
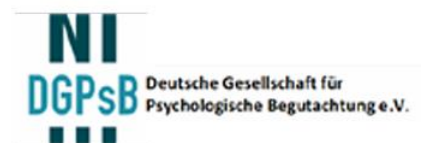
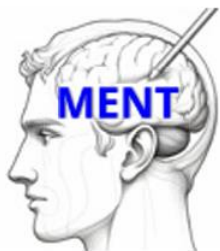


# Abstract Book

7<sup>th</sup> European Conference on Symptom Validity Assessment (SVA)

10<sup>th</sup> – 12<sup>th</sup> October, 2024

 Polignano a Mare, Italy



## Contents

<b>Congress Committees</b> .....	<b>4</b>
<b>Symposia Abstracts</b> .....	<b>6</b>
Attention and Validity Assessment .....	7
Validity testing in children with ADHD (Nicholls).....	8
Performance validity testing of adults referred for the assessment of adult ADHD (Fuermaier).....	9
Attention-based performance validity assessment in chronic solvent-induced encephalopathy (Van Vliet et al.) .....	10
Brain electrical activity in symptom and performance validity assessment in ADHD (Albrecht et al.) .....	12
Validity Assessment: Detection, Feedback and Recommendations .....	13
Feedback on SVT/ PVT failure: An experimental demonstration (Roor et al.).....	14
Interviewing forensic inpatients on their biased responding (Van Helvoort et al.) .....	16
Introducing a new instrument for the detection of feigned hallucinations (Peters) .....	17
Eight recommendations for advancing the field (and a question) (Dandachi-FitzGerald & Merckelbach) .....	18
Culture and Validity Assessment.....	19
Ethnoracial and gender effects on decision-making about malingering/ feigning (Akca).....	20
Cross-cultural applicability of the MENT (Morel) .....	21
Cross-cultural applicability of IOP-29 and SIMS (Puente-López).....	22
Help! The rest of the world is on my doorstep! (Brockhaus).....	23
Validity Assessment and Biases: Concealment, Practitioners' Biases, and Other Issues .....	24
Detecting deception at single item level in questionnaire: The TF-IDF-based approach (Melis et al.) .....	25
Symptom concealment and its association with defensive functioning (De Page) .....	26
Uncovering social desirability in responses: novel approaches to psychological assessment (Bosco et al.) .....	27
Derivation of a novel PAI discriminant function for the detection of concealed psychopathology (Kurtz) .....	28
Variation in social desirability by assessment purpose: A comparative study (Mazza et al.) .....	29
Addressing bias in DSM criteria for malingering suspicion (Akca) .....	30
Memory and Validity Assessment.....	31
Applying the verifiability approach to detect malingered COVID-19 symptoms (Ribatti et al.) .....	32
Analysis of response latencies and error rates to unexpected questions (Melis et al.) .....	33
Distinguishing true from false autobiographical memories (Battista et al.).....	34
<b>Poster Abstracts</b> .....	<b>35</b>
Classification accuracy of the IOP-29-M across administration languages in Bilinguals (Crişan et al.) .....	36
Unveiling crime-related amnesia using a visual version of Forced Choice Test (Zago et al.) .....	37
Once is enough! An analogue study on repeated validity assessment in adults with ADHD (Dong et al.) .....	38
The Groningen Effort Test for assessing performance validity in early retirement claimants (Teßmann et al.) .....	39
The utility of attention-based performance validity testing (Kneidinger et al.) .....	40
Performance validity assessment in a clinical children and adolescent sample (Raasch et al.) .....	41
Assessing memory over-reporting with the Forgetfulness Assessment Inventory (Bayrhammer-Savel & Lehrner) .....	42
Does coaching weaken the accuracy of symptoms (SVT) and performance (PVT) validity tests? (Pasqualini et al.) .....	43
Dutch insurance physicians' practices and beliefs on using validity tests: A survey study (Giles et al.) .....	44

