

Concordance of the MMPI-2-RF Infrequent Responses Scale and TOMM in Detecting Overstated Pathology

Samuel T. Gontkovsky, Psy.D., and Charles J. Golden, Ph.D.

Abstract

This study examined the co-occurrence of response invalidity as measured by the Infrequent Responses (F-r) scale of the Minnesota Multiphasic Personality Inventory-2 Restructured Form (MMPI-2-RF) and the Test of Memory Malingering (TOMM) in 479 participants classified into forensic, quasi-forensic, and acquired brain injury groups. Although participant scores on the MMPI-2-RF F-r scale and TOMM were significantly correlated, concordance of response invalidity in the form of overstated pathology was only 13%. Comparable rates of potential feigning were found across groups on the MMPI-2-RF, but a larger percentage of participants in the forensic group obtained scores suggestive of possible feigning on both Trial 2 and the Retention Trial of the TOMM. Multivariate analyses also revealed that participants in the forensic group scored significantly below the quasi-forensic and brain injury groups on both Trial 2 and the Retention Trial of the TOMM. No significant mean score differences on the MMPI-2-RF F-r scale were found among groups. Results have meaningful implications for neuropsychological practitioners, suggesting limited concordance of symptom and performance validity as assessed by these instruments and highlighting potential concerns of performance invalidity associated with possible feigning in individuals referred for outpatient neuropsychological testing secondary to court order.

Introduction

Validity of data is the foundation for accurate interpretation of assessment results and appropriate determination of recommendations. Evaluation of response validity, both symptom and performance, has long been a critical component of forensic neuropsychological examinations but is increasingly recognized as important even in non-forensic neuropsychological evaluations. For example, recent research indicates that nonvalid performance validity testing is common among patients with multiple sclerosis referred for neuropsychological evaluation and reflects contributions from disability status and depressive symptoms (Galioto, Dhima, Berenholz, & Busch, 2020). The inclusion of validity tests alongside other measures, therefore, has become the standard of care in professional neuropsychological practice (Larrabee, 2012).

Although invalid test performance may reflect malingering, response styles are numerous and complex, with subtle differences distinguishing certain response styles from others (see Rogers & Bender, 2018). Further, failed response validity testing may occur in several contexts and

situations. Numerous factors that have been shown to influence test results include but are not limited to emotional distress, physical pain, fatigue, sleepiness, medication side effects, boredom, health beliefs, and language proficiency (Greher & Wodushek, 2017; Henry et al., 2018; Lippa, 2018).

Overstated pathology is common in forensic examinations. As described by Rogers and Bender (2018), overstated pathology may be categorized as malingering, factitious presentations, or feigning. According to their classifications, factitious presentations involve a desire to assume a sick role and require some unspecified internal motivation on the part of the individual. The key distinction between malingering and feigning involves the existence of external incentives. Feigning does not make assumptions about an individual's goals of deliberate fabrication or gross exaggeration of psychological or physical symptoms. It is introduced as a category of overstated pathology because standardized measures of response styles in the form of psychological tests have not been validated to assess specific motivations, and therefore cannot be used to establish malingering.

Extensive research has been conducted on response validity, but it remains unclear as to the cooccurrence of symptom invalidity and performance invalidity, particularly across groups differing in age, education, legal status, disability status, diagnosis, and numerous other variables. The aims of this exploratory investigation were: 1) To examine the concordance of symptom and performance validity indicators between two commonly used psychological measures; and 2) To compare overstated pathology (feigning) as measured by validity indicators across three groups, expected to exhibit varying motivation to suggest elevated psychological symptoms or impaired neurocognitive functioning.

Method

Participants. Participants for this retrospective archival study were selected from individuals referred for neuropsychological testing at a university-based outpatient psychology clinic located in the southeastern United States. For inclusion, participants were required to have completed the measures under investigation and to have been referred for testing either 1) by court order, 2) secondary to a request for school or work accommodations, or 3) following an acquired brain injury. Individuals mandated to complete neuropsychological testing by court order were provided with the name and contact information of the clinic as one of multiple options for completing the required examination. These participants mandated for testing by court order comprised the forensic group. Patients seen for neuropsychological testing secondary to a request for accommodations comprised a quasi-forensic group. The final group of participants consisted of patients referred for neuropsychological testing after an acquired brain injury. The final sample included 479 participants. Table 1 provides demographic characteristics for the participants. No significant differences were found among the groups above for gender, race, handedness, or education, but the forensic group was significantly older than the quasi-forensic and brain injury groups.

Measures. *Minnesota Multiphasic Personality Inventory-2-Restructured Form.* The Minnesota Multiphasic Personality Inventory-2 Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008) is a 338-item theoretically guided revision of the MMPI-2. Although using the same pool of items as the MMPI-2, the construction of this revision eliminated item overlap among scales and reduced validity concerns of prior versions of the scale. For this study, T scores for the Infrequent Responses (F-r) scale of MMPI-2-RF were examined. The F-r scale consists of 32 items designed to detect unusual or infrequent responses in the normative population, with higher scores

suggesting overreporting of symptoms. Elevations on the scale are considered a more general indicator of overstated pathology, sensitive to the overreporting of psychological, cognitive, or somatic symptoms (Ben-Porath, 2013). Based on the findings of Wygant et al. (2009), T scores on the F-r scale ≥ 90 were utilized in this study to suggest symptom invalidity.

Table 1

Sample Characteristics

	Total <i>n</i> = 479	Forensic <i>n</i> = 92	Quasi-forensic <i>n</i> = 112	Brain injury <i>n</i> = 275	χ^2 (<i>df</i>)	<i>p</i>
Frequency						
Gender					2.83 (4)	.590
Male	214	43	55	116		
Female	263	49	56	158		
Other	2	0	1	1		
Race					10.40 (6)	.110
White	270	57	65	148		
Black	62	7	19	36		
Latino/Hispanic	112	24	24	64		
Other	35	4	4	27		
Handedness					3.37 (4)	.500
Right	418	81	95	242		
Left	59	10	16	33		
Mixed	2	1	1	0		
		<i>M (SD)</i>			<i>F (df)</i>	<i>p</i>
Age (years)	33.16 (13.07)	37.29 (14.64) ^a	31.86 (12.61)	32.31 (12.46)	5.85 (2)	.003
Education (years)	13.86 (2.38)	13.97 (2.62)	13.88 (2.13)	13.82 (2.41)	0.13 (2)	.880

^aSignificantly different from Quasi-Forensic Group and Acquired Brain Injury Group.

Test of Memory Malingering. The Test of Memory Malingering (TOMM; Tombaugh, 1996) is a 50-item recognition test consisting of line drawings of common objects presented over two learning trials, each followed by a test of object recognition. An optional Retention Trial, which may be administered 15 minutes following trial two, consists of recognition only without a learning trial. Raw scores for the TOMM range from 0-50, with lower scores reflecting an increased likelihood of invalidity. Recommended guidelines for performance invalidity are raw scores below 45 on Trial 2 or the Retention Trial.

Procedure. Participants completed the MMPI-2-RF and TOMM as part of a battery of neuropsychological measures. Administration and scoring of tests were completed according to standardized procedures by doctoral students in clinical psychology under the supervision of a senior neuropsychologist. Statistical analyses were executed using IBM SPSS Statistics.

Results

Descriptive statistics were calculated to examine sample scores as a whole and participants by group on the F-r scale of the MMPI-2-RF and TOMM (Table 2). Pearson correlation coefficients were calculated to assess the relationships among the three trials of the TOMM and the MMPI-2-RF F-r scale for the sample. The F-r scale was significantly associated with all TOMM trials (Trial 1, $r = -.13$, $p = .004$; Trial 2, $r = -.14$, $p = .002$; Retention, $r = -.21$, $p < .001$).

Frequency analyses revealed that 7.3% ($n = 35$) of the total sample generated scores on the MMPI-2-RF F-r scale suggestive of symptom invalidity. Of these participants, seven were from the forensic group, nine were from the quasi-forensic group, and 19 were from the brain injury group. Considering these numbers in terms of relative percentages by group, 7.6% (7/92) of the forensic group, 8% (9/112) of the quasi-forensic group, and 6.9% (19/275) of the brain injury group demonstrated symptom invalidity in the form of potential feigning. In terms of concordance with performance invalidity as measured by the TOMM, 11.4% and 17.1% of these 35 participants obtained scores suggesting performance invalidity as measured by the second trial and the Retention Trial of the TOMM, respectively. It also should be noted that the four participants obtaining scores below 45 on Trial 2 of the TOMM also obtained scores below 45 on the Retention Trial of the TOMM.

Table 2

Sample Test Scores and Test Comparisons

	<i>M (SD)</i>				<i>F (df)</i>	<i>p</i>
	Total	Forensic	Quasi-forensic	Brain injury		
MMPI-2-RF Scale					1.35 (2)	.260
F-r	62.91 (17.91)	63.98 (17.14)	64.81 (19.80)	61.78 (17.33)		
TOMM					2.73 (6)	.012
Trial 1	46.79 (4.31)	46.26 (5.82)	47.38 (2.84)	46.72 (4.21)	1.79	.170
Trial 2	49.30 (2.67)	48.55 (5.04) ^a	49.71 (1.09)	49.39 (1.80)	5.17	.006
Retention	49.32 (2.66)	48.45 (4.85) ^a	49.79 (.89)	49.41 (1.97)	7.10	.001

^aSignificantly different from Quasi-Forensic Group and Acquired Brain Injury Group

In contrast, 4.2% ($n = 20$) of the total sample obtained scores on Trial 2 of the TOMM suggestive of performance invalidity. Of these participants, eight were from the forensic group, one was from the quasi-forensic group, and 11 were from the brain injury group. Thus, 8.7% (8/92) of the forensic group, < 1% (1/112) of the quasi-forensic group, and 4% (11/275) of the brain injury group demonstrated performance invalidity in the form of feigning as measured by Trial 2 of the TOMM. In terms of concordance with symptom invalidity measured by the MMPI-2-RF, 20% of these 20 participants generated scores suggesting performance invalidity as measured by the F-r scale. The mean MMPI-2-RF F scale score for these 20 participants was 74.05 ($SD = 24.71$), with four participants scoring greater than 100.

Only 3.5% ($n = 17$) of the total sample obtained scores on the Retention Trial of the TOMM suggestive of performance invalidity. Of these participants, nine were from the forensic group, one was from the quasi-forensic group, and seven were from the brain injury group. Thus, 9.8% (9/92) of the forensic group, < 1% (1/112) of the quasi-forensic group, and 2.5% (7/275) of the brain injury group demonstrated performance invalidity in the form of feigning as measured by the Retention Trial of the TOMM. In terms of concordance with symptom invalidity as measured by the MMPI-2-RF, 35% of these 17 participants generated scores suggesting performance invalidity measured by the F-r scale. The mean MMPI-2-RF F scale score for these 17 participants was 83.71 ($SD = 30.31$), with five participants scoring greater than 100.

Overall, 11.3% ($n = 54$) of the sample generated at least one score suggestive of overstated pathology. Of these participants, 15 were from the forensic group, ten were from the quasi-forensic group, and 29 were from the brain injury group. Considering these numbers in terms of relative percentages by group, 16.3% (15/92) of the forensic group, 8.9% (10/112) of the quasi-forensic group, and 10.5% (29/275) of the brain injury group demonstrated response invalidity in the form of overstated pathology. Examining the frequencies of participants with more than one score suggestive of feigning, 2.9% ($n = 14$) of the sample as a whole had at least two scores suggestive of overstated pathology, and < 1% ($n = 4$) of the total sample had all three scores suggestive of overstated pathology. It should be noted that only 2.5% ($n = 12$) of the total sample obtained invalid scores (< 45) on both the second and Retention Trial of the TOMM. The overall concordance of symptom invalidity and performance invalidity based on a T score of ≥ 90 on the MMPI-2-RF F-r scale in conjunction with at least one invalid score of < 45 on either the second or Retention Trial of the TOMM was 13% (7/54).

An analysis of variance and a multivariate analysis of variance were computed to compare group differences in mean MMPI-2-RF F-r and TOMM scores, respectively. Given the significant differences in age between the forensic group and other groups, correlations were calculated between age and all test variables to determine the potential influence of age on scores; no significant relationships were found, and therefore age was not used as a covariate in these analyses. Results of these computations are presented in Table 2. No significant mean score differences were found among groups on the MMPI-2-RF F-r scale. Significant mean score differences were found between groups, however, on both the second and Retention Trial of the TOMM, with the forensic group scoring lower than both quasi-forensic and brain injury groups.

Discussion

Contrasted with previously reported estimated base rates of probable malingering and symptom exaggeration (Mittenberg, Patton, Canyock, & Condit, 2002), results of this investigation suggest that overstated pathology in the form of feigning is a relatively uncommon occurrence in outpatient neuropsychological testing; approximately 11% of our sample obtained at least one score indicating possible response invalidity. Only 7.3% of our sample generated MMPI-2-RF F-r scale scores suggestive of symptom invalidity, and only 4.2% and 3.5% of our sample obtained scores on the second or Retention Trial of the TOMM respectively, suggesting performance invalidity. Although the F-r scale score was significantly negatively correlated with all TOMM scores, of the 54 participants in our study having a profile with at least one validity indicator suggestive of feigning, the concordance rate between the MMPI-2-RF and TOMM was only 13%. This particular finding may indicate that the tests are measuring different constructs, as was suggested by McCaffrey, O'Bryant, Ashendorf, and Fisher (2003). Indeed, although the F-r scale is purported to be a more general indicator of overstated pathology and sensitive to the

overreporting of various symptoms, including those in the cognitive domain (Ben-Porath, 2013), the TOMM is more specific to memory. Another consideration is that individuals undergoing neuropsychological testing simply do not over-pathologize across all aspects of functioning but rather may focus on one or more specific symptoms. As noted by Sharf, Rogers, Williams, and Henry (2017), F-r also tends to be relatively ineffective in distinguishing feigners from genuine patients with certain diagnoses, which may contribute to the rates found in this investigation.

Although comparable rates of overstated pathology were found among groups on the MMPI-2-RF, a larger percentage of participants in the forensic group demonstrated potential feigning on both Trial 2 and the Retention Trial of the TOMM in comparison to the quasi-forensic and brain injury groups. Participant litigation status also emerged in multivariate analyses as a significant variable in determining performance invalidity as measured by mean scores on the TOMM, with the forensic group scoring significantly lower on the second or Retention Trial of the TOMM as compared with the quasi-forensic and brain injury groups. This particular finding is more in line with the results of Mittenberg et al. (2002) with respect to higher rates of symptom exaggeration seen among individuals in litigation and compensation-seeking cases relative to medical or psychiatric cases.

This study has several important limitations, including its differential prevalence design. The participants, although homogeneously grouped by referral, were also heterogeneous in terms of presenting problem, diagnosis, and other features. We had no independent means for classifying participants according to various factors that may have influenced scores on the measures, such as degree of emotional distress or level of cognitive functioning. Given the archival nature of this investigation, we could not control for the various other factors mentioned earlier that could have impacted response validity during the testing process, such as physical pain, sleepiness, and medication side effects. Future research should address these limitations and include other concurrent measures of response validity. Additional research also is needed to examine rates of symptom invalidity in quasi-forensic cases, those not involved in litigation or seeking compensation but requesting some form of non-monetary accommodation. However, from an ecologically valid perspective, our sample represents referrals commonly seen in a general outpatient neuropsychological testing clinic, and findings provide data that may be useful for both general and forensic neuropsychological practitioners.

About the Authors

Samuel T. Gontkovsky, Psy.D. is with the Department of Psychology, Twin Valley Behavioral Health Care in Columbus, OH, USA. Correspondence concerning this article should be addressed to Dr. Gontkovsky at [sgontkovsky\(at\)hotmail.com](mailto:sgontkovsky(at)hotmail.com).

Charles J. Golden, Ph.D. is with the College of Psychology, Nova Southeastern University in Fort Lauderdale, FL, USA.

Author Note

This project received no financial support from outside funding agencies. Relevant ethical guidelines were followed throughout the project. This retrospective study utilizing archival data was conducted according to the 1964 Helsinki declaration and its subsequent amendments and the ethical principles of the American Psychological Association (2002).

References

- American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist, 57*, 1060-1073. <https://doi.org/10.1037/0003-066X.57.12.1060>
- Ben-Porath, Y. S. (2013). Forensic applications of the Minnesota Multiphasic Personality Inventory-2 Restructured Form. In R. P. Archer and E. M. A. Wheeler (Eds.), *Forensic use of clinical assessment instruments* (pp. 63-107). New York, NY: Routledge.
- Ben-Porath, Y. S., & Tellegen, A. (2008). *Minnesota Multiphasic Personality Inventory-2 Restructured Form: Manual for administration, scoring, and interpretation*. Minneapolis, MN: University of Minnesota Press.
- Galioto, R., Dhima, K., Berenholz, O., & Busch, R. (2020). Performance validity testing in multiple sclerosis. *Journal of the International Neuropsychological Society, 26*, 1028-1035. <https://doi.org/10.1017/S1355617720000466>
- Greher, M. R. & Wodushek, T. R. (2017). Performance validity testing in neuropsychology: Scientific basis and clinical application—A brief review. *Journal of Psychiatric Practice, 23*, 134-140. <https://doi.org/10.1097/PRA.0000000000000218>
- Henry, G. K., Heilbronner, R. L., Suhr, J., Gornbein, J., Wagner, E., & Drane, D. L. (2018). Illness perceptions predict cognitive performance validity. *Journal of the International Neuropsychological Society, 24*, 735-745. <https://doi.org/10.1017/S1355617718000218>
- Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society, 18*, 625-630. <https://doi.org/10.1017/s1355617712000240>
- Lippa, S. M. (2018). Performance validity testing in neuropsychology: A clinical guide, critical review, and update on a rapidly evolving literature. *The Clinical Neuropsychologist, 32*, 391-421. <https://doi.org/10.1080/13854046.2017.1406146>
- McCaffrey, R. J., O'Bryant, S. E., Ashendorf, L., & Fisher, J. (2003). Correlations among the TOMM, Rey-15, and MMPI-2 validity scales in a sample of TBI litigants. *Journal of Forensic Neuropsychology, 3*, 45-53. https://doi.org/10.1300/J151v03n03_03
- 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. <https://doi.org/10.1076/jcen.24.8.1094.8379>
- Rogers, R., & Bender, S. D. (2018). *Clinical assessment of malingering and deception* (4th ed.). New York, NY: The Guilford Press.
- 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*, 441-455. <https://doi.org/10.1007/s10862-017-9590-1>
- Tombaugh, T. N. (1996). *Test of Memory Malingering (TOMM)*. Toronto, Ontario: Multi Health Systems.

Wygant, D. B., Ben-Porath, Y. S., Arbisi, P. A., Berry, D. T. R., Freeman, D. B., & Heilbronner, R. L. (2009). Examination of the MMPI-2 Restructured Form (MMPI-2-RF) validity scales in civil forensic settings: Findings from simulation and known group samples. *Archives of Clinical Neuropsychology*, 24, 671-680. <https://doi.org/10.1093/arclin/acp073>