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

(Article begins on next page)

19 November 2024

Abstract

This study tested whether combining the Inventory of Problems – 29 (IOP-29) with the Test of Memory Malingering (TOMM) would increase sensitivity in the detection of experimentally feigned mental health problems, compared to using either test alone. Additionally, it also evaluated (a) the effects of administration order of these two tests and (b) the cultural and linguistic applicability of these tests to a European Portuguese population. The IOP-29 and TOMM were administered to a community sample of 100 nonclinical, adult volunteers from Portugal, with the instruction to feign mental health problems. Half were instructed to feign mild traumatic brain injury (mTBI) symptoms, half were instructed to feign major depression. Administration order had no effects on the tests' scores, and both measures produced excellent sensitivity values, ranging from .82 to .98 for the TOMM, and from .88 to 1.00 for the IOP-29, when using standard a-priori cut-off scores. More importantly, combining the results of TOMM with those of IOP-29 notably increased sensitivity compared to using either test alone. This study thus supports the use of the IOP-29 together with the TOMM in multi-method symptom validity assessments and provides initial evidence that both tests can be used also in Portugal.

Keywords: IOP-29; TOMM; Order; Malingering; Brain Injury; Depression.

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

A crucial issue in forensic psychological assessment is the determination of whether a given symptom presentation is credible or not, particularly when there is some potential for reinforcement, compensation or secondary gain (Bush, Heilbronner, & Ruff, 2014). To that goal, it has been suggested that practitioners should always try to include in their evaluations multiple response bias measures, possibly taken from different types of tests (Boone, 2009; Bush et al., 2005; Heilbronner et al., 2009; Iverson, 2006; Larrabee, 2008). The underlying assumption is that different tools may tap different feigning strategies, so that using multiple, diverse tests might provide incremental validity compared to using one test alone or two similar measures using the same method or feigning strategies. Statistically, the lower the correlation between any two tests is associated with the potential for greater incremental validity and better prediction. In addition, because of the large amount of variance shared by measures using the same method relative to hetero-method tests (Campbell & Fiske, 1959), using hetero-method tests is preferable. For assessing credibility of mental health and cognitive complaints in forensic, high-stakes cases this logic would typically translate to a preference for using a performance validity with a self-report, symptom validity test over using one test only, two performance validity tests only, or two self-report, symptom validity tests.

Among available performance validity tests, one of the most widely utilized and researched is the Test of Memory Malingering (TOMM; Tombaugh, 1996, 1997), which consists of a series of two-choice, discrimination, memory recognition tasks. The main reason for its

2

efficacy in discriminating valid versus invalid cognitive symptom presentations is that most feigners do not realize that even brain-injured patients typically perform quite well on simple recognition tasks. Also, many individuals feigning mental health problems mistakenly believe that such severe memory problems might occur with mental health disorders. As such, feigners tend to show poorer performances and to obtain worse results when compared to bona fide patients (Larrabee, 2007). The validity of the TOMM as a valuable measure of poor or suboptimal effort is supported by a wide body of published studies (e.g., Greve et al., 2008; Rogers, 2008; Wisdom et al., 2012), and an important study by Gervais et al. (2004) on 519 disability claimants suggested that its classification accuracy might be even superior to that of other similar performance validity tests such as the Word Memory Test (WMT; Green, Allen, & Astner, 1996).

A different category of instruments often used to evaluate the credibility of symptom reports in forensic assessment is that of self-report, symptom validity tests. These measures essentially rely on the assumption that as opposed to bona fide patients, feigners might not know what exact symptoms are associated with a given psychopathological condition, and therefore they might tend to endorse atypical, rare, bizarre or extreme symptoms more than bona fide patients do (Rogers & Bender, 2018). In Western countries, the most widely utilized stand-alone, self-report, symptom validity test is probably the Structured Inventory of Malingered Symptomatology (SIMS; Smith & Burger, 1997; Widows & Smith, 2005), a 75-item, true/false, questionnaire covering a broad spectrum of improbable symptoms (Dandachi-FitzGerald, Ponds, & Merten, 2013; Martin, Schroeder, & Odland, 2015). According to a recent meta-analysis, the SIMS might possess adequate sensitivity, but suboptimal specificity when standard cut-off scores (i.e., ≥ 15 and ≥ 17) are employed (van Impelen et al., 2014).

Using the IOP-29 with the TOMM

Incorporating the belief that multiple detection strategies and maximizing incremental validity among items all through its development would improve validity, a new, brief, largely self-report, self-administered measure comprised of 29 items was recently introduced. Named the Inventory of Problems – 29 (IOP-29; Viglione, Giromini & Landis, 2017), its purpose is to evaluate the credibility of various symptom presentations by intermixing a few performance validity items with symptom validity items and, most importantly, with some other items addressing the examinee's ability to cope with his/her problems and with the testing situation. The IOP-29 is also designed as an "omnibus test," i.e., a test that addresses psychosis/schizophrenia, depression, post-traumatic stress symptomology and cognitive disorders, in addition to mixtures of these as seen in clinical and forensic practice.

Initial empirical evidence strongly supports the utility and validity of the IOP-29. In their developmental research, Viglione et al. (2017) found that its classification accuracy was similar to that of the longer and more complex Personality Assessment Inventory (PAI; Morey, 1991, 2007) and Minnesota Multiphasic Personality Inventory (MMPI–2; Butcher et al., 2001; Green, 1991) when assessing various symptom presentations, and perhaps better than that of the TOMM when evaluating depression-related complaints. More recently, Giromini, Viglione, Pignolo and Zennaro (2018) conducted an independent, clinical comparison, simulation study with 216 bona fide patients and 236 experimental feigners (malingering experimental paradigm), and found that the classification accuracy of the IOP-29 compared favorably to that of the two-and-a-half times longer SIMS. Indeed, when considering the entire sample ($N = 452$) the receiver operator characteristic curve (area under the curve; AUC) was .89 ($SE = .02$) for the IOP-29 vs. .83 ($SE = .02$) for the SIMS; Cohen's d effect sizes were 1.93 for the IOP-29 vs. 1.39 for the SIMS.

Using the IOP-29 with the TOMM

Despite these initial, promising results in these first two publications, the performance of the IOP-29 in multi-method symptom validity assessment has not yet been investigated. Given its inclusion of both performance validity and symptom validity items, it is currently unknown whether using the IOP-29 in conjunction with another test would increase the overall classification accuracy compared to using either test alone. Indeed, both Viglione et al. (2017) and Giromini et al. (2018) compared the validity of the IOP-29 against that of other tools, but they did so without evaluating whether combining the results of the IOP-29 with those from the other tools used in their studies would yield any incremental validity. Furthermore, but not less importantly, if one wanted to use the IOP-29 together with another test, at this time there would be no indication as to what tool one should administer first. In fact, more generally, despite the increasing importance attributed to the multi-method symptom validity assessment, the possible influence of administration order on response bias measures has not been sufficiently investigated (e.g., Bigler, 2012; Ryan et al., 2010).