Equivalency of in-person versus remote assessment of the PAI, IOP-29, and IOP-M (Pignolo & Giromini) .....	45
Feigning judgments and misunderstanding the negative predictive power of symptom validity tests (Niesten et al.) ..	46
Say what? The interpreter effect: Impact of interpreters on symptom and performance validity scores (Alsop) .....	47
Uncovering drug use using mouse kinematics: Validity assessment through symptom sentence analysis (Melis et al.)	48
The Tangram Task: first results of a newly developed symptom validation test (Riedl & Exner).....	50
<b>Keynote</b> .....	<b>51</b>
Malingering in psychological injury and law: New developments (Young).....	52
<b>Roundtable</b> .....	<b>53</b>
Validity assessment in Europe: Where are we heading? (Merten et al.).....	54
<b>Workshops</b> .....	<b>55</b>
The Morel Emotional Numbing Test: A validity tool for ensuring the accuracy of PTSD assessments (Morel) .....	56
Assessment of negative response distortion with the personality assessment inventory (PAI) (Kurtz).....	57
Multivariate models of performance and symptom validity tests: Conceptual and practical considerations (Erdodi)...	58
Assessing symptom/ performance validity with the Inventory of Problems – 29 and its Memory Module (Giromini) ..	59
The Self-Report Symptom Inventory (SRSI): Five years after its publication (Merten) .....	60

## Congress Organizers

[Luciano Giromini](#)

Italy

[Anselm Fuermaier](#)

The Netherlands

[Domenico Laera](#)

Italy

[Harald Merckelbach](#)

The Netherlands

## Scientific Program Committee

[Luciano Giromini](#)

Italy

[Anselm Fuermaier](#)

The Netherlands

[Brechtje Dandachi-FitzGerald](#)

The Netherlands

[Irena Boskovic](#)

The Netherlands

[Andrea Plohmann](#)

Switzerland

[Esteban Puente-López](#)

Spain

[Claudia Pignolo](#)

Italy

[Vicki Hall](#)

United Kingdom

[Antonietta Curci](#)

Italy

[Thomas Merten](#)

Germany

[Harald Merckelbach](#)

The Netherlands

## Poster Committee

[Vicki Hall](#)

United Kingdom

[Claudia Pignolo](#)

Italy

[Brechtje Dandachi-FitzGerald](#)

The Netherlands

[Jeroen Roor](#)

The Netherlands

## Organizational Assistance

[Janina Teßmann](#)

The Netherlands

[Veronica Tatti](#)

Italy

## Organizational Partner

[Studio RiPsi](#)

Italy

# Symposia Abstracts

# Attention and Validity Assessment



## Validity testing in children with ADHD

**Christopher Nicholls**

The Nicholls Group, Scottsdale, AZ, USA; Adjunct Faculty, Department of Psychology, Arizona State University, Tempe, AZ, USA

### **Abstract**

College students and adults who present for evaluations of Attention-Deficit/Hyperactivity Disorder (ADHD) have been found to often fail validity testing. Our research has also found that younger children who fail one or more standalone objective performance validity tests (PVT) also report significantly more symptoms than those who pass PVTs. Although some question the utility of objective cognitive testing in the diagnosis of ADHD, this conclusion may also be secondary to not considering the role of performance validity, and some researchers have concluded that embedded PVTs are less sensitive than standalone measures. Our research has found that ADHD children who fail standalone PVTs obtain substantially lower scores on various components of traditional computerized continuous performance tests. We have also found that parent ratings of ADHD symptoms don't correlate with objective testing when measures of performance validity are not considered, suggesting that parents may be insensitive to poor effort on their children's part during such testing. More recently, the Nesplora immersive virtual reality tests of executive functions have been found to offer a more ecologically valid assessment of ADHD in the "real life" setting of a school classroom. We find that children who failed common standalone PVTs scored lower on Nesplora Attention Kids aula scores than those who passed, but not to a significant degree. Nevertheless, those children who failed the embedded validity measures of aula scored significantly below ADHD children, as compared with those who passed, on the subtests of attention, vigilance and motor activity. We propose that because ADHD is not a unitary phenomenon, and that because subtypes of ADHD suggest impairments in differing brain regions and executive functions, the analysis of children's symptom and validity test results must be considered from a dimensional perspective, including gender, developmental factors, and the nature of the skill sets being assessed.



