

ABSTRACT BOOK

8th European Conference on Symptom Validity Assessment (SVA)

8–10 October 2026

Málaga, Spain



Contents

Conference Committees	4
Symposia Abstracts.....	5
Technology & Test Security (chair: Anselm Fuermaier)	6
Technology and implications for test security in clinical and forensic practice (Fuermaier)	7
Self-administered remote neuropsychological testing: Suitable alternative to in-person assessment? (Teßmann et al.).....	8
The differential effects of sophisticated AI coaching on symptom and performance validity (Crişan & Voin).....	10
Managing test security in the digital age: A case for abstinence (Erdodi & Olla).....	11
Applying Malingering Frameworks and Multidimensional Evidence: Case Studies in Symptom Validity (chairs: Vicki Hall & Michael Chafetz)	12
Careful review of records: At the heart of validity analysis (Chafetz)	13
From atypical test scores to non-credible presentation: A case study of multidimensional formulation (Hall)	14
Two cases of invalid performance in the memory clinic; multidimensional validity assessment and feedback (Roor)	15
When formal tests of validity are normal within the context of clear potential exaggeration (Shindell)	16
Advances in the Assessment of Symptom and Performance Validity (chair: Esteban Puente-López)	17
Psychometric difficulties in detecting malingering in cases of PTSD or developmental trauma (Garrido Gaitán)	18
Performance validity in women who are victims of intimate partner violence (Hidalgo Solarte & Pérez García)	19
Systematic literature review on symptom validity testing in criminal forensic contexts (Silva)	20
The utility of response time in forced-choice recognition PVTs: A systematic review and meta-analysis (Braw)	21
Testing the limits of validity indicators: Test coaching in negative and positive response bias assessment (Pignolo et al.) ...	22
Clinical Tests Meet Symptom Validity: A New Signal or Just an Echo? (chair: Irena Bošković)	23
The evolution of embedded SVTs using U.S. veteran samples (Shura)	24
New methods for symptom validity testing in whiplash associated disorders (Puente-López)	25
Embedded performance and symptom validity indicators in early retirement claimants – a comparison (Lehrner).....	26
Embedded validity indicators in the PCL-5 for detecting feigned intimate partner violence-related PTSD (Ales & Giromini) ..	27
Keynote & Special Presentation	28
Symptom validity assessment with the MMPI: Lessons learned and future directions (Ben-Porath)	29
Building the future of the SVA community: Sustainability, Psychological Injury and Law, and more (Giromini).....	30
Spotlights	31
Distorted symptom presentation and its consequences (Dandachi-FitzGerald).....	32
How to defend against attacks on PVT and SVT data in forensic cases (Boone)	33
Changes in validity test scores after COVID: Reflection of general distress or something else? (Harrison et al.)	34
Distinction without a difference: Do SVTs differentiate symptom presentations? (Ingram)	35
Roundtable.....	36
Present and future validity assessment in European countries (Merten).....	37
Poster Abstracts	38
Are feigning screens “fit to stand trial”? (Krishnan et al.)	39
Assessing symptom credibility in PTSD: A multi-method simulation study in an Italian sample (Bosco et al.).....	40

Associations between SVTs and PVTs and MMPI-3 measures of personality disorders (Holmes et al.)	41
Associations between the IOP-29, IOP-M, and MCI in non-head injury disability sample (Ross et al.)	42
Conceptualizing “good old days” bias from a dialectical behavior therapy perspective (Ben-Porath & Ben-Porath)	44
Convergence among psychometric findings, record-based inconsistencies, and medication adherence (Tatti et al.)	45
Criteria-based validity assessment in legal cases involving pension and accident insurance (Klett et al.)	46
Do post-pandemic increases in AD/HD symptom validity indicator violations differ by gender? (Davin et al.)	48
Do post-pandemic increases in PAI symptom validity indicator violations differ by gender? (Davin et al.)	49
Evaluating continuous performance tests as embedded measures of performance validity (Toptas et al.)	50
Extremely elevated self-reported autism symptoms in postsecondary assessment (Davin et al.)	52
Feasibility of eye-tracking-based indicators of performance validity in digital cognitive testing (Schneider)	53
MMPI item endorsements can predict multidimensional demographic variables and validity scales (Clemons et al.)	54
MMPI-2 symptom validity patterns in U.S. veterans (Childers et al.)	56
MMPI-2 vs. PAI-A validity scales profiles in a case study involving a functional neurological disorder (Nicholls)	58
Outcomes of symptom overreporting and cognitive underperformance in ADHD assessments (Raasch et al.)	59
Personality assessment inventory CB-SOS as a predictor of performance validity test failure (Harrison & Armstrong)	60
PVT use in neuropsychological assessment in the UK and Republic of Ireland (O’Brien et al.)	61
Strategies for faking good: Concealed psychosis in UAP reports (Costopoulos et al.)	62
Symptom over- and underreporting in forensic inpatients and undergraduates (Helvoort et al.)	63
The benefits of being ill (Wise & Beezhold)	64
The state of symptom validity assessment in North America (Matthew et al.)	69
Use of the Morel Emotional Numbing Test (MENT) in the detection of feigned psychopathology (Pasqualini et al.)	71
Validation of the Blockwise Attention Test (BAT) “Palamedes” (Pals)	72
Validation of the Groningen Effort Test (GET) on Italian samples in a simulation design (Losavio et al.)	73

Conference Committees

CONGRESS ORGANIZERS

Anselm Fuermaier

The Netherlands

Luciano Giromini

Italy

SCIENTIFIC PROGRAMME COMMITTEE

Anselm Fuermaier

The Netherlands

Luciano Giromini

Italy

Esteban Puente-López

Spain

Brechje Dandachi-FitzGerald

The Netherlands

Vicki Hall

United Kingdom

Irena Bošković

Serbia

Janina Teßmann

The Netherlands

Iulia Crişan

Romania

Andrea Plohm

Switzerland

Johann Lehrner

Austria

Thomas Bodner

Austria

Thomas Merten

Germany

Harald Merckelbach

The Netherlands

POSTER COMMITTEE

Jeroen Roor (chair)

The Netherlands

ORGANIZATIONAL PARTNER

Studio RiPsi

Italy

SYMPOSIA ABSTRACTS

The following abstracts are presented as part of the symposium programme.

Technology & Test Security

Chair: Anselm Fuermaier



Technology and implications for test security in clinical and forensic practice

Anselm B.M. Fuermaier

Department of Clinical and Developmental Neuropsychology, University of Groningen, the Netherlands

Abstract

Neuropsychological testing plays a critical role in detecting cognitive impairment, monitoring functional changes over time, and guiding diagnostic and treatment decisions. A foundational assumption in this process is that test scores accurately reflect an individual's current cognitive functioning, and that psychological questionnaires rely on truthful and accurate symptom reporting. At the same time, rapid advances in technology are reshaping the field of neuropsychological and psychological assessment. The growth of teleneuropsychology, remote and hybrid assessment models, digital test delivery platforms, self-administered online tools, and artificial intelligence (AI) - assisted applications has broadened the scope and reach of forensic and clinical practice. These innovations offer unprecedented opportunities to enhance accessibility, especially in situations where traditional in-person testing is impractical. However, these technological shifts introduce profound challenges for upholding test security, maintaining psychometric validity, and ensuring ethical standards in both clinical and forensic settings. The widespread availability of digital information increases the risk that examinees may access test content or strategies to manipulate outcomes. Emerging evidence indicates that AI tools can enhance individuals' ability to feign cognitive or psychological symptoms, yet their effectiveness has still to be explored. As AI models continue to evolve, this issue requires precaution and rigorous scientific investigation.

This symposium includes three presentations from outstanding scholars who address the topic of technology and test security from different perspectives, including the potential and limitations of self-administered remote neuropsychological assessment (Teßmann), the effects of AI coaching on validity assessment (Crişan), and a case for abstinence from online digital assessment practices (Erdodi).

Self-administered remote neuropsychological testing: Suitable alternative to traditional in-person assessment?

Janina Teßmann¹, Luciano Giromini², & Anselm B. M. Fuermaier¹

¹Department of Clinical and Developmental Neuropsychology, University of Groningen, Groningen, the Netherlands

²Department of Psychology, University of Turin, Turin, Italy

Abstract

Objectives: Neuropsychological tests are ideally performed under the supervision of a qualified examiner. However, constraints such as accessibility issues or staff shortages can make in-person testing unfeasible. The presented study investigated whether self-administered online cognitive testing could serve as a suitable alternative to in-person assessments.

Methods: The sample consisted of 395 individuals ($M_{\text{age}} = 25.0$ years): 252 undergraduates ($M_{\text{age}} = 20.0$ years) and 143 community members ($M_{\text{age}} = 33.7$ years). Participants completed the Adaptive Working Memory Test (SPAN) and The Attention and Concentration Test (TACO) remotely without any supervision. Additionally, participants performed one symptom validity test and two performance validity tests (PVTs), namely the Inventory of Problems-29 (IOP-29) with its memory module (IOP-M) and the Fifteen-Item Test (FIT). Upon completion, participants reported information about the assessment context (i.e., input device, presence of others, background noise, interruptions) and personal states (i.e., motivation, effort, tiredness, prior-night sleep, phone engagement, time pressure). It was tested whether the study sample performed comparably to the normative sample and whether PVT failure was related to lower cognitive performance. Lastly, associations between contextual factors and cognitive performance were explored.

Results: The study sample scored higher than the normative sample on measures of attention but slightly lower on working memory tasks. In total, 1.3% of the participants showed elevated symptom validity scores, 5.6% failed one of the PVTs, and none of them failed both. There were no differences in cognitive test scores between individuals who failed one PVT and individuals who passed both PVTs. Moreover, only perceived time pressure was meaningfully associated with decreased attentional performance.

Conclusions: The results tentatively suggest that cognitive tests performed remotely without supervision can provide valid clinical data. Thus, this approach could complement in-person testing, simplify cognitive screening, and facilitate repeated assessments, thereby reducing resource requirements in

clinical practice. However, individuals with pronounced cognitive impairment may be more affected by suboptimal test conditions. Future research should therefore examine self-administered online cognitive testing in these populations.

The differential effects of sophisticated AI coaching on symptom and performance validity tests: A replication of Voin et al. (2025)

Iulia Crișan & Samira Voin

Department of Psychology, West University of Timișoara, Timișoara, Romania

Abstract

The present study replicated previous findings regarding the ability of symptom validity tests (SVTs) and performance validity tests (PVTs) to detect symptom-coached and AI-coached experimental feigning. Compared with the original study, we increased the sample size, added an additional SVT, and refined the instructions for the AI-coached group by using an AI-generated information protocol, following Fuermaier and Niesten (2025). The total sample consisted of 195 undergraduates and community participants (84 community members and 111 undergraduates; 127 [65.1%] females; M Age = 23.86, SD = 8.50; M Education = 13.92, SD = 2.09). Participants were assigned to an AI-assisted experimental feigning group ($n = 62$), a symptom-coached experimental feigning group ($n = 64$), or a control group ($n = 69$). The two feigning groups were instructed to feign a mild traumatic brain injury (mTBI) to evade work responsibilities. All participants completed the Inventory of Problems-29 and its memory module (IOP-29 and IOP-29-M), the Forgetfulness Assessment Inventory (FAI), and the Memory of Objects and Digits and Examination of Memory Malingering Test (MODEMM).

Results indicated differences between the two feigning groups on SVTs, but not on PVTs. AI-coached participants obtained lower IOP-29 False Disorder Score (FDS) values ($U = 1428.5$, $Z = -2.71$, $p < .05$) and lower FAI scores ($U = 1183$, $Z = -3.91$, $p < .001$) than symptom-coached participants. In contrast, no significant differences were observed on the IOP-29-M or MODEMM indicators. All validity indicators successfully discriminated between experimental feigners and control participants at previously established cutoffs. However, SVT cutoffs had to be lowered in the AI-coached group to achieve acceptable sensitivity levels (.52–.88). These findings suggest that sophisticated AI coaching may reduce the detection accuracy of SVTs, whereas it does not appear to have a significant impact on PVT performance.