The Current Study

The current study sought to contribute to symptom validity assessment literature in multiple ways. First, we intended to investigate whether using the IOP-29 together with the TOMM would yield any incremental validity compared to using either test alone. More precisely, we focused on sensitivity and tested how many feigned cases, among those undetected by the TOMM, could be detected by the IOP-29, and vice versa. Second, we evaluated the possible influence of administration order on IOP-29 and TOMM scores. Third, since we conducted our study in Portugal, we also indirectly tested the cross-cultural applicability and generalizability of both the IOP-29 and the TOMM.

Method

Prior to initiating data collection, the research project received formal approval by the pertinent ethical review board, and a Portuguese version of the IOP-29 was developed. Next, the IOP-29 and TOMM were administered, together with a few additional measures, to a Portuguese sample of experimental feigners. All were instructed to take three tests as if they intended to convince the examiner that they suffered mental illness (malingering experimental paradigm): half feigned neuropsychological impairment associated with mild traumatic brain injury (mBTI), and half feigned major depression following a work-related accident causing physical pain. These two conditions were chosen as they are often the focus of symptom validity assessment (Bigler, 2012; Dandachi-FitzGerald et al., 2013; Etherton et al., 2005; Gervais et al., 2001; Iverson et al., 2001; Mittenberg et al., 2002; Rainville et al., 1997; Rogers, 2008; Rogers & Bender, 2018; Young, 2014) and because they are among the disorders for which the IOP-29 is designed. Furthermore, while meta-analytic research indicates that there is a medium-sized association between depression and cognitive dysfunction (e.g., Zakzanis, Leach, & Kaplan, 1998), the literature on whether or not symptom and/or performance validity measures would be similarly sensitive to feigning of depression versus mTBI-related symptoms is still very scarce. Data analysis thus examined the TOMM and IOP-29 scores, focusing on classification accuracy, incremental validity, administration order effects, and possible differences between depression-related versus mTBI-related presentations.

Participants

Participants were 100 adult volunteers from Porto, Portugal, whose mother tongue was European Portuguese. More than half, i.e., 61, were students from four different universities; the remaining 49 were friends or acquaintances of student participants. Ages ranged from 19 to 56

(M = 28.3; SD = 8.5), years of education ranged from 12 to 25 (M = 16.6; SD = 2.9), and about

three quarters, i.e., 77, were women. Only a small minority, i.e., 15, were married or cohabiting.

None of these demographic variables associated significantly with any of the response

bias scores under investigation, with the sole exception of years of education, which correlated

.24 (p = .02) with one of the TOMM variables under investigation. Post-hoc analyses performed

by entering years of education as covariate (or nuisance variable) in the ANOVAs described

below, however, led to the same conclusions as those presented below in the Results section.

That is, the moderate correlation observed between years of education and one of the TOMM

scores has no effects on the main research questions addressed by this article.

Measures

The two key measures investigated in this study are the IOP-29 and TOMM.

Additionally, participants were also administered the Symptom Checklist 90-Revised (SCL-90-

R; Derogatis, 1977, 1994), and the Levenson Self-Report Psychopathy Scale (LSRP; Levenson,

Kiehl, & Fitzpatrick, 1995). The SCL-90-R is a 90-item, self-report questionnaire addressing

various psychopathological symptoms; in this study, it was used to distract the participants for

about 15 minutes before administering the last (i.e., retention) trial of the TOMM, consistent

with standard TOMM procedures (Tombaugh, 1996, 1997). The LSRP is another brief, self-

report questionnaire, which was administered at the end of the experimental procedure, after

debriefing, to evaluate the extent to which the presence of sociopathic or psychopathic traits

could affect one's own ability to feign a psychiatric or cognitive disorder. However, the current

article focuses on the TOMM and IOP-29 only, so that we only report relevant information

concerning these two tests. Analyses of the SCL-90 and LSRP scores will be presented

elsewhere.

Inventory of Problems – 29 (IOP-29; Viglione, Giromini, & Landis, 2017). The IOP-29 is a brief, 29-item, self-report test designed to assist practitioners to evaluate the credibility of symptom presentations related to various psychiatric or cognitive disorders. Most of its items (i.e., 27 out of 29) offer three response options: "true," "false," and "doesn't make sense." The two, remaining items are open-ended performance-validity questions, one a calculation problem and the second a reasoning problem. By analyzing the responses given to these 29 items, the official website of the IOP-29 (www.iop-test.com) calculates the False Disorder Probability Score (FDS), a probability value derived from a logistic regression equation that compares the responses of the test-taker against those provided by a group of bona fide patients, and those provided by a group of experimental feigners (Viglione et al., 2017). The higher the FDS, the lower the credibility of the symptom presentation, with zero and one being the minimum and maximum possible values respectively. According to Viglione et al. (2017) and Giromini et al. (2018), with no a-priori expectations, a cut-off score of FDS ≥ .50 would ensure the best balance between sensitivity and specificity. Conversely, to maximize sensitivity, for example when using the IOP-29 for screening purposes, a cut-off scores of FDS ≥ .30 or FDS ≥ .15 may be adopted to seek for expected sensitivity values of .90 and .95, respectively. To maximize specificity, for example when the costs associated with false positive findings are higher than those associated with false negative findings, cut-off scores of FDS ≥ .65 or FDS ≥ .70 may be expected to yield, respectively, specificity values of .90 and .95.

To produce a European Portuguese version of the IOP-29, we followed the recommended translation-back translation procedure (Brislin, 1980; Geisinger, 2003; Van de Vijver & Hambleton, 1996) focusing on both linguistic and cultural adaptation, as done with all the IOP-29 translations. This first draft was then used in a pilot research study with a few adult,

Portuguese volunteers to evaluate its readability and possible problems. Based on the feedback received at this step, the final, European Portuguese version of the IOP-29 was considered to be highly consistent with original, American version.

Test of Memory Malingering (TOMM; Tombaugh, 1996, 1997). The TOMM is probably the most frequently utilized and researched test to measure suboptimal effort (Sharland & Gfeller, 2007). As outlined above, it is comprised of 50 stimuli depicting line drawings of common objects, each of which is shown to the test-taker for three seconds. Two, forced-choice, dichotomous recognition trials, i.e., Trial 1 and Trial 2, are presented immediately after showing the stimuli. A third forced-choice recognition trial, i.e., the Recognition Trial, is then administered after a delay of about 15 minutes. The number of correct responses to Trial 2 and the number of correct responses to the Retention Trial are the two key variables to measure suboptimal effort: the lower those numbers, the lower the credibility of the symptom report. As a cutoff, a score of ≤ 44 correct for the 50 presentations on Trial 2 and/or the Retention Trial is considered to be an invalid or non-credible performance.

According to standard administration guidelines (Tombaugh, 1996, 1997), when the test-taker scores 45 or higher on Trial 2, the Retention Trial does not need to be administered (early termination criterion). So, for research purposes, experimental feigners with a Trial 2 score ≤ 44 would be classified as TOMM "true positives" (i.e., feigners classified as feigners), whereas those with a Trial 2 score > 44 would be classified as TOMM "false negatives" (i.e., feigners classified as honest test-takers). On the other hand, some authors have recently suggested that administering the Retention Trial to everyone, regardless of his or her score on Trial 2, might reduce the risk of false negative results (Greve & Bianchini, 2006). Though it is currently unclear whether this decrease of false negative classifications may in fact also be associated with

an increase in false positive classifications (Booksh et al., 2007; Greve & Bianchini, 2006; Wisdom et al., 2012), the current study administered the Retention Trial to all participants, to contribute to understanding whether this third trial should or should not be optionally used.

To our knowledge, only one published study (in Portuguese language) tested the TOMM in Portugal, with a sample of inmates (Mota et al., 2008). In fact, in an interesting paper by Merten et al. (2013), concerning European practitioners working in the field of symptom validity assessment, only two studies were authored by Portuguese psychologists (i.e., Martins & Martins, 2010; Simões et al., 2010). One focused on the WMT, the other one on the Rey 15 item test (Rey, 1958); neither used the TOMM. Furthermore, except for a couple of unpublished dissertations, our literature search also did not retrieve any other Portuguese studies with the TOMM. Accordingly, the current study could be the first article published in an international journal reporting data on the sensitivity of the TOMM with a Portuguese sample.

Procedures

Recruitment occurred mainly via convenience, or snowball sampling. Specifically, students attending various classes from four Portuguese universities were informed about the possibility to participate in a psychological research investigating feigning of psychiatric or cognitive disorders. Those who volunteered were then asked to also spread the word to their friends and acquaintances. Additionally, some faculty members also contributed to the recruitment of participants by spreading the word among students and collaborators via email. Although participation was completely voluntary, and no financial reward was provided, all prospective participants were informed that if they were interested, they could receive feedback regarding their performance as experimental feigners once data collection was completed.

Inclusion criteria required: (a) having Portuguese citizenship; (b) being a native European Portuguese speaker (Brazilian Portuguese speaking people could not participate); (c) denying experiencing head trauma; and (d) denying a history of major depression. Individuals who met these inclusion criteria and decided to volunteer were then asked to sign an informed consent form and were given a document detailing the instructions to follow for the study. Specifically, participants assigned to the mTBI group (half of the total) were instructed to take the tests as if they wanted to convince the examiner that they were experiencing symptoms associated with mTBI. To facilitate feigning (see Rogers & Gillard, 2011; Viglione et al., 2001), they were asked to read the description of an accident in which a person slipped while walking down a flight of stairs, striking his head against a cement step, and subsequently remaining unconscious for about three minutes. Participants assigned to the depression group (the other half of the total) were instructed to take the tests as if they wanted to convince the examiner that they were experiencing symptoms associated with major depression. In this case, the scenario to facilitate feigning (Rogers & Gillard, 2011; Viglione et al., 2001) involved a work-related accident in which an employee fell on a freshly mopped floor and hit his/her tailbone, which led him/her to experience intense pain, difficulties at work, and ultimately, major depression. These two scenarios were adapted and translated from Viglione et al. (2017). As in previous studies, both vignettes included cautions about overdoing the portrayal of the disorder because they would be detecting as faking (Viglione et al., 2001)

Half of the participants within both the mTBI and depression groups were administered the TOMM first, and the IOP-29 second. The other halves were administered the IOP-29 first and the TOMM second. As might be done in a real evaluation and to meet the time delay requirements of the retention trial of the TOMM, the SCL-90-R was administered after the

second trial and before the retention trial. Upon completion of IOP-29, TOMM, and SCL-90,

they were instructed to discontinue the role of an experimental feigner to complete in an honest

fashion a brief demographic form (which was used also to verify that the exclusion criteria were

met), and the LSRP. Lastly, participants were inquired about their experience and feigning

strategies, to ensure that they had recalled and followed the instructions.

Data Analysis

Our main goals were (a) to evaluate whether combining the results of IOP-29 and TOMM

would provide incremental validity compared to using either test alone, and (b) to examine

potential administration order effects on IOP-29 and TOMM scores. Prior to addressing

incremental validity, we tested whether administration order and/or type of symptom

presentation (i.e., neuropsychological impairment versus major depression) would influence

IOP-29 and TOMM scores, to decide whether or not we could perform our incremental validity

analyses by collapsing all available data. We thus first performed a series of three two-way

ANOVAs, with administration order (TOMM first versus IOP-29 first) and symptom

presentation (mTBI versus depression) as between-subject factors, and TOMM Trial 2, TOMM

Retention, and IOP-29 FDS scores as our dependent variables. Next, we examined the sensitivity

of both the TOMM and IOP-29 scores and investigated whether the false negatives produced by

the TOMM would be detected by the IOP-29, and vice versa. Lastly, we also examined the

correlation between TOMM and IOP-29 scores, to evaluate the degree to which the two tests are

redundant with each other.

<div align="center">

Results

</div>

Descriptive statistics concerning all three scores under investigation (i.e., TOMM Trial 2,

TOMM Retention, and IOP-29 FDS) are reported in Table 1, split by symptom presentation (m-

TBI versus depression) and administration order (TOMM first versus IOP-29 first). In terms of

significance testing, neither the interaction effect (administration order X symptom presentation)

nor the main effect of administration order were statistically significant for any of the dependent

variables under investigation, all $p \geq .11$. When considering the entire sample, Cohen's d effect

sizes characterizing the differences in the scores obtained in the two different administration

orders were $d = .08$ for TOMM Trial 2, $d = .13$ for TOMM Retention, and $d = .17$ for IOP-29

FDS. According to standard benchmarks, these values may be characterized as *small* or *very*

small effect sizes (Cohen, 1988). Conversely, the main effect of symptom presentation was

statistically significant for all scores under investigation, with $F(1, 96) = 7.04$, $p = .01$, $d = .53$,

for TOMM Trial 2, $F(1, 96) = 6.56$, $p = .01$, $d = .51$, for TOMM Retention, and $F(1, 96) = 6.26$,

$p = .01$, $d = .50$, for IOP-29 FDS. Thus, subsequent analyses collapsed all data with different

administration orders, but inspected the two subsamples assigned to the mTBI- versus

depression-related symptom presentations both separately and combined.

Sensitivity analyses are summarized in Table 2. For the TOMM, Trial $2 \leq 44$ is the

classic cut-off recommended in the manual of the test (Tombaugh, 1996, 1997). Additionally,

because this study administered the Retention Trial to all participants and not just to those who

scored ≤ 44 on Trial 2 (Greve & Bianchini, 2006), sensitivity data for TOMM Retention Trial

and for the combination of Trial 2 and Retention Trial (i.e., TOMM Trial $2 \leq 44$ or TOMM

Retention ≤ 44) are presented too. These values allow the appreciation of the extent to which

sensitivity increases if the Retention Trial is administered to all cases, rather than just to those

who score below 45 on Trial 2. As for the IOP-29, FDS $\geq .50$ is the standard cut-off to be used if

there are no a-priori expectations (Giromini et al., 2018; Viglione et al., 2017). Additionally,

since our study is basically a sensitivity study, by the fact that it does not include any patient

13

data, Table 2 also reports sensitivity data associated with cut-off scores of FDS ≥ .30 and FDS ≥ .15. As noted above, according to Giromini et al. (2018) these scores should yield sensitivity levels of .90 and .95, respectively.

Examination of Table 2 reveals that both tests achieved excellent sensitivity values, ranging from .82 to .98 for the TOMM, and from .88 to 1.00 for the IOP-29. Interestingly, the number of cases above versus below the selected cut-off scores did not significantly change from one symptom presentation to another. Indeed, *Phi* statistics testing the association between correctness of classifications (correct versus wrong) and type of symptom presentation (mTBI or depression) were: *Phi* = .18, *p* = .06, for TOMM Trial 2 ≤ 44; *Phi* = .16, *p* = .10, for TOMM Retention ≤ 44; *Phi* = .10, *p* = .31, for TOMM Trial 2 ≤ 44 or TOMM Retention ≤ 44; *Phi* = .15, *p* = .14, for IOP-29 FDS ≥ .50; and *Phi* = .06, *p* = .56, for IOP-29 FDS ≥ .30 (for IOP-29 FDS ≥ .15, *Phi* was not calculated as no cases scored below FDS = .15). This pattern of findings indicates that although the average scores of the TOMM and IOP-29 significantly differed from one symptom presentation to another, the overall number of correct versus wrong classifications did not statistically differ when looking at mTBI-related versus depression-related symptom presentations.

Next, incremental validity was inspected by examining the extent to which the cases misclassified by one test (false negatives) were correctly classified by the other test. More in detail, to evaluate the incremental validity of the IOP-29 over the TOMM, we calculated the proportion of cases correctly classified by the IOP-29 among all cases that were misclassified (false negatives) by the TOMM. As shown in Table 3, depending on the sample and criterion under consideration, the IOP-29 was able to detect 67% to 100% of the cases misclassified by the TOMM. Noteworthy, when considering the entire sample, the number of false negatives

produced by the TOMM was reduced by 75%, 82%, or 92%, depending on the TOMM criterion under consideration, when using the standard IOP-29 cut-off of FDS ≥ .50. Along the same lines, to evaluate the incremental validity of the TOMM over the IOP-29, we calculated the proportion of cases that were correctly classified by the TOMM among all the cases that were misclassified (false negatives) by the IOP-29. As shown in Table 4, these proportion values ranged from 50% to 100%. Taken together, these findings indicate that feigning individuals who passed one test were typically detected by the other test.

Statistically, incremental validity is greatest when the correlation between predictors is low. When considering the entire, combined sample, correlations of IOP-29 FDS to TOMM Trial 2 and TOMM Retention were, respectively, $r = -.10$, $p = .34$, and $r = -.11$, $p = .29$. With the mTBI subsample, these correlations were $r = -.09$, $p = .53$, and $r = -.14$, $p = .33$; with the depression subsample they were $r = -.30$, $p = .04$, and $r = -.24$, $p = .09$. All these values may be characterized as *small* or *small* to *medium* effect sizes (Cohen, 1988). Visual inspection of the scatterplots depicted in Figure 1 also indicates that using the TOMM together with the IOP-29 would yield better classifications compared to using one test alone.

Lastly, Table 5 reports on the sensitivity levels that may be achieved by combining the information from the TOMM with that from the IOP-29. Depending on the sample and criterion under consideration, these values range from sensitivity = .98 to sensitivity = 1.00. Noteworthy, combining the results of TOMM Trial 2 with those of the IOP-29 improved sensitivity more than did combining the results of TOMM Trial 2 with those of TOMM Retention. Indeed, in the first case, sensitivity raised from .88 to .99 (considering the entire sample and the standard FDS cut-off score of ≥ .50); in the second case it raised from .88 to .96 (again considering the entire

sample size). This finding is in line with Greve and Bianchini's (2006) conclusion that the costs of not administering the Retention Trial are minimized by the inclusion of another test.

Discussion

The current study was conducted to evaluate the extent to which the recently developed, Inventory of Problems – 29 (IOP-29; Viglione, Giromini & Landis, 2017) could provide useful, nonredundant information when included in multi-method symptom validity assessment. We administered the IOP-29 and the Test of Memory Malingering (TOMM; Tombaugh, 1996, 1997) to 100 nonclinical, adult, Portuguese, participants instructed to feign mental health problems associated with mTBI ($n = 50$) or major depression ($n = 50$). Our findings reveal that combining the TOMM with the IOP-29 increased sensitivity compared to using either test alone, regardless of what tool was administered first. Furthermore, they also provide some additional, empirical evidence that both measures may be applicable cross-culturally with minimal variations from one country to another.

The key finding of the current study is that using the TOMM together with the IOP-29 remarkably reduced the number of false negatives compared to using either test alone. Indeed, depending on the cut-off scores under consideration, 50% to 100% of IOP-29 false negatives failed the TOMM, and 67% to 100% of the TOMM false negatives failed the IOP-29. Only one case out of one hundred had a TOMM Trial 2 score greater of 44 and an IOP-29 FDS lower than (but not far from) .50, and none of those who failed TOMM Trial 2 had an IOP-29 FDS lower than .30. Thus, using the TOMM in combination with the IOP-29 notably increased sensitivity compared to using either test alone. Importantly, there were minimal variations between one symptom presentation to another, so that these conclusions likely apply both to mTBI-related and to depression-related evaluations.

Despite the growing consensus that multiple measures should be used when evaluating the credibility of symptoms in forensic assessment, so far research has dedicated little attention to the effects of administration order on performance and/or self-report symptom validity tests (Bigler, 2012). Consistent with this limited, available empirical research (e.g., Ryan et al., 2010; Zu & Tulsky, 2000), our study showed that different administration orders led to very similar results. Pending future replications and additional research addressing specificity, these findings thus do not substantiate any worries about possible administration order effects for the TOMM and IOP-29.

In terms of generalizability of our findings, it is important to highlight that our sensitivity outcomes are comparable to (or perhaps even slightly better than) those observed in previous studies evaluating the sensitivity of TOMM or IOP-29 individually, in other cultural contexts. For example, Jelicic et al. (2011) reported on a Dutch sample of experimental feigners who took the TOMM and SIMS with the instruction to simulate mental health problems associated with brain injury. One of the samples included in their research ($n = 30$) was given a brief scenario about an accident causing brain injury, a succinct description of the key symptoms of mTBI, and a caution warning that extreme portrayals would result in non-believable outcomes, like we did in our study. With that sample, using TOMM Trial 2 ≤ 44 as cut-off yielded a sensitivity of .87, and using TOMM Trial 2 ≤ 44 or TOMM Retention ≤ 44 yielded a sensitivity of .97. In our study very similar values were observed, with sensitivity being equal to .94 and .97, respectively, when considering the mTBI subsample. Similar conclusions may be drawn also when considering other similar studies, such as those conducted by Rees et al. (1998). Along the same lines, previous IOP-29 research in the US and in Italy has demonstrated that the IOP-29 FDS tends to produce sensitivity values of .80, .90, and .95 when using cut-off values of FDS ≥ .50,

17

FDS ≥ .30, and FDS ≥ .15, respectively (Viglione et al., 2017; Giromini et al., 2018). In our

study, similarly high – or perhaps even slightly higher – sensitivity values were observed for

these same cut-off scores, with small variations from one subsample to the other. Interestingly,

the fact that the sensitivity of the IOP-29 might be even slightly higher than originally

hypothesized by Viglione et al. (2017) and Giromini et al. (2018) is consistent also with a more

recent Italian study by Giromini et al. (in press).

In their initial research leading up to the development of the IOP-29, Viglione et al.

(2017) reported that the IOP-29 performed slightly better than the TOMM when assessing bona

fide versus feigned (via malingering experimental paradigm) depression, with AUCs being .90

(SE = .04) for the IOP-29 FDS and .82 (SE = .05) for TOMM Trial 2. To some extent, our study

confirms that the IOP-29 might produce a slightly higher sensitivity than the TOMM when

assessing depression-related symptoms with standard, a-priori established, cut-off scores. Indeed,

when using IOP-29 FDS ≥ .50 and TOMM Trial 2 ≤ 44, sensitivity was .96 for the IOP-29 and

.82 for the TOMM. On the other hand, with these same cut-off scores, the TOMM produced a

slightly higher sensitivity than the IOP-29 when assessing mTBI-related presentations,

sensitivity = .94 versus sensitivity = .88. The TOMM and IOP-29 thus appear to complement

each other particularly well also from this point of view.

In previous TOMM research, it was observed that using the Retention Trial in all cases,

rather than only if one fails Trial 2 (i.e., with Trial 2 ≤ 44), may increase the sensitivity of the

TOMM (Greve & Bianchini, 2006). In that same research, however, examination of scores from

150 patients with traumatic brain injury and 150 chronic pain patients revealed that using another

test, in addition to the TOMM, may compensate for the loss in sensitivity associated with not

administering the Retention Trial. In our study, Greve and Bianchini's (2006) consideration finds

full support. Indeed, using TOMM Trial 2 without the Retention Trial combined with IOP-29 data yielded even higher sensitivity (.99 versus .96, within the entire sample) than combining TOMM Trial 2 with TOMM Retention Trial information. Given that administering the IOP-29 takes on average five minutes or less, in terms of costs-benefits ratio, administering the TOMM in its standard format (i.e., administering Retention only if TOMM Trial 2 is ≤ 44) together with the IOP-29 might be preferable over administering all three trials of the TOMM to all cases, without using the IOP-29.

Limitations and Future Perspectives

We note that several methodological limitations characterize this study which necessitate it be understood as exploratory and preliminary. The most obvious one is the lack of a patient contrast group with bona fide mTBI or major depression, which did not allow us to investigate specificity. Future studies with clinical comparison samples are therefore needed to evaluate whether the increased sensitivity obtained by combining the information from TOMM and IOP-29 might have any notable costs in terms of reduced specificity. Also, our experimental feigners had a relatively high education level, which may limit the generalizability of our findings, given that real-life malingerers are likely less educated than were our participants. Along the same lines, our using experimental feigners rather than real-life malingerers might limit external validity. Indeed, although there is some evidence that college students might be able to simulate neuropsychological problems from a mTBI in a fairly believable manner (Haines & Norris, 2001), real-life patients feigning cognitive disorders could in fact score higher on the TOMM and lower on the IOP-29 than our experimental simulators did. Another important limitation is that we did not offer any monetary rewards for feigners able to avoid detection (Rogers & Gillard, 2011). Although our debriefing procedures suggested that our participants did

understand the instructions and did make an effort to successfully feign and avoid detection, future replications with external incentives would be highly beneficial. Furthermore, future studies should attempt to include some validity checks to determine if the participant is trying to deceive or is simply performing poorly. Additionally, the fact that our administrators were not blind to the different instructions to feign (depression vs. mTBI symptoms) might have influenced the outcome, too. Lastly, it should also be noted that our article primarily focuses on incremental validity, highlighting the importance of using multiple tools and scales to evaluate symptom validity, so to tap different feigning strategies and maximize classification accuracy. Nevertheless, to broaden our understanding of the potential usefulness of the IOP-29, future studies should also focus on other important psychometric properties, such as concurrent or predictive validity, and test-retest stability. In particular, it would be very useful if future studies tested the correlation of symptom validity tests (SVT) such as the Morel Emotional Numbing Test for Posttraumatic Stress Disorder (MENT; Morel, 1998) or SIMS to the SVT items of the IOP-29, and the correlation of performance validity tests (PVT) such as the Word Memory Test (WMT; Green et al., 1996), Medical Symptom Validity Test (MSVT; Green, 2004), Victoria Symptom Validity Test (VSVT; Slick et al., 1997), or Rey 15-item Memorization Test (RMT; (Lezak, 1995) to its PVT items.

Conclusion

Despite the limitations listed above, our investigation is still the first to show that: (a) using the TOMM together with the IOP-29 notably increases sensitivity compared to using either test alone; (b) administration order does not notably affect TOMM and IOP-29 scores; (c) the TOMM and IOP-29 achieve excellent sensitivity values also when used with a Portuguese population.