## **Performance validity testing of adults referred for clinical evaluation of adult ADHD**

**Anselm B.M. Fuermaier**

Department of Clinical and Developmental Neuropsychology, Faculty of Behavioral and Social Sciences, University of Groningen, Groningen, The Netherlands

### **Abstract**

Cognitive deficits in attention and executive control are core features of attention-deficit/hyperactivity disorder (ADHD) in adulthood, which requires adequate assessment. Many, if not all, assessment settings make use of standardized self- and other-report rating scales, and/or neuropsychological performance tests, in order gain a comprehensive and valid picture of the cognitive functioning of the client. However, the neuropsychological assessment is complicated by non-trivial base rates of symptom overreporting and cognitive underperformance of adults referred for a clinical evaluation of adult ADHD. In this talk, I will present evidence defining the role of a neuropsychological assessment of adults referred for a clinical evaluation of adult ADHD. My talk will focus on attention deficits, the utility of attention tests for embedded validity testing, and the need for stand-alone attention-based performance validity tests. Data from our working group will be shown to empirically elucidate the differentiation between symptom and performance validity assessment and their role in clinical assessment batteries. I will conclude with a broader perspective of attention-based performance validity testing in near- and far-related mental health conditions.

## Validity of the Groningen Effort Test in patients with suspected chronic solvent-induced encephalopathy

Fabienne I. M. van Vliet<sup>1,3</sup>, Henrita P. van Schothorst<sup>2</sup>, Birgit H.P.M. Donker-Cools<sup>1,4</sup>, Frederieke G. Schaafsma<sup>1,4</sup>, Rudolf W.H.M. Ponds<sup>3</sup>, & Gert J. Geurtsen<sup>3</sup>

<sup>1</sup> Department of Public and Occupational Health, Amsterdam Public Health research institute, Amsterdam University Medical Centres, Amsterdam, location AMC, The Netherlands

<sup>2</sup> Department of Psychology, Faculty of Social and Behavioural Sciences, University of Amsterdam, Amsterdam, The Netherlands

<sup>3</sup> Department of Medical Psychology, Amsterdam University Medical Centres, location AMC and VUmc, Amsterdam neuroscience, Amsterdam, The Netherlands

<sup>4</sup> Research Centre for Insurance Medicine: collaboration between AMC-UMCG-UWV-VUmc, Amsterdam, The Netherlands

### Abstract

**Introduction.** Most performance validity tests (PVTs) are memory-based. Although memory loss is the most prevalently presented cognitive complaint, attention complaints are also frequently seen. By using memory-based PVTs only, attention-based invalid performance may be missed. For that reason the Groningen Effort Test (GET) was developed, which represents attention and visuospatial-perception.

A clinical population in which invalid performance occurs frequently is suspected chronic solvent-induced encephalopathy (CSE). CSE is caused by long-term, mostly work-related, exposure to organic solvents and can lead to cognitive impairments. Memory loss and attention problems are the most prevalent complaints in patients with suspected CSE.

**Objective.** This study aimed to validate the GET in patients with suspected CSE using the criterion standard of 2PVTs, determine the diagnostic accuracy for GET and the coherence with neuropsychological assessment.

**Methods.** Sixty patients with suspected CSE were included. The GET was compared to the criterion standard of 2PVTs based on the Test of Memory Malingering and the Amsterdam Short Term Memory Test. Comparative analyses were done between 2PVTs, GET and neuropsychological tests.

**Results.** The frequency of invalid performance using the GET was significantly higher compared to the criterion of 2PVTs (51.7% versus 20.0% respectively;  $p < 0.001$ ). For the GET-index, sensitivity was 75%

and specificity 54%. No significant correlation between PVTs and neuropsychological tests were found.

**Conclusion.** The GET showed significantly more invalid performance compared to the 2PVTs criterion suggesting a high number of false positives. The minimum norm of specificity for PVTs of >90% was not met. Further, we did not find significant correlations between the GET, 2PVTs and memory, attention and visuospatial tasks. However, development of attention-based PVT is still warranted, since comparison between attention-based PVT and memory-based PVT is a subject of discussion. For now, caution for using the GET in a clinical population such as CSE is advised.

## Symptom and Performance Validity in ADHD – Is the Imitation Game Harder if Brainelectrical Activity is Concerned?

Björn Albrecht<sup>1</sup>, Anselm Fuermaier<sup>2</sup>, Jonathan Haarmann<sup>1</sup>, Anna Westenfelder-Gil<sup>1</sup>, Johanna Kneidinger<sup>1</sup>, Hanna Christiansen<sup>1</sup>, & Mira-Lynn Chavanon<sup>1</sup>

<sup>1</sup> Clinical Child and Adolescent Psychology Department of Psychology, Philipps-University Marburg, Marburg, Germany

<sup>2</sup> Department of Clinical and Developmental Neuropsychology, Faculty of Behavioral and Social Sciences, University of Groningen, Groningen, The Netherlands

### Abstract

Invalid symptom reports and performance in tasks assessing symptoms of attention deficit hyperactivity disorder (ADHD) may be a rather common, but frequently overlooked complication in clinical assessments (Fuermaier et al. 2021). Feigning, but also avoiding a diagnosis of ADHD, may be motivated by seeking possible secondary disorder benefits or compensation for disadvantages as well as avoiding stigmatization or losses – all of which may compromise proper diagnosis, treatment and support.

Aim of this multi-level study is exploring how healthy participants simulate or instead avoid ADHD symptoms in neuropsychological assessments, how simulation may be recognized in the Groningen Effort Test (GET, a performance validity test) as well as Flanker-Task (FTA) performance and brainelectrical activity as endophenotypes of ADHD (Albrecht et al. 2008, McLoughlin et al. 2009).

The current study is still ongoing and we aim a sample size large enough for detecting medium effects with conventional significance level ( $\alpha=.05$ ) and power ( $1-\beta=.80$ ) between a sample of healthy subjects that simulates vs. avoids difficulties associated with ADHD in behavioral ratings and neuropsychological assessments using a pre-post manipulation repeated measure design.

Preliminary data suggests that those simulating show slower, more variable and more error-prone responses in FTA and GET. In addition, response-locked brainelectrical activity was also compromised, in particular the error positivity (Pe) associated with affective assessment of committed errors. Simulants may be discriminated with high accuracy from Avoiders by means of the Groningen Effort Test Index, Flanker-Task performance or Pe-amplitude.

Implications of these findings and further evidence differentiating healthy simulants from patients with ADHD will be discussed.

# Validity Assessment: Detection, Feedback and Recommendations



## A Sympathetic Approach to Feigning is Preferable Compared to a Confrontational Approach to Impact Subsequent Testing: An Experimental Demonstration

Jeroen Roor<sup>1,2</sup>, Fleur Bloem<sup>3</sup>, & Brechje Dandachi-FitzGerald<sup>3,4</sup>

<sup>1</sup> Department of Medical Psychology, VieCuri Medical Center, Venlo, The Netherlands

<sup>2</sup> School for Mental Health and Neuroscience, Maastricht University, Maastricht, The Netherlands

<sup>3</sup> Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands

<sup>4</sup> Faculty of Psychology, Open University, Heerlen, The Netherlands

### Abstract

**Objective.** The aim of the current study was to explore whether confrontational versus sympathetic feedback differ in their effectiveness to influence both symptom over-reporting and underperformance on cognitive testing.

**Method.** Two experimental groups were given a case vignette about a social security disability claimant and were instructed to emphasize with the role described and feign symptoms. Pre- and post-experimental checks were performed to examine whether the participants understood and executed these instructions. One experimental group was provided with confrontational feedback, and the other with sympathetic feedback. All participants subsequently completed the Self-Report Symptom Inventory (SRSI) and the Verbal Learning Test (VLT). Responses of the two experimental feedback groups ( $n = 15$  each) were compared to a control group ( $n = 15$ ) of participants who responded honestly.

**Results.** A total of 12/30 (40.0%) of the participants in the two intervention groups stopped feigning after feedback, with no significant group differences between the confrontational ( $n = 5$ ) and sympathetic feedback intervention ( $n = 7$ ) (Fisher's Exact Test,  $p = .710$ ). Additionally, Kruskal-Wallis testing found significant group (confrontational feedback, sympathetic feedback, versus control participants) differences for both genuine and pseudo SRSI symptom reporting and for VLT performance. SRSI scores remained higher and VLT performance lower for the feedback groups compared to the control participants. However, pairwise comparisons revealed significant group differences in favor of the sympathetic feedback approach to exaggeration compared to the confrontational feedback approach.

**Conclusion.** Sympathetic and confrontational feedback are equally (in)effective in getting instructed feigners to respond honestly. Next, our findings confirm that feigning has residual effects. And above all, that a sympathetic feedback approach to exaggeration seems preferable compared to

a confrontational approach. Further research is needed to examine various forms of feedback and other approaches to manage patients' test-taking efforts before, during, or after (neuro)psychological assessment.