Keywords: malingering, feigning, Artificial Intelligence, coaching, validity testing

Managing test security in the digital age: A case for abstinence

Laszlo A. Erdodi^{1,2} & Joelle Olla¹

¹Psychology Department, University of Windsor, Windsor, Ontario, Canada

²Star UBB Institute, Department of Psychology, Babeş-Bolyai University, Cluj-Napoca, Romania

Abstract

Protecting neuropsychological tests from unauthorized access is a fundamental duty of all clinicians. Historically, maintaining test security has been a straightforward task: restrict access to hard copies of stimulus booklets, record forms, technical manuals, and scoring templates to qualified personnel who are ethically and legally bound to protect the content against unauthorized disclosure. In the digital age when remote/virtual assessment is becoming widespread due to the combination of convenience, increased social acceptance, and availability of digital infrastructure, the task has become more complex. Concurrently, security threats to digitally stored information (including test items, administration, and scoring protocols) have multiplied and continue to evolve (possibly faster than safety measures). We argue that the only fail-safe approach to protecting sensitive information from accidental disclosure or cyberattacks is bypassing the digital highway and performing assessments the old-fashioned way (face-to-face, paper-and-pencil, hard copies, and records stored in secure physical location).

Applying Malingering Frameworks and Multidimensional Evidence: Case Studies in Symptom Validity

Chair: Vicki Hall



Careful review of records: At the heart of validity analysis

Michael Chafetz

Independent Practice, New Orleans, Louisiana, United States

Abstract

A careful review of records gives the examiner a foundation of objective facts for opinion to answer the referral questions in a forensic case and for a determination of whether the case meets criteria for invalidity. A review of records is important in addressing the unreliability of claimant self-report, particularly because human memory is changeable over time. Every part of one's exam must be scientifically defensible, and particular records support or refute claims of brain injury, give more relevant reasons as to other causes of cognitive decline, and provide the basis for marked discrepancies and compelling inconsistencies regarding the current claims versus the history of the case. This workshop will show the use of records for scientific objectivity regarding the determination of invalidity in contested cases.

From atypical test scores to a non-credible presentation: A case study in multidimensional formulation

Vicki Hall

Neuromindworks, Sutton Coldfield, West Midlands, United Kingdom

Abstract

Failed validity tests, from a psychometric perspective, indicate that an examinee's test scores are atypical relative to an appropriate bona fide normative sample. However, atypical performance alone does not establish that a presentation is invalid. Invalidity means that the examinee's presentation is "not fully explained by brain dysfunction; is not reasonably attributable to moderating or confounding factors (e.g., age, education, fatigue, psychological conditions); and is substantially worse than, or qualitatively different from, patterns seen in genuine neurological disorders" (Heilbronner et al., 2009). Consequently, validity test findings require interpretation within the broader context of a multidimensional clinical formulation rather than in isolation. The diagnostic framework proposed by Sherman et al. (2020) provides a structured method for integrating multiple independent sources of evidence when evaluating invalidity and malingering. Beyond interpreting the psychometric significance of failed validity tests, the framework also assists in determining whether the evidence supports a broader non-credible presentation that generalises beyond the test scores to the examinee's history, behaviour, reported symptoms, and everyday functioning. This case study demonstrates how validity test findings can be interpreted within this multidimensional framework.

Two cases of invalid performance in the memory clinic; multidimensional validity assessment and feedback strategies

Jeroen J. Roor

Department of Medical Psychology, VieCuri Medical Center, Venlo, the Netherlands
Maastricht University, Maastricht, the Netherlands

Abstract

The outpatient memory clinic serves the (early) detection, differential diagnostic clarification and treatment of patients with cognitive disorders. Mostly in elderly patients with a specific focus on dementia. Although biological staging and biomarkers are introduced in the updated Alzheimer's disease dementia (ADD) criteria, at its core, the diagnostic algorithm remains clinical. Neuropsychological assessment plays a crucial role in examining cognitive and behavioral decline – one of the major clinical diagnostic pillars of ADD.

In this brief presentation, two cases from the outpatient memory clinic are presented. The first case concerns a woman of middle age who was already diagnosed with and treated for mild cognitive impairment (MCI). A new diagnostic work-up revealed clear inconsistencies between the disabilities and symptoms reported by the patient, and (intact) level of daily functioning. Also, medical record review showed that the patient had some parts of her medical records deleted by the hospital's medical administration service. Specifically, parts that might have shed a different light on the origins of the symptoms that led to an MCI diagnosis. The second case concerns a man of middle age who presented with significant loss of memory functioning. The differential diagnosis of the treating physician consisted of early ADD. However, the acute and recent onset of the cognitive and behavioral symptoms were inconsistent with the gradual onset that is typical for ADD. This recent onset was corroborated by information from the medical file and by collateral information provided by a relative, altering the diagnostic differential hypotheses. Both cases showcase the importance of a multidimensional approach to validity assessment. Finally, invalid performance was communicated to both patients. The utilized feedback strategies - based upon two published feedback models – and their impact on the diagnostic process will also be discussed.

When formal tests of performance and symptom validity are normal within the context of clear potential exaggeration.

Steve Shindell

Georgia Neuropsychology, Atlanta, Georgia, United States

Abstract

Formal tests of performance and symptom validity are an essential part of any evaluation, but given the prior wide exposure of these devices to many patients involved in medio/legal procedures, the base rates of failure on these instruments changes over time, and the clinician needs to continue to use other aspects of the evaluation to ensure an accurate diagnosis. Examination of attorney's websites sometimes actually name the specific tests that will be potentially given, and data will be shown in this presentation as to how often the patient's attorney either discusses this with their client as well as actually practices taken the tests with them prior to an examination.

The adage "Clinicians diagnose, tests do not diagnosis" is vital in any analysis, and while validity tests are a vital aspect of our examination, clinicians utilize a number of other factors in coming to their diagnostic conclusion.

Examinations involve a complete review of all available documents, a complete history of the patient, a thorough examination of the actual stressors (emphasizing on the behavioral aspects of the patient at the time of the stressors as well as the timeline of events), as well as formal measures of validity. It is important for the clinician to not be laser focused on the binomial question of PTSD or NOT but rather be open to all possible diagnoses at the onset and use their training to eliminate inapplicable diagnoses throughout the examination until they form their diagnostic conclusion. Formal tests of validity are simply an additional means of carrying on this process.

This paper will discuss the role of the clinician, the realities of the exposure of various validity measures, and show case studies where the clinician can use the rest of the evaluation to come to an accurate conclusion.

Advances in the Assessment of Symptom and Performance Validity

Chair: Esteban Puente-López



Psychometric difficulties in detecting malingering in cases of complex Post-Traumatic Stress Disorder or developmental trauma

Elena Garrido Gaitán

Universitat Autònoma de Barcelona, Barcelona, Spain

Abstract

Repeated and prolonged childhood trauma severely alters child development, with long-term effects on personality. Given the insufficiency of classic Post-Traumatic Stress Disorder (PTSD) to explain these alterations, the diagnosis evolved toward Complex PTSD (C-PTSD), focusing on identity and relational capacities, although it was not initially formulated from a purely developmental perspective. Consequently, in 2005, the concept of Developmental Trauma Disorder (DTD) emerged as a specific entity for childhood; its core lies in continuous stressors that cause attachment trauma, directly impacting the child's development, bonds, and personality. This multidimensional symptom constellation is often misdiagnosed during adolescence or adulthood, as it does not fit traditional diagnostic manuals. This clinical dispersion leads to the overmedication and overdiagnosis of victims and, in the forensic field, hinders the establishment of a unified causal link between victimization and the assessed symptoms. Furthermore, it generates severe psychometric difficulties in forensic evaluation, since the complexity of the trauma elevates validity scales on standard personality tests. This extreme response—which is inherently a cry for help—is often misinterpreted as malingering or exaggeration, thereby increasing the risk of discrediting valid symptomatology in court. Thus, an analysis of 12 cases of C-PTSD and developmental trauma in adults is presented, who were victims of continuous childhood violence by their caregivers or attachment figures. The analysis focuses on the elevation of validity scales in classic measures of personality and psychopathology, providing more comprehensive evaluation strategies suited for analyzing malingering or exaggeration in these types of cases, with C-PTSD and DTD being necessary diagnoses to understand traumatology.

The evaluation of performance validity in women who are victims of intimate partner violence

Miguel Pérez García & Isabela Hidalgo Solarte

Mind, Brain and Behaviour Research Center (CIMCYC), Granada, Spain

University of Granada, Granada, Spain

Abstract

Intimate partner violence (IPV) is a major public health and social problem that can be associated with significant psychological, neuropsychological, and physical consequences. Neuropsychological assessment may contribute to the objective characterization of cognitive sequelae in women who have experienced IPV. However, in forensic contexts, establishing the validity of cognitive performance is essential to ensure the appropriate interpretation of assessment findings. Despite the widespread use of performance and symptom validity tests in other medico-legal populations, their accuracy and applicability in women who are victims of IPV remain insufficiently investigated. At this conference, a research proposal will be presented whose objective is to evaluate the accuracy of the classification of various performance indicators in this population. An observational cross-sectional study with a simulation design will include three groups: women who have experienced IPV, women without a history of IPV or psychological or neuropsychological impairment instructed to simulate symptoms without specific preparation, and women under the same conditions who will additionally receive information about the symptoms to be simulated and a warning regarding excessive symptom presentation. Participants will complete the Inventory of Problems–29 (IOP-29), Inventory of Problems–Memory (IOP-M), Forgetfulness Assessment Inventory (FAI) and Coin in Hand Test–Extended Version (CIH-Extended). Between-group differences will be examined using one-way ANOVA with Bonferroni post-hoc comparisons, while classification accuracy will be assessed through sensitivity, specificity, positive and negative likelihood ratios, and predictive values. Hierarchical logistic regression analyses will also examine the incremental validity of the measures. This study is expected to contribute to the development of empirically supported procedures for assessing performance validity in women who have experienced IPV, with potential implications for forensic neuropsychological assessment.

Systematic literature review on symptom validity testing in criminal forensic contexts

Emmanuel Silva

University of Coimbra, Coimbra, Portugal

Abstract

Symptom validity tests (SVTs) are widely used in psychological assessment to safeguard data validity, particularly in forensic settings. However, compared to performance validity tests, the empirical literature on SVTs remains relatively scarce. Given the lack of systematic reviews on SVTs in criminal forensic settings, we systematically examined the representation of SVTs in criminal forensic and related research literature, which measures and cutoffs were used, how many were included in assessment protocols, and how fail rates varied across common referral questions. The review followed PRISMA guidelines and the PICO framework. A total of 92 studies met the inclusion criteria, resulting in a pooled sample of 73,013 participants. Assessment protocols included an average of 2.92 SVTs, however, 31 of the 92 studies used only a single SVT. We found that embedded SVTs were the most frequently used validity measures. Average SVT fail rates were generally consistent with the literature but varied widely across SVTs, cutoff scores, and referral questions. The high variability in cutoff scores for embedded validity scales appeared to contribute substantially to the variance in fail rates. Additionally, statistically significant effects of sociodemographic and other person-related variables on SVT results were reported in 25% of the reviewed studies. These findings highlight the need for wider dissemination of information regarding both the importance of SVTs in psychological assessment and current best-practice guidelines for their use. At the same time, they indicate growing awareness of and interest in SVT research in recent years.

The utility of response time in forced-choice recognition performance validity tests: A systematic review and meta-analysis

Yoram Braw

University of Ariel

Abstract

Performance validity tests (PVTs) are essential in clinical and forensic neuropsychological evaluations. While traditional PVTs rely primarily on accuracy scores, response time (RT) has garnered attention as a supplementary validity marker. This systematic review and meta-analysis evaluated the utility of RT metrics in forced-choice recognition memory PVTs (FCRM-PVTs), contrasted RT with accuracy metrics, and examined key moderators. Following PRISMA and Cochrane guidelines, searches across five major databases identified 31 eligible studies ($N=3,096$) examining measures such as the VSVT, WRMT-Words, WMT, and ASTM. Twenty studies were quantitatively pooled using random-effects models, and 11 were synthesized narratively. Across all FCRM-PVTs, non-credible performance was consistently characterized by slower response times. In the Victoria Symptom Validity Test (VSVT), large RT effects emerged across both simulation and clinical known-groups designs. Although accuracy yielded larger group differences on difficult items, it exhibited substantial between-study heterogeneity, whereas RT demonstrated more stable and predictable effects. On easy items, RT and accuracy performed comparably. Qualitative synthesis indicated that intra-individual RT variability is elevated in simulators but shows reduced discriminability in clinical populations where genuine neurological impairment also causes cognitive slowing. Combining RT with accuracy offered the greatest incremental value in borderline or indeterminate cases. Overall, RT serves as a robust supplementary validity indicator, supporting its use as an adjunct within a multi-method assessment framework rather than as a stand-alone criterion.

Testing the limits of validity indicators: Test coaching in negative and positive response bias assessment

Claudia Pignolo¹, Sara Pasqualini^{1,2}, Veronica Tatti¹, & Luciano Giromini¹

¹Department of Psychology, University of Turin, Turin, Italy

²Department of Human Neurosciences, Sapienza University of Rome, Rome, Italy

Abstract

Simulation studies frequently provide participants with symptom coaching; however, real-world examinees may also acquire information about how validity indicators operate. Methodological recommendations therefore encourage more informed simulation paradigms, including test coaching. Providing detailed information about validity-detection strategies to members of the general population, however, raises ethical and test-security concerns. To address this issue, we recruited 238 psychology students enrolled in a master's-level Psychological and Personality Assessment course, providing an ethically appropriate context for intensive test coaching. In one session, students first responded honestly (T0), received detailed test coaching on Positive Response Bias (PRB), and subsequently attempted to conceal depressive symptoms (T1). PRB was assessed using the General Inventory of Behaviour, Symptoms, and Opinion (GIBSO) and the 16-item Balanced Inventory of Desirable Responding (BIDR-16). One week later, students first feigned depression following symptom coaching (T1) and then repeated the task after receiving detailed test coaching on Negative Response Bias (NRB) in addition to symptom coaching (T2). NRB was assessed using the Inventory of Problems-29 (IOP-29) and Inventory of Problems-Memory (IOP-M). Following PRB test coaching, GIBSO and BIDR-16 total scores differed significantly from honest responding ($d = -1.33$ and -1.40 , respectively). In the NRB condition, test coaching significantly attenuated IOP-29 FDS scores ($d = 0.87$) and increased IOP-M number of correct answers ($d = -0.58$). Nevertheless, sensitivity was 60% for the IOP-29 and 77% for the GIBSO. These findings suggest that detailed knowledge of validity-detection strategies can weaken some validity indicators without necessarily enabling respondents to evade detection.

Clinical Tests Meet Symptom Validity: A New Signal or Just an Echo?

Chair: Irena Bošković



The evolution of embedded SVTs using U.S. veteran samples

Robert D. Shura

Department of Veterans Affairs (VA) Health Care System, Salisbury, North Carolina, United States

VA Mid-Atlantic Mental Illness Research, Education, and Clinical Center, United States

Wake Forest University School of Medicine, Winston-Salem, North Carolina, United States

Abstract

Until recently, symptom validity tests (SVTs) have predominantly existed in the form of dedicated, free-standing measures or as scales that are a part of longer, broad-band self-report measures. There may be situations in research or clinical contexts in which the added length and/or cost of these SVTs precludes their use. Identifying embedded SVT indices within shorter, more common symptom questionnaires has been a gap in SVT evolution and addresses the limitations of longer measures. The purpose of this presentation is to describe the development of embedded SVTs based on several samples of U.S. veterans. SVTs using a variety of detection strategies will be discussed across the Neurobehavioral Symptom Inventory, PTSD Checklist for the DSM-5, Beck Depression Inventory-II, Beck Anxiety Inventory, and the briefer Patient Health Questionnaire-9 and General Anxiety Disorder-7. Detection strategies span indiscriminate symptom endorsement (Symptom Severity scales), rare/quasi-rare symptoms (Rare Item scales), and extreme symptoms (Extreme Symptom scales). Scale creation and cutoff scores across various groups will be presented. Finally, strengths and limitations of these embedded indices will be discussed.

Keywords: embedded validity indices, symptom questionnaires, veteran, overreporting

New methods for symptom validity testing in whiplash associated disorders

Esteban Puente-López

Department of Psychology, Universidad de Valladolid, Valladolid, Spain

Abstract

Whiplash-Associated Disorders (WAD) are among the most challenging conditions to assess because several of their symptoms are not objectively observable using current diagnostic methods and cannot be adequately quantified or evaluated. Also, to date, no method has been established to efficiently assess the validity of self-reported symptoms in individuals with WAD. The purpose of this presentation is to examine the ability of several biopsychosocial psychometric self-report instruments—including the Brief Pain Inventory (BPI), Neck Disability Index (NDI), SF-36 Health Survey, Beck Depression Inventory-II (BDI-II), Beck Anxiety Inventory (BAI), and Brief Illness Perception Questionnaire (BIPQ)—to discriminate between patients diagnosed with WAD following a motor vehicle accident and non-clinical participants instructed to simulate symptoms. A simulation design was employed with 630 participants: 200 non-clinical controls instructed to respond honestly, 201 instructed malingerers, and 229 WAD clinical outpatients. The strengths and limitations of these measures for assessing symptom validity in WAD will be discussed.

Keywords: malingering, feigning, neck injury, symptom validity test, simulation design

Embedded performance (TAP-Alertness, Cognitrone-S8, VVMT, FACE) and symptom validity (BDI-II, FAI) indicators in Austrian early retirement claimants – a comparison

Johann Lehrner

Department of Neurology, Medical University Vienna, Vienna, Austria

Abstract

Background: Neuropsychological assessments are essential in the evaluation of cognitive functions and symptoms. Performance validity tests (PVTs) and Symptom validity tests (SVTs) are being developed to evaluate the credibility of neuropsychological test performance and to determine whether reported symptoms are authentic. These tools can be embedded within standard cognitive assessments or administered as freestanding measures. The goal of the present study was to compare the embedded performance (TAP-Alertness, Cognitrone-S8, VVMT, FACE) and symptom validity (BDI-II, FAI) indicators in Austrian early retirement claimants.

Methods: A retrospective, cross-sectional design, regarding the assessment of Austrian early retirement claimants was performed. The classification into underperformance vs. no underperformance and over-reporting vs. no overreporting was carried out using specific standard procedures and defined criteria.

Results: Specific results for baseline underperformance and baseline over-reporting will be reported. Area under the Curve (AUC), specificity, sensitivity, Youden-Index (YI), PPV, NPV, LR+ and LR- for each measure will be reported separately.

Conclusions: In summary, our data show that TAP-Alertness, Cognitrone-S8, VVMT-S8, FACE, BDI-II, FAI may be useful embedded indicators for identifying underperformance and over-reporting in ERC albeit with different sensitivities. However, more independent research is needed.

Conflict of Interest:

The VVMT and the FAI can be obtained from www.psimistri.com.

Johann Lehrner is CEO of psimistri GmbH which is the owner of www.psimistri.com.

Embedded validity indicators in the PCL-5 for detecting feigned intimate partner violence-related PTSD: Comparative and incremental validity evidence

Francesca Ales & Luciano Giromini

Department of Psychology, University of Turin, Turin, Italy

Abstract

Symptom Validity Tests (SVTs) typically use dedicated, free-standing measures, yet common symptom questionnaires may also provide embedded signals of overreporting. This may be especially valuable in intimate partner violence (IPV) evaluations, where genuine posttraumatic stress and depressive symptoms overlap with presentations exaggerated for external gain. This study examined whether symptom-severity patterns on the PTSD Checklist for DSM-5 (PCL-5) and Center for Epidemiologic Studies Depression Scale (CES-D) differentiate honest and coached-feigned IPV presentations, using the Inventory of Problems-29 (IOP-29) as an SVT benchmark and evaluating coaching format. In an online study, 1,056 Italian women participated: 383 reporting IPV responded honestly, whereas 673 without IPV feigned related problems after receiving identical information as text or video. Feigners obtained higher scores on the IOP-29 False Disorder Probability Score ($M = .73$ vs. $.31$, $d = 1.96$), CES-D ($M = 34.2$ vs. 26.9 , $d = .95$), and PCL-5 ($M = 55.1$ vs. 40.4 , $d = 1.13$), all $ps < .001$. At $FDS \geq .50$, IOP-29 sensitivity was 84.2% and specificity was 79.4%. Video coaching modestly reduced FDS scores relative to text coaching ($M = .70$ vs. $.75$, $d = .25$, $p < .01$), with sensitivity remaining 80.1% versus 87.8%, respectively. Although all measures differentiated groups, separation was greatest for the IOP-29. Taken together, these findings support a multimethod approach in which routinely administered symptom measures may provide validity-relevant information alongside a dedicated SVT, while offering a basis for evaluating their incremental contribution to classification in IPV assessments.

Keywords: intimate partner violence, malingering, symptom validity testing, embedded validity indicators, IOP-29

KEYNOTE & SPECIAL PRESENTATION ABSTRACTS

The following abstracts are presented as part of the keynote and special presentation programme.

Symptom validity assessment with the MMPI: Lessons learned and future directions

Yossef S. Ben-Porath

Kent State University, Kent, Ohio, United States

Abstract

Symptom validity assessment (SVA) has been an integral part of the MMPI instruments since the original MMPI was first introduced in 1943. After outlining the conceptual framework that underlies SVA with the MMPI and defining key terms, the address will review the evolution of SVA with the MMPI instruments and discuss future directions salient to both practice and research. Key points will include the distinction between symptom validity testing and symptom validity assessment, two very different uses of the term “validity”, threats to protocol validity, mapping the MMPI-3 Validity Scales onto threats to protocol validity, consideration of confounds when interpreting MMPI-3 Validity Scale scores, and use of the MMPI-3 Validity Scales across cultures and languages.

Building the future of the SVA community: Conference sustainability, Psychological Injury and Law, and new ways to stay connected

Luciano Giromini

Department of Psychology, University of Turin, Turin, Italy

Abstract

At the 2024 European Conference on Symptom Validity Assessment (SVA), the presenter, who co-organized both that conference and the current meeting, discussed some of the practical challenges of maintaining a recurring international conference without the support of a formal society or professional organization. These include securing initial resources, managing contracts and other administrative responsibilities, and maintaining continuity between conferences. To address these challenges, the presenter proposed exploring a formal connection between the SVA community and an existing professional organization. More specifically, he argued that such an arrangement could provide a sustainable organizational home for future conferences while avoiding the administrative burden of establishing a new society, and could create opportunities for continued interaction between conferences, including discussion of emerging issues in symptom and performance validity assessment, webinars, and other initiatives reflecting the interests of the community. When discussing this possibility, the presenter also noted that the journal *Psychological Injury and Law* publishes many influential articles on topics central to the SVA conference and community, but currently lacks a formal “scientific home” where its authors, editorial board members, and readers can regularly connect and exchange ideas. Accordingly, he suggested that a new organizational space for the SVA community might also serve as a scientific home for *Psychological Injury and Law*. Attendees at the 2024 conference were invited to express their views on each of these ideas, and all received clear support. This presentation will provide a follow-up to those discussions, updating the audience on the steps undertaken between SVA-24 and SVA-26. In doing so, some concrete options for strengthening the long-term sustainability and organizational infrastructure of the SVA community will be presented, together with an update on *Psychological Injury and Law* and its potential relationship with these developments. Participants will be invited to discuss these options and help identify practical next steps that could be pursued shortly after the conference.

SPOTLIGHT ABSTRACTS

The following abstracts are presented as part of the spotlight presentation programme.

Distorted symptom presentation and its consequences

Brechje Dandachi-FitzGerald

Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, the Netherlands

Faculty of Psychology, Open University, Heerlen, the Netherlands

Abstract

Poor symptom validity is often viewed as primarily associated with forensic assessment, malingering, or secondary gain. This perspective overlooks the broader phenomenon of distorted symptom presentations and its relevance to routine clinical practice. Poor symptom validity can take different forms, including careless responding, underreporting, overreporting, and underperformance, and may arise for different reasons. In this spotlight presentation, I will first address several common misconceptions about poor symptom validity. I will then discuss how symptom and performance validity tests can help identify different forms of symptom distortion but do not, by themselves, establish its cause. Finally, I will consider the consequences of distorted symptom presentation for clinical psychological assessment and treatment. Together, these considerations highlight the need for further systematic research on the different forms and causes of distorted symptom presentation, their consequences for clinical practice, and how clinicians can address poor symptom validity constructively.

How to defend against attacks on PVT and SVT data in forensic cases

Kyle B. Boone

Independent Practice, Torrance, California, USA

Abstract

In the last 30 years there has been an explosion in the development and validation of performance validity tests (PVTs) as well as publication of new versions of psychological inventories (such as the MMPI-3) that contain validity scales addressing credibility of report of physical and cognitive symptoms, as well as psychiatric symptoms (symptom validity tests; SVTs). However, despite the demonstrated effectiveness of PVTs and SVTs in identifying noncredible presentations, forensic neuropsychologists routinely encounter attempts by opposing attorneys and experts to undermine test findings. This presentation will explore nine common attacks on PVTs and SVTs, and will provide testimony examples regarding how to address these criticisms. In addition to exploring direct attacks on PVTs and SVTs, the lecture will also examine other methods used by opposing counsel to undermine accuracy of our testimony, including “yes-no” question formats, and will provide recommendations as to how experts can respond.

Changes in validity test scores after COVID: Reflection of general distress or something else?

Allyson G. Harrison, Nathan Davin, & Emma Jamieson

Queen's University, Kingston, Ontario, Canada

Abstract

Objective: This study examined whether rates of symptom and performance validity test (SVT, PVT) failure among assessment-seeking postsecondary students changed during and after the COVID-19 pandemic relative to pre-pandemic levels, and whether such changes co-occurred with increased general psychological distress (GPD).

Methods: Archival data were analyzed from 1,076 students assessed for possible attention-related disorders between 2018 and 2024 at a regional university-based assessment center. Participants completed multiple symptom and performance validity measures, a self-report measure of Attention-Deficit/Hyperactivity Disorder (ADHD), and a higher-order measure of general psychological distress. Students were grouped by assessment timing: pre-COVID (2018–March 2020), during COVID (March 2020–August 2022), or post-COVID (September 2022–September 2024).

Results: Failure rates on most PVTs did not differ significantly across time periods, indicating overall stability in performance validity, with one dyslexia-related validity measure showing higher failure rates post-COVID. On a personality assessment, students assessed during and after COVID reported significantly higher Negative Impression Management scores, lower Positive Impression Management scores, and greater GPD. Rates of severe GPD increased from 23% pre-COVID to 38% during and after COVID. Failure on ADHD-specific SVTs also increased significantly post-COVID, indicating higher rates of non-credible ADHD symptom reporting despite stable performance validity.

Discussion: Since the onset of COVID-19, postsecondary students have demonstrated heightened psychological distress alongside increased non-credible self-reporting, particularly for ADHD symptoms. These findings reflect parallel trends rather than a direct causal relationship and underscore the importance of incorporating both symptom and performance validity testing when interpreting self-reported symptoms in clinical and psychoeducational assessments.

Distinction without a difference: Do SVTs differentiate symptom presentations?

Paul Ingram

Department of Psychological Sciences at Texas Tech University, Lubbock, Texas

Abstract

Over-reporting is a form of response bias that involves exaggeration or fabrication of a client's symptoms or impairment. Symptom validity tests (SVTs) are self-report measures designed to detect over-reporting. SVTs can act stand-alone (e.g., Memory Complaints Inventory) or as part of a broadband instrument (e.g., The MMPI and PAI), and they generally assess endorsement of infrequent or improbable symptoms. Many practice guidelines and scale development histories have focuses on over-reporting in specific symptom domains (Rogers & Bender, 2018). Accordingly, assessment has seen the proliferation of SVTs targeting differing symptom domains (e.g., cognitive, somatic, psychological) across different contexts. While some studies and meta-analyses conclude the notion that SVTs differentially measure symptom presentations, others suggest that scales perform within a standard over-reporting effect range across symptom groups and settings. I will review and discuss recent and emerging research examining if SVTs differentially assess symptom presentation across embedded and standalone measures.

ROUNDTABLE ABSTRACT

The following abstract is presented as part of the Roundtable.

Present and future of validity assessment in European countries

Thomas Merten

Department of Neurology, Vivantes Klinikum im Friedrichshain, Berlin, Germany

Abstract

Following up on the 2024 roundtable event, conference participants are invited to discuss their thoughts on the state of the art and the future of symptom and performance validity assessment. To this end, there will also be two structured individual presentations. The 2024 idea of developing European standards for validity assessment will be picked up and elaborated, featuring the presentation of an initial draft and a discussion on how to drive this project forward.

POSTER ABSTRACTS

The following abstracts are presented as part of the poster presentation programme.

Are feigning screens 'fit to stand trial'?

Nishant Krishnan^{1,2}, Michael Trood¹, Janet Ruffles¹, Grant A. Blake¹, & James R. P. Ogloff¹

¹Centre for Forensic Behavioural Science, Swinburne University of Technology, Melbourne, Victoria, Australia

²Victorian Institute of Forensic Mental Health (Forensicare), Melbourne, Victoria, Australia

Abstract

Competency or fitness to stand trial represents the most common psycho-legal referral in common law jurisdictions. Yet, there remains little consensus regarding which symptom validity tools (SVTs) most effectively discriminate between defendants who feign psychopathology and those who do not. We systematically reviewed and meta-analysed the extent to which three established SVTs: the Atypical Presentation (ATP) Scales, Miller Forensic Assessment of Symptoms Test (M-FAST), and Structured Inventory of Malingered Symptomatology (SIMS), discriminated between feigners and non-feigners within the context of fitness evaluations. Fourteen research reports yielding 20 effect sizes and comprising 1,929 participants were examined. Findings showed that the pooled effect sizes for the M-FAST ($k = 13$, $g = 2.75$, 95% CI [2.18, 3.32]), ATP Scales ($k = 5$, $g = 1.96$, 95% CI [1.37, 2.54]), and SIMS ($k = 2$, $g = 2.99$, 95% CI [2.50, 3.48]) were all very large, far exceeding conventional benchmarks for effect size interpretation in feigning research. Owing to a paucity of data, further analyses of the SIMS were not feasible. Among the subscales of the ATP Scales and M-FAST, those incorporating two or more detection strategies demonstrated the strongest discriminative performance (ATP-Psychotic and Rare Combinations, respectively), whereas scales relying principally on clinical observation (e.g., Reported versus Observed) performed comparatively less well. Relevantly, the M-FAST and ATP Scales demonstrated utility as screening tools, yielding pooled sensitivity estimates of .87 and .91, respectively. Beyond discriminant and classification accuracy performance, independent meta-analyses of feigning prevalence across all three tools converged at a base rate of approximately 25%, suggesting that as many as one in four defendants referred for a fitness evaluation may exaggerate or overreport symptoms of psychopathology. Implications for the selection, interpretation, and development of SVTs in fitness to stand trial evaluations will be presented.

Assessing symptom credibility in PTSD: A multi-method simulation study in an Italian sample

Bosco, F., Bardo, C., Barchielli, B., & Roma, P.

Department of Human Neuroscience, Sapienza University of Rome, Rome, Italy

Abstract

Introduction: In forensic and medico-legal settings psychologists are often required to assess the psychological consequences of traumatic events, where the potential for secondary gain raises a substantial risk of symptom feigning. Post-Traumatic Stress Disorder is particularly vulnerable to this risk, as its diagnosis relies largely on self-reported symptoms and widely known diagnostic criteria. To address this, forensic practice increasingly combines broad-spectrum inventories with specific symptom validity tests, yet evidence on their comparative and combined performance in detecting feigned PTSD within the Italian context remains limited.

Methods: Using a simulation design, 52 participants instructed to respond honestly were compared with 54 instructed to feign PTSD on the basis of a clinical vignette describing a plausible traumatic event and its core symptoms. The groups did not differ in age, gender or education. Participants completed two broad-spectrum instruments, the MMPI-3 and the Trauma Symptom Inventory-2 (TSI-2), together with three symptom validity tests, the Inventory of Problems-29 (IOP-29), the Structured Inventory of Malingered Symptomatology (SIMS) and the Self-Report Symptom Inventory (SRSI).

Results: Simulators scored markedly higher than honest respondents on all MMPI-3 overreporting scales, with large effect sizes (Cohen's d up to 2.20) and strong discrimination (AUC .87 to .93), with Fs, RBS, FBS and Fp performing best; clinical scales showed a generalized and diffuse elevation. The TSI-2 ATR scale also discriminated well (AUC .945). The three symptom validity tests reached comparable accuracy, with the SIMS total (AUC .921), the IOP-29 FDS (AUC .909) and the SRSI scales (AUC .91 to .92) all separating simulators from honest respondents with large effect sizes.

Discussion: The convergence across the MMPI-3, the TSI-2, the IOP-29, the SIMS and the SRSI provides preliminary evidence that feigned PTSD can be reliably detected in an Italian sample and supports the value of a multimethod approach to symptom validity assessment.

Associations between SVTs and PVTs and MMPI-3 measures of traditional and contemporary conceptualizations of personality disorders

Kaci F. Holmes¹, Adriana Ross², Roger Gervais^{2,3}, & Yossef S. Ben-Porath¹

¹Department of Psychological Science, Kent State University, Kent, Ohio, United States

²Neurobehavioural Associates

³University of Alberta, Edmonton, Alberta, Canada

Abstract

Associations between personality disorders and functional impairment are well established. Functional impairment related to occupational functioning in particular has implications for personal injury and disability assessments in which psychologists must determine the extent of an individual's impairment in connection with work-related physical or psychological injuries. Disability claimants have an inherent incentive to appear impaired in these assessments. Best practices for disability evaluations include the use of validated psychological tests to assess psychological functioning and measures of symptom and performance validity (SVTs and PVTs) to identify efforts to overreport symptoms or feign cognitive impairment. The current study evaluates associations between MMPI-3-based personality disorder measures and SVTs and PVTs in a sample of disability claimants. Our sample included 301 disability claimants undergoing comprehensive psychological evaluations following claims of work-related disability. They were, on average, roughly 46 years old, about half (52%) identified as male, and most of the sample (98%) were referred for complaints related to depression (59.3%) and anxiety (37.1%). Finally, most were referred by the Workers' Compensation Board (94.4%). Race/ethnicity was unavailable. SVTs included (i) MMPI-3 Validity Scales, (ii) the Detailed Assessment of Posttraumatic Stress (DAPS) Positive Bias (PB) and Negative Bias (NB) scales, (iii) the Inventory of Problems- 29 (IOP-29) False Disorder Probability (FDS) Score, (iv) the Battery of Health Inventory -2 Defensiveness and Self-Disclosure Scales, and (v) the Modified Somatic Perception Questionnaire (MPSQ) raw score total. Our PVTs included the Word Memory Test and the Inventory of Problems-Memory (IOP-M) score. We found moderate to large (i.e., $> |.30|$) associations between some SVTs and PVTs and MMPI-3 measures of personality disorder-related dysfunction.

Associations between the IOP-29, IOP-M, and MCI in a non-head injury disability sample

Adriana Ross¹, Dana Doucette¹, Jessica Scott¹, Kaci F. Holmes², & Roger O. Gervais^{1,3}

¹Neurobehavioural Associates

²Department of Psychological Science, Kent State University, Kent, Ohio, United States

³University of Alberta, Edmonton, Alberta, Canada

Abstract

This study is an investigation of the association between the Inventory of Problems-29 (IOP-29) and its memory module (IOP-M), and the Memory Complaints Inventory (MCI; Green, 2003) in a sample of non-head injury disability claimants ($N = 333$) with financial incentives. The sample was 52% male, with a mean age of 45 years and comprised predominantly of worker's compensation claimants (96%). The most frequent diagnoses were depression (61%) and anxiety disorders/PTSD (36%).

The IOP-29 (Viglione & Giromini, 2020) is a freestanding symptom validity test (SVT) and the IOP-M (Giromini, Viglione, et al., 2020) is a performance validity test (PVT) derived from the content of the IOP-29. The MCI (Green, 2003) is a 58-item self-report questionnaire assessing common subjective memory complaints categorized under nine scales: GMP – General Memory Problems, NIP – Numerical Information Problems, VSMP – Visuospatial Memory Problems, VMP – Verbal Memory Problems, PIM – Pain Interferes with Memory, MIW – Memory Interferes with Work, and three “implausible” memory complaints scales, IRM – Impairment of Remote Memory, ACB – Amnesia for Complex Behaviour, and AAB – Amnesia for Antisocial Behaviour. We also calculated a mean MCI score. Elevated memory complaints are associated with symptom exaggeration and failure on the Word Memory Test (WMT; Green, 2003).

We began by examining associations between the IOP-29 FDS (False Disorder Probability Score) and the MCI scales. As expected, the FDS was positively correlated with the MCI scales, with the strongest associations noted for Mean MCI (.489), PIM (.445), and MIW (.426). The IOP-M score was negatively associated with all MCI scales, most notably mean MCI (-.369), AAB (-.366), NIM (-.328), and ACB (-.304).

We also conducted ROC analyses to examine the effectiveness of the IOP-29 FDS, IOP-M, and MCI scales to predict performance on an independent PVT, the WMT. The most effective measure was mean MCI ($AUC = .789$), followed by MIW ($AUC = .785$) and VSMP ($AUC = .780$). IOP-M was also effective predicting WMT failure ($AUC = .768$). Using more conservative WMT failure criteria (IR or DR $\leq 77.5\%$) proposed by

Rohling et al. (2024), the most effective measures were mean MCI (AUC = .835), IOP-M (AUC = .822), MIW (AUC = .814) and ACB (AUC = .805).

Results of this study support the utility of the IOP-29 and the IOP-M as effective measures of symptom validity and performance validity in psychological assessment of non-head injury disability claimants.

Conceptualizing “good old days” bias from a dialectical behavior therapy (DBT) perspective

Denise D. Ben-Porath¹ & Yossef S. Ben-Porath²

¹John Carroll University, University Heights, Ohio, United States

²Kent State University, Kent, Ohio, United States

Abstract

The "good old days" phenomenon refers to the tendency for people to remember the past as being better than it was, often emphasizing positive experiences while minimizing or forgetting negative ones. In the context of response bias research, it has been described by Bošković et al (2026) as a distorted retrospective, self-perception that one's past functioning was exceptionally high and psychopathology free when compared with current self-report. Bošković and colleagues found evidence of good old days bias among both genuinely injured and experimentally feigning individuals, noting that its presence alone cannot differentiate between the two groups, however the magnitude of the distorted recollection may inform symptom validity assessment. These authors conceptualized the “good old days” phenomenon from a cognitive psychology perspective as one that involves memory distortions stemming from the reconstructive nature of autobiographical memory. In this poster, we advance a conceptualization of the phenomenon from a Dialectical Behavior Therapy (DBT) perspective, with an emphasis on the role played by emotional dysregulation in these cognitive distortions. Specifically, emotional dysregulation may bias autobiographical memory, such that individuals experiencing current stress, may perceive the present as worse than it objectively is, and the past as more positive than it was. The interaction between emotional dysregulation and current environmental stressors can lead to a positively distorted recollection of the past and a negatively distorted perception of current psychological functioning. These distortions can in turn negatively bias current symptom reporting, manifesting as negative response bias. Implications for symptom validity assessment are discussed.

Convergence among psychometric findings, record-based inconsistencies, and biologically verified medication adherence in a sample of medico-legal disability evaluatees suspected of noncredible responding

Veronica Tatti¹, Sara Pasqualini^{1,2}, Paolo Boselli³, Nicola Molteni³, Valentina Verzura Robbiani³, Claudia Pignolo¹, & Luciano Giromini¹

¹Department of Psychology, University of Turin, Turin, Italy

²Department of Human Neurosciences, University of La Sapienza, Rome, Italy

³Centro Peritale per le Assicurazioni Sociali (CPAS), Massagno, Ticino, Switzerland

Abstract

Malingering is defined as the intentional production of false or grossly exaggerated psychological or physical symptoms to obtain external incentives. In forensic and insurance contexts, its assessment is essential due to the significant economic and social costs associated with this phenomenon. The combined use of collateral information, such as compliance with pharmacological treatment, alongside the administration of symptom and performance validity tests (i.e., SVTs/PVTs), is fundamentally important in detecting response style. Non-compliance with pharmacological treatment is commonly associated with an increased risk of malingering or feigning psychopathology. Within this framework, the Inventory of Problems–29 (IOP-29) appears particularly suitable due to its statistical robustness and high classification accuracy. A retrospective criterion-group study is currently being conducted on an ecologically valid forensic sample of $n = 50$ individuals undergoing evaluation of their residual work capacity. The study aims to examine the association between the IOP-29 False Disorder Probability Score (FDS) and pharmacotherapy compliance, objectively measured through blood tests. The sample includes four groups: participants who correctly adhered to pharmacological treatment (i.e., compliant group), those who did not take their medication (i.e., non-compliant group), those who took it in excessive dosage (i.e., over-dosed group), and those who took it in insufficient dosage (i.e., under-dosed group). It is hypothesized that FDS scores will be significantly higher ($\geq .50$) among participants with low compliance (non-compliant, over-dosed, and under-dosed) compared to those with high compliance. Findings are expected to provide preliminary evidence for the relationship between pharmacological adherence and response style, contributing to the refinement of malingering assessment in forensic settings.

Keywords: forensic assessment, malingering, IOP-29, classification accuracy, SVTs

Criteria-based validity assessment in legal cases involving pension and accident insurance

Klett, N.^{1,2,4}, Dohrenbusch, R.³, Siegmann, E. M.², Schütz, A.⁴, Kornhuber, J.², Röhner, J.⁴, Keller, F.⁵, Capito, E. S.¹, & Grömer, T.^{1,2}

¹Practice Clinic for Neurology, Psychiatry, Psychosomatic Medicine and Psychotherapy, Bamberg, Germany

²Department of Psychiatry and Psychotherapy, Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen, Germany

³University of Bonn, Bonn, Germany

⁴University of Bamberg, Bamberg, Germany

⁵Thüringer Landessozialgericht, Erfurt, Germany

Abstract

Background: Medical expert witness assessments (MEWAs) evaluate case validity using information derived from both psychometric and non-psychometric sources. Although symptom validity tests (SVTs) and performance validity tests (PVTs) are widely used, many established thresholds originate from analogue or known-groups validation studies, and their applicability to MEWAs remains uncertain. Previous research established a Criteria-Based Validity Assessment (CVA) as a multimodal framework for evaluating case plausibility. However, it remained unclear whether CVA criteria and psychometric validity indicators of typical tests yield comparable results across different legal contexts, such as pension and accident insurance evaluations.

Objective: This study examined whether CVA, SVT, and PVT results differ in pension and accident insurance cases.

Methods: A total of 721 MEWAs (572 pension; 149 accident) were analyzed. Trained raters evaluated CVA criteria, while response and performance validity was assessed with the Structured Inventory of Malingered Symptomatology (SIMS) and the Amsterdam Short-Term Memory test (ASTM). Two-component beta-binomial mixture models were used to identify empirically derived CVA plausibility thresholds.

Results: Mixture modeling consistently identified a distinct bimodal distribution of conspicuous CVA criteria across both pension and accident cases, supporting the differentiation of plausible and implausible case presentations. A threshold of ≥ 4 conspicuous CVA criteria provided the optimal classification of case plausibility. Pension claimants exhibited significantly higher SIMS scores and lower

ASTM scores than accident claimants. CVA-defined implausibility rates did not differ between pension and accident cases.

Conclusions: The findings support the use of CVA as a multimodal framework for integrating validity information in different legal contexts. However, CVA did not appear to capture context-specific response biases. The presence of ≥ 4 conspicuous criteria provides conservative support for a classification of CVA-defined case implausibility. Cases with ≤ 2 conspicuous CVA criteria may be considered plausible, whereas cases with three criteria warrant individualized examination because feigning cannot be ruled out.

Do post-pandemic increases in AD/HD symptom validity indicator violations differ by gender, and are any differences explained by psychological distress?

Nathaniel Davin, Allyson G. Harrison, Emma Jamieson, & Beth Pollock

Regional Assessment & Resource Centre, Queen's University Kingston, Ontario, Canada

Abstract

Background: Recent research has demonstrated an increase in Attention Deficit/Hyperactivity Disorder (AD/HD) symptom validity indicator (SVI) scores following the COVID-19 pandemic. With post-pandemic increases in AD/HD diagnoses among women, it remains unclear whether there are gender differences in SVI scores and violations, and whether any observed differences are explained by general psychological distress (GPD).

Methods: An archival review of 454 files examined scores obtained by men and women on three symptom validity tests (SVTs) derived from the Conners' Adult ADHD Rating Scales (CAARS), Exaggeration Index (EI), CAARS Infrequency Index (CII), and Dissimulation score (DISS), across pre- and post-pandemic time periods (2016-2020 and 2022-2026). A MANOVA was conducted to examine the effects of gender and time period (TP) on SVI performance. A subsequent MANCOVA controlled for GPD. A binary logistic regression compared rates of validity cut-score violations across groups and analyses were performed to control for GPD.

Results: Significant main effects were observed for gender and TP, with no significant Gender $\times$ TP interaction. Women and post-pandemic participants showed higher SVI scores and higher rates of validity cut-score exceedance across measures after excluding inconsistent responders. However, after controlling for GPD, these effects were largely non-significant, except for increased CII violations in the post-pandemic group and a small Gender $\times$ TP interaction on the EI.

Conclusions: Results confirm an increase in SVI violations post-pandemic, particularly in women. Elevated general psychological distress may account for these increases. As such, routine assessment of symptom validity remains important in AD/HD evaluations given increasing referral rates and the potential risk of diagnostic error when self-report validity is not considered.

Do post-pandemic increases in PAI symptom validity indicator violations differ by gender, and are any differences explained by psychological distress?

Nathaniel Davin, Allyson G. Harrison, & Beth Pollock

Regional Assessment & Resource Centre, Queen's University Kingston, Ontario, Canada

Abstract

Background: Rates of general psychological distress (GPD), including anxiety and depression, have increased following the COVID-19 pandemic, with women disproportionately affected. Prior research also shows post-pandemic increases in ADHD-specific symptom indicator (SVI) scores associated with higher GPD, raising the possibility that elevated distress may influence SVI performance more broadly. The present study examined whether similar post-pandemic patterns are observed on Personality Assessment Inventory (PAI) SVIs, and whether any differences by time period or gender are explained by GPD.

Methods: An archival review of differences by Gender, women ($n = 624$) and men ($n = 389$), and by Time Period (TP), 2016-2020 ($n = 675$) and 2022-2026 ($n = 338$), on three Personality Assessment Inventory (PAI) SVIs: Negative Impression Management (NIM), Malingering (MAL), and Positive Impression Management (PIM) were examined. Differences were analysed using MANOVA, followed by MANCOVA controlling for GPD. Binary logistic regression analyses were performed on dichotomous pass/fail SVI variables before and after controlling for GPD.

Results: Significant main effects of Gender and TP were found, with no significant interaction. Women scored higher on NIM and MAL and lower on PIM, while post-pandemic participants scored higher on NIM and lower on PIM. These effects were no longer significant after controlling for GPD. Using dichotomous outcomes, women showed higher SVI failure rates on NIM and MAL, and post-pandemic participants similarly showed higher failure rates on NIM and MAL. Again, all significant effects disappeared after controlling for GPD.

Conclusions: Findings suggest that psychological distress may underlie both post-pandemic increases and gender differences in PAI SVIs. These findings raise important questions about the interpretation of self-reported psychological symptoms during periods of widespread psychological strain, particularly in women.

Evaluating continuous performance tests as embedded measures of performance validity in ADHD assessments: A systematic review and meta-analysis

Pinar Toptas^{1,2,3}, Tycho J. Dekkers^{2,3,4}, Yvonne Groen¹, Geraldina F. Gaastra¹, Dick de Waard¹, & Anselm B.M. Fuermaier¹

¹Department of Clinical and Developmental Neuropsychology, University of Groningen, the Netherlands

²Accare Child Study Center, Groningen, the Netherlands

³Department of Child and Adolescent Psychiatry, University Medical Center Groningen, the Netherlands

⁴Levvel, Child and Adolescent Psychiatry, Amsterdam, the Netherlands

Abstract

Background: Noncredible performance occurs in a non-trivial minority of adults assessed for ADHD and cannot be reliably detected by clinical judgment alone. Continuous Performance Tests (CPTs) are widely used to assess attention and response control and can also serve as embedded performance validity indicators (EVIs). Prior findings on their validity are promising but inconsistent, warranting quantitative synthesis.

Objective: To evaluate the diagnostic utility of CPTs in differentiating credible from noncredible cognitive performance in adults undergoing ADHD assessment, and to examine the moderating effects of study design (criterion vs simulation) and sample type (student vs nonstudent).

Methods: We conducted a systematic review and meta-analysis of studies using a CPT variant. Included studies applied either a simulation or criterion-group design to classify credible versus noncredible performance. Eighteen studies ($n = 3,021$; 67 effect sizes) met inclusion criteria. Eight studies adopted a simulation design and ten criterion-group studies. Weighted mean effect sizes were calculated using Hedges' g , and study design and sample type were examined as moderators.

Results: Across CPT variants, CPTs significantly distinguished credible from noncredible performance ($g = 0.73$), though heterogeneity was high. According to field-specific criteria for validity research, this magnitude is small (Rogers, 2018), underscoring the need for cautious interpretation and avoidance of single-test decisions. Effect sizes were nearly twice as large in simulation-group studies ($g = 0.94$) compared with criterion-group studies ($g = 0.55$), indicating that design affects observed utility and that both designs offer complementary information. Student and nonstudent samples did not differ significantly.

Conclusions: CPTs are useful as embedded validity indicators in adult ADHD assessments but should not be used in isolation to infer noncredible performance. Consistent with current practice standards, CPT-

based EVIs are best interpreted alongside multiple embedded and stand-alone validity measures to reduce misclassification and to better distinguish invalid performance from genuine cognitive impairment.

Extremely elevated self-reported autism symptoms in postsecondary assessment: Associations with symptom validity and psychological distress

Nathaniel Davin, Allyson G. Harrison, & Beth Pollock

Regional Assessment & Resource Centre, Queen's University Kingston, Ontario, Canada

Abstract

Background: Psychologists are increasingly encountering postsecondary students who report unusually high autism spectrum disorder (ASD) symptoms on self-report measures. The Social Responsiveness Scale–Second Edition (SRS-2) is widely used in adult assessments but lacks embedded symptom validity indicators, raising concerns about interpretation of extreme scores.

Methods: This study examined associations between SRS-2 Total scores, psychological distress, and symptom/performance validity indicators in a sample of $n=167$ postsecondary students evaluated at a regional assessment centre. Symptom validity indices (SVIs) from the Conners' Adult ADHD Rating Scales (CAARS), Personality Assessment Inventory (PAI), Behaviour Rating Inventory of Executive Function–Adult (BRIEF-A), and a battery of performance validity tests (PVTs) were analyzed using correlations, partial correlations controlling for distress, hierarchical regression, and receiver operating characteristic (ROC) analyses.

Results: SRS-2 scores were strongly associated with general psychological distress (GPD; $r = .76$) and symptom validity indicators, particularly the CAARS Dissimulation ($r = .67$), Infrequency ($r = .64$), and Exaggeration Index ($r = .58$). After controlling for GPD, associations with CAARS SVIs remained moderate ($r \approx .45-.50$), whereas associations with PAI validity indicators were substantially reduced. The Dissimulation Index, reflecting endorsement of atypical, low base-rate dissociative experiences, remained a significant predictor after controlling for GPD. In hierarchical regression, GPD accounted for most variance in SRS-2 scores ($R^2 = .67$), while CAARS Dissimulation ($\Delta R^2 = .08, p = .003$) and Infrequency ($\Delta R^2 = .03, p = .048$) SVIs contributed unique variance. ROC analyses showed good-to-excellent discrimination of elevated SRS-2 scores by symptom validity indicators (AUCs = .75–.92), whereas PVT failure was not predictive (AUC = .58, ns).

Conclusions: Findings suggest that extremely elevated SRS-2 scores reflect GPD and endorsement of atypical, low base-rate experiences rather than ASD-specific symptomatology alone. Results underscore the importance of incorporating symptom validity indicators when interpreting self-reported ASD symptoms in adults.

Feasibility of eye-tracking-based indicators of performance validity in digital cognitive testing

Charlotte Schneider

Department of Neurology, Medical University Vienna, Vienna, Austria

Abstract

Digital cognitive assessments for Alzheimer's disease (AD) enable standardised, scalable testing and support early detection. However, self-administered formats reduce behavioural observation and may limit detection of insufficient effort, underperformance, or overreporting. Performance validity tests (PVTs) rely mainly on outcome scores and provide limited insight into test-taking behaviour. Eye-tracking offers continuous, objective measures of gaze behaviour and may add indicators of engagement and response credibility, however its use within digital cognitive testing for validity assessment is not well established. This study evaluates the feasibility of integrating eye-tracking into a digital cognitive test battery and examines whether eye-tracking metrics relate to performance patterns relevant to performance validity. Ten healthy adults completed the tablet-based International Neurocognitive Profile (INCP). Eye movements and pupil data were recorded using mobile eye-tracking glasses. Processed metrics included fixation patterns, gaze distribution, and measures related to task engagement. Feasibility outcomes included data quality, wear acceptance, and stability of extracted measures. Analyses are exploratory. Data collection is complete, and analyses are ongoing. Initial findings indicate that integration of eye-tracking into digital cognitive testing is feasible. Preliminary patterns suggest that gaze behaviour varies across tasks and may relate to performance differences. Eye-tracking in digital cognitive testing appears feasible and may provide behavioural markers relevant to detecting underperformance, non-compliance. Further research is needed to evaluate its utility alongside established PVT approaches in clinical populations.

MMPI item endorsements can predict multidimensional demographic variables and validity scales

Paul A. Clemons^{1,2}, Mason C. Juday³, Ralf Dohrenbusch⁴, & Robbi Brockhaus⁵

¹PAC Advisory (Independent Investigator), Medford, Massachusetts, United States

²Galiana, LLC (Principal and Sole Member), Medford, Massachusetts, United States

³Independent Data and Software Consultant, Berlin, Germany

⁴Institut für Psychologie, Psychotherapeutische Hochschulambulanz, Fachbereich Psychologische Begutachtung, Bonn, Germany

⁵Institut für Psychologische Diagnostik und Begutachtung, Duisburg, Germany

Abstract

Symptom validity tests (SVTs) help examiners infer motivation when evaluating individual self-descriptions, particularly when multiple SVTs with empirical foundations are used in combination. The Minnesota Multiphasic Personality Inventory (MMPI) family of personality instruments has impressive empirical foundations for validity, and each version of the instrument has SVT scales. Predictive models in MMPI studies have mainly focused on clinical scales, with comparable studies for validity scales lagging. To help bridge this gap, we report an integrated predictive modeling approach using a diverse well-annotated data set. We illustrate by predicting multidimensional demographic variables (e.g., gender) and validity scales (e.g., Md, MMDS) using MMPI item endorsements. **Subjects and Methods:** We ingested a source database from diverse settings in Germany and Switzerland (7,630 records). We tagged incoming data with structured metadata and controlled vocabularies to support data analysis. We extracted 3,568 records with MMPI-2 item endorsements whose subjects were examined in the contexts of forensic assessment, independent medical examination, therapeutic treatment, or experimental malingering. We also captured 31 demographic variables, 22 MMPI-2 validity scales, and 81 other scales from these records. We devised a diversity approach to select a demographic test data stratum (~1/8) forbidden to take part in model training. We built a predictive modeling framework for demographic variables (e.g., sex, age, education) and validity scales using item endorsements. We compared model performance using all MMPI-2 items with models using subsets of items. We compared models using MMPI-2-RF items to predict validity scales developed for MMPI-2 across demographic segments. We used feature-selection algorithms (Relieff, RRrelieff) to rank the importance of individual items in making predictions. **Results:** Classification models using all MMPI-2 items are highly predictive of gender

(94%) with predictable item importance scores. We therefore compared model performance using two subsets of MMPI-2 items: those included or excluded from MMPI-2-RF. Taking either subset degraded model performance, but each subset is still highly predictive of gender (>85%). We also asked whether MMPI-2-RF items can predict scales developed for MMPI-2. Considering changes in use of gender-based norms, we compared models using all subjects to models built separately for each gender. Regression model performance using MMPI-2-RF to predict Md ($r^2=0.94$) or MMDS ($r^2=0.97$) was not substantially affected when separate models were built for males (Md: 0.94; MMDS: 0.96) and females (Md: 0.93; MMDS: 0.96). We are investigating whether systematic biases in these predictions can be linked to other demographic variables, and whether combining Md and MMDS as a joint prediction target will improve performance. To combine scales, we built generalized probabilistic models of multidimensional SVT scales using published criterion-group data and assigned membership likelihoods to each subject. Outlook: We ingested and carefully characterized a diverse set of >3K well-annotated MMPI-2 item endorsement sets. While our study made use of response-bias expectations from German criterion groups, this work suggests a multidimensional framework to inform international norms for SVT scales. The combination of our data set, metadata discipline, and principled modeling framework allows new questions to be rapidly asked and answered. We encourage direct communication with the authors to explore the feasibility of questions of interest.

MMPI-2 symptom validity patterns in U.S. veterans

Childers, L. G.¹, Streit, Z.¹, Oehlert, M. E.², Sumerall, S. W.³, & Ingram, P. B.¹

¹Department of Psychological Sciences at Texas Tech University, Lubbock, Texas

²Independent researcher

³Private practice

Abstract

Background: Accurately assessing symptom validity is challenging, particularly given that bone fide pathology and other factors can influence construct validity. SVTs are undergoing rapid theoretical and applied refinement in broadband self-report instruments. One trend examines how SVTs differentiate from one another according to design or intended scope of content, rather than functioning to detect shared, higher-order validity patterns. This study explores shared variance patterns across over- and under-reporting SVTs on the MMPI-2.

Methods: The study used retrospective, nationally sampled records analyzed in 2013, drawn from the VA's central computer infrastructure and research database via a VA approved IRB-study. Descriptive patterns, and correlations between, are reported for raw scale scores of the MMPI-2 validity scales were compared using Cohen Q analyses across a nationally sample ($N = 2,966$), subdivided into three groups to test replicability.

Objectives: This study examines SVT patterns on the MMPI-2. The MMPI-2 has unique scales whose methodology is no longer standard in MMPI-2-RF/3 response style scales (e.g., Fptsd [disorder specific dissimulation approach]). Comparing these results to MMPI-2-RF/3 SVTs also provides insight into how validity scale psychometric relationships changed during the MMPI revision process, and may offer insight into scale construction method influence on domain detection.

Conclusions: Correlations had generally equivalent functioning, suggesting low fluctuations in magnitude differences between scales and groups. Correlations were stable across the samples. Fb was more strongly associated with F than Fp and S converged more strongly with K than L and was largely redundant. These findings support MMPI-2-RF/3 exclusions of these scales for Veterans. Measures of pathology and stress (e.g., CS7, RC7, PTSD Score, Combat experience), underscoring sensitivity to bona fide pathology in SVTs. Implications for SVT research and clinical use will be presented.

Table 1*Cohen's q Effect Sizes for MMPI-2 Validity Scale Correlations Across Three Randomized Subsamples*

		Group 1 vs. Group 2	Group 1 vs. Group 3	Group 2 vs. Group 3
MMPI-2 Scale Pair		q	q	q
Superlative				
Self-Presentation (S)	F	-0.04	-0.04	0.00
	Fp	-0.03	-0.04	-0.01
	Fb	-0.03	-0.01	0.03
	L	0.00	-0.05	-0.05
	K	-0.05	-0.04	0.01
Infrequency Back (Fb)				
F	F	0.06	0.10	0.04
	Fp	0.02	0.07	0.05

Note. $N = 2,966$ (Group 1 $n = 988$, Group 2 $n = 989$, Group 3 $n = 989$). Cohen's q effect size at small ($|.1$ to $< .3|$), medium ($|.3$ to $< .5|$), and large ($|\geq .5|$) (Cohen, 1988). F = Infrequency; Fp = Infrequency Psychopathology; L = Lie; K = Correction; S = Superlative Self-Presentation; F-K = Dissimulation.

MMPI-2 vs. PAI-A validity scales profiles in a case study involving a Functional Neurological Disorder

Christopher J. Nicholls

The Nicholls Group, Scottsdale, Arizona, United States

Abstract

The term Functional Neurological Disorder (FND) refers to neurological symptoms that are incompatible with better understood neurological or medical conditions, and include psychogenic nonepileptic seizures (PNES), functional movement disorders, disorders of sensory and motor functioning, and functional cognitive disorders. FND conditions are distinct from feigning, factitious disorders and malingering. They occur worldwide and represent the second most common referrals to neurologists, after headaches. FND symptoms occur in more individuals in the United States than those with multiple sclerosis, brain tumors and motor neuron diseases, combined.

This poster presents the case of a 32-year-old woman with no premorbid neurological or psychiatric impairment who was diagnosed with PNES following her being involved in a motor vehicle accident. The patient has since undergone two independent neuropsychological evaluations, both of which documented intact (cognitive) performance validity testing and neither of which identified significant cognitive impairments. Administration of the MMPI-2 in her first evaluation nevertheless produced validity scale (F, Fp and FBS) elevations three standard deviations above the normative mean, and a clinical scale “conversion V” disorder symptom profile. In contrast, PAI-Plus administration during the second evaluation produced entirely within normal limits scores across both clinical and validity scales.

Interpretation of the obtained differences in validity and clinical scale elevations on two commonly used personality assessment questionnaires is discussed from the perspective of advances in symptom validity assessment and conceptual understanding of the functional neurological disorder continuum. This case study suggests the need for caution against rapid/unquestioning dismissal of patient complaints based upon symptom validity scale elevations from older assessment tools, and provides evidence supporting the need to use assessment tools that reflect technical advances and current scientific knowledge.

Long term outcomes of symptom overreporting and cognitive underperformance in adult ADHD assessments

Emily Raasch^{1,2}, Hanna Christiansen¹, Johanna Kneidinger¹, Björn Albrecht¹, & Anselm B. M. Fuermaier²

¹Department of Clinical Child and Adolescent Psychology and Psychotherapy, Philipps-University Marburg, Marburg, Germany

²Department of Clinical and Developmental Neuropsychology, University of Groningen, Groningen, the Netherlands

Abstract

Invalid responding can bias clinical judgement, particularly in adult attention deficit/hyperactivity disorder (ADHD), where rates of symptom overreporting and cognitive underperformance are elevated. Performance validity tests (PVTs) and symptom validity tests (SVTs) are commonly used to assess response credibility, yet their relevance for long-term clinical outcomes in ADHD remains unclear. The present study examined whether PVT and SVT scores obtained during diagnostic assessment are associated with long-term clinical outcomes in adults with ADHD. Data were drawn from an outpatient cohort of adults diagnosed with ADHD ($N= 181$). Participants completed standardised diagnostic assessment including two memory-based PVTs, one attention-based PVT, and two embedded SVTs. Long-term clinical outcomes, including psychotherapy use, pharmacotherapy, and diagnostic re-evaluation, were assessed via self-report approximately one year later. Associations were examined using correlational analyses. No significant associations were observed between PVT and SVT scores and long-term clinical outcomes. In contrast, baseline ADHD symptom ratings showed small and domain-specific associations with treatment-related outcomes. Future research should examine how validity indicators influence decision-making and treatment processes in adult ADHD assessment, using longitudinal and multi-method designs with more sensitive outcome measures.

Personality Assessment Inventory Cognitive Bias Scale of Scales (CB-SOS) as a predictor of performance validity test failure in postsecondary psychoeducational assessment

Allyson G. Harrison & Irene T. Armstrong

Regional Assessment & Resource Centre, Queen's University Kingston, Ontario, Canada

Abstract

Background: The Cognitive Bias Scale of Scales (CB-SOS-1-3) are Personality Assessment Inventory (PAI)-derived indicators intended to identify exaggerated cognitive symptom reporting and potential invalid neuropsychological performance. Although prior studies have suggested that CB-SOS may predict performance validity test (PVT) failure better than traditional PAI validity indicators, research remains limited, particularly in postsecondary settings.

Methods: Archival data were examined from students referred for psychoeducational assessment at a university-affiliated postsecondary clinic between 2008 and 2026. Participants completed the PAI and at least two PVTs. Careless or random PAI profiles (elevated Infrequency or Inconsistency scores) were excluded, leaving 691 students (mean age = 22.0 years, $SD = 6.5$; 57.1% female). Standalone PVTs included the Word Memory Test, Medical Symptom Validity Test, Victoria Symptom Validity Test, Test of Memory Malingering, and b Test. Embedded measures included Reliable Digit Span and the TOVA Symptom Exaggeration Index. Participants were classified as invalid if they failed either one or more PVTs or two or more PVTs.

Results: Fifty-four participants (7.8%) failed two or more PVTs, and 222 (32.1%) failed one or more. Participants failing two or more PVTs demonstrated significantly higher CB-SOS scores than those passing PVTs, with medium effect sizes observed for CB-SOS-1 (Hedges $g = -0.54$) and CB-SOS-3 ($g = -0.55$). ROC analyses demonstrated modest classification accuracy, with AUCs ranging from .64 to .68 for prediction of two or more PVT failures. Traditional PAI validity indicators showed smaller effects. Sensitivity of recommended CB-SOS cutoffs was low, although specificity was high.

Conclusions: Results provide partial support for the CB-SOS as a marker of noncredible cognitive symptom presentation in postsecondary psychoeducational assessment. CB-SOS indices demonstrated meaningful group differences and modest predictive utility, although classification accuracy was insufficient for standalone clinical decision-making. Findings support the use of CB-SOS scales as adjunctive indicators within a comprehensive validity assessment framework.

PVT use in neuropsychological assessment in the UK and Republic of Ireland

Marie Claire O'Brien, Rachael Parry, Sonja Soeterik, & Louise Warren

Neurolink Psychology Expert Witness Services, London, United Kingdom

Abstract

There has been growing recognition of the need for PVT use in all neuropsychometric assessment, yet the current use in practice, interpretation and reporting of PVTs is poorly understood in the UK and Republic of Ireland.

Introduction: Central to all clinical formulation and opinion on the nature of a patient's neuropsychological profile, is being able to rely on the validity of the scores obtained in an assessment. Research has shown that performance can be compromised for many reasons and detailed assessment of performance is needed in all assessment.

Methods: 130 Psychologists working with adults clinically and/or medico-legally in the Republic of Ireland and the UK (England, Wales, Scotland and Northern Ireland) completed an online survey. Recruitment was via professional bodies, networks and snowballing.

Results:

- The number of standalone and embedded PVTs used in clinical and medicolegal assessments and the factors that influence their use
- The ways respondents manage the PVT security in reporting
- The range of methods used to investigate performance validity
- What standalone and embedded PVTs are in use in clinical and medicolegal assessments
- When in the assessment the PVTs are administered
- Instructions and information given to patients about PVTs
- How respondents dealt with issues of validity during the assessment and in reporting
- The criteria employed to determine performance invalidity and the factors interpreted as associated with failure and how this is reported

Discussion:

- Change in PVT use since McCarter et al., 2009.
- Similarities and difference to the Northern American context.
- Practice as reported versus professional guidance.

Strategies for faking good: Concealed psychosis in UAP reports

Costopoulos, J., Plescow, K., Minor, S., Wesley, M., & May, B.

Florida Institute of Technology, Melbourne, Florida, United States

Abstract

1. To identify new strategies for detecting faking good, this scoping review examined reports of Unidentified Anomalous Phenomena (UAP) that are the product of psychiatric symptoms. Some reports may be obviously contaminated by psychosis, while others may attempt to conceal it to present as credible witnesses, the aim being to encourage the government to investigate their experiences. Improved detection strategies would allow non-credible accounts, by those faking good, to be identified within the many thousands of UAP reports logged and investigated annually by government agencies.

2. Following PRISMA-ScR guidance, searches were undertaken using OpenAlex, EuropePMC, PsycINFO, PubMed, and PsyArXiv using 87 terms including UAP and related experiences, and 151 terms denoting psychotic and related characteristics applied to title, abstract, and keyword fields. This returned approximately 190,000 articles post de-duplication.

3. Programmatic and manual screening based on title and abstract resulted in 110 articles retained for full-text analysis. Article eligibility was framed using the Peters et al. (2020, 2024) guidance for scoping reviews. Traits noted in these reported experiences were analyzed for characterological features of UAP perceptions resulting from psychotic experiences.

4. Preliminary results of relevant literature indicated that psychotic experiences relevant to UAP are notably reported with a greater degree of specific details of the experience, high confidence in the authenticity of their report, and beliefs of personalization regarding the event. Developing strategies to identify traits associated with faking good to conceal psychotic origins of the sighting are needed to rapidly distinguish credible and non-credible UAP reports.

5. While there is significant research on faking good, it is primarily centered around standardized personality assessment. This research explores specific traits and strategies of faking good in those concealing psychotic symptoms, enriching the SVA literature with possible new approaches for detection of this specific set of symptoms and disorders.

Symptom over- and underreporting are relatively stable behavioral opposites in forensic patients and undergraduates

Daniël van Helvoort¹, Henry Otgaar^{1,2}, Chijs van Nieuwenhuizen^{3,4}, & Harald Merckelbach¹

¹Clinical Psychological Science, Maastricht University, Maastricht, the Netherlands

²Leuven Institute of Criminology, Catholic University of Leuven, Leuven, Belgium

³GGzE Institute for Mental Health Care, Eindhoven, the Netherlands

⁴Tranzo, Scientific Center for Care and Wellbeing, Tilburg University, Tilburg, the Netherlands

Abstract

Objective: Research on symptom validity has often linked symptom over- and underreporting to personality traits, but usually in cross-sectional designs and rarely in relation to one another. However, before attributing trait-like properties to these phenomena, a prerequisite question should be addressed: are these distorted response patterns stable over time?

Methods: We conducted a test-retest study in forensic inpatients and undergraduates (all $ns \geq 64$). Participants twice completed stand-alone measures of symptom over- and underreporting, as well as inattentive responding – as this could obscure or spuriously inflate associations – and Big Five personality domains. They completed these measures across a six- to ten-week interval. Forensic patients completed the assessment at intake, and again near their first treatment evaluation, when treatment-related consequences could be more salient.

Results: Inattentive responding was non-trivial in forensic patients, at approximately 15%, but low in students. Symptom overreporting rates were similar in forensic patients and students, at approximately 5%, whereas underreporting occurred in approximately 15–20% of forensic patients and 5–10% of students. Across both samples, symptom overreporting and underreporting were relatively stable behavioral opposites. Participants who overreported or underreported showed patterned differences in self-reported personality traits. In a diagnostic subsample ($n = 36$), borderline personality disorder was linked to symptom overreporting, whereas narcissistic personality disorder was associated with symptom underreporting.

Conclusions: The results suggest that both traits and contextual factors shape how accurately people report their symptoms. Clinically, they support assessing inattentive responding, overreporting, and underreporting together when interpreting validity in forensic settings.

The benefits of being ill

M.E. Jan Wise¹ & Julian Beezhold²

¹Central and North West London NHS Foundation Trust, London, United Kingdom

²Norwich Medical School, University of East Anglia, Norwich, United Kingdom

Abstract

Whilst the roots of a welfare system in the United Kingdom (UK) can be traced back to the poor laws of the 16th century, which established a framework for assisting the impoverished, the modern welfare state began to take shape with the publication of the Beveridge Report in 1942. Since then, a cradle-to-grave welfare system has existed with health care for all from the inception of the NHS in 1952; Aneurin Bevan the then Minister for Health saying, 'Illness is neither an indulgence for which people have to pay, nor an offence for which they should be penalised, but a misfortune the cost of which should be shared by the community'!

Whilst the bureaucracy, nomenclature and quantity of support have changed across the decades, in essence there remain a comprehensive framework designed to provide financial assistance and support to individuals and families in need. It aims to alleviate poverty, promote social equity, and ensure a basic standard of living citizens. The main benefits are:

1. Universal Credit (UC) was introduced in 2013 a means tested benefit that replaced six prior benefits or tax credits streamlining the process with claimants. It is designed to support low-income individuals and families incorporating elements such as housing and child support, repair costs and support for those with disabilities. It aims to incentivise work while allowing recipients to keep a portion of their earnings as they transition into employment.
2. Job Seekers Allowance (JSA) is available for individuals who are actively seeking work with 2 components, one is based on national insurance contributions. The other is an income-based allowance in which means-tested claimants must meet specific eligibility criteria including actively looking for work and being available to take up employment.
3. Employment and Support Allowance (ESA) provides financial support to individual who are unable to work due to illness or disability. Similar to JSA, ESA has both contribution-based and income-based components. Claimants undergo assessment to determine their ability to work and receive appropriate support based on their needs.

4. Personal Independence Payment (PIP) is designed to help individuals with the extra cost associated with living with a disability or long-term health condition. 'PIP' helps with the extra cost of living if someone has a long-term physical or mental health condition, or disability, and difficulty doing certain everyday tasks or getting around because of your condition. One can get PIP even if one is working, has savings or receives most other benefits. It has a daily living component and a mobility component. Eligibility to turn through assessment evaluate a person's condition impacts the daily life.
5. Housing Benefit assists individuals and families with rental costs. It is primarily aimed at low-income tenants whether in the private rental sector or social housing it is largely been replaced by universal credit but remains available for some claimants especially those and supported accommodation.
6. Child Benefit a universal payment to families with children, paid regardless of income, aimed at mitigating the costs of raising children. However higher earners are subject to tax changes that reduce or even eliminate the benefit.

The welfare benefit system has faced numerous reforms particularly since the introduction of austerity measures following the 2008 financial crisis. UK broadsheets have written about issues with the system. *The Guardian* covered the House of Lords Economic Affairs Committee report, on which 2 former Treasury Permanent Secretaries and a Chancellor of the Exchequer sit, stating that major reform was needed to address the rising social and fiscal costs of disability benefits.

The Telegraph wrote how those with 'self-diagnosed mental health conditions' can access up to £70,000 a year on the 'Access to Work' scheme', with a spend of £259 million in the financial year ending April 2024. People can apply for a maximum of £69,260 each year – twice the average wage – to help pay for adjustments, and do not have to be medically diagnosed with a health condition or disability to be eligible for support. Mental health was cited as the primary reason behind 22,000 of the 67,000 payment approvals in the last financial year, with learning difficulties the second most common condition.

Attempts to contain costs have begun with the Department of Work and Pensions addressing a rise in claims related to support workers – including ADHD coaching – reducing the hourly costs that can be reimbursed under the scheme from £450 to £205. The Prime Minister of the UK is reported to have said that the benefits system is unsustainable.

Whilst arguments rally back and forth about the statistics of validity tests, and their use clinically, whether critical, or supportive a cry against them has been that it is impossible that the number of those feigning in a clinical population could be so high.

Nonetheless Edwards *et al* reported that 'A meta-analysis of PVTs performed in a range of neurological conditions including FND found failure on single tests to be quite common across conditions, especially in people with dementia and moderate-to-severe traumatic brain injury. Although some people with FND do fail PVTs, the evidence does not suggest that they fail more often than people with other disorders.'

It has been shown that a substantial portion of patients (up to 42%) have covert motives for obtaining secondary gains associated with their patient status, e.g., financial support, help or attention from others, stimulant medication, work or study related privileges, or evasion of responsibilities. Less than 10% shared their expectations with the psychiatrist.

In the UK experts opine that there is insufficient obvious external incentive, especially when compared to the impairments described. We thought it would be useful to examine what were the potential obvious external incentives for feigning or exaggerating a condition that significantly impairs quality of life, in England & Wales.

As of the Autumn of 2024 'PIP' claimants get the standard rate if they score between eight and 11 points for their daily living needs in the PIP test. The enhanced rate applies if they score 12 points or more. Receiving PIP means the benefits cap does not apply, and if there is a limited capability for work related activity there are more benefits. These would include Housing Benefit, JSA, Income Support, Working Tax Credit, ESA and Pension Credit - if they get the PIP daily living component. Free prescription, travel support, and a reduction in Council Tax banding (a local tax set according to the value of the property) are additional follow-on benefits received by virtue of being granted PIP.

As of October 2024, there were 3.6 million claimants entitled to PIP in England and Wales. 37% received the highest level of award, for both daily living and mobility components. Sites exist that provide guidance and training for PIP applicants, to improve the chances of receiving benefits. The reported success rates for a range of somatoform, dissociative, and related disorders is listed in government statistics.

Disorder	Percent of claims succeeding
Cognitive Disorder – other Type not known [Not Dementia or due to CVA]	80.1%
Conversion disorder	79.1%
Somatoform Disorders - Other / Type Not Known	65.7%
Pain Syndromes - Chronic - Other / Type Not Known	64.8%
Fibromyalgia	62.7
Non-Epileptic Attack Disorder (Pseudoseizures)	59.2%
Munchausen Syndrome	57.1%
Back Pain	56.2%
Chronic Fatigue Syndrome	54.8%
Factitious Disorders - Other / Type Not Known	52%

Table 1: Frequency of success of applications for PIP

There are other benefits, such as NHS Prescriptions - free when an individual is on PIP, saving about £114.50. Travel benefits materialise, such as a Transport for London (TfL) travelcard for Londoners, where a Zone 1-4 travelcard is valued at £2,456. A reduction in Council Tax of one band, or a 17% reduction if on Band A. The difference between the national averages for Band C (£1,809) and Band D (£2,171) is £362.

Thus, for an individual with higher rate benefits, and savings of less than £6,000 excluding the value of their domicile, the benefits they receive are not insignificant.

Benefit	Annual Value (£)
Higher Rate maximal benefit of £529.30 per week Including Universal Credit, Council Tax, new-style ESA which gives NICs and PIP	27,523.60
TfL travel card zone 1-4	2,456
NHS prescription	114.50
Council Tax band reduction	362
Total	30,456.50

Table 2: Value of Social Benefits

The Office for National Statistics stated that the median gross annual earnings for full-time employees were £37,430 in April 2024. Using the HMRC (the office that collects revenue and taxes for the government) tax calculator, after-tax they receive £30,471, not significantly different from the benefits receivable for individuals receiving higher rate benefits for disabilities and mobility issues.

Thus, impairments assessed as eligible for higher rate PIP, 37% of awards in 2024, leads to a benefits package 98.8% of the median earnings of full-time employees.

In the opinion of some these financial benefits would be viewed as an obvious external incentive; Criteria A in the Sherman et al (2020) model for analysing exaggeration or feigning. This would raise a range of substantial issues, including diagnostic – such as with Factitious Disorder which requires that external incentives are not the motivation.

Summary: In our analysis we have avoided trying to quantify the value of hard to estimate aspects, such as the provision of informal care, or non-monetary benefits such as the surrender of social obligations, e.g., employment, caring for others, or the value of receiving care, attention or sympathy from others.

The statistics from public sites show that there are substantial direct and indirect financial benefits associated with the provision of welfare in England & Wales. When less tangible benefits are added it seems to be incontrovertible that being ill-enough to receive benefits in England & Wales is associated with obvious external incentives.

The state of symptom validity assessment in North America: Trends in procedures and prevalence estimates

Taylor Matthew², Katerine Aminot^{1,2}, Tara Ryan³, & Alicia Nijdam-Jones^{1,2}

¹Simon Fraser University, Burnaby, British Columbia, Canada

²University of Manitoba, Winnipeg, Manitoba, Canada

³Oregon State Hospital, Salem, Oregon, United States

Abstract

The assessment of non-credible performance, often referred to as malingering or symptom validity, is one of the most commonly considered issues in forensic mental health services (McLaughlin & Kan, 2014). Given the difficulty of identifying non-credible performance within forensic contexts, it is important for clinicians to have accurate estimates of non-credible performance in practice. However, commonly cited estimates are derived from older studies across various psycholegal contexts, ranging from 8% to 30% of cases (Mittenberg et al., 2002; Young, 2015). Therefore, updating prevalence estimates and assessment practices is crucial to understanding how symptom validity assessment has changed over the past decade, and whether non-credible performance has become more common in practice.

The present mixed-methods study surveyed 118 American and Canadian licensed forensic psychologists to examine symptom validity assessment procedures and prevalence estimates. Replicating several studies in Europe and other regions (e.g., Martin et al., 2025; Merten et al., 2023; Mittenberg et al., 2002; Puente-Lopez et al., 2023), participants estimated the prevalence of non-credible performance in their practice, identified the evidence that most strongly raised suspicion of non-credible performance, and described their assessment procedures.

Preliminary results revealed higher estimated rates of non-credible performance than previously reported, although estimates ranged widely across clinicians and contexts (e.g., disability claims: $M=36.2\%$, $SD=27.0$; criminal litigation: $M=25.2\%$, $SD=26.48$). Participants estimated that, on average, 44.5% of referrals were at risk for non-credible performance. The strongest indicator of suspected non-credible performance was inconsistency between clinical presentation and assessment findings. Specific assessment procedures will be discussed.

These findings highlight the limitations of prevalence estimates. Although estimated rates of non-credible performance were higher than those reported in previous studies, substantial variability existed

across clinicians. Characterizing current assessment practices and prevalence estimates provides an important point of reference for future research, but clinician-estimated prevalence should be interpreted cautiously.

Use of the Morel Emotional Numbing Test (MENT) in the detection of feigned psychopathology

Sara Pasqualini^{1,2}, Veronica Tatti¹, Claudia Pignolo¹, & Luciano Giromini¹

¹Department of Psychology, University of Turin, Turin, Italy

²Department of Human Neurosciences, University of La Sapienza, Rome, Italy

Abstract

Malingering is defined as the fabrication or exaggeration of physical or psychological symptoms for the purpose of obtaining external incentives, such as financial compensation or reduced legal responsibility. Its prevalence in forensic settings ranges from 5% to 40%, representing a substantial social and economic burden. This research represents the first Italian validation of the Morel Emotional Numbing Test (MENT), a Performance Validity Test (PVT) specifically designed to detect feigned Post-Traumatic Stress Disorder (PTSD), aimed at examining its diagnostic accuracy and discriminative ability. A simulation design was conducted with 400 non-clinical Italian adults. The study aimed to evaluate the MENT's ability to accurately classify honest respondents and experimental simulators across four experimental conditions. Furthermore, its incremental validity over a battery of Symptom Validity Tests (SVTs) and Performance Validity Tests (PVTs) was examined. Participants were randomly assigned to one of four groups: (a) simulators instructed to feign PTSD ($n = 100$), (b) simulators instructed to feign mild traumatic brain injury (mTBI; $n = 100$), (c) simulators instructed to feign autism ($n = 100$), and (d) honest respondents instructed to answer truthfully ($n = 100$). Results revealed a strong discriminative ability of the MENT, with sensitivity values above .60 and specificity values above .90 across all conditions (PTSD, mTBI, autism, and honest respondents). These findings suggest that the MENT is a promising and highly specific measure for detecting feigned psychopathology, thereby enhancing the accuracy and efficiency of forensic psychological assessment.

Keywords: malingering, forensic assessment, MENT, classification accuracy, PVTs

Validation of the Blockwise Attention Test (BAT) “Palamedes”

Ingo Pals

BG Trauma Center Murnau, Murnau, Germany

Health Innovation University Berlin (HIU), Berlin, Germany

Abstract

Background: Symptom validity assessment is a cornerstone of neuropsychological diagnostics (Larrabee, 2012; Merten et al., 2022). While conventional performance validity tests (PVTs) predominantly utilize memory-based forced-choice paradigms, attention-focused indicators remain critically underrepresented. Though structured as a concentration task, the BAT serves as a covert validity indicator detecting non-authentic response behaviour. This pilot study provides a primary empirical validation of the instrument.

Methods: The computerized BAT operationalizes covert ease via a visual interference paradigm requiring participants to name the presented letter or number while ignoring the respective distractor under stringent temporal constraints (exposure: 200–400 ms; Hiscock & Hiscock, 1989). This induces an intentional ceiling effect in authentic performers while simulating a challenging attention test. Across 4 blocks (160 trials), audio-recorded verbal responses were collected from 87 participants stratified into three cohorts (each $n = 29$): clinically diagnosed with Attention-Deficit/Hyperactivity Disorder (ADHD), an instructed simulation group, and a matched healthy control baseline.

Results: The simulation cohort exhibited significantly lower success rates than baseline and ADHD groups ($p < .001$). ROC analysis demonstrated high discriminability ($AUC = 0.994$; $CI_{95\%} [0.983, 1.000]$), confirmed via bootstrapping (5,000 resamples; $CI_{95\%} [0.979, 1.000]$). Driven by the Youden Index, the optimal threshold was established at a 91.3% success rate (sensitivity and specificity: 96.6%).

Conclusions: This initial piloting offers promising evidence for the BAT (“Palamedes”) as a covert performance validity indicator. However, a methodological limitation inherent to this pilot stage is the reliance on an instructed simulation design, which may not fully capture the nuanced and heterogeneous presentation of real-world malingering. Given the well-documented prevalence of non-credible performance in mild traumatic brain injury populations (Larrabee, 2003), future validation trajectories must systematically evaluate the BAT’s diagnostic accuracy within known-groups designs involving actual clinical patients undergoing forensic or medico-legal assessments.

Validation of the Groningen Effort Test (GET) on Italian samples in a simulation design

Losavio F.¹, Fuermaier A.B.M.², & Curci A.¹

¹Department of Education, Psychology, Communication, University of Bari 'Aldo Moro', Bari, Italy

²Department of Clinical and Developmental Neuropsychology, University of Groningen, Groningen, the Netherlands

Abstract

In clinical and forensic settings, assessing the validity of symptomatology is crucial, as secondary motivations may lead individuals to fabricate, exaggerate or minimize symptoms. Malingering is defined as the intentional production of false or exaggerated symptoms for secondary gain. The literature supports multimethod assessment protocols integrating Performance Validity Test (PVTs) and Symptom Validity Test (SVTs). This study presents an Italian validation of the Groningen Effort Test (GET), an attention-based PVT that represents an innovation for the Italian context compared with gold-standard measures focused on memory tasks. The GET employs a forced-choice embedded-figures paradigm (96 items) and provides two computerized composite indices (ADHD and CVA) through the Vienna Test System. A total of 200 adults were randomly assigned to one of three experimental conditions: a control group ($N = 50$), a naive simulator group ($N = 100$) instructed to feign ADHD without guidance, and a coached simulator group ($N = 50$) trained to feign ADHD plausibly. Participants also completed the Test of Memory Malingering (TOMM) and the Self-Report Symptom Inventory (SRSI). Experimental condition significantly affected both GET composite indices (ADHD and CVA; $F = 17.816$ – 17.46 ; $p < .001$), with instructed simulators scoring significantly higher than controls. The GET indices were strongly correlated with the TOMM ($r = .681$ – $.635$; $p < .001$) and moderately correlated with the SRSI ($r = .334$ – $.298$; $p < .001$). ROC analyses showed good-to-excellent discrimination between simulators and controls ($AUC = .858$ – $.891$; $p < .001$), with no significant difference between naive and coached simulators ($AUC = .588$ – $.554$; $p < .066$ – $.281$). The GET demonstrated promising psychometric properties as a novel attention-based PVT and appears relatively resistant to coaching effects. These findings extend the evidence for the GET beyond its original validation, contribute to evaluating its cross-cultural applicability and represent a step toward establishing the instrument's cultural invariance.