Using the IOP-29 with the TOMM

References

Bigler, E. D. (2012). Symptom validity testing, effort, and neuropsychological assessment. *Journal of the International Neuropsychological Society, 18*, 632–642.

Booksh, R. L., Aubert, M. J., & Andrews, S. R. (2007). Should the retention trial of the Test of Memory Malingering be optional? A reply. *Archives of Clinical Neuropsychology, 22*, 87–89.

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

Brislin, R. W. (1980). Translation and content analysis of oral and written material. In H.C. Triandis & J.W. Berry (Eds.), *Handbook of cross-cultural psychology (Vol.1)* (pp. 389–444). Boston, MA: Allyn & Bacon.

Bush, S. S., Heilbronner, R. L., & Ruff, R. M. (2014). Psychological assessment of symptom and performance validity, response bias, and malingering: Official position of the Association for Scientific Advancement in Psychological Injury and Law. *Psychological Injury and Law, 7*, 197–205.

Bush, S. S., Ruff, R. M., Troster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., et al. (2005). Symptom validity assessment: Practice issues and medical necessity. *Archives of Clinical Neuropsychology, 20*, 419–426.

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S.,Tellegen, A., & Dahlstrom, W. G. (2001). *Minnesota Multiphasic Personality Inventory–2 (MMPI-2): Manual for administration and scoring (rev. ed.).* Minneapolis, MN: University of Minnesota Press.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the

multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81-105.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale, NJ:

Erlbaum.

Dandachi-FitzGerald, B., Ponds, R. W. H. M., & Merten, T. (2013). Symptom validity and

neuropsychological assessment: a survey of practices and beliefs of neuropsychologists in

six European countries. *Archives of Clinical Neuropsychology, 28*, 771–783.

Derogatis, L. R. (1977). *The SCL-90 Manual I: Scoring, administration and procedures for the

SCL-90*. Baltimore, MD: Johns Hopkins University School of Medicine, Clinical

Psychometrics Unit.

Derogatis, L. R. (1994). *SCL90-R: Administration, scoring and procedures manual*.

Minneapolis, MN: National Computer Systems

Etherton, J. L., Bianchini, K. J., Greve, K. W., & Ciota, M. A. (2005). Test of Memory

Malingering performance is unaffected by laboratory-induced pain: implications for

clinical use. *Archives of Clinical Neuropsychology, 20*, 375–384.

Geisinger, K. F. (2003). Testing and assessment in cross-cultural psychology. In J. R. Graham, J.

A. Naglieri, & I. B. Weiner (Eds.), *Handbook of psychology (Vol. 10)*. Assessment

Psychology (pp. 95–118). New Jersey, NJ: John Wiley & Sons.

Gervais, R. O., Green, P., Allen, L. M., & Iverson, G. L. (2001). Effects of coaching on symptom

validity testing in chronic pain patients presenting for disability assessments. *Journal of

Forensic Neuropsychology, 2*, 1–19.

Gervais, R., Rohling, M., Green, P., & Ford, W. (2004). A comparison of WMT, CARB and TOMM failure rates in non-head injury disability claimants. *Archives of Clinical Neuropsychology, 19 (4)*, 475–487.

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*, doi:10.1007/s12207-018-9314-1

Giromini, L., Viglione, D.J., Pignolo, C. & Zennaro, A. (in press). An Inventory of Problems – 29 (IOP-29) sensitivity study investigating feigning of four different symptom presentations via malingering experimental paradigm. *Journal of Personality Assessment.* (manuscript accepted for publication).

Green, P. (2004). *Medical Symptom Validity Test for Windows: User's manual and program*. Edmonton, Alberta, Canada: Green's.

Green, P., Allen, L. M., & Astner, K. (1996). *The Word Memory Test: A user's guide to the oral and computer-administered forms, US version 1.1*. Durham, NC: CogniSyst.

Green, R. L. (1991). *MMPI-2/MMPI: An interpretative manual*. Boston, MA: Allyn & Bacon.

Greve, K. W., & Bianchini, K. J. (2006). Should the retention trial of the Test of Memory Malingering be optional? *Archives of Clinical Neuropsychology, 21*, 117–119

Greve, K., Ord, J., Curtis, K., Bianchini, K., & Brennan, A. (2008). Detecting malingering in traumatic brain injury and chronic pain: A comparison of three forced-choice symptom validity tests. *The Clinical Neuropsychologist, 22 (5)*, 896–918.

Haines, M.E., & Norris, M.P. (2001). Comparing student and patient simulated malingerers' performance on standard neuropsychological measures to detect feigned cognitive deficits. *Clinical Neuropsychologist, 15*, 171–182

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Participants, C. (2009). American Academy of Clinical Neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist, 23*, 1093–1129.

Iverson, G. L., King, R. J., Scott, J. G., & Adams, R. L. (2001). Cognitive complaints in litigating patients with head injuries or chronic pain. *Journal of Forensic Neuropsychology, 2*, 19–30.

Iverson, G. L. (2006). Ethical issues associated with the assessment of exaggeration, poor effort, and malingering. *Applied Neuropsychology, 13*, 77-90.

Jelicic, M., Ceunen, E., Peeters, M. J. V., & Merckelbach, H. (2011). Detecting coached feigning using the Test of Memory Malingering (TOMM) and the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of Clinical Psychology, 67*, 850–855.

Larrabee, G.J. (2007). *Assessment of malingered neuropsychological deficits*. New York: Oxford University Press.

Larrabee, G.J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist, 22*, 666–679. doi:10.1080/13854040701494987

Levenson, M., Kiehl, K., & Fitzpatrick, C. (1995). Assessing psychopathic attributes in a noninstitutionalized population. *Journal of Personality and Social Psychology, 68*, 151-158.

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

Martins, M., & Martins, I. P. (2010). Memory malingering: Evaluating WMT criteria. *Applied
Neuropsychology, 17*, 177-182.

Martin, P. K., Schroeder, R. W., & Odland, A. P. (2015). Neuropsychologists' validity testing
beliefs and practices: a survey on North American professionals. *The Clinical
Neuropsychologist, 29*, 741–776.

Merten, T., Dandachi-FitzGerald, B., Hallc, V., Schmandd, B. A., Santamaríae, P., & González-
Ordif, H. (2013). Symptom validity assessment in European countries: Development and
state of the art. *Clínica y Salud, 24*, 129–138.

Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering
and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology, 24*,
1094–1102.

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

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

Mota, M., Simões, M. R., Amaral, L., Dias, I., Luís, D., Pedrosa, C., Maior, F. S., & Silva, I.
(2008). Test of Memory Malingering (TOMM): Estudos de validação numa amostra de
reclusos. *Psiquiatria, Psicologia & Justiça, 2*, 23-41.

Rainville, J., Sobel, J. B., Hartigan, C., & Wright, A. (1997). The effect of compensation
involvement on the reporting of pain and disability by patients referred for rehabilitation
of chronic low back pain. *Spine, 22*, 2016–2024.

Rees, L. M., Tombaugh, T. N., Gansler, D. A., & Moczynski, N. P. (1998). Five validation experiments of the Test of Memory Malingering (TOMM). *Psychological Assessment, 10(1)*, 10.

Rey, A. (1958). *L'examen clinique en psychologie [The clinical examination in psychology]*. Paris: Presses Universitaires de France.

Rogers, R. (2008). *Clinical assessment of malingering and deception*. New York: Oxford University Press.

Rogers, R. & Bender, D. (2018). *Clinical assessment of malingering and deception*. New York, NY: Guilford Press.

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.

Ryan, J. J., Glass, L. A., Hinds, R. M., & Brown, C. N. (2010). Administration order effects on the Test of Memory Malingering. *Applied Neuropsychology, 17*, 246–250.

Sharland, M. J., & Gfeller, J. D. (2007). A survey of neuropsychologists' beliefs and practices with respect to the assessment of effort. *Archives of Clinical Neuropsychology, 22*, 213–223.

Simões, M., Sousa, L., Duarte, P., Firmino, H., Pinho, M. S., Gaspar, N., … França, S. (2010). Avaliação da simulação ou esforço insuficiente com o Rey 15-Item Memory Test (15-IMT): Estudos de validação em grupos de adultos idosos [Validation of simulation or insufficient effort with the Rey 15-IMT: Validation studies in aged adult groups]. *Análise Psicológica, 28*, 209–226.

Slick, D., Hopp, G., Strauss, E., & Thompson, G. B. (1997). *Victoria Symptom Validity Test*. 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.

Tombaugh, T. N. (1996). *Test of Memory Malingering (TOMM)*. New York: Multi Health Systems.

Tombaugh, T. N. (1997). The Test of Memory Malingering (TOMM): Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment, 9*, 260–268.

Van de Vijver, F., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist, 1*, 89–99.

van Impelen, A., Merckelbach, H., Jelicic, M., & Merten, T. (2014). The Structured Inventory of Malingered Symptomatology (SIMS): A systematic review and meta-analysis. *The Clinical Neuropsychologist, 28*, 1336–1365.

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*, 534-544.

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*, 237–250.

Widows, M. R., & Smith, G. P. (2005). *SIMS-Structured Inventory of Malingered*

Symptomatology. Professional manual. Lutz, FL: Psychological Assessment Resources.

Wisdom, N. M., Whitney, L. B., Chen, D. K., & Collins, R. L. (2012). The use of all three Test

of Memory Malingering trials in establishing the level of effort. *Archives of Clinical*

Neuropsychology, 27, 208-212.

Young, G. (2014). *Malingering, feigning, response bias in psychiatric/psychological injury*.

Dordrecht: Springer SBM.

Zakzanis, K. K., Leach, L., & Kaplan, E. (1998). *Neuropsychological differential diagnosis*.

Visse, The Netherlands: Swets & Zeitlinger.

Zhu, J., & Tulsky, D. S. (2000). Co-norming the WAIS-III and WMS-III: Is There a Test-Order

Effect on IQ and Memory Scores? *The Clinical Neuropsychologist, 14*, 461-467.

Table 1. TOMM and IOP-29 Scores by Symptom Presentation and Administration Order.

	mTBI-related Subsample			Depression-related Subsample			Entire Sample		
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Administration Order: TOMM 1st, IOP-29 2nd									
TOMM Trial 2	25	29.0	10.1	25	35.3	9.6	50	32.1	10.3
TOMM Retention	25	27.5	11.4	25	35.8	10.3	50	31.7	11.6
IOP-29 FDS	25	0.74	0.19	25	0.84	0.12	50	0.79	0.16
Administration Order: IOP-29 1st, TOMM 2nd									
TOMM Trial 2	25	31.2	9.0	25	34.7	8.5	50	32.9	8.9
TOMM Retention	25	32.0	9.9	25	33.9	7.7	50	33.0	8.8
IOP-29 FDS	25	0.78	0.22	25	0.86	0.17	50	0.82	0.20

Table 2. TOMM and IOP-29 Sensitivity by Symptom Presentation.

	mTBI-related Subsample ($n = 50$)	Depression-related Subsample ($n = 50$)	Entire Sample ($N = 100$)
TOMM Only			
TOMM Trial 2 $\leq$ 44	.94	.82	.88
TOMM Retention $\leq$ 44	.94	.84	.89
TOMM Trial 2 $\leq$ 44 or TOMM Retention $\leq$ 44	.98	.94	.96
IOP-29 Only			
IOP-29 FDS $\geq$ .50	.88	.96	.92
IOP-29 FDS $\geq$ .30	.96	.98	.97
IOP-29 FDS $\geq$ .15	1.00	1.00	1.00

Table 3. Incremental Validity of IOP-29 over TOMM, by Symptom Presentation.

	mTBI-related Subsample	Depression-related Subsample	Entire Sample
TOMM Trial 2 > 44 &			
IOP-29 FDS ≥ .50	3 (100%)	8 (89%)	11 (92%)
IOP-29 FDS < .50	0 (0%)	1 (11%)	1 (8%)
TOMM Trial 2 > 44 &			
IOP-29 FDS ≥ .30	3 (100%)	9 (100%)	12 (100%)
IOP-29 FDS < .30	0 (0%)	0 (0%)	0 (0%)
TOMM Retention > 44 &			
IOP-29 FDS ≥ .50	2 (67%)	7 (88%)	9 (82%)
IOP-29 FDS < .50	1 (33%)	1 (13%)	2 (18%)
TOMM Retention > 44 &			
IOP-29 FDS ≥ .30	3 (100%)	8 (100%)	11 (100%)
IOP-29 FDS < .30	0 (0%)	0 (0%)	0 (0%)
TOMM Trial 2 > 44 & TOMM Retention > 44 &			
IOP-29 FDS ≥ .50	1 (100%)	2 (67%)	3 (75%)
IOP-29 FDS < .50	0 (0%)	1 (33%)	1 (25%)
TOMM Trial 2 > 44 & TOMM Retention > 44 &			
IOP-29 FDS ≥ .30	1 (100%)	3 (100%)	4 (100%)
IOP-29 FDS < .30	0 (0%)	0 (0%)	0 (0%)

Table 4. Incremental Validity of TOMM over IOP-29, by Symptom Presentation.

	mTBI-related Subsample	Depression-related Subsample	Entire Sample
IOP-29 FDS < .50 &			
TOMM Trial 2 ≤ 44	6 (100%)	1 (50%)	7 (88%)
TOMM Trial 2 > 44	0 (0%)	1 (50%)	1 (13%)
IOP-29 FDS < .50 &			
TOMM Retention ≤ 44	5 (83%)	1 (50%)	6 (75%)
TOMM Retention > 44	1 (17%)	1 (50%)	2 (25%)
IOP-29 FDS < .50 &			
TOMM Trial 2 ≤ 44 or TOMM Retention ≤ 44	6 (100%)	1 (50%)	7 (88%)
TOMM Trial 2 > 44 & TOMM Retention > 44	0 (0%)	1 (50%)	1 (13%)
IOP-29 FDS < .30 &			
TOMM Trial 2 ≤ 44	2 (100%)	1 (100%)	3 (100%)
TOMM Trial 2 > 44	0 (0%)	0 (0%)	0 (0%)
IOP-29 FDS < .30 &			
TOMM Retention ≤ 44	2 (100%)	1 (100%)	3 (100%)
TOMM Retention > 44	0 (0%)	0 (0%)	0 (0%)
IOP-29 FDS < .30 &			
TOMM Trial 2 ≤ 44 or TOMM Retention ≤ 44	2 (100%)	1 (100%)	3 (100%)
TOMM Trial 2 > 44 & TOMM Retention > 44	0 (0%)	0 (0%)	0 (0%)

Table 5. Sensitivity Associated with Combination of TOMM and IOP-29 Data, by Symptom Presentation.

	mTBI-related Subsample ($n = 50$)	Depression-related Subsample ($n = 50$)	Entire Sample ($N = 100$)
Combination of TOMM & IOP-29			
TOMM Trial 2 ≤ 44 or IOP-29 FDS ≥ .50	1.00	.98	.99
TOMM Retention ≤ 44 or IOP-29 FDS ≥ .50	.98	.98	.98
TOMM Trial 2 ≤ 44 or TOMM Retention ≤ 44 or IOP-29 FDS ≥ .50	1.00	.98	.99
TOMM Trial 2 ≤ 44 or IOP-29 FDS ≥ .30	1.00	1.00	1.00
TOMM Retention ≤ 44 or IOP-29 FDS ≥ .30	1.00	1.00	1.00
TOMM Trial 2 ≤ 44 or TOMM Retention ≤ 44 or IOP-29 FDS ≥ .30	1.00	1.00	1.00

Figure 1. Graphical Representation of TOMM and IOP-29 Classifications, by Symptom Presentation
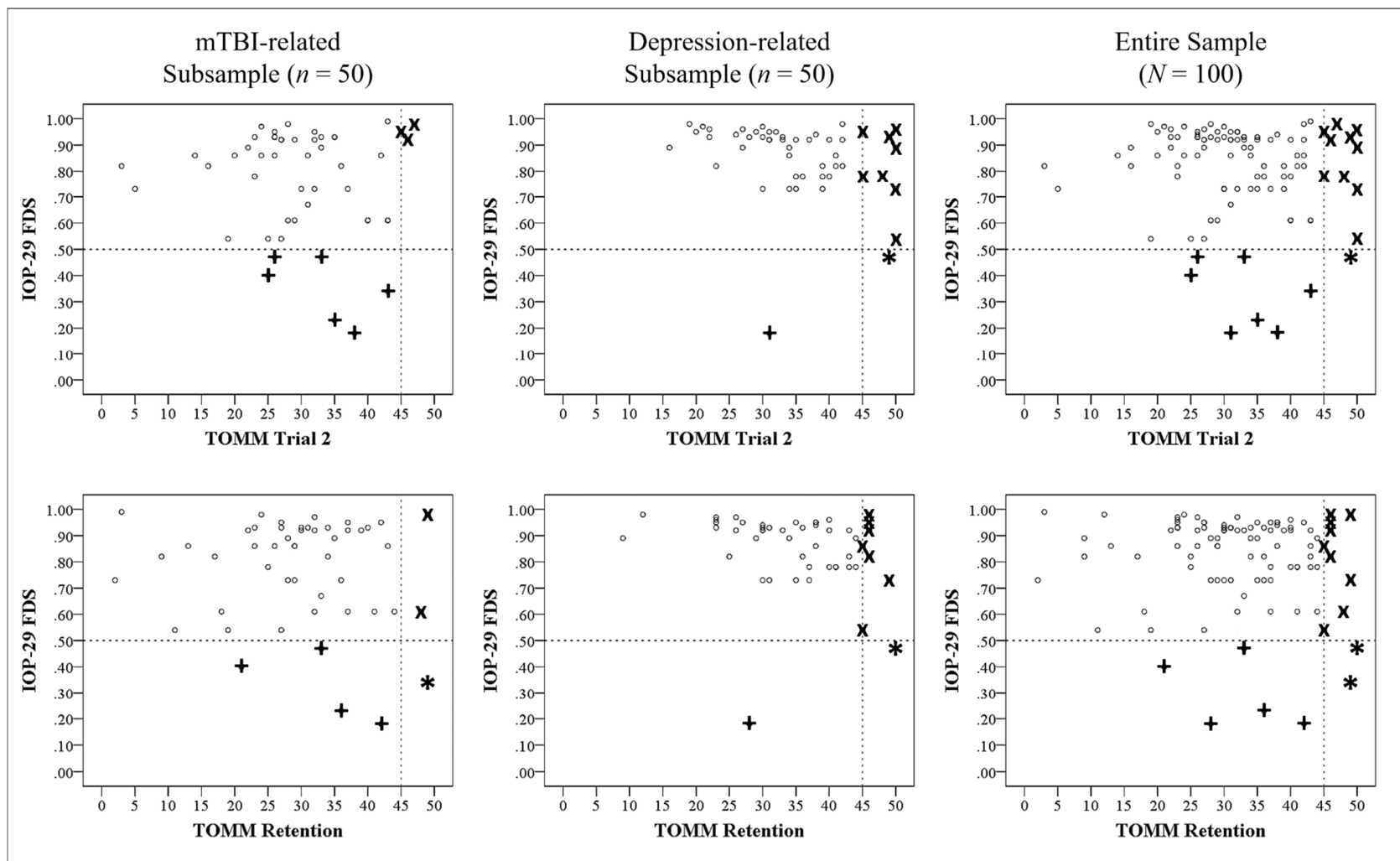
Note. False negatives on TOMM only are highlighted by the symbol "x"; false negatives on IOP-29 only are highlighted by the symbol "+"; false negatives on both tests are highlighted by the symbol "*". For TOMM, false negatives are cases with scores ≥ 45, for IOP-29 false negatives are cases with scores $< .50$.