## Interviewing Forensic Inpatients in Cases of Response Bias: Identifying the Antecedents

Daniël van Helvoort<sup>1,2</sup>, Harald Merckelbach<sup>1</sup>, Chijs van Nieuwenhuizen<sup>2,3</sup>, Henry Otgaar<sup>1,4</sup>, & Brechje Dandachi-FitzGerald<sup>1,5</sup>

<sup>1</sup> Clinical Psychological Science, Maastricht University, Maastricht, The Netherlands

<sup>2</sup> GGzE, Institute for Mental Health Care, Eindhoven, The Netherlands

<sup>3</sup> Tranzo, Scientific Center for Care and Wellbeing, Tilburg University, Tilburg, The Netherlands

<sup>4</sup> Leuven Institute of Criminology, Catholic University of Leuven, Leuven, Belgium

<sup>5</sup> Clinical Psychology, Open University of the Netherlands, Heerlen, The Netherlands

### Abstract

In clinical practice, validity tests are increasingly used as a standard means to assess the credibility of symptom presentations. Yet, guidelines on how to discuss validity test outcomes with patients and how to identify antecedents of symptom distortions are largely lacking. Thus, clinicians often feel at a loss in drawing hypotheses about practical implications in cases of poor symptom validity. We asked Dutch forensic inpatients at intake (within 2 weeks after admission) and around their first treatment plan evaluation (+/- 6-8 weeks later) to complete questionnaires that included validity scales that tap into inattentive responding and over- and underreporting. Next, patients who exceeded established cut-offs on one or more validity scales were administered a novel semi-structured interview aimed at identifying antecedents of poor symptom validity. In the interview they were asked to comment on their supposed treatment and social support needs by going over a so-called “checklist of circumstances”, followed by questions about some of their specific validity scale responses. Thereafter, the overall response pattern was labeled as unusual and the patients were encouraged to explore underlying reasons. Inattentive responders tended to emphasize boredom or tiredness at testing, and the comprehensibility or phrasing of questions. Overreporting patients alluded to alexithymia, social attention, medication, and/or accommodative benefits. Cases of underreporting related to a lack of illness insight, stigma, distrust, or self- and/or other-deception. Overall the interviews provided valuable information as to the attitude, motivation and responsiveness of the forensic patients, and provided a means to identify and discuss potential treatment avenues (e.g. psychoeducation; experiential interventions; a focus on motivation and boundary conditions). These findings illustrate that in order to optimize the utility of validity tests in clinical practice, conceptual and practical studies are a welcome addition to current research themes.



## **The Assessment of Hallucination Symptoms (AHS): Introducing a New Instrument for the Detection of Feigned Hallucinations**

**Maarten J. V. Peters**

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

### **Abstract**

Feigning of psychotic symptoms seems to be quite common in forensic practice, with hallucinations being the preferred symptom. If one looks into the literature, the detection of these feigned hallucinations demands significant time and effort. Methods mainly focus on detailed exploration of the features of hallucinations in a clinical interview (e.g., Resnick & Knoll, 2018). Up until now, no specific symptom validity questionnaire existed in speeding up this process. In this presentation, we will introduce a newly developed 55-items questionnaire: the Assessment of Hallucination Symptoms (AHS). This self-report questionnaire, based on earlier findings by Resnick & Knoll (2018), asks for 21 genuine hallucination and 34 atypical hallucination indicators that are scored on a correct-incorrect format. In this presentation we will present the first findings on diagnostic accuracy of the AHS, in comparing this questionnaire to the SIMS and SRSI. Furthermore, its robustness against coaching and genuine psychopathology were examined in a known-groups comparisons with in total 94 non-clinical participants (instructed to feign and coached for feigning hallucinations), 27 forensic psychiatric patients and 28 voice-hearers. The first results show that the AHS has the potential to differentiate feigners from non-clinical responders. Furthermore, acceptable reliability indicators and promising results were found related to differentiating non-coached feigners from genuine patients. In this presentation we will further elaborate on these preliminary findings.

## **Eight Recommendations for Advancing the Field (and a Question)**

**Brechje Dandachi-FitzGerald<sup>1,2</sup> & Harald Merckelbach<sup>1</sup>**

<sup>1</sup> Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands

<sup>2</sup> Faculty of Psychology, Open University, Heerlen, The Netherland

### **Abstract**

1. **Strive for Open Science:** Foster transparency and collaboration.
2. **Revise test nomenclature:** Ensure clarity and precision in test naming.
3. **Invest in Psychometric Training:** Equip yourself and your peers with robust assessment skills.
4. **Pioneer causal models:** Conceptualize, theorize, