Using the MMPI-2-RF, IOP-29, IOP-M, and FIT in the In-Person and Remote Administration Formats: A Simulation Study on Feigned mTBI

Accepted for publication: https://doi.org/10.1177/10731911241235465

Assessment

Using the MMPI-2-RF, IOP-29, IOP-M, and FIT in the In-Person and Remote Administration Formats: A Simulation Study on Feigned mTBI

Journal:	*Assessment*
Manuscript ID	ASMNT-23-0054.R2
Manuscript Type:	Original Research Article
Keywords:	MMPI-2-RF, IOP-29, Fifteen Item Test, mTBI, Teleassessment, malingering

Using the MMPI-2-RF, IOP-29, IOP-M, and FIT in the In-Person and Remote Administration Formats: A Simulation Study on Feigned mTBI

DETECTING FEIGNED MTBI

Abstract

Our study compared the impact of administering Symptom Validity Tests (SVTs) and Performance Validity Tests (PVTs) in in-person versus remote formats and assessed different approaches to combining validity test results. Using the MMPI-2-RF, IOP-29, IOP-M, and FIT, we assessed 164 adults, with half instructed to feign mild traumatic brain injury (mTBI) and half to respond honestly. Within each subgroup, half completed the tests in person, and the other half completed them online via videoconferencing. Results from 2 by 2 ANOVAs showed no significant effects of administration format on SVT and PVT scores. When comparing feigners to controls, the MMPI-2-RF RBS exhibited the largest effect size (d = 3.05) among all examined measures. Accordingly, we conducted a series of 2-step hierarchical logistic regression models by entering the MMPI-2-RF RBS first, followed by each other SVT and PVT individually. We found that the IOP-29 and IOP-M were the only measures that yielded incremental validity beyond the effects of the MMPI-2-RF RBS in predicting group membership. Taken together, these findings suggest that administering these SVTs and PVTs in-person or remotely yields similar results, and the combination of MMPI and IOP indexes might be particularly effective in identifying feigned mTBI.

Keywords: MMPI-2-RF; IOP-29; Fifteen Item Test; mTBI; Teleassessment; malingering.

DETECTING FEIGNED MTBI

The assessment of mild traumatic brain injury (mTBI) is a complex process that necessitates a thorough evaluation of the patient's symptoms and medical history (Bigler, 2014, 2015). In these assessments, it is crucial for assessors to consider the possibility that the examinee may exaggerate the severity of their problems or intentionally perform worse than their actual capability on neuropsychological tests or cognitive tasks (Sherman et al., 2020). Various methods can be employed for this purpose, including interviews, reviews of medical records, behavioral observations of the patient, and psychological testing. Regarding the latter, two types of tests are particularly valuable: Symptom Validity Tests (SVTs), which assess the credibility of self-reported psychological problems, and Performance Validity Tests (PVTs), which evaluate the credibility of observed performance on cognitive tasks (Larrabee, 2012).

In the assessment of mTBI, employing multiple SVTs and multiple PVTs is crucial for a comprehensive assessment of the credibility of neuropsychological profiles (Sherman et al., 2020). Indeed, as emphasized by Boone (2009), the presence of negative response bias in neuropsychological assessments varies over time, so assessors should monitor symptom and performance validity throughout their assessments. Furthermore, but relatedly, not everyone who exaggerates their psychological problems underperforms on cognitive tasks and vice versa (Shura et al., 2021). Indeed, although there is some overlap between symptom and performance validity, SVTs and PVTs often yield conflicting results (Giromini et al., 2020; Sabelli et al., 2021). From another perspective, this also means that the presence of credible results on one or more SVTs or PVTs should not be considered as evidence to counter or nullify the fact that one or more SVT and/or PVT results are in the noncredible range (Jennette et al., 2021). For all these reasons, the use of multiple SVTs and PVTs is likely to be beneficial to obtain a more comprehensive picture of the credibility of the mTBI-related problems presented.

Consistent with these considerations, the American Academy of Clinical Neuropsychology (AACN) recently released a consensus statement emphasizing that when evaluating the credibility of a given clinical or forensic presentation, multiple SVTs and multiple PVTs should be included in the assessment battery (Sherman et al., 2020; Sweet et al., 2021). In line with revised criteria for malingering by Sherman et al. (2020), this recent AACN consensus statement also underscored that the SVTs and PVTs selected for a given evaluation provide non-redundant information, i.e., that each validity test included in the assessment battery provide a unique contribution to the assessment of the credibility of presented psychological problems (Sherman et al., 2020; Sweet et al., 2021).[1] While this recommendation is important for maximizing incremental validity at the battery-wise level (Hunsley & Meyer, 2003), from a practical and applied perspective, a consequence of selecting non-redundant measures is that the SVTs and PVTs administered may provide discordant results. And unfortunately, to date there is very little agreement on how assessors should interpret the results of SVTs and PVTs that disagree with each other.

Combining the Results of Multiple SVTs and PVTs

Different authors have discordant opinions about how the results of multiple SVTs and PVTs should be considered to make a final determination on the overall credibility of a given presentation. Some believe that the professional should count the number of validity test failures (i.e., the number of SVTs and/or PVTs with results in the noncredible range), and if more than a given number of failures are observed, then the presentation should be considered noncredible. For example, Sherman, Slick, and Iverson (2020) recommend as a general indication that a presentation is likely invalid if two or more failures are observed; Larrabee, Rohling, and Meyers (2019) contend that a presentation is likely invalid if three or more failures are observed. In contrast to this position,

[1] More specifically, Sherman et al. (2020) proposed that validity tests that "tap the same item pool or consist of derived scores from the same items would not be considered independent" (p. 748).

other authors recommend adopting a more dimensional approach when assessing the overall credibility of presented problems. For example, Erdodi (2019) introduced the "EI model," an algorithm designed to capture both the frequency and the severity of validity check failure(s) by aggregating a number (typically 5 to 9) of individual PVTs into a single number summary of performance validity (see Cutler et al., 2022; Erdodi, 2021, 2022). In some agreement with this approach, Giromini, Young, and Sellbom (2022) recently opined that the number of failures is probably not that important when assessing symptom and/or performance validity, as it is likely to be the psychometric quality of the SVTs and PVTs administered that matters most. A few years earlier, Chafetz (2020) had also suggested that the number of failures is probably not that important, but he put forth a substantially different reason, namely that when people lie, they do not necessarily lie about everything they say, so one validity check failure might be enough to invalidate a given clinical presentation (Chafetz, 2020).

To complicate matters further, note that different contexts may require different guidelines and different procedures for integrating the results of multiple symptom and performance tests. For example, SVTs are typically used and considered more important when assessing the credibility of presented symptoms of various psychiatric disorders such as depression, post-traumatic stress disorder (PTSD), or schizophrenia (Fox & Vincent, 2020; Giromini, Viglione, et al., 2020). Conversely, PVTs are typically used and considered more important when assessing the credibility of presented cognitive or neuropsychological problems, such as in the case of mild traumatic brain injury (mTBI) or attention-deficit hyperactivity disorder (ADHD) (De Boer et al., 2022; White et al., 2012, 2022). As such, the results of SVTs may be more relevant when assessing the credibility of presented psychiatric symptoms, whereas PVTs may be more informative when assessing the credibility of brain damage and/or neuropsychological problems. Nonetheless, there is evidence that both types of validity checks could be useful in either of these two broad assessment contexts

(Banovic et al., 2022; Gegner et al., 2020; Green et al., 2012; Giromini, Viglione, et al., 2020; Pivovarova et al., 2009; Sabelli et al., 2021), and yet there is still little guidance on how to optimally integrate the results of multiple SVTs and PVTs administered in either context.

Assessing Symptom and Performance Validity in Teleassessment

While telemedicine was introduced by Dwyer (1973) in the 1970s, telepsychology and teleneuropsychology have been effectively used since the 1980s, well before the COVID-19 pandemic outbreak (e.g., Batastini et al., 2020; Cullum & Grosch, 2013; Cullum et al., 2014). Nevertheless, the use of telepsychology by practitioners increased during the COVID-19 pandemic (e.g., Bernhard et al., 2021; Daffern et al., 2021; Hammers et al., 2020). Indeed, following the spread of COVID-19, a number of guidelines were published to help psychologists navigate the uncertain terrain of physical distancing (e.g., American Psychological Association, 2020; Bilder et al., 2020; Chenneville & Schwartz-Mette, 2020; Farmer et al., 2020; Pliskin et al., 2020; Wright & Raiford, 2021; Wright et al., 2020), and, accordingly, test manufacturers offered guidelines on how to administer their tests remotely and implement them on online platforms (e.g., Pearson and PAR). While some authors expressed a general caution about the use of remote psychological assessment in forensic settings where the stakes are high (Corey & Ben-Porath, 2020; Goldenson & Josefowitz, 2021), others suggested that it is unlikely that a *Daubert* challenge would exclude evidence collected via remote assessment, as long as practitioners acknowledge potential technical and practical challenges associated with remote assessment (Heilbrun, 2022; Recupero, 2022).

Research in telepsychology has evolved over the past 30 years, and there is currently a fair amount of research on the psychometric equivalence of computer-based and paper-and-pencil administration formats for some widely used measures such as the Minnesota Multiphasic Personality Inventory – Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008/2011) (e.g., Finger & Ones, 1999; Forbey & Ben-Porath, 2007; Menton et al., 2019; Pinsoneault, 1996; Roper et

al., 1995). However, this research has mainly been conducted using the in-person administration method, while very few studies have investigated whether administering the same test(s) on site (in-person) versus remotely leads to significantly different results. Indeed, the question of whether assessing symptom and performance validity in person versus remotely yields significantly different results has been addressed only sparingly (e.g., Agarwal et al., 2023; Giromini, Pignolo, et al., 2021; Reeves et al., 2022). Thus, it is probably only a matter of time before the courts will engage in a legal debate over whether teleassessment constitutes a significant and problematic deviation from standard testing protocols (Carroll, 2020; Drogin, 2020; Heilbrun, 2022; Kois et al., 2020; Levy, 2020; Recupero, 2022). And this issue is particularly relevant at this time because, although the practice of teleassessment has been fueled by the recent spread of the COVID-19 pandemic, it would be unrealistic to believe that it will cease once the pandemic has fully subsided (Drogin, 2020; Heilbrun, 2022).

This Study

Our study aimed to make a dual contribution to the literature on the use of SVTs and PVTs, pursuing two objectives of equal significance. First, given the increasing prevalence of reliance on teleassessment in forensic psychology (Drogin, 2020; Heilbrun, 2022), this study aimed to investigate whether administering a series of SVTs and PVTs in person *versus* remotely would produce significantly different results. Additionally, due to the limited understanding of how to effectively combine and integrate results from multiple SVTs and PVTs when assessing mTBI-related presentations, our second goal was to investigate the battery-wise effectiveness of different approaches to combining the results of multiple SVTs and PVTs. More in detail, this study sought to contribute to the empirical literature by identifying the SVT/PVT with the best predictive power for assessing the credibility of presented symptoms of mTBI and evaluating the incremental validity of adding other selected SVTs/PVTs, one at a time, in a regression model. Lastly, we also explored

several algorithms that could be used to integrate the results of multiple SVTs and PVTs when assessing the overall credibility of the clinical presentation.

Method

Below we report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. Interested researchers may contact the first author to receive the data set used for the statistical analyses.

Participants

The research project was reviewed and approved by the Institutional review board (IRB) of the University of Turin, Italy. Eligibility criteria required that participants were able to read and sign an informed consent form written in Italian, were at least 18 years old, did not have a traumatic brain injury and had not received a diagnosis of it in the past, did not have a psychiatric disorder and had not received a diagnosis of it in the past, were not enrolled in and did not have a psychology degree. All these inclusion criteria were explicitly stated in the flyers and social media advertisements inviting potential participants to volunteer and then reviewed with our research assistants when data collection began. Additionally, participants were informed of the voluntary nature of the study, and each participant was required to sign an informed consent form before participation. Exclusion criteria were limited to instances of careless or random responding in the MMPI-2-RF, specifically defined as CNS ≥ 15 (raw), VRIN-r ≥ 80T, and/or TRIN-r ≥ 80T.

Simulation studies have widely varying sample sizes, typically ranging from a minimum of 20 to 30 cases to a few hundred cases per group. In addition, a power analysis revealed that a sample size of 128 participants was required to be able to detect a medium-size effect ($f = 0.25$) for the interaction between condition (control *versus* mTBI feigner) and administration format (in-person *versus* remote). Thus, because we anticipated that approximately 20% to 30% of recruited participants might provide invalid data, we aimed for a minimum number of 40 cases per subgroup

(i.e., at least 40 in-person controls, 40 remote controls, 40 mTBI in-person feigners, and 40 remote mTBI feigners), with the goal of obtaining an initial sample of at least 160 participants in total. Following these general guidelines and applying the inclusion and exclusion criteria, the research assistants (RAs) responsible for recruiting the participants assembled a final sample of 164 individuals, divided into four subgroups: 41 controls and 42 experimental mTBI feigners with the in-person administration, and 41 controls and 40 experimental mTBI feigners with the remote administration. Table 1 shows the demographic composition of the sample.

[Insert Table 1 here]

Procedure

Potential participants were recruited through advertising the study on social media, distributing flyers at various locations on the university campus, utilizing word of mouth, and employing snowball sampling. Half of the participants were instructed to meet with our RAs via video conference (e.g., using Zoom or Webex), while the other half were directed to meet the RA in a private room (e.g., in a library or in a room designated for running experiments on the university campus). For the participants who met via video conference, the RAs were asked to determine if there were any problems with the connection immediately after the call began – however, no such problems were observed in any case.

After participants gave their written informed consent and confirmed that they met all eligibility criteria, they were then randomly assigned to either the mTBI feigner group or the control group. Participants in the mTBI feigner group were asked to put themselves in the position of a person who had been hit by another car while driving and was suffering from mTBI as a result of this car accident. They were told that because the other driver was at fault in the accident, their insurance company would cover the cost of any physical damage and give them additional money if the psychological tests administered confirmed that they were actually suffering from mTBI. To help

them feign in a credible manner, they were given a list of neuropsychological symptoms that characterize the condition of mTBI and informed that their presentation would not be considered credible if they described their symptoms in an overly dramatic manner.[2] In addition, to motivate them to feign mTBI credibly, they were also informed that the best feigner, i.e., the person who most credibly portrayed the symptoms of mTBI,[3] would receive a shopping voucher worth 30 euros. As for the control group, participants in this condition were simply asked to answer all psychological tests truthfully and honestly. To encourage attentive and active participation, participants in the control group were informed that a randomly selected participant in this group would receive a shopping voucher worth 30 euros at the end of the study.

Of the 179 participants who had initially indicated their intent to participate to the study, four were excluded from the sample because they were not actually eligible to participate in this study, six showed careless or random responding in the MMPI-2-RF, three had technical problems while taking the one of the instruments (i.e., the Fifteen Item Test; FIT), and two failed the post-test manipulation check (see below), so their data were ultimately excluded from data analysis, resulting in a final valid sample size of 164. More specifically, 82 were assigned to the control condition and 82 to the mTBI feigner condition. Of the 82 controls, 41 completed the tests online (remotely) and 41 completed them on-site (in-person); of the 82 who feigned mTBI, 40 completed the tests online (remotely) and 42 completed them on-site (in-person).

All participants were then administered (a) the MMPI-2-RF, (b) the Inventory of Problems – 29 (IOP-29; Viglione & Giromini, 2020) with its memory module (IOP-M; Giromini et al., 2020), and (c) the FIT (both its free recall and its recognition trials; Boone, 2002; Lezak, 1995; Rey, 1941),

[2] To avoid providing resources that could be utilized to engage in ethically questionable behavior, the specific instructions employed in this study for feigning mTBI are not publicly disclosed. However, they can be obtained upon reasonable request by contacting the corresponding author.

[3] Operationalized as having the lowest average Z-sum of all SVT and PVT results, calculated after multiplying individual PVT results by -1

in a randomized order. Before the administration of each test, each participant was reminded what to do during the test, i.e., whether to feign an mTBI or answer honestly. After completing all tests, participants in the experimental group (mTBI feigner group) were asked to give up the role of a person with an mTBI. At this point, all participants were given a demographic form and a post-test manipulation check that asked them whether they had answered honestly or feigned an mTBI while completing the tests. As mentioned earlier, two individuals did not answer this post-test manipulation check correctly, so they were excluded from the analyses.

It should be noted that, except for the FIT, which was presented in a custom-built PDF format, all tests were administered and scored according to standard procedures, utilizing the official administration and scoring platforms. For the remote administration subgroups, this means that participants were constantly monitored using a synchronous (live) teleconferencing application so as to comply with existing guidelines for conducting psychological teleassessments (e.g., Drogin, 2020; Menton, Corey, & Ben-Porath, 2022).

Measures

Minnesota Multiphasic Personality Inventory-2 Restructured Form (MMPI-2-RF). The MMPI-2-RF (Ben-Porath & Tellegen, 2008/2011) is a widely used (Neal & Grisso, 2014) broadband personality inventory consisting of 338 true-false items designed to assess personality and psychopathology. Five of the nine validity scales are intended to assess negative response bias or overreporting. These five SVTs embedded in the MMPI-2-RF are labelled Infrequent Responses (F-r), Infrequent Psychopathology Responses (Fp-r), Infrequent Somatic Responses (Fs), Symptom Validity Scale (FBS-r), and Response Bias Scale (RBS). The F-r and Fp-r measure overreporting of general psychopathology; the Fs measures overreporting of somatic symptoms; the RBS measures overreporting of memory complaints; and the FBS-r reflects overreporting of unusual combination(s) of noncredible cognitive and somatic symptoms. Two recent meta-analytic studies

have shown that the five embedded SVTs of the MMPI-2-RF are highly effective in discriminating credible from noncredible/feigned cognitive impairment, mental illness, and medical complaints in both simulation and criterion-group study designs (Ingram & Ternes, 2016; Sharf et al., 2017). As the MMPI-3 is not available in Italy yet, we opted for the Italian version of the MMPI-2-RF, distributed by Giunti Psychometrics. Its inclusion in our study stems from the widely acknowledged and empirically supported use of its embedded SVTs (Burchett & Bagby, 2022; Giromini et al., 2022).

The Inventory of Problems – 29 (IOP-29). The IOP-29 (Viglione & Giromini, 2020), with only 29 items, is the shortest of the currently available, empirically researched, and psychometrically sound free-standing SVTs (Giromini, Young et al., 2022). The responses that the examinee gives to these 29 items are processed by an algorithm to generate the False Disorder probability Score (FDS), the chief feigning index of the IOP-29. The higher the FDS, the lower the credibility of the presentation. The IOP-29 was introduced only relatively recently, in 2017 (Viglione, Giromini, & Landis, 2017). Nevertheless, its validity and effectiveness have already been demonstrated in numerous countries, summarized in a quantitative literature review (Giromini, Young et al., 2022) and a bivariate diagnostic test accuracy meta-analysis (Puente-López et al., 2023a). Both these review articles concluded that the IOP-29 is a highly effective tool for discriminating between credible and noncredible symptom presentations, although additional research on its effectiveness in criterion-group studies would be beneficial. In our research, we utilized the Italian adaptation of the IOP-29, which is accessible on the test website (https://www.iop-test.com/) and has been thoroughly examined in Italian contexts (e.g., Roma et al., 2023). The selection of the IOP-29 was guided by its apparent psychometric superiority over the potentially more commonly used Structured Inventory of Malingered Symptomatology (SIMS; Smith & Burger, 1997), as shown in multiple studies (Boskovic et al., 2020; Giromini et al., 2018; Puente-López et al., 2023b).

The Inventory of Problems – M (IOP-M). The IOP-M (Giromini, Viglione, et al., 2020) is a 34-item PVT designed to be administered immediately after the IOP-29 to assesses the credibility of a memory-related performance, so that the higher the number of correct responses, the higher the credibility of the performance. To date, relatively few studies have examined the IOP-M, and most of them used a simulation design and did not include cognitively impaired or mentally ill controls. However, five simulation studies as well as a criterion-group study found some promising results on the validity and effectiveness of the IOP-M (Banovic et al., 2021; Bosi et al., 2022; Carvalho et al., 2021; Erdodi et al., 2023; Gegner et al., 2021; Holcomb et al., 2022; Šömen et al., 2021). For this study, we employed the Italian version of the IOP-M, accessible from the official test website (https://www.iop-test.com/). Its inclusion was motivated by its utility in providing a swift assessment of performance validity when used in conjunction with the IOP-29. Furthermore, unlike other established PVTs like the Test of Memory Malingering (TOMM; Tombaugh, 1996) or Word Memory Test (WMT; Green, 2003), which lack validation and are not commercially available in Italy, the IOP-M has undergone extensive investigation within the Italian context (Giromini et al., 2020).

The Fifteen Item Test (FIT). The FIT (Lezak, 1995; Rey, 1941) is one of the oldest free-standing PVTs for discriminating between credible and noncredible memory-related complaints. It presents the examinee with a visual input containing 15 symbols and a request to memorize them. The more symbols that are correctly reproduced, the higher the credibility of the performance. In the late 1990s, numerous research studies indicated that the FIT had excellent specificity (when studies excluded cases with intellectual disability or dementia) but relatively low sensitivity (for a meta-analysis, see Reznek, 2005). Therefore, to improve sensitivity, Boone et al. (2002) introduced an additional recognition trial, yielding a measure of the overall credibility of the performance (see, for example, Green et al., 2016). The higher this value, the higher the credibility of the performance. We included the FIT in our study because it is the only freely available PVT in the Italian context,

and because despite its psychometric weaknesses, it is commonly used in real-world assessments, both in Italy (Giromini, Pasqualini, et al., 2022) and elsewhere (e.g., Sullivan et al., 2006; White et al., 2016).

Data Analysis

We performed a series of 2 by 2 ANOVAs with condition (control *versus* mTBI feigner) and administration format (in-person *versus* remote) as between-subject factors and the relevant MMPI-2-RF, IOP-29, IOP-M, and FIT scores as dependent variables. Next, we examined the classification accuracy of all these measures by considering a number of a-priori determined cutoff scores. Specifically, for the MMPI-2-RF and IOP-29, we considered the range of cutoff scores analyzed in Sharf et al. (2017; see Table 6, p. 451) and Viglione and Giromini (2020; see Table 3, p. 47), respectively. For the IOP-M and FIT, the standard cutoff scores suggested by Giromini, Viglione, et al., (2020) and Boone et al. (2002), respectively, were examined.

A series of 2-step hierarchical logistic regressions were then performed to assess incremental validity. More specifically, the dependent variable in these analyses was group membership (0 = control; 1 = mTBI feigner); the scale that proved most effective in distinguishing the experimental from the control group was entered in the first step; and each of the other measures, one at a time, was entered in the second step. The purpose of these analyses was to determine which of the other measures might provide incremental validity when entered after the most effective score of the battery under study. In real-world contexts, the outcomes of these analyses could guide the selection of individual components for an assessment battery, especially when validity testing needs to be conducted under volume pressures and within a tight timeline. Additionally, from a statistical perspective, the choice to enter one variable at a time, rather than all simultaneously, aimed at mitigating collinearity and related issues. Finally, we examined the battery-wise effectiveness of a

variety of different approaches to combining the results of the SVTs and PVTs included in the current study.

Results

Sample composition and preliminary analyses. Before conducting the planned statistical analyses, we evaluated whether there were differences in the demographic variables between the four subgroups. The results of two 2 by 2 ANOVAs with condition (control *versus* mTBI feigner) and administration format (in-person versus remote) as between-subjects factors indicated that the four subgroups were well balanced in terms of age (main effect of condition: $F_{(1, 160)} = 1.05$, $p = .31$; main effect of administration format: $F_{(1, 160)} = 0.02$, $p = .89$; interaction effect: $F_{(1, 160)} = 0.39$, $p = .53$) and years of education (main effect of condition: $F_{(1, 160)} = 0.97$, $p = .33$; main effect of administration format: $F_{(1, 160)} = 0.19$, $p = .66$; interaction effect: $F_{(1, 160)} = 1.01$, $p = .32$). Similarly, gender did not differ across the four subgroups, $\chi^2_{(3)} = 6.29$, $p = .10$.

Administration Format and Detection of Overreporting. Next, we tested the effectiveness of the selected measures to detect feigning (control *versus* feigning) and to perform similarly in different administration settings (in-person *versus* remote). Results of a series of 2 by 2 ANOVAs revealed that there were no significant interaction effects (condition by administration format) for any of the scores of the MMPI-2-RF ($F_{(1,160)} \leq 1.18$, $p \geq .28$), of the two IOP modules ($F_{(1,160)} \leq 2.08$, $p \geq .15$), and of the two FIT trials ($F_{(1,160)} \leq 0.03$, $p \geq .87$). The main effects of administration format were also not significant for any of these scores ($F_{(1,160)} \leq 3.55$, $p \geq .06$). In contrast, the main effects of condition were statistically significant for all scores in the analyses ($F_{(1,160)} \geq 22.86$, $p < .001$), as expected. Descriptive statistics for all scores examined are presented in Table 2, subdivided by condition (control *versus* mTBI feigner) and administration format (in-person *versus* remote).

[Insert Table 2 here]

Cohen's *d* and *AUC* values for the main effects of condition are reported in Table 3. The most effective indicator was the MMPI-2-RF RBS (*d* = 3.05, *AUC* = .97), followed by the IOP-29 FDS (*d* = 2.58, *AUC* = .95), the MMPI-2-RF FBS-r (*d* = 2.37, *AUC* = .95), the IOP-M (*d* = 2.20, *AUC* = .92), and the MMPI-2-RF Fs (*d* = 2.04, *AUC* = .93). The other indicators yielded relatively less optimal results. Indeed, according to Rogers et al.'s (2003) characterization of Cohen's *d* values for simulation studies, all these indicators achieved "*very large*" (i.e., ≥ 1.75) effect sizes. In contrast, the MMPI-2-RF F-r achieved a "*large*" (i.e., ≥ 1.25) effect size, the MMPI-2-RF Fp-r and the combined score of the FIT (recall & recognition) achieved "*moderate*" (i.e., ≥ .75) effect sizes, and the free recall score of the FIT yielded a less optimal effect size of *d* = .74.

[Insert Table 3 here]

Classification Accuracy. Subsequently, we examined classification accuracy of all the selected measures. Because there were no significant interaction effects between condition and administration format, classification accuracy analyses were conducted after combining the data obtained from the in-person and the remote administration formats. Sensitivity, specificity, and overall correct classification values for all a-priori identified cut scores are presented in Table 4. Considering that the control group consisted of nonclinical individuals, it should not surprise that the specificity was above .90 for all inspected cut scores (Giromini et al., 2022; Sweet et al., 2021), with the exception of the most liberal, screening cut score of the IOP-29, i.e., FDS ≥ .30, which yielded a specificity of .85. As for sensitivity, it was excellent (i.e., > .70; Erdodi et al., 2014; Vickery et al., 2001) for the most liberal (≥ 80) cut score of the MMPI-2-RF RBS, for the standard (≥ .50) and liberal (≥ .30) cut scores of the IOP-29 FDS, and for the standard cut score of the IOP-M (< 30). It was more than satisfactory (i.e., between .50 and .70; Erdodi et al., 2014; Vickery et al., 2001) for the MMPI-2-RF cut offs of Fs ≥ 90, Fs ≥ 80, FBS-r ≥ 80, and RBS ≥ 90, and for the conservative (≥ .65) cut score of the IOP-29 FDS. Although the sensitivity of the FIT increased

from .16 to .44 when considering not only its free recall but also its recognition trial, the FIT showed less optimal sensitivity results overall compared with the symptom validity indicators of the MMPI-2-RF, IOP-29, and IOP-M.

[Insert Table 4 here]

Incremental Validity. Finally, we evaluated which measures might provide incremental validity when entered after the most effective score of the battery and how those measures can be implemented together in detecting non-credible mTBI presentations. Because the MMPI-2-RF RBS was the most effective scale for discriminating the experimental from the control group, we then performed a series of hierarchical logistic regression analyses aimed at predicting group membership (0 = control; 1 = mTBI feigner) by entering the MMPI-2-RF RBS in the first step, and each of the other measures, one at a time, in the second step. The results of these analyses showed that the IOP-29 FDS and the IOP-M were the only scores that significantly added incremental validity to the models after entering the MMPI-2-RF RBS in the first step (Table 5). Accordingly, we explored multiple combinations of scores to detect feigning of mTBI symptoms .

[Insert Table 5 here]

Based on Table 4, for the MMPI-2-RF RBS, the a-priori selected cut score that maximized overall correct classification was ≥ 80; for the IOP-29 FDS, it was ≥ .50. For the IOP-M, < 30 is the only available cut score, at this time. As such, to determine whether a given presentation is valid/credible or invalid/noncredible, nine possible combinations of results using these particular cut scores were examined. These nine approaches / invalidity determination rules and their associated battery-wise classification accuracy are presented in Table 6. Two solutions yielded the same and the highest overall correct classification value (.95). One sets that a presentation is invalid/noncredible if the MMPI-2-RF RBS ≥ 80T and/or both the IOP scores are in the noncredible range (i.e., FDS ≥ .50 and IOP-M < 30). In other words, the highest classification

accuracy is attained when a presentation is deemed noncredible based on a noncredible score on the MMPI-2-RF RBS alone, noncredible results on both IOP tests, or the simultaneous noncredible scores on both the MMPI-2-RF RBS and IOP tests. The other solution with the highest overall correct classification value (.95) sets that a presentation is invalid/noncredible if the IOP-M < 30 and/or both the MMPI-2-RF RBS ≥ 80T and IOP-29 FDS ≥ .50. That is, the symptom presentation should be deemed as noncredible when the IOP-M is invalid, when both the MMPI-2-RF RBS and IOP FDS are invalid, or when all the three measures are simultaneously invalid.

[Insert Table 6 here]

Discussion

Currently, little is known about whether administering a validity test on-site (in-person) or remotely will produce significantly different results. In addition, there is little agreement on how the assessor should ideally combine the results of multiple SVTs and PVTs. To address these gaps, we administered various SVTs and PVTs to 164 adults, with half feigning mTBI and half responding honestly, either in person or online. The results of our statistical analyses can be summarized as follows: First, no significant effects of administration format on SVT and PVT scores were found. Second, the MMPI-2-RF Response Bias Scale (RBS) showed the largest effect size (d = 3.05) among all SVTs and PVTs in the study when comparing simulators and controls. Third, subsequent regression analyses showed that only the IOP-29 and the IOP-M exhibited incremental validity beyond the MMPI-2-RF RBS in predicting group membership.

Using SVTs and PVTs in Teleassessment

To our knowledge, this study is the first to demonstrate the equivalency of the MMPI validity scales in detecting feigned mTBI when administered on-site (in person) or remotely (via videoconference) with participants from the general population. Indeed, while numerous studies indicated the equivalence of in-person paper-and-pencil and in-person computerized administration

of the MMPI (Finger & Ones, 1999; Forbey & Ben-Porath, 2007; Menton et al., 2019; Pinsoneault, 1996; Roper et al., 1995), only two recent studies, Agarwal et al. (2023) and Reeves et al. (2022), specifically explored the effects of remote *versus* in-person administration. However, Agarwal et al. (2023) focused on students, and Reeves et al. (2022) did not test the interaction effect between condition and administration format, as individuals in the control (honest) condition completed the MMPI-3 in person. Thus our study extends existing MMPI literature by examining the interaction effect and analyzing data from the general population, rather than a student sample.

In a similar vein, regarding the IOP-29, a previous study found comparable effectiveness in detecting feigned mental health problems when administered remotely rather than in person (Giromini, Pignolo, et al., 2021). However, participants in the remote administration condition in that study completed the IOP-29 without direct contact, such as via videoconference, with the examiner. Moreover, they had not been administered the IOP-M or the FIT. Therefore, our study also contributes to the existing IOP and FIT literature by investigating the IOP-29, IOP-M, and the FIT, and by ensuring constant participant monitoring through synchronous (live) teleconferencing applications in alignment with established guidelines for conducting psychological teleassessments (e.g., Drogin, 2020; Menton, Corey, & Ben-Porath, 2022).

Taken together, the absence of differences between remote and in-person administration formats for any of the examined SVTs and PVTs holds particular relevance in the current landscape, given the anticipated rise in the popularity of forensic teleassessment in the coming years (Drogin, 2020; Heilbrun, 2022; Menton et al., 2022; Wright et al., 2020).

Combining the Results of Multiple SVTs and PVTs

Regarding the effectiveness of administered SVTs and PVTs in the classification of symptom and performance validity, the current study is consistent with accumulating evidence that the validity scales of the MMPI-2-RF are very good at detecting feigned mTBI. In fact, among all SVTs and

PVTs examined in this study, the most effective indicator was the MMPI-2-RF RBS (d = 3.05). The other measures that yielded very large effect sizes were the IOP-29 FDS (d = 2.58), the MMPI-2-RF FBS-r (d = 2.37), the IOP-M (d = 2.20), and the MMPI-2-RF Fs (d = 2.04). Remarkably, three SVTs (i.e., MMPI-2-RF RBS, IOP-29 FDS, and MMPI-2-RF FBS-r) achieved larger effect sizes than the most effective PVT (i.e., IOP-M), challenging the traditional emphasis on PVTs over SVTs in assessing the credibility of cognitive or neuropsychological problems (De Boer et al., 2022; White et al., 2012, 2022). Of course, it is possible that the results would have been different if we had included other, more empirically consolidated and perhaps more effective PVTs such as the TOMM, the WMT, the Word Choice Test (WTC; Pearson, 2009), or the Victoria Symptom Validity Test (VSVT; Slick, Hopp, Strauss, & Thompson, 1997). However, given that a recent criterion-groups study found that the IOP-M outperformed the TOMM in detecting performance invalidity (Holcomb et al., 2022), these results still suggest that certain SVTs could be at least as effective as some PVTs in identifying noncredible mTBI symptoms.

The results summarized in Table 5 and Table 6 show that the MMPI-2-RF RBS could be particularly useful when interpreted together with the IOP-29 and IOP-M. To some extent, this finding is consistent with current recommendations to use validity tests that are not overly redundant with each other when assessing symptom and performance validity (Sherman et al., 2020; Sweet et al., 2021). Indeed, the RBS has some item overlap with the other overreporting scales of the MMPI-2-RF (Burchett & Bagby, 2022). In contrast, the IOP-29 was designed and developed with the very purpose to look at symptom validity from a different perspective than the MMPI overreporting scales, so it is unlikely to be overly redundant with them (Giromini, Carfora Lettieri et al., 2019; Viglione et al., 2017). Similarly, as the IOP-M focuses on performance rather than symptom validity, the IOP-M is also unlikely to be overly redundant with the MMPI-2-RF overreporting scales (De Boer et al., 2023; Obolsky et al., 2023). In line with these considerations,

Table 5 results indicate that only the IOP-29 and the IOP-M demonstrated statistically significant incremental validity when added after the MMPI-2-RF RBS in logistic regression models predicting group membership. Other MMPI-2-RF overreporting scales and the FIT did not achieve statistically significant incremental validity beyond the effects of the MMPI-2-RF RBS.

In summary, these findings suggest that the MMPI-2-RF RBS emerges as the most effective measure for detecting feigned mTBI, and the only validity tests that could enhance battery-wise precision when considered in combination with the MMPI-2-RF RBS score were the IOP instruments. Looking ahead, considering the substantial similarity between the validity scales of the MMPI-2-RF and those of the MMPI-3, it is highly likely that these considerations will also apply to future studies using the increasingly popular MMPI-3. Indeed, the correlations between the validity scales in the MMPI-2-RF and their counterparts in the MMPI-3 are in the very high range, with $r \geq .95$ (Ben-Porath & Tellegen, 2020), and the scale generating the largest effect size in this study, the RBS, contains the exact same 28 items found in the MMPI-3 counterpart (Burchett & Bagby, 2022).

From a broader perspective, our findings underscore the importance for professionals to assess both symptom and performance validity using multiple SVTs and PVTs when evaluating the credibility of mTBI-related clinical presentations (Sherman et al., 2020; Sweet et al., 2021). While the presence of a score in the credible range for one or more validity tests should not be interpreted as conclusive evidence of the presentation's credibility – given the variability of negative response bias over time and across psychological domains (Boone, 2009) – equally, a single result in the noncredible range is likely insufficient to declare the profile invalid (Giromini et al., 2022). Thus, to reconcile conflicting results from different tests, assessors must adopt a comprehensive approach, examining the entire file. Indeed, since different measures offer unique perspectives on the same individual (Sabelli et al., 2021), this holistic evaluation is essential for evidence-based and empirically sound decisions. To some extent, this conclusion aligns with Messerly et al. (2021), emphasizing that

combining results from multiple validity tests enhances classification accuracy, provided these tests within the battery exhibit strong psychometric properties (which may not be the case for the FIT). And as a general heuristic, it appears that many professionals are relatively confident in concluding that a clinical presentation is not credible only when two or more validity tests of the same type show results in the noncredible range (e.g., two or more SVTs or two or more PVTs).

Final Remarks

In considering the results of the present study, however, some limitations should be highlighted. First, and most importantly, because our study did not include a clinical sample of patients with mTBI, the results presented in this article artificially inflate estimates of specificity (Erdodi, Green, et al., 2019; Giromini, Young, & Sellbom, 2022; Rogers & Benders, 2018): In real-world patient samples, it is likely that all validity tests examined in this study will show less positive results. Second, as with all simulation studies, the ecological validity of our study may be questioned, as there is no guarantee that real-world feigners use the same strategies to feign mTBI used by our healthy participants who were experimentally asked to overreport the symptoms (Rogers & Benders, 2018). Third, since our study was conducted in Italy, and our participants had a relatively high level of average education, the generalizability of our findings to other cultural contexts and individuals with lower education remains an empirical question that requires further research. Fourth, as mentioned previously, performance validity was assessed in this study by using a PVT that has often been criticized for its suboptimal sensitivity, i.e., the FIT (Bailey et al., 2018; Morse et al., 2013; Reznek, 2005; Whitney et al., 2008), and another PVT that is relatively new and thus has been poorly researched so far, i.e., the IOP-M. Additional research using more consolidated PVTs such as the TOMM, WMT, WCT, or VSVT would therefore be beneficial. Fifth, most of our eligibility criteria, such as those requiring that participants had not experienced an mTBI, had not received a diagnosis

of a psychiatric disorder, and were not enrolled in a psychology degree, relied on self-reported information that could not be verified with certainty.

Despite these and possible other limitations, our study nevertheless has the merit of (a) being the first to empirically investigate whether administration of the MMPI-2- RF, IOP-29, IOP-M, and FIT in person or remotely leads to significantly different results and (b) providing some valuable information on the battery-wise effectiveness of different approaches to combining the results of multiple SVTs and PVTs. Although additional research replications using newer (e.g., the MMPI-3 rather than the MMPI-2-RF) and more empirically consolidated (e.g., the TOMM rather than the FIT) measures would be beneficial in the future, the current study shows that the examined SVTs and PVTs should operate similarly well when used in person versus remotely, and that the combination of the MMPI with the IOP might be particularly effective in detecting feigned mTBI.

References

Agarwal, L. P., Keen, M. A., Morris, C. S., & Ingram, P. B. (2023). Contrasting MMPI-3 validity scale effectiveness differences across in-person and telehealth administration procedures. *Psychological Assessment, 35*(11), 925–937. https://doi.org/10.1037/pas0001258

American Psychological Association. (2020). Telehealth guidance by state during COVID-19. Retrieved May 1, 2023 from https://www.apaservices.org/practice/clinic/covid-19-telehealth-state-summary

Bailey, K. C., Soble, J. R., & O'Rourke, J. J. F. (2018). Clinical utility of the Rey 15-item Test, recognition trial, and error scores for detecting noncredible neuropsychological performance in a mixed clinical sample of veterans. *The Clinical Neuropsychologist, 32(1)*, 119–131. https://doi.org/10.1080/13854046.2017.1333151

Banovic, I., Filippi, F., Viglione, D.J., Scrima, F., Zennaro, A., Zappalà, A., & Giromini, L. (2021). Detecting Coached Feigning of Schizophrenia with the Inventory of Problems – 29 (IOP-29) and Its Memory Module (IOP-M): A Simulation Study on a French Community Sample. *International Journal of Forensic Mental Health*, [Epub ahead of print]. https://doi.org/10.1080/14999013.2021.1906798

Batastini, A. B., Pike, M., Thoen, M. A., Jones, A. C., Davis, R. M., Escalera, E. (2020). Perceptions and use of videoconferencing in forensic mental health assessments: A survey of evaluators and legal personnel. *Psychology, Crime & Law, 26*(6), 593-613. https://doi.org/10.1080/1068316X.2019.1708355

Ben-Porath, Y. S. & Tellegen, A. (2008/2011). *MMPI-2-RF manual for administration, scoring, and interpretation*. Minneapolis: University of Minnesota Press.

Ben-Porath, Y. S., & Tellegen, A. (2020). *MMPI-3 technical manual*. Minneapolis: University of Minnesota Press.

Bernhard, P. A., McDowell, L., & Vincent, G. M. (2021). Forenci practionioners' ude and perceptions of telepsychology before and during COVID-19. *Law and Human Behavior, 45*(5), 468-480. https://doi.org/10.1037/hb0000464t

Bigler, E. D. (2014). Effort, symptom validity testing, performance validity testing and traumatic brain injury. Brain Injury, 28(13–14), 1623–1638. https://doi.org/10.3109/02699052.2014.947627

Bigler, E. D. (2015). Neuroimaging as a biomarker in symptom validity and performance validity testing. Brain Imaging and Behavior, 9(3), 421–444. https://doi.org/10.1007/s11682-015-9409-1

Bilder, R. M., Postal, K. S., Barisa, M., Aase, D. M., Cullum, C. M., Gillaspy, S. R. et al. (2020). Inter Organizational Practice Committee recommendations/guidance for teleneuropsychology in response to the COVID-19 pandemic. *Archives of Clinical Neuropsychology, 35*(6), 647–659.

Boone K. B. (2009). The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *The Clinical neuropsychologist, 23(4)*, 729–741. https://doi.org/10.1080/13854040802427803

Boone, K. B., Salazar, X., Lu, P., Warner-Chacon, K., & Razani, J. (2002). The Rey 15-Item Recognition Trial: A technique to enhance sensitivity of the Rey 15-Item Memorization Test. *Journal of Clinical and Experimental Neuropsychology, 24(5)*, 561–573. https://doi.org/10.1076/jcen.24.5.561.1004

Boskovic, I., Akca, A. Y. E., & Giromini, L. (2022). Symptom coaching and symptom validity tests: An analog study using the structured inventory of malingered symptomatology, self-report symptom inventory, and inventory of problems-29. Applied *Neuropsychology: Adult*. Advance online publication. https://doi.org/10.1080/23279095.2022.2057856

Bosi, J., Minassian, L., Ales, F., Akca, A. Y. E., Winters, C., Viglione, D. J., ... & Giromini, L. (2022). The sensitivity of the IOP-29 and IOP-M to coached feigning of depression and mTBI: An online simulation study in a community sample from the United Kingdom. *Applied Neuropsychology: Adult,* [Epub ahead of print]. https://doi.org/10.1080/23279095.2022.2115910

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

Carroll, A. (2020). Forensic mental-health assessments after coronavirus disease 2019: Will telehealth lead us to trade psychological depth for convenience? *Medicine, Science and the Law, 60(3)*, 169–171. https://doi.org/10.1177/0025802420940618.

Carvalho, L., Reis, A., Colombarolli, M.S., Pasian, S.R., Miguel, F.K., Erdodi, L.A., Viglione, D.J., & Giromini, L. (2021). Discriminating Feigned from Credible PTSD Symptoms: a Validation of a Brazilian Version of the Inventory of Problems-29 (IOP-29). *Psychological Injury and Law, 14*, 58-70. https://doi.org/10.1007/s12207-021-09403-3

Chafetz, M. D. (2020). Deception is different: Negative validity test findings do not provide "evidence" for "good effort. *The Clinical Neuropsychologist*, Epub Ahead of Print. https://doi.org/10.1080/13854046.2020.1840633

Chenneville, T., & Schwartz-Mette, R. (2020). Ethical considerations for psychologists in the time of COVID-19. *American Psychologist, 75*(5), 644–654. https://doi.org/10.1037/amp0000661

Corey, D. M., & Ben-Porath, Y. S. (2020). Practical guidance on the use of the MMPI instruments in remote psychological testing. *Professional Psychology, Research and Practice, 51*(3), 199–204. https://doi.org/10.1037/pro0000329

Cullum, C. M., & Grosch, M. C. (2013). Special considerations in conducting neuropsychology assessment over videoteleconferencing. In K. Myers & C. L. Turvey (Eds.), *Telemental health: Clinical, technical, and administrative foundations for evidence-based practice* (pp. 275–293). Elsevier. https://doi.org/10.1016/B978-0-12-416048-4.00014-2

Cullum, C. M., Hynan, L., Grosch, M., Parikh, M., & Weiner, M. (2014). Teleneuropsychology: Evidence for Video Teleconference-Based Neuropsychological Assessment. *Journal of the International Neuropsychological Society, 20*(10), 1028-1033. doi:10.1017/S1355617714000873

Cutler, L., Greenacre, M., Abeare, C. A., Sirianni, C. D., Roth, R. M., & Erdodi, L. (2022). Multivariate models provide an effective psychometric solution to the variability in classification accuracy of D-KEFS Stroop performance validity cutoffs. *The Clinical Neuropsychologist*. (Epub Ahead of Print). https://doi.org/10.1080/13854046.2022.2073914

Daffern, M., Shea, D. E., & Ogloff, J. R. P. (2021). Remote Forensic Evaluations and Treatment in the Time of COVID-19: An International Survey of Psychologists and Psychiatrists. *Psychology, Public Policy, and Law, 27*(3), 354-369. https://doi.org/10.1037/law0000308

De Boer, A. B., Phillips, M. S., Barwegen, K. C., Obolsky, M. A., Rauch, A. A., Pesanti, S. D., ... & Soble, J. R. (2022). Comprehensive analysis of MMPI-2-RF symptom validity scales and performance validity test relationships in a diverse mixed neuropsychiatric setting. *Psychological Injury and law*, Epub Ahead of Print. https://doi.org/10.1007/s12207-022-09467-9

Drogin, E. Y. (2020). Forensic mental telehealth assessment (FMTA) in the context of COVID-19. *International Journal of Law and Psychiatry, 71*, 101595. https://doi.org/10.1016/j.ijlp.2020.101595

Dwyer, T.F. (1973). Telepsychiatry: Psychiatric consultation by interactive television. *American Journal of Psychiatry, 130*, 865-869.

Erdodi, L. A. (2019). Aggregating validity indicators: The salience of domain specificity and the indeterminate range in multivariate models of performance validity assessment. *Applied Neuropsychology: Adult, 26(2)*, 155-172. https://doi.org/10.1080/23279095.2017.1384925.

Erdodi, L. A. (2021). Five shades of gray: Conceptual and methodological issues around multivariate models of performance validity. *NeuroRehabilitation, 49(2)*, 179-213. doi: 10.3233/NRE-218020

Erdodi, L. A. (2022). Multivariate models of performance validity: The Erdodi index captures the dual nature of non-credible responding (continuous and categorical). *Assessment,* (Epub Ahead of Print). Doi: 10731911221101910.

Erdodi, L. A., & Abeare, C. A. (2020). Stronger together: The Wechsler Adult Intelligence Scale – Fourth Edition as a multivariate performance validity test in patients with traumatic brain injury. *Archives of Clinical Neuropsychology, 35(2)*, 188-204. https://doi.org/10.1093/arclin/acz032/5613200

Erdodi, L. A., Green, P., Sirianni, C., & Abeare, C. A. (2019). The myth of high false positive rates on the Word Memory Test in mild TBI. Psychological Injury and Law, 12(2), 155–169. https://doi.org/10.1007/s12207-019-09356-8

Erdodi, L. A., Kirsch, N. L., Lajiness-O'Neill, R., Vingilis, E., & Medoff, B. (2014). Comparing the Recognition Memory Test and the Word Choice Test in a mixed clinical sample: Are they equivalent? *Psychological Injury and Law, 7(3)*, 255–263.

Erdodi, L., Calamia, M., Holcomb, M., Robinson, A., Rasmussen, L., & Bianchini, K. (2023). M is For Performance Validity: The IOP-M Provides a Cost-Effective Measure of the Credibility of Memory Deficits during Neuropsychological Evaluations. Journal of Forensic Psychology Research and Practice, Epub Ahead of Print. https://doi.org/10.1080/24732850.2023.2168581

Farmer, R. L., McGill, R. J., Dombrowski, S. C., McClain, M. B., Harris, B., Lockwood, A. B., & Stinnett, T. A. (2020). Teleassessment with children and adolescents during the coronavirus (COVID-19) pandemic and beyond: Practice and policy implications. *Professional Psychology: Research and Practice, 51*(5), 477–487. https://doi.org/10.1037/pro0000349

Finger, M. S., & Ones, D. S. (1999). Psychometric equivalence of the computer and booklet forms of the MMPI: A meta-analysis. *Psychological Assessment, 11(1),* 58–66. https://doi.org/10.1037/1040-3590.11.1.58

Forbey, J. D., & Ben-Porath, Y. S. (2007). Computerized adaptive personality testing: A review and illustration with the MMPI-2 Computerized Adaptive Version. Psychological Assessment, 19(1), 14–24. https://doi.org/10.1037/1040-3590.19.1.14

Fox, K.A., & Vincent, J.P. (2020). Types of Malingering in PTSD: Evidence from a Psychological Injury Paradigm. *Psychological Injury and Law 13*, 90–104 (2020). https://doi.org/10.1007/s12207-019-09367-5

Gegner, J., Erdodi, L.A., Giromini, L., Viglione, D.J., Bosi, J. & Brusadelli, E. (2021). An Australian study on feigned mTBI using the Inventory of Problems – 29 (IOP-29), its Memory Module (IOP-M), and the Rey Fifteen Item Test (FIT). *Applied Neuropsychology: Adult*, [Epub ahead of Print]. https://doi.org/10.1080/23279095.2020.1864375

Giromini, L., & Viglione, D. J. (2022). Assessing negative response bias with the Inventory of Problems-29 (IOP-29): A quantitative literature review. *Psychological Injury and Law, 15(1)*, 79-93. https://doi.org/10.1007/s12207-021-09437-7

Giromini, L., Barbosa, F., Coga, G., Azeredo, A., Viglione, D. J., & Zennaro, A. (2020a). Using the inventory of problems - 29 (IOP-29) with the test of memory malingering (TOMM) in symptom validity assessment: A study with a portuguese sample of experimental feigners. *Applied Neuropsychology: Adult*, 27, 504-516. https://doi.org/10.1080/23279095.2019.1570929

Giromini, L., Carfora Lettieri, S. C., Zizolfi, S., Zizolfi, D., Viglione, D. J., Brusadelli, E., Zennaro, A. (2019). Beyond rare-symptoms endorsement: A clinical comparison simulation study using the minnesota multiphasic personality inventory-2 (MMPI-2) with the inventory of problems-29 (IOP-29). *Psychological Injury and Law, 12*, 212-224. https://doi.org/10.1007/s12207-019-09357-7

Giromini, L., Pasqualini, S., Corgiat Loia, A., Pignolo, C., Di Girolamo, M., & Zennaro, A. (2022). A survey of practices and beliefs of Italian psychologists regarding malingering and symptom validity assessment. *Psychological injury and law, 15(2)*, 128-140. https://doi.org/10.1007/s12207-022-09452-2

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

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

Giromini, L., Viglione, D. J., Zennaro, A., Maffei, A., & Erdodi, L. (2020). SVT meets PVT: development and initial validation of the Inventory of Problems – Memory (IOP-M. *Psychological Injury and Law, 13(3)*, 261–274. https://doi.org/10.1007/s12207-020-09385-8

Giromini, L., Young, G., & Sellbom, M. (2022). Assessing Negative Response Bias Using Self-Report Measures: Introducing the Special Issue. *Psychological Injury and Law, 15*, 1-21. https://doi.org/10.1007/s12207-022-09444-2

Goldenson, J., & Josefowitz, N., (2021). Remote Forensic Psychological Assessment in Civil Cases: Considerations for Experts Assessing Harms from Early Life Abuse. *Psychological Injury and Law, 14*, 89-103. https://doi.org/10.1007/s12207-021-09404-2

Green, C. M., Kirk, J. W., Connery, A. K., Baker, D. A., & Kirkwood, M. W. (2014). The use of the Rey 15-Item Test and recognition trial to evaluate noncredible effort after pediatric mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology, 36(3)*, 261-267. https://doi.org/10.1080/13803395.2013.879096

Green, D., Rosenfeld, B., Belfi, B., Rohlehr, L., & Pierson, A. (2012). Use of measures of cognitive effort and feigned psychiatric symptoms with pretrial forensic psychiatric patients. *International Journal of Forensic Mental Health, 11(3)*, 181–190. https://doi.org/10.1080/14999013.2012.723665

Green, P. (2003). *Green's Word Memory Test for Microsoft Windows*. Edmonton, Alberta: Green's Publishing Inc.

Hammers, D. B., Stolwyk, R., Harder, L., & Cullum, C. M. (2020). Provision of clinical teleneuropsychology services prior to and during COVID-19. *The Clinical Neuropsychologist, 34*(7-8), 1267-1283. doi: 10.1080/13854046.2020.1810323

Heilbrun, K. (2022). Psychological Testing in Forensic Contexts Conducted Remotely. *Journal of the American Academy of Psychiatry and the Law, 50*(4) 529-532. https://doi.org/10.29158/JAAPL.220083-22

Holcomb, M., Pyne, S., Cutler, L., Oikle, D. A., & Erdodi, L. A. (2022). Take Their Word for It: The Inventory of Problems Provides Valuable Information on Both Symptom and Performance Validity. *Journal of Personality Assessment*, [Epub ahead of print]. https://doi.org/10.1080/00223891.2022.2114358

Hunsley, J., & Meyer, G. J. (2003). The incremental validity of psychological testing and assessment: conceptual, methodological, and statistical issues. *Psychological Assessment, 15(4)*, 446–455. https://doi.org/10.1037/1040-3590.15.4.446

Ingram, P. B., & Ternes, M. S. (2016). The detection of content-based invalid responding: A meta-analysis of the MMPI-2 Restructured Form's (MMPI-2-RF) over-reporting Validity Scales. *The Clinical Neuropsychologist, 30(4)*, 473–496.

Kois, L. E., Cox, J., & Peck, A. T. (2020). Forensic E-Mental Health: Review, research priorities, and policy directions. *Psychology, Public Policy, and Law*, Advance online publication. https://doi.org/10.1037/law0000293

Jennette, K. J., Williams, C. P., Resch, Z. J., Ovsiew, G. P., Durkin, N. M., O'Rourke, J. J. F., Marceaux, J. C., Critchfield, E. A., & Soble, J. R. (2022). Assessment of differential neurocognitive performance based on the number of performance validity tests failures: A cross-validation study across multiple mixed clinical samples. *The Clinical neuropsychologist, 36(7)*, 1915–1932. https://doi.org/10.1080/13854046.2021.1900398

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of International Neuropsychological Society, 18(4)*, 625–630.

Larrabee, G. J., Rohling, M. L., & Meyers, J. E. (2019). Use of multiple performance and symptom validity measures: Determining the optimal per test cutoff for determination of invalidity, analysis of skew, and inter-test correlations in valid and invalid performance groups. *The Clinical Neuropsychologist, 33(8)*, 1354-1372.

Levy, M. I. (2020). Virtual forensic psychiatric practice: A lawyer's guide. Forensic psychiatric associates medical corporation. Retrieved January 30, 2021 from https://fpamed.com/virtual-forensic-psychiatricpractice-a-lawyers-guide/

Lezak, M. D. (1995). *Neuropsychological assessment (3rd ed.)*. Oxford University Press.

Menton, W. H., Corey, D. M., & Ben-Porath, Y. S. (2022). Evidence for the comparability of local and remote administrations of the MMPI-2-RF in police candidate evaluations. *Psychological Assessment, 34(1)*, 98–104. https://doi.org/10.1037/pas0001088

Menton, W. H., Crighton, A. H., Tarescavage, A. M., Marek, R. J., Hicks, A. D., & Ben-Porath, Y. S. (2019). Equivalence of laptop and tablet administrations of the Minnesota Multiphasic Personality Inventory-2 Restructured Form. *Assessment, 26(4),* 661–669. https://doi.org/10.1177/1073191117714558

Messerly, J., Soble, J. R., Webber, T. A., Alverson, W. A., Fullen, C., Kraemer, L. D., & Marceaux, J. C. (2021). Evaluation of the classification accuracy of multiple performance validity tests in a mixed clinical sample. *Applied neuropsychology. Adult, 28(6)*, 727–736. https://doi.org/10.1080/23279095.2019.1698581

Morse, C. L., Douglas-Newman, K., Mandel, S., & Swirsky-Sacchetti, T. (2013). Utility of the Rey-15 recognition trial to detect invalid performance in a forensic neuropsychological sample. *The Clinical Neuropsychologist, 27(8)*, 1395–1407. https://doi.org/10.1080/13854046.2013.832385

Neal, T. M. S., & Grisso, T. (2014). Assessment practices and expert judgment methods in forensic psychology and psychiatry: An international snapshot. *Criminal Justice and Behavior, 41(12)*, 1406–1421. https://doi.org/10.1177/0093854814548449.

Obolsky, M. A., Resch, Z. J., Fellin, T. J., Cerny, B. M., Khan, H., Bing-Canar, H., ... & Soble, J. R. (2022). Concordance of Performance and Symptom Validity Tests Within an Electrical Injury Sample. *Psychological Injury and Law*, Epub Ahead of Print. https://doi.org/10.1007/s12207-022-09469-7

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

Pinsoneault, T. B. (1996). Equivalency of computer-assisted and paper-and-pencil administered version of the Minnesota Multiphasic Personality Inventory-2. *Computers in Human Behavior, 12(2)*, 291–300. https://doi.org/10.1016/0747-5632(96)00008-8

Pivovarova, E., Rosenfeld, B., Dole, T., Green, D. & Zapf, P. (2009). Are Measures of Cognitive Effort and Motivation Useful in Differentiating Feigned from Genuine Psychiatric Symptoms?. *International Journal of Forensic Mental Health, 8(4)*, 271-278. https://doi.org/10.1080/14999011003635514

Pliskin, N. H., Puente, A. E., Morgan, J. M., & Gillaspy, S. R. (2020). Neuropsychological and psychological testing during COVID-19. Retrieved May 1, 2023 from https://www.apaservices.org/practice/clinic/covid-19-neuropsychological-psychological-testing

Puente-López, E., Pina, D., López-Nicolás, R., Iguacel, I., & Arce, R. (2023a). The Inventory of Problems–29 (IOP-29): A systematic review and bivariate diagnostic test accuracy meta-analysis. *Psychological Assessment* (Epub Ahead of Print). https://doi.org/10.1037/pas0001209

Puente-López, E., Pina, D., Rambaud-Quiñones, P., Ruiz-Hernández, J. A., Nieto-Cañaveras, M. D., Shura, R. D., Alcazar-Crevillén, A., & Martinez-Jarreta, B. (2023b). Classification accuracy and resistance to coaching of the Spanish version of the Inventory of Problems -29 and the Inventory of Problems - Memory: a simulation study with mTBI patients. *The Clinical neuropsychologist, 1–25*. Advance online publication. https://doi.org/10.1080/13854046.2023.2249171

Recupero, P. R. (2022). Daubert Considerations in Forensic Evaluations by Telepsychiatry. *Journal of the American Academy of Psychiatry and the Law, 50*(4), 517-528. https://doi.org/10.29158/JAAPL.210161-21

Reeves, C. K., Brown, T. A., & Sellbom, M. (2022). An examination of the MMPI-3 validity scales in detecting overreporting of psychological problems. *Psychological Assessment, 34(6)*, 517–527. https://doi.org/10.1037/pas0001112

Rey, A. (1941). L'examen psychologique dans les cas d'encephalopathie traumatique [Psychological examination in cases of traumatic encephalopathy]. *Archives de Psychologie, 28*, 286–340.

Reznek, L. (2005). The Rey 15-item memory test for malingering: A meta-analysis. *Brain Injury, 19(7)*, 539–543. https://doi.org/10.1080/02699050400005242

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

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*, 160-177. https://doi.org/10.1177/1073191103010002007

Roma, P., Giromini, L., Burla, F., Ferracuti, S., Viglione, D. J., & Mazza, C. (2019). Ecological validity of the Inventory of Problems-29 (IOP-29): an Italian study of court-ordered, psychological injury evaluations using the Structured Inventory of Malingered Symptomatology (SIMS) as criterion variable. *Psychological Injury and Law, 13*, 57-65. https://doi.org/10.1007/s12207-019-09368-4

Roma, P., Giromini, L., Sellbom, M., Cardinale, A., Ferracuti, S., & Mazza, C. (2023). The ecological validity of the IOP-29: A follow-up study using the MMPI-2-RF and the SIMS as criterion variables. *Psychological assessment, 35(10)*, 868–879. https://doi.org/10.1037/pas0001273

Roper, B. L., Ben-Porath, Y. S., & Butcher, J. N. (1995). Comparability and validity of computerized adaptive testing with the MMPI-2. *Journal of Personality Assessment, 65*, 358–371. https://doi.org/10.1207/s15327752jpa6502_10

Sabelli, A. G., Messa, I., Giromini, L., Lichtenstein, J. D., May, N., & Erdodi, L. A. (2021). Symptom versus performance validity in patients with mild TBI: Independent sources of non-credible responding. *Psychological Injury and Law, 14(1)*, 17–36. https://doi.org/10.1007/s12207-021-09400-6

Sharf, A. J., Rogers, R., Williams, M. M., & Henry, S. A. (2017). The effectiveness of the MMPI-2-RF in detecting feigned mental disorders and cognitive deficits: A meta-analysis. *Journal of Psychopathology and Behavioral Assessment, 39(3)*, 441–455.

Sherman, E. M. S., Slick, D. J., & Iverson, G. L. (2020). Multidimensional malingering criteria for neuropsychological assessment: A 20-year update of the malingered neuropsychological dysfunction criteria. *Archives of Clinical Neuropsychology, 35(6)*, 735–764. https://doi.org/10.1093/arclin/acaa019.

Shura, R. D., Yoash-Gantz, R. E., Pickett, T. C., McDonald, S. D., & Tupler, L. A. (2021). Relations among performance and symptom validity, mild traumatic brain injury, and posttraumatic stress disorder symptom burden in postdeployment veterans. *Psychological injury and law, 14(4)*, 257-268. https://doi.org/10.1007/s12207-021-09415-z

Slick, D., Hopp, G., Strauss, E., & Thompson, G. B. (1997). *VSVT: Victoria symptom validity test (Version 1.0)*. Odessa, FL: Psychological Assessment Resources.

Smith, G. P., & Burger, G. K. (1997). Detection of malingering: Validation of the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of the American Academy on Psychiatry and Law, 25*, 180–183.

Šömen, M.M., Lesjak, S., Majaron, T., Lavopa, L., Giromini, L., Viglione, D.J., & Podlesek, A. (2021). Using the Inventory of Problems-29 (IOP-29) with the Inventory of Problems Memory (IOP-M) in Malingering-Related Assessments: a Study with a Slovenian Sample of

Experimental Feigners. *Psychological Injury and Law*, [Epub ahead of print]. https://doi.org/10.1007/s12207-021-09412-2

Sullivan, K., Lange, R. T., & Dawes, S. (2006). Methods of detecting malingering and estimated symptom exaggeration base rates in Australia. *Journal of Forensic Neuropsychology, 4(4),* 49–70. https://doi.org/10.1300/J151v04n04_04

Sweet, J. J., Heilbronner, R. L., Morgan, J. E., Larrabee, G. J., Rohling, M. L., Boone, K. B., Kirkwood, M. W., Schroeder, R. W., Suhr, J. A., & Conference Participants (2021). American Academy of Clinical Neuropsychology (AACN) 2021 consensus statement on validity assessment: Update of the 2009 AACN consensus conference statement on neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist, 35(6)*, 1053-1106. https://doi.org/10.1080/13854046.2021.1896036

Tombaugh, T. N. (1996). *Test of memory malingering (TOMM)*. New York, NY: Multi Health Systems.

Vickery, C. D., Berry, D. T. R., Inman, T. H., Harris, M. J., & Orey, S. A. (2001). Detection of inadequate effort on neuropsychological testing. *Archives of Clinical Neuropsychology, 16(1)*, 45–73. doi: 10.1016/S0887-6177(99)00058-X.

Viglione, D. J., Giromini, L., & Landis, P. (2017). The development of the Inventory of Problems–29: A brief self-administered measure for discriminating bona fide from feigned psychiatric and cognitive complaints. *Journal of Personality Assessment, 99(5)*, 534–544. http://dx.doi.org/10.1080/00223891.2016.1233882

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

White, A. J., Batchelor, J., Pulman, S., & Howard, D. (2012). The role of cognitive assessment in determining fitness to stand trial. *International Journal of Forensic Mental Health, 11(2)*, 102–109. https://doi.org/10.1080/14999013.2012.688091

White, D. J., Ovsiew, G. P., Rhoads, T., Resch, Z. J., Lee, M., Oh, A. J., & Soble, J. R. (2022). The divergent roles of symptom and performance validity in the assessment of ADHD. *Journal of Attention Disorders, 26(1)*, 101–108. https://doi.org/10.1177/1087054720964575

Whitney, K. A., Hook, J. N., Steiner, A. R., Shepard, P. H., & Callaway, S. (2008). Is the Rey 15-Item Memory Test II (Rey II) a valid symptom validity test?: Comparison with the TOMM. *Applied Neuropsychology: Adult, 15(4),* 287–292. https://doi.org/10.1080/09084280802325215

Wright, A. J., & Raiford, S. E. (2021). Essentials of psychological teleassessment. Wiley.

Wright, A. J., Mihura, J. L., Pade, H., & McCord, D. M. (2020). Guidance on psychological tele-assessment during the COVID-10 crisis. Retrieved January 30, 2021 from https://www.apaservices.org/practice/reimbursement/health-codes/testing/tele-assessment-covid-19

Table 1. Demographic composition of the sample (N = 164).

		Control			mTBI Feigner		
		In-person (n = 41)	Remote (n = 41)	Total (n = 82)	In-person (n = 42)	Remote (n = 40)	Total (n = 82)
Age (years)	*M*	33.1	31.5	32.3	29.6	30.6	30.1
	SD	15.4	13.9	14.6	12.9	13.0	12.9
Education (years)	*M*	16.5	16.7	16.6	17.2	16.7	17.0
	SD	3.3	2.0	2.7	1.9	2.5	2.2
Gender							
	Male	11	15	26	22	19	41
	Female	29	26	55	20	21	41
	Other	1	0	1	0	0	0

Table 2. Descriptive statistics by condition and administration format: MMPI-2-RF, IOP-29, IOP-M, and FIT scores.

		Control			mTBI Feigner		
		In-person (n = 41)	Remote (n = 41)	Total (n = 82)	In-person (n = 42)	Remote (n = 40)	Total (n = 82)
MMPI-2-RF							
VRIN-r	*Min*	35	35	35	39	35	35
	Max	65	69	69	65	77	77
	M	51.3	50.4	50.9	51.9	52.7	52.2
	SD	6.7	8.6	7.7	7.3	10.6	9.0
TRIN-r	*Min*	52	52	52	52	52	52
	Max	70	73	73	76	79	79
	M	58.2	56.9	57.5	58.5	60.1	59.3
	SD	5.3	5.7	5.5	7.2	6.9	7.0
F-r	*Min*	41	41	41	47	47	47
	Max	73	95	95	119	121	121
	M	52.5	54.6	53.5	76.1	75.7	75.9
	SD	7.8	11.5	9.8	16.6	19.1	17.7
Fp-r	*Min*	41	41	41	41	46	41
	Max	77	82	82	138	103	138
	M	53.7	52.1	52.9	66.9	66.6	66.7
	SD	9.4	10.3	9.8	17.7	15.7	16.6
Fs	*Min*	43	43	43	54	54	54
	Max	66	96	96	131	131	131
	M	54.3	56.5	55.4	87.4	89.4	88.4
	SD	7.5	13.9	11.2	21.4	18.3	19.9
FBS-r	*Min*	36	33	33	41	50	41
	Max	72	80	80	111	97	111
	M	52.0	55.2	53.6	81.0	80.3	80.7
	SD	9.7	12.4	11.2	12.0	11.2	11.5
RBS	*Min*	44	34	34	47	60	47
	Max	77	93	93	116	113	116
	M	54.4	53.7	54.0	88.7	87.4	88.1
	SD	8.8	11.2	10.0	12.5	12.0	12.2
IOP-29							
FDS	*Min*	.03	.04	.03	.15	.18	.15
	Max	.67	.54	.67	.97	.98	.98
	M	.18	.16	.17	.65	.72	.68

		Control			mTBI Feigner		
		In-person (n = 41)	Remote (n = 41)	Total (n = 82)	In-person (n = 42)	Remote (n = 40)	Total (n = 82)
	SD	.16	.12	.14	.26	.22	.24
IOP-M							
# of correct	*Min*	30	30	30	15	10	10
	Max	34	34	34	34	34	34
	M	33.0	33.3	33.1	24.6	23.3	24.0
	SD	1.0	1.0	1.0	5.7	5.9	5.8
FIT							
Recall	*Min*	11	9	9	0	2	0
	Max	15	15	15	22	15	22
	M	14.4	13.4	13.9	12.1	11.3	11.7
	SD	1.3	2.1	1.8	4.1	3.5	3.8
Recall & Recognition	*Min*	13	15	13	8	6	6
	Max	30	30	30	34	30	34
	M	27.6	26.4	27.0	21.1	20.1	20.6
	SD	3.8	4.2	4.1	7.0	7.1	7.0

Note. MMPI-2-RF = Minnesota Multiphasic Personality Inventory-2-Restructured Form; VRIN-r = Variable Response Inconsistency; TRIN-r = True Response Inconsistency; F-r = Infrequent Responses; Fp-r = Infrequent Psychopathology Responses; Fs = Infrequent Somatic Responses; FBS-r = Symptom Validity; RBS = Response Bias; IOP-29 = Inventory of Problems-29; FDS = False Disorder probability Score; IOP-M = inventory of Problems-Memory; FIT = Fifteen Item Test.

Table 3. Main effect of condition: Cohen's d and Area Under the Curve (AUC) values.

	Cohen's *d*	95% C.I.		AUC	95% C.I.	
		Lower	Upper		Lower	Upper
MMPI-2-RF						
F-r	1.56	1.21	1.91	.88	.83	.93
Fp-r	1.01	0.68	1.33	.78	.71	.85
Fs	2.04	1.66	2.42	.93	.89	.97
FBS-r	2.37	1.97	2.77	.95	.92	.98
RBS	3.05	2.60	3.50	.97	.95	1.00
IOP-29						
FDS	2.58	2.16	2.99	.95	.93	.98
IOP-M						
# of correct	2.20	1.81	2.59	.92	.87	.97
FIT						
Recall	0.74	0.42	1.06	.68	.59	.76
Recall & Recognition	1.11	0.78	1.44	.79	.72	.86

Note. MMPI-2-RF = Minnesota Multiphasic Personality Inventory-2-Restructured Form; F-r = Infrequent Responses; Fp-r = Infrequent Psychopathology Responses; Fs = Infrequent Somatic Responses; FBS-r = Symptom Validity; RBS = Response Bias; IOP-29 = Inventory of Problems-29; FDS = False Disorder probability Score; IOP-M = inventory of Problems-Memory; FIT = Fifteen Item Test.

Table 4. Classification accuracy of a-priori identified cut scores.

		Spec.	Sens.	OCC
MMPI-2-RF				
F-r				
	≥ 100	1.00	.12	.56
	≥ 110	1.00	.06	.53
	≥ 120	1.00	.01	.51
Fp-r				
	≥ 80	.98	.16	.57
	≥ 90	1.00	.07	.54
	≥ 100	1.00	.05	.52
Fs				
	≥ 80	.96	.62	.79
	≥ 90	.96	.54	.75
	≥ 100	1.00	.29	.65
FBS-r				
	≥ 80	.99	.60	.79
	≥ 90	1.00	.20	.60
	≥ 100	1.00	.02	.51
RBS				
	≥ 80	.99	.82	.90
	≥ 90	.99	.56	.77
	≥ 100	1.00	.18	.59
IOP-29				
FDS				
	≥ .30	.85	.85	.85
	≥ .50	.94	.78	.86
	≥ .65	.99	.66	.82
IOP-M				
# of correct				
	< 30	1.00	.79	.90
FIT				
Recall				
	< 9	1.00	.16	.58
Recall & Recognition				
	< 20	.93	.44	.68

Note. MMPI-2-RF = Minnesota Multiphasic Personality Inventory-2-Restructured Form; F-r = Infrequent Responses; Fp-r = Infrequent Psychopathology Responses; Fs = Infrequent Somatic Responses; FBS-r = Symptom Validity; RBS = Response Bias; IOP-29 = Inventory of Problems-29; FDS = False Disorder probability Score; IOP-M = inventory of Problems-Memory; FIT = Fifteen Item Test.

DETECTING FEIGNED MTBI

Table 5. Incremental validity analyses.

Variables Entered		χ^2			B	
Step 1	Step 2	Model 1	Model 2	Δ	Model 1	Model 2 [a]
MMPI-2-RF RBS	MMPI-2-RF F-r	161.9**	164.5**	2.55	0.21**	0.25** / -0.06
MMPI-2-RF RBS	MMPI-2-RF Fp-r	161.9**	162.3**	0.34	0.21**	0.21** / -0.02
MMPI-2-RF RBS	MMPI-2-RF Fs	161.9**	165.5**	3.58	0.21**	0.18** / 0.05
MMPI-2-RF RBS	MMPI-2-RF FBS-r	161.9**	165.6**	3.64	0.21**	0.16** / 0.07
MMPI-2-RF RBS	IOP-29 FDS	161.9**	173.5**	11.61**	0.21**	0.17** / 5.35**
MMPI-2-RF RBS	IOP-M (# of correct)	161.9**	187.5**	25.59**	0.21**	0.18** / -0.75**
MMPI-2-RF RBS	FIT (recall)	161.9**	162.4**	0.46	0.21**	0.20** / -0.08
MMPI-2-RF RBS	FIT (recall & recognition)	161.94**	163.5**	1.53	0.21**	.020** / -0.08

Notes. MMPI-2-RF = Minnesota Multiphasic Personality Inventory-2-Restructured Form; F-r = Infrequent Responses; Fp-r = Infrequent Psychopathology Responses; Fs = Infrequent Somatic Responses; FBS-r = Symptom Validity; RBS = Response Bias; IOP-29 = Inventory of Problems-29; FDS = False Disorder probability Score; IOP-M = inventory of Problems-Memory; FIT = Fifteen Item Test.

[a] Values on the left of the slash refer to the variable entered at Step 1; values on the right of the slash refer to the variable entered at Step 2.

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

Table 6. Battery-wise classification accuracy for different determination rules.

Criterion to establish invalidity/noncredibility	Spec.	Sens.	OCC
One positive result			
(MMPI-2-RBS ≥ 80) *or* (IOP-29 FDS ≥ .50) *or* (IOP-M < 30)	.93	.94	.93
Two positive results			
((MMPI-2-RBS ≥ 80) *&* (IOP-29 FDS ≥ .50)) *or* ((MMPI-2-RBS ≥ 80) *&* (IOP-M < 30)) *or* ((IOP-29 FDS ≥ .50) *&* (IOP-M < 30))	1.00	.87	.93
Three positive results			
(MMPI-2-RBS ≥ 80) *&* (IOP-29 FDS ≥ .50) *&* (IOP-M < 30)	1.00	.59	.79
Specific combinations of results			
(MMPI-2-RBS ≥ 80) *&* ((IOP-29 FDS ≥ .50) *or* (IOP-M < 30))	1.00	.78	.89
(IOP-29 FDS ≥ .50) *&* ((MMPI-2-RBS ≥ 80) *or* (IOP-M < 30))	1.00	.77	.88
(IOP-M < 30) *&* ((MMPI-2-RBS ≥ 80) *or* (IOP-29 FDS ≥ .50))	1.00	.77	.88
(MMPI-2-RBS ≥ 80) *or* ((IOP-29 FDS ≥ .50) *&* (IOP-M < 30))	.99	.90	.95
(IOP-29 FDS ≥ .50) *or* ((MMPI-2-RBS ≥ 80) *&* (IOP-M < 30))	.94	.88	.91
(IOP-M < 30) *or* ((MMPI-2-RBS ≥ 80) *&* (IOP-29 FDS ≥ .50))	1.00	.89	.95

Notes. MMPI-2-RF = Minnesota Multiphasic Personality Inventory-2-Restructured Form; RBS = Response Bias; IOP-29 = Inventory of Problems-29; FDS = False Disorder probability Score; IOP-M = inventory of Problems-Memory; Spec. = Specificity; Sens. = Sensitivity; OCC = Overall Correct Classification.