assessment* (Vigilione, Giromini, & Landis, 2017). The current article, in contrast to this previous publication, is the first to report on the full length, 181-item version of the IOP. Accordingly, before describing the development and validation of its first and most important scale, the IOP-FDS, we summarize the theoretical and empirical background of the IOP itself.~~

Background for the Inventory of Problems (IOP)

Comment [LG3]: R1, Issue 1: “I recommend deleting the two paragraphs in Introduction, on page 2 and on pages 5-6, that address the IOP-29. Brief mention of the existence of a short form can be placed in the Method section or in the Results section, as the FDS based on the full IOP will be compared to the IOP-29. Including so much discussion of the IOP-29 in the Introduction gives the reader the false impression that this short form is an important part of the study when in fact it plays a minor role in the third study”

Comment [LG4]: REPLY: As suggested, we deleted these paragraphs, and removed the section on the IOP-29 from the Introduction.

INTRODUCING THE IOP-FDS

3

A leading principle in developing the first version of the IOP (Viglione & Landis, 1994), which was mentioned by Rogers (1998) in his second edition of his book on malingering and deception, was that an omnibus feigning test should incorporate multiple detection strategies. Based on the literature at the time (Rogers, 1988) and experience with feigning in the military and practice (see for example, Viglione, Fals-Stewart, & Moxham, 1995) 27 different strategies and 245 corresponding items were formulated for the first version. The reviews of the literature and experience in practice revealed that scales with symptom oriented, keyed true, answered true items were highly redundant with one another and that they dominated most of the feigning measures available at the time. Thus, our item-development strategy focused on potential incremental validity over such items and our pilot research focused on examining whether or not they (a) differentiated bona fide patients and honest probationers from feigners, and (b) whether they added incremental validity beyond keyed true answered true symptom items. This initial research led to many item and detection strategy revisions and refinements and to the development of a second IOP version, comprised of 162 items. Then, based on the research and experience accumulated with the first and second version of the IOP, we again pruned, revised, and added items to create the third and current version of the IOP, which includes 181 items.

Throughout this long developmental period, several doctoral dissertations and other research projects were conducted specifically to test our detection strategies and to refine the IOP item pool. In fact, prior to conducting the three studies described below, about 1,000 participants had already been tested with one of the three developmental versions of the IOP. These research efforts focused on the contrast between patients and feigners, with little concern for discriminating patients from honest non-patients since creating such a scale is easily done but largely irrelevant to the goal of helping examiners in the field to opine whether a given

Comment [LG5]: R1, Issue 2: "Some discussion of where the 181 items of the IOP came from is warranted in the Introduction. How were the items generated? References are made to different versions of the IOP. Was the item pool narrowed or expanded? How were items selected prior to the analyses reported here?"

Comment [LG6]: REPLY: As recommended, we added some information about the genesis and rationale of the test items, how they were revised with 3 versions of the test, and the number of items of each version. We believe that these additions and the information already included in the next few paragraphs should provide the requested background for the reader.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

INTRODUCING THE IOP-FDS

4

psychiatric presentation is bona fide or feigned. Moreover, to minimize extraneous variance and confounds, in most of these studies, we matched or minimized demographic differences and used community rather than college samples. In some cases, feigners were instructed to fake the symptoms or history of the patient with whom he or she was matched. In the great majority of the cases, feigners were instructed “not to overdo it” (Viglione et al., 2001), so that they would more closely resemble successful malingerers in real life situations and to diminish artificially large effect sizes (Rogers & Bender, 2013; Rogers & Gillard, 2011).

The final, 181-item version of the IOP, is administered electronically. In addition to the classic, self-report or symptom validity items (e.g., “I feel terribly sad every day.” [Keyed *True*] “Sometimes, others help me to feel OK.” [Keyed *False*]),¹ the IOP also includes performance validity items in the form of easy cognitive problems, including calculation (e.g., “150 – 50 = ?”), memory (e.g., the test-taker is asked to recall content or simple pictures that were introduced earlier in the test), reasoning (e.g., “A tree is to a forest, as a lightning is to a thunderstorm.”), and other pattern recognition and cognitive items. Additionally, a few Likert-scale symptom ratings, for example, “Rate your problems with depression on a 1 to 7 scale where 1 = No problem7 = Unbearable,” are included. Some of the items were inclusively worded affirmations, such as “It’s killing me,” as we thought that such pronouns and vague language (for example, “it” or “my problem”) might capture a wide variety of false complaints. Of course, only those strategies and item-formats that were supported in the pilot research with the first two IOP versions were retained in the final IOP.

In addition to the prototypical “*True*” versus “*False*” response options, most IOP self-report items offer a third response option, “*Doesn’t Make Sense*.” This is because in our

¹ Test items quoted demonstrate the principles described and closely resemble items in the test the structure but are not themselves necessarily included in the test.

INTRODUCING THE IOP-FDS

5

developmental research leading up to the final version of the test, simulators more readily endorsed oddly worded items with pathological content. That is, rather than reading carefully, simulators perhaps ignored odd wording because they were scanning for pathological content to decide whether or not to endorse it. Conversely, patients and control participants did not always understand these items, and thus tended to choose the response option *Doesn't Make Sense* more frequently. With clearly written items, in contrast, endorsing the *Doesn't Make Sense* response option seemed to reflect an attempt to feign cognitive impairment or express an uncooperative intent, behavior that is sometimes adopted by real life malingerers. Indeed, for these items, simulators tended to choose *Doesn't Make Sense* more frequently than did bona fide patients.

Another key feature of the IOP is that some of its items address test-taking behaviors and experiences. Indeed, we wrote items to capture the behavioral dramatization of symptoms seen in clinical interviews (e.g., "My hands are shaking uncontrollably during this test." [Keyed *True*] "This test does not cover enough of my problems." [Keyed *True*]). Likewise, we hypothesized that item content which incorporated (a) externalization of responsibility for one's woes, symptoms, and predicament while minimizing one's ability to improve them (e.g., "Sometimes, I can think about things that make me feel better." [Keyed *False*]) and (b) refusal to admit qualified positive attributes (e.g., "I feel attractive sometimes." [Keyed *False*]) could be particularly effective. Pilot research with the IOP versions one and two confirmed these expectations.

A final distinctive feature of the IOP presented here is that its computerized administration allows recording of item/response latencies, i.e., the reaction time between presentation of an item and the response via computer key. Using this information, we hypothesized that the interaction between the specific content of an item and the response latency

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

INTRODUCING THE IOP-FDS

6

to that item also could contribute to discriminating between bona fide and feigned mental problems. Thus, we evaluated latencies and interactions between latencies and the *True, False,* and *Doesn't Make Sense* response alternatives.

The IOP-29

As noted above, a brief version of the IOP, i.e., the IOP-29 was introduced in a recent publication (Viglione et al., 2017). This measure demonstrated validity in distinguishing bona fide cases from feigned demonstrations of mild traumatic brain injury, psychosis/schizophrenia, PTSD, and depression. In addition, the IOP-29 effectively discriminated feigned, mixed PTSD, depression, and psychosis presentations from controls among offenders with mental health problems treated in the community. Despite its being comprised of only 29 items, the IOP-29 performed similarly to the Minnesota Multiphasic Personality Inventory (MMPI-2 [R. L. Green, 1991]; MMPI-2 RF [Ben Porath & Tellegen, 2008]) and Personality Assessment Inventory (PAI; Morey, 2007) and perhaps better than the Test of Memory Malingering (TOMM; Tombaugh, 1996) as an all-purpose feigning detection tool.²

The Current Studies on the IOP-FDS

A brief derivative of the IOP, comprised of 29 of the 181 items of the IOP, and named “IOP-29,” was recently introduced in the *Journal of Personality Assessment* (Viglione, Giromini, & Landis, 2017). The current article, in contrast to this previous publication, is the first to report on the full length, 181-item version of the IOP. In the present paper, we describe the development and validation of the 181-item IOP’s focal feigning scale, the IOP-FDS. The validity of this new scale as a measure of feigning is then compared with that of the IOP-29 and, for a small subset of data, with that of the PAI.

² The IOP-29 items are included in 181-item IOP and within the IOP-FDS.

Comment [LG7]: R2, Issue 1: “The separate section on the IOP-29 is likely unnecessary. The needed information can be integrated into the section on the IOP.”

Comment [LG8]: REPLY: As suggested, in line also with Reviewer #1’s first comment, we deleted this section.

INTRODUCING THE IOP-FDS

7

Materials and Methods**Sampling Procedures and Participants**

The consolidated sample utilized to develop and validate the IOP–FDS is the same one used to scale and cross-validate the IOP-29 (Viglione et al., 2017): It combines data from six dissertation studies (Abramsky, 2005; Connell, 2004; McCullaugh, 2011; O’Brien, 2004; Pizitz, 2001; Woods, 2008).³ More specifically, it contains 275 patients or offenders on probation taking the third version of the IOP with standard instructions contrasted to 274 volunteers or offenders on probation taking the IOP with the instruction to feign a psychiatric and or cognitive disorders.

Honest respondents were 38 patients suffering neuropsychological deficits (Pizitz, 2001), 89 patients affected by schizophrenia or psychosis (O’Brien, 2004; Woods, 2008), 40 volunteers suffering from PTSD (Connell, 2004), 44 patients with depression (Abramsky, 2005), and 64 adult offenders on probation being treated for mental health or substance abuse (McCullaugh, 2011). Simulators were 211 non-clinical adult volunteers instructed to either feign (a) neuropsychological deficits ($n = 37$; Pizitz, 2001), (b) schizophrenia or psychosis ($n = 90$; O’Brien, 2004; Wood, 2008), (c) PTSD ($n = 39$; Connell, 2004), or (d) depression ($n = 44$; Abramsky, 2005), as well as 64 adult offenders on probation instructed to feign a mixture of neuropsychological, depressive, and PTSD symptoms (McCullaugh, 2011). Thus, the data are relevant to the four target diagnostic categories of the IOP – cognitive/neuropsychological impairment, psychosis, PTSD, and depression.

Details on procedures and demographic composition of the sample and sample sources are described in Viglione et al. (2017). Briefly, all participants were adult volunteers, and heterogeneous regarding gender, age, racial characterization, education, and marital status. All

³ Item selection for the IOP-29 was done with previously collected samples.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

INTRODUCING THE IOP-FDS

had signed an informed consent form prior to being enrolled in the study. To maximize external validity, all simulators were provided with a brief scenario or vignette aimed at improving their feigning abilities, and were instructed not to produce excessively dramatic or severe symptom presentations, or else their performances would easily be detected as fake or feigned (Rogers & Bender, 2013; Rogers & Gillard, 2011, Viglione et al., 2001). Within each of the data sources under consideration, honest respondents (patients and offenders on probation) and simulators did not differ from each other on any important, demographic variables (Viglione et al., 2017).

Approach to Scale Construction & Validation

To apply these data sources to the development and validation of the IOP-FDS, we randomly split this combined sample ($N = 549$) into three⁴ subsamples: Item Selection Subsample 1 ($n = 301$); Scaling Subsample 2 ($n = 148$); and Cross-validation Subsample 3 ($n = 100$). In parsing the composite sample, we allotted more participants to statistical procedures with less power. Specifically, the Item Selection Subsample 1 required the largest number of participants and the most power as we worked with both the answers and the response latencies for the individual 181 items of the IOP. The Scaling Subsample 2 was used to derive the equations for combining these two indicators (answers and latencies) into our final IOP-FDS and therefore, required fewer participants. The Cross-validation Subsample 3 was the smallest because it was used to cross-validate only one variable (i.e. the final IOP-FDS). Although the assignment of IOP's to the three subsamples was done randomly, this randomization was stratified by dissertation and equally between honest patient versus simulator. As a result, each subsample contained approximately the same proportion of individuals from each dissertation. (See Table 1).

⁴ The reader should endeavor not to confuse the *three versions* of the IOP with the *three subsamples* or research studies presented in this paper. The current paper presents three subsamples with the third and final version of the IOP.

INTRODUCING THE IOP-FDS

9

Study 1 – Item Selection

Research Questions and Analysis Plan

Study 1 aimed at developing two independent subcomponents, or indicators, to be included in the IOP–FDS. More specifically, we intended to develop a feigning indicator based on the responses given to the IOP, and a feigning indicator using the latencies or reaction times for each item. We thus planned two sets of analyses. First, we tested the Phi association of each item/response combination (e.g., *True* to item 1; *False* to item 1; etc.) to group membership (dummy code, with 0 = honest respondent; 1 = experimental simulator), to select the items that would demonstrate the strongest associations. These items would then be combined to generate our first feigning indicator. Next, we examined all individual patterns of response latencies (e.g., reaction time to endorse *True* on item 1; reaction time to endorse *False* on item 1; etc.) and their point bi-serial correlations to group membership, to select the latency patterns that would demonstrate the strongest associations. These would then be used to generate our second feigning indicator.

The first indicator, named “Item-Based Feigning” indicator, would thus be a multi-method scale derived from self-report items, after inspecting all keyed *True* (T) and *False* (F) items, all cognitive items (Cog), all *Doesn’t Make Sense* (D) responses⁵, and all the a-priori determined, item pairs (Pairs), i.e., contradictory responses for two related items, (e.g. answering *True* to “My worst time is when I get up” yet *False* to “My best time is early in the day.”) The second feigning indicator scale, “Latency Feigning,” would instead be derived from response latencies of individual items.

⁵ There are six items in which the “D” response alternative corresponds to some version of “Do not have that problem.” Three address denials of hearing voices, one denial of depression, and one denying of being punished physically as a child.

Comment [LG9]: R2, Issue 2: “In study 1, the analysis plan states that you “tested the association” between items. Please include the analysis that was performed. If there was a “cut point” that was used, it would be helpful to include that as well.”

Comment [LG10]: REPLY: We clarified the type analyses conducted for both indicators, i.e., Phi associations for the Item-Based indicator, and point bi-serial correlations for the Latency indicator. We considered, but eventually decided not to report the cut points here, as they are detailed below, in the dedicated subsections. In our opinion, repeating them also here would make it difficult for the reader to understand the general approach we followed.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

INTRODUCING THE IOP-FDS

10

Development of Item-Based Feigning Indicator. Two principles guided the development of the Item-Based Feigning and Latency Feigning indicators. First, we wanted to select (i.e., assign points to) item/response combinations that would perform similarly well from one data source, or diagnostic target, to another. Second, to be selected or given an extra point, an item/response combination should provide incremental validity over the other, already available or selected, item/response combinations in the indicator. According to the first of these principles, we selected the IOP-29 item/response combinations as the core of the IOP-FDS, and assigned them the same item/response weights established for the IOP-29 main scale. These item/response combinations, indeed, have already demonstrated to be applicable to multiple symptom presentations in various data sets both in the U.S. (Viglione et al., 2017) and in Italy (Giromini et al., 2018), with virtually no shrinkage from one sample to another. As such, they represent an already established, fully validated measure of credibility of a wide range of different symptom presentations. Next, we inspected all item/response combinations from all 181 IOP items, and gave one point (or extra point) to those item/response combinations that correlated (*Phi* correlations) with group membership (0 = honest respondent; 1 = simulator) with an effect size of at least $r = .30$ (i.e., *medium*; Cohen, 1992) in each and every one of the five data sources, i.e., head injury (Pizitz, 2001), psychotic (O’Brien, 2004; Woods, 2008), PTSD (Connell, 2004), depressed (Abramsky, 2005), and offenders on probation (McCullaugh, 2011). This procedure terminated with the creation of our first, “in-progress,” Item-Based Feigning Indicator.

According to our second guiding principle, we then used an iterative procedure, or algorithm, to further refine this in-progress index, and assign extra points to any item/response combinations that would demonstrate incremental validity. More specifically, this in-progress

Comment [LG11]: R1, Issue 3: “What is the rationale for selecting the item/response combinations from the IOP-29 first, rather than sampling the entire 181-item domain of the full IOP?”

Comment [LG12]: REPLY: We have clarified that the item/response combinations making up the IOP-29 have been widely validated in various samples, more than any other items in the 181-item IOP.

INTRODUCING THE IOP-FDS

11

index was then further improved step-by-step, or item-by-item, by using partial correlations, with the final version being the Item-Based Feigning Indicator. That is, at each step, the item/response combination with the highest partial correlation with group membership, after partialing out the in-progress index, was examined. If its average partial correlation across the data sources was (condition one) greater than .10 (*small* effect; Cohen, 1992), and (condition two) all data sources correlated (partial correlations) in the same direction, this item/response combination was added to the index and then applied to the next partial correlation. This repetitive procedure terminated when no more partial correlations met both these conditions.

Development of Latency Feigning Indicator. After developing our Item-Based Feigning Indicator, we next focused on patterns of response latencies. The purpose was to investigate whether, compared to the examinee's average latency to all the IOP items, taking a longer or shorter time to answer any given individual test item would be related to feigning of mental and/or cognitive disorders. For example, we speculated that ~~when confronted with the second item of an item pair, a simulator might take a longer time than an honest respondent to answer, because some additional time might be spent to remember what his/her answer was to the first item of the pair. Conversely, we postulated that~~ when compared to honest patient and offender respondents, **simulators might be faster to answer *True* to some keyed *True* items, in that honest respondents might be more resistant than simulators to admit or reveal that they suffer from certain psychopathological problems (e.g., sexual problems).**

To limit the potential impact of outliers (i.e., extremely long item latencies), we truncated the maximum latency interval for each IOP item to the 95th percentile of the combined, patient/forensic group. Next, because latencies are typically positively skewed and non-normal, we applied a square root transformation to all these trimmed, latency values. These values were

Comment [LG13]: R1, Issue 4: "I don't follow why the latency patterns are evaluated in the context of a response to the second item of a pair (page 10). Are these the a priori contradictory pairs referred to on page 9? The reference to "the second item of an item pair" on page 10 suggests it is latency to "contradictory" pairs, but why is latency to the second item in each pair measured, rather than latency to the first item? I am not following something here."

Comment [LG14]: REPLY: Per this comment, we removed the confusing passage, and added some detail to the other example listed here.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

INTRODUCING THE IOP-FDS

12

then standardized, based on z-scores from the combined, patient and offender on probation group, to place all of these latency variables on the same unit of measurement. Lastly, via multiple regression analyses, each item’s latency was predicted based on the particular individual’s average latency for all other IOP items, and the residuals between predicted and observed latencies were used as our target latency parameter. This latency could be understood as the particular item’s characteristic after controlling for the average reaction time for the entire IOP.

Because we anticipated that latencies would produce smaller effect sizes in predicting feigning than would the IOP items, which were specifically designed to achieve that goal, we reduced the cut-off or threshold for item selection to $r = .10$ (*small* effect size; Cohen, 1992). Thus, based on our two guiding principles described above, reaction-time based variables were included in our final Latency Feigning Indicator if they (a) produced a correlation with group membership of at least $|r| = .10$, or (b) produced an average partial correlation with group membership of at least $r = .10$, after partialing out the Item-Based Feigning. If both conditions (a) and (b) were satisfied, the variable was double-weighted. The Latency Feigning Indicator was then calculated as the arithmetic mean of all selected individual item latencies.

Results

Based on item selection and item weight procedures detailed above, the Item-Based Feigning Indicator ultimately included 60 items and 33 item pairs, for a total of 79 different IOP items (some of the individual test items included in the indicator were also part of one or more item pairs). The average *Phi* correlation of these 79 items to group membership was .35 (*SD* = .13). Importantly, many of these items received multiple weights. Indeed, the Item-Based Feigning Indicator can potentially range from 0 to 143 points. More in detail, 75 of these points

Comment [LG15]: R1, Issue 5: "At the end of Study 1, the paper should be clearer with regard to how many of the 181 items are used to compute the FDS. It includes "60 items and 33 item pairs." Do these items overlap or are there 126 different items used to calculate the FDS?"

Comment [LG16]: REPLY: We have largely revised this entire section, to provide the readers with the specific information requested in this comment. Thanks.

INTRODUCING THE IOP-FDS

13

came from IOP-29 item/response combinations and/or from items with *Phi* correlations $\geq .30$ in each of the five data sources mentioned above; 68 came from the iterative partial correlation procedures described in the above sections. Honest responders had an overall mean of 61.6 points ($min = 55$; $max = 72$; $SD = 3.7$), whereas simulators had an overall mean of 76.7 points ($min = 60$; $max = 88$; $SD = 5.6$); this difference is statistically significant, with a very large effect size, $t(259.6) = 27.7$, $p < .001$, $d = 3.2$. Point bi-serial correlation of Item-Based Feigning Indicator to group membership was .848, $p < .001$. *AUC* was .971 ($SE = .010$).

The final Latency Feigning Indicator included 63 items, four of which received double weights. Of these 63 items, 25 were included also in the Item-Based Feigning Indicator, whereas 38 were not, so that to calculate both the indicators (i.e., the Item-Based and Latency Feigning Indicators) a total of 117 items ($= 79 + 38$) would be needed. The average point bi-serial correlation of these 63 reaction time-related variables to group membership was .15 ($SD = .04$). Honest responders had a mean Latency Feigning Indicator value of -.11 ($min = -.48$; $max = .26$; $SD = .19$), simulators had mean value of .12 ($min = -.47$; $max = .61$; $SD = .14$). The difference between the two groups is statistically significant, with a large effect size, $t(278.9) = 12.2$, $p < .001$, $d = 1.4$. The point bi-serial correlation of the Latency Feigning Indicator to group membership was $r = .576$, $p < .001$. Perhaps more importantly, partial correlation with group membership after removing the effects of the Item-Based Feigning Indicator was .209, $p < .001$, which is quite impressive, given the high association between Item-Based Feigning and group membership. As desired, the Latency Feigning Indicator incremented over Item-Based Feigning Indicator.

Study 2 – Scaling

Research Questions and Analysis Plan

Comment [LG17]: R1, Issue 3: "A table reporting the phi correlations for all 181 items would obviously be rather large, but why not summarize these discrimination metrics in the text for the final selected items?"

Comment [LG18]: REPLY: As suggested, we have summarized these statistics and descriptions in the text.

Comment [LG19]: R2, Issue 4: "It would be helpful to include descriptive data (mean, SD, range, etc.) for the two indicators. While much of the information is clear, what the actually scoring or numbers that one would get from the indicators in still unclear. Some of this information is presented in the discussion, but would be nice to include in the results section"

Comment [LG20]: REPLY: Good point. We have added this information.

Comment [LG21]: R1, Issue 5: "latencies are calculated with 63 items, but it is not clear whether and how many of these items are also used as item-level indicators. Ultimately, this reader is curious about how many of the 181 items are not used to calculate the FDS"

Comment [LG22]: REPLY: As noted above, we have largely revised this paragraph, to add the information requested.

Comment [LG23]: R2, Issue 3: "Please clarify the range of scores for the Latency Feigning Indicator."

Comment [LG24]: REPLY: As suggested, we have reported key descriptive statistics concerning this indicator, including the range, mean and standard deviation within each sample.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

INTRODUCING THE IOP-FDS

14

The goal of Study 2 was to test whether the Item-Based Feigning and Latency Feigning indicators would continue to predict group membership when inspecting a new, independent sample, i.e., the Scaling Subsample 2 ($n = 148$). Perhaps more importantly, Study 2 also aimed at testing whether, also within this independent sample, the Latency Feigning Indicator would continue to contribute to predict group membership after controlling for the effects of the Item-Based Feigning Indicator. Indeed, the ultimate goal of Study 3 was to combine the two aforementioned indicators to optimize prediction. As such, we planned to enter our two newly-developed feigning indicators into a logistic regression with group membership as the outcome variable. The regression equation derived from this analysis would then be used to generate the formula to calculate the IOP – FDS score, which would thus be expressed as a probability score ranging between one (certainly a simulator) and zero (certainly not a simulator).

Results

The logistic regression was statistically significant, $\chi^2(2) = 79.73, p < .001$. Nagelkerke R^2 was .555. Both the Item-Based Feigning and Latency Feigning predictors significantly and uniquely contributed to the model. Item-Based Feigning produced an $exp(B)$ greater than 10,000, $p < .001$; the Latency Feigning produced an $exp(B)$ of 63.78, $p = .015$. The IOP-FDS was thus calculated using the regression equation formula obtained from this model. Thus, its score varies from 0.00 to 1.00, reflecting the likelihood of drawing a particular IOP-FDS score from a group of experimental feigners versus the group with bona fide disorders, when the base rate of feigners is 50%. For example, an IOP-FDS probability score of .75 would suggest a 3 to 1 odds that that IOP-FDS came from a person attempting to feign a mental or cognitive disorder, rather than from a bona fide patient.

Study 3 – Cross-validation

Comment [LG25]: R1, Issue 6: "At the end of Study 2, it would be useful to describe the characteristics of the FDS metric that is generated from the logistic regression equation. Some description is brought up in the second paragraph of the Discussion section, but it seems out of place here."

Comment [LG26]: REPLY: As suggested, we have added important information on the metric of the FDS also here, at the end of Study 2.

INTRODUCING THE IOP-FDS

15

Research Questions and Analysis Plan

The aim of Study 3 was to test the validity (as a measure of feigning) of the IOP-FDS with another, independent subsample ($n = 100$). We thus inspected receiver-operator characteristic curves (*AUC*'s), diagnostic efficiency statistics, as well as point-biserial correlations with group membership, and Cohen's d effect sizes. Because the IOP-FDS score was designed to be superior to the IOP-29 and to compete with available feigning tools, we also compared the validity of the IOP-FDS to the IOP-29 and PAI. The PAI was selected because a limited number of PAI data were available in Subsample 3.

Results

The IOP-FDS was tested with the independent, Cross-validation Subsample 3, which included 50 honest respondents and 50 simulators. The point bi-serial correlation of IOP-FDS to group membership was $.797, p < .001$; *AUC* was $.950 (SE = .020)$. The difference between the IOP-FDS of patients vs. simulators corresponded to a Cohen's d of 2.61. Figure 1 shows the distribution of IOP-FDS among honest respondents and simulators.

Diagnostic efficiency results can be found in Table 2. Because positive predictive power (PPP), negative predictive power (NPP), and overall correct classification (OCC) largely depend on the base rates of the conditions under consideration (Meehl & Rosen, 1955), in addition to the .50 base rate of this subsample, we also inspected other common base rates, i.e., .25, .10, and .05, adjusting PPP, NPP, and OCC formulas accordingly (Streiner, 2003). Furthermore, we also inspected various IOP-FDS cut-offs, as they obviously affect diagnostic efficiency statistics, too. When using a cut-off of .50, which matches the base rate of all three subsamples, and the optimal base rate for the use of diagnostic tests, sensitivity is .90, specificity is .80 and the OCC, assuming a base rate of .50, is .87.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Validity for Individual Diagnostic Categories. Because the IOP–FDS was designed to detect feigning across various mental and/or cognitive disorders, we next investigated its validity within each of the five data sources under investigation. As seen in Table 3, the *AUC* values ranged from .88 to 1.00, with a mean of .93 (*SD* = .06). Sensitivity ranged from .86 to 1.00 (*M* = .90, *SD* = .06), and specificity ranged from .57 to 1.00 (*M* = .78, *SD* = .16). Although the size of the individual samples within Subsample 3 was rather small, the IOP–FDS worked fairly well with all data sources. The best performance was observed within the forensic (McCullaugh, 2011) and depression (Abramsky, 2005) data sources; whereas the worst one was within the head injury subgroup (Pizitz, 2001). It is difficult to say whether these differences are associated with substantive, replicable differences connected to type of disorder or are due to variability with the small subsample sizes, which ranged from 14 to 32.

Comparative Validity. For the entire Cross-validation Subsample 3, we examined the IOP-29 score, to inspect whether the IOP–FDS incremented over the IOP-29. Whereas the IOP–FDS produced a point bi-serial correlation with group membership of .797, *p* < .001, the IOP-29 index produced a point bi-serial correlation of .687, *p* < .001. We tested whether the difference between these two correlations was statistically significant using procedures described by Meng, Rosenthal, and Rubin (1992), and found that indeed it was statistically significant, *z* = 3.518, *p* < .001. *AUC* was .950 (*SE* = .020) for the IOP–FDS, and .901 (*SE* = .032) for the IOP-29. Although both IOP–FDS and IOP-29 produced very promising results, the IOP–FDS performed significantly better than the IOP-29.

In addition to the IOP, 40 individuals in the Cross-validation Subsample 3 had also previously been administered the Personality Assessment Inventory (PAI; Morey, 1991, 1996, 2007). Sixteen were from the psychosis sample (Woods, 2008) and 24 were from forensic

INTRODUCING THE IOP-FDS

17

settings (McCullough, 2011), with half being feigners. Consistent with previous research (for example, Archer et al., 2006; Blanchard et al., 2003; Edens, Poythress, & Watkins-Clay, 2007; Hawes & Bocaccini, 2009), we selected the Personality Assessment Inventory (PAI) Negative Impression Management (NIM; Morey, 1996), Malinger Index (MAL; Morey, 1996), and Rogers Discriminant Function (RDF; Rogers, Sewell, Morey, & Ustad, 1996) for PAI validity comparisons. The results of these analyses, reported in Table 4, indicate that within this small sample, the IOP-FDS outperformed all three of the selected PAI scales. In fact, using Meng et al. (1992) procedures, we found that the point bi-serial correlation value produced by the IOP-FDS was significantly greater than that produced by the PAI – NIM, $z = 4.835, p < .001$, PAI – MAL, $z = 5.880, p < .001$, and PAI – RDF, $z = 4.941, p < .001$. The mean *AUC* for the three PAI scales was .79, whereas this value was .99, nearly perfect, for the IOP – FDS.

Discussion

This paper presents the development and initial validation of the Inventory of Problems – False Disorder Scale (IOP-FDS), the focal scale of the Inventory of Problems (IOP). After a development period involving approximately 1,000 participants and two previous versions, and the creation of a brief, derivative version (the IOP-29; Viglione et al. 2017), we examined the third and final IOP version. The current studies encompass (a) 275 patients or offenders on probation taking the IOP genuinely, and (b) 274 adult volunteers or offenders on probation taking the IOP with the instruction to feign psychiatric and/or cognitive disorders. With the independent Cross-validation Subsample 3 ($n = 100$), using a probability cutoff core of .50 and considering a base rate of .50, sensitivity was .90, specificity was .80, positive predictive power was .82, negative predictive power was .89, and overall correct classification was .85. The validity of the IOP-FDS (as a measure of feigning) was demonstrated using a variety of data

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

INTRODUCING THE IOP-FDS

18

from different conditions and contexts, including head injury, psychotic, PTSD, depressed, and forensic samples, with minimal differences from one data source to another.

The IOP–FDS features a useful and understandable metric, a probability score, to help examiners in the field to opine whether a given presentation is bona fide or feigned. The IOP–FDS score varies from zero to one, representing the probability that the tested individual resembles feigners of psychiatric or cognitive disorders, when the base rate is 50%. For example, a score of .90 suggests that the tested individual is attempting to fake a disorder with approximately 90% probability whereas a score of .50 suggests about a 50% chance. To explain the interpretation of this IOP–FDS probability score, one might imagine two equal stacks of IOP–FDS score outputs, one stack from feigners and one from bona fide patients. Let us say for the case in question, the IOP–FDS score is .80. This score would estimate the probability of such a score coming from the feigning stack to be approximately 80%.

Compared to classically adopted T-scores derived from comparisons to community samples, this IOP–FDS value more directly links the examinee’s score to a theoretical likelihood that he or she is presenting a false condition. To some degree, this approach minimizes the problem produced by the variety of commonly suggested T-score cutoffs across samples, settings, disorders and studies. In practice, it is difficult to be certain of which cut-off to apply to any given, single case (Rogers & Bender, 2013). Only one probability score is produced by the IOP–FDS across various disorders. This score can be easily re-calculated to correspond to other base rates (Streiner, 2003) but the initial research suggests that cutoffs do not need to change across disorders, e.g. psychosis/schizophrenia vs. PTSD.

The iterative methodological approach we adopted throughout the studies leading up to the IOP–FDS allowed it to perform similarly across psychiatric and cognitive conditions, as

INTRODUCING THE IOP-FDS

19

shown in Table 3. This was accomplished by our testing and refining items and strategies across two previous IOP versions, selecting and weighting the IOP–FDS items in the three current, independent, subsamples based on their efficacy and generalizability. Based on research designed to address depression, psychosis, post-traumatic reactions, and cognitive and neuropsychological conditions, as well as combinations thereof, the IOP–FDS should provide excellent coverage for disorders involved with forensic and high stakes determinations. This may be related to a conscious effort in the IOP development to address how complaints are presented rather than the symptoms themselves. Pending further research support, the IOP–FDS has potential to be used across a variety of situations, from homogeneous to diverse symptom presentations. This supports the utility and ecological validity of the IOP in real-life practice, given that malingerers typically present with a multitude of problems rather than with symptoms belonging to a single and specific diagnostic category. Further refinements and developments of the instrument, however, should probably attempt to also develop new diagnosis-specific scales, aimed at providing incremental validity over the current IOP–FDS. For instance, future IOP research could try to develop a depression-specific feigning scale by selecting or double-weighting those items that increment over the IOP–FDS when discriminating between genuine and feigned depression. The 64 IOP items that did not enter in the 117-item, IOP-FDS equation might be particularly suitable to this purpose, as they had received some empirical validation in past IOP research, but their detection efficacy likely does not generalize very well from one condition to another, given that they were not selected for the IOP-FDS.

Computerization of the IOP allows the recording and analysis of item latency patterns. In the subsample of Study 1, we found that certain latency patterns were associated with the participant belonging to the simulator versus the control groups. Within the independent

Comment [LG27]: R1, Issue 5: "Should the excluded items be deleted in further revisions of the IOP? Does the IOP have other intended uses beyond the identification of feigning mental disorders?"

Comment [LG28]: REPLY: We have clarified that the 64 items that did not take part in the IOP-FDS might be used in an attempt to develop diagnosis-specific scales. As for the other intended uses beyond the identification of feigning mental disorders, below we mention that the IOP might also contribute to investigate specific psychopathological symptoms, when the feigning scales suggest that the profile is rather credible, and for identifying feigning strategies.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

INTRODUCING THE IOP-FDS

20

Subsample 2, the so-derived Latency Feigning indicator demonstrated statistically significant, incremental validity over the Item-Based Feigning indicator. Given these results, and because it is based on within-test patterns rather than individual item latencies, we anticipate that the validity of this measure will generalize across samples and situations, regardless of an individual's average reading speed, for example. On the other hand, because the Latency Feigning indicator is used to calculate the IOP-FDS, a computer (or tablet) is evidently needed to administer the IOP and record reaction times, together with an administration and scoring software. Readers interested in conducting research with the 181-item IOP are thus encouraged to contact the first or second author of this article in order to obtain it.

In a meta-analysis of MMPI-2 scales in the detection of feigned mental disorders, after inspecting 65 feigning studies and 11 diagnostic studies, Rogers, Sewell, Martin, and Vitacco (2003) reported that F was the most effective scale, producing a Cohen's *d* effect size of 2.21, whereas Fb produced a *d* of 1.62. A recent meta-analysis of PAI scales conducted by Hawes and Bocaccini (2009) revealed that Cohen's *ds* were 1.48 and 1.59 for NIM, 1.15 and 1.00 for MAL, and 1.13 and 1.65 for RDF in uncoached and coached malingering studies respectively. In comparison, during our cross-validation research with Subsample 3, which consisted of a series of five clinical comparison simulation studies, the IOP-FDS produced a larger Cohen's *d* effect size of 2.61. Moreover, in our Study 3, with a small sample of 20 patients and 20 simulators, the IOP-FDS outperformed the Personality Assessment Inventory (PAI; Morey, 1991, 1996, 2007) scales: Negative Impression Management (NIM; Morey, 1996), Malingering Index (MAL; Morey, 1996), and Rogers Discriminant Function (RDF; Rogers, Sewell, Morey, & Ustad, 1996) scales. All in all, these findings lend initial support to the use of the IOP-FDS as a valid tool to detect feigned mental or cognitive problems across a variety of conditions and situations.

Comment [LG29]: R1, Issue 7: "How will the interested reader obtain the IOP for use in practice or research? Evidently, a computer is needed to administer the test, and probably to score it as well, but this issue is not addressed in the paper. Are the programs going to be made available through the authors?"

Comment [LG30]: REPLY: Good point. We have added this information in the text of the Discussion. Thanks.

INTRODUCING THE IOP-FDS

21

Although both the IOP-FDS and the IOP-29 yielded encouraging results, the longer IOP version might be preferable over the IOP-29 for a couple of reasons. First, in our independent, Cross-validation Subsample 3, the IOP-FDS significantly outperformed the IOP-29 ($r = .797$ vs. $r = .687$; $z = 3.518$, $p < .001$). Second, future research with the 181-item IOP version might allow for a deeper understanding of the strategies used by each test-taker. Indeed, a number of “descriptive scales” that aim to provide information on how an examinee with a high IOP-FDS score attempts to feign mental health or cognitive problems are currently under development. These include externalizing responsibility for problems, attributing exaggerated symptoms to a trauma, or poor effort on cognitive items. In addition, for those with low IOP-FDS feigning scores who appear to be presenting bona fide complaints, additional diagnostic scales and a measure of cognitive processing are being developed.

Future research might also attempt to address some of the limitations of the current studies. First, in our report, the validity of the IOP-FDS (as a measure of feigning) was tested with clinical comparison, simulation studies only. Additional field research with known-group comparisons of suspected malingers and bootstrapping comparisons are therefore needed to establish ecological validity (Rogers & Cruise, 1998). Moreover, the independent, cross-validation Subsample 3 was comprised of only 100 individuals, which is a modest sample size, raising questions concerning the generalizability of our findings. In this regard, however, it should be noted that the shrinkage from Subsample 2 to Subsample 3 was virtually nonexistent, which makes it reasonable to assume that the IOP-FDS will likely perform similarly well in future studies. Furthermore, only about a quarter of our participants suffered from or feigned cognitive or neuropsychological disorders, so additional research would be beneficial with this group. Lastly, the generalizability of our findings to non-U.S. contexts also deserves additional

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

INTRODUCING THE IOP-FDS

22

research. Accordingly, we are currently researching Italian, Chinese, German, Dutch, and Portuguese versions of the test. Regardless of these limitations, the IOP-FDS demonstrated strong validity in this study as a multipurpose test of feigning, with similar detection efficacy across a variety of clinical conditions without changing the recommended cutoff.

Comment [LG31]: R2, Issue 5: "The last phrase of the discussion says "without changing the recommended cutoff". It would be helpful to clarify what this means. Is there a recommended cutoff?"

Comment [LG32]: REPLY: We agree with the reviewer that that sentence was confusing. Different situations obviously require different considerations – and possibly different cut-offs – based on the cost-benefit associated with the specific case being assessed. As such, we have reworded the final sentence, not to enter into the details concerning the possible cut-off scores to be used in practice.

References

- Abramsky, A. B. (2005). *Assessment of test behaviors as a unique construct in the evaluation of malingered depression on the inventory of problems: Do test behaviors add significant variance beyond problem endorsement strategies?* Unpublished doctoral dissertation, California School of Professional Psychology, San Diego, CA.
- Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment*, 87, 84-94.
- Blanchard D. D., McGrath, R. E., Pogge, D L., Khadivi, A. (2003). A Comparison of the PAI and MMPI-2 as predictors of faking bad in college students. *Journal of Personality Assessment*, 80(2), 197-205
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.
- Connell, K. (2004). *Detecting simulated versus genuine posttraumatic stress disorder*. Unpublished doctoral dissertation, California School of Professional Psychology, San Diego, CA.
- Edens, J. F., Poythress, N. G., Watkins-Clay. M. M., (2007). Detection of Malingering in Psychiatric Unit and General Population Prison Inmates: A Comparison of the PAI, SIMS, and SIRS. *Journal of Personality Assessment*, 88(1):33-42, doi: 10.1080/00223890709336832
- Hawes, S., & Boccaccini, M.T. (2009). Detection of overreporting of psychopathology on the Personality Assessment Inventory: A meta-analytic review. *Psychological Assessment*, 21, 112-124. doi: 10.1037/a0015036

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

INTRODUCING THE IOP-FDS 24

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, Advance Online Publication, March 26*, doi: <https://doi.org/10.1007/s12207-018-9314-1>

Landis, P. E. (1996). *Detection of simulated posttraumatic stress disorder: A validation study of the Inventory of Problems*. Unpublished doctoral dissertation, California School of Professional Psychology, San Diego, CA.

McCullaugh, J. M. (2011). *The convergent and ecological validity of the Inventory of Problems with a community-supervised, forensic sample*. Unpublished doctoral dissertation, California School of Professional Psychology, San Diego, CA.

McDougall, A. (1996). *Rorschach indicators of simulated schizophrenia*. Unpublished doctoral dissertation, California School of Professional Psychology, San Diego, CA.

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin*, 52, 194-216.

Meng, X. L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, 111, 172-175.

Morey, L. (1991). *Professional manual for the Personality Assessment Inventory*. Odessa, FL: Psychological Assessment Resources

Morey, L. (1996). *An interpretive guide to the Personality Assessment Inventory (PAI)*. Odessa, FL: Psychological Assessment Resources.

Morey, L.C. (2007). *Personality Assessment Inventory Professional Manual*, second edition. Odessa, FL: Psychological Assessment Resources.

INTRODUCING THE IOP-FDS

25

- O'Brien, S. M. (2004). *An investigation into the incremental value of test-dependent malingering of schizophrenia*. Unpublished doctoral dissertation, California School of Professional Psychology, San Diego, CA.
- Pizitz, T. D. (2001). *Detection of malingered mild head injury using the tripartite conceptual model of malingering and the inventory of problems*. Unpublished doctoral dissertation, California School of Professional Psychology, San Diego, CA.
- Rogers, R. (1988). *Clinical assessment of malingering and deception*. New York: Guilford Press.
- Rogers, R. (1998). *Clinical assessment of malingering and deception* (2nd ed.). New York: Guilford Press.
- Rogers, R. & Bender, S. D. (2013). Evaluation of malingering and related response styles. In I. B. Weiner (Ed.-in-Chief), J. R. Graham & J. A. Naglieri (Vol. Eds.), *Comprehensive Handbook of Psychology: Assessment Psychology* (2nd Edition Vol. 10, pp. 517-540). Hoboken, NJ: John Wiley & Sons.
- Rogers, R. & Cruise, K. R. (1998). Assessment of malingering with simulation designs: Threats to external validity. *Law and Human Behavior*, 22 (3), 273-285.
- 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). Hoboken, NJ: Wiley.
- Rogers, R., Sewell K. W., Martin M. A., & Vitacco M. J. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI-2 and malingering. *Assessment*, 10(2):160-77.
- Rogers, R., Sewell, K. W., Morey, L. C., & Ustad, K. L. (1996). Detection of feigned mental disorders on the Personality Assessment Inventory: A discriminant analysis. *Journal of Personality Assessment*, 67, 629-640

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

INTRODUCING THE IOP-FDS 26

Streiner, D. L. (2003). Diagnosing Tests: Using and Misusing Diagnostic and Screening Tests. *Journal of Personality Assessment, 81*, 209-219.

Tombaugh, T.N. (1996). *Test of Memory Malingering*. Toronto, Canada: Multi-Health.

Viglione, D.J., Fals Stewart, W. & Moxham, E. (1995). Maximizing internal and external validity in MMPI malingering research: A study in a military population. *Journal of Personality Assessment, 65*, 502-513.

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, 5*, 534-544. doi.org/10.1080/00223891.2016.1233882

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, 5*, 534-544. doi.org/10.1080/00223891.2016.1233882

Viglione, D. J. & Landis, P. (1994). The Development of an Objective Test for Malingering. Paper presented at the biennial meeting American Psychology-Law Society, Santa Fe, New Mexico.

Viglione, D. J., Wright, D., Dizon, N. T., Moynihan, J. E., DuPuis, S., & Pizitz, T. D. (2001). Evading detection on the MMPI-2: Does caution produce more realistic patterns of responding? *Assessment, 8*(3), 237-250.

Wood, S. M. (2008). *Unique contributions of performance and self-report methods in the detection of malingered psychotic symptoms*. Unpublished doctoral dissertation, California School of Professional Psychology, San Diego, CA.

INTRODUCING THE IOP-FDS

27

For Peer Review Only

INTRODUCING THE IOP-FDS

Table 1. Composition of the Study Subsamples and Data Sources.

	Pizitz (2001) Head Injury	O'Brien (2004) Psychosis	Woods (2008) Psychosis	Connell (2004) PTSD	Abramsky (2005) Depression	McCullaugh (2011) Forensic Sample
Item Selection Subsample 1 (<i>n</i> = 341)						
Honest responders	21	24	25	22	24	35
Simulators	20	25	25	21	24	35
Total	41	49	50	43	48	70
Scaling Subsample 2 (<i>n</i> = 148)						
Honest responders	10	12	12	11	12	17
Simulators	10	12	12	11	12	17
Total	20	24	24	22	24	34
Cross-Validation Subsample 3 (<i>n</i> = 100)						
Honest responders	7	8	8	7	8	12
Simulators	7	8	8	7	8	12
Total	14	16	16	14	16	24
Composite Sample, all data (<i>N</i> = 549)						
Honest responders	38	44	45	40	44	64
Simulators	37	45	45	39	44	64
Total	75	89	90	79	88	128

INTRODUCING THE IOP-FDS

29

Table 2. Diagnostic Efficiency Statistics of the IOP-FDS within Validation Subsample 3 ($n = 100$)

Cut-off	Sensitivity	Specificity	<i>Base Rate = .50^a</i>			<i>Base Rate = .25</i>			<i>Base Rate = .10</i>			<i>Base Rate = .05</i>		
			PPP	NPP	OCC	PPP	NPP	OCC	PPP	NPP	OCC	PPP	NPP	OCC
0.05	1.00	0.16	0.54	1.00	0.58	0.28	1.00	0.37	0.12	1.00	0.24	0.06	1.00	0.20
0.10	1.00	0.34	0.60	1.00	0.67	0.34	1.00	0.51	0.14	1.00	0.41	0.07	1.00	0.37
0.20	0.98	0.56	0.69	0.97	0.77	0.43	0.99	0.67	0.20	1.00	0.60	0.11	1.00	0.58
0.50	0.90	0.80	0.82	0.89	0.85	0.60	0.96	0.83	0.33	0.99	0.81	0.19	0.99	0.81
0.80	0.78	0.96	0.95	0.81	0.87	0.87	0.93	0.92	0.68	0.98	0.94	0.51	0.99	0.95
0.90	0.58	0.98	0.97	0.70	0.78	0.91	0.88	0.88	0.76	0.96	0.94	0.60	0.98	0.96
0.95	0.44	1.00	1.00	0.64	0.72	1.00	0.84	0.86	1.00	0.94	0.94	1.00	0.97	0.97

^a Current sample base rate.

INTRODUCING THE IOP-FDS

30

Table 3. Validity of the IOP–FDS by Data Source (Cross-Validation Subsample 3, $n = 100$).

	Pizitz (2001) Head Injury ($n = 14$)	O’Brien (2004) & Woods (2008) Psychosis ($n = 32$)	Connell (2004) PTSD ($n = 14$)	Abramsky (2005) Depression ($n = 16$)	McCullaugh (2011) Forensic Sample ($n = 24$)
IOP – FDS					
Sensitivity	.86	.88	.86	1.00	.92
Specificity	.57	.75	.71	.88	1.00
PPP (base rate = .50)	.67	.78	.75	.89	1.00
NPP (base rate = .50)	.80	.86	.83	1.00	.92
OCC (base rate = .50)	.71	.81	.79	.94	.96
Correlation with group membership ^a	.69**	.70**	.73**	.93**	.96**
AUC	.88*	.91**	.88*	1.00**	1.00**

^a Point bi-serial correlation with group membership (dummy code: 0 = honest respondent; 1 = simulator); * $p < .05$; ** $p < .01$.

INTRODUCING THE IOP-FDS

31

Table 4. Comparative Validity of the IOP-FDS against the PAI ($n = 40$).

	Point bi-serial correlation with group membership	<i>AUC</i>
IOP – FDS	.92**	.99**
PAI – NIM	.63**	.85**
PAI – MAL	.35*	.70*
PAI – RDF	.54**	.83**

Note: This is a subgroup from Cross-Validation Subsample.

* $p < .05$; ** $p < .01$.

Figure 1. Distribution of IOP–FDS among Honest Respondents ($n = 50$) and Simulators ($n = 50$) of Cross-Validation Subsample 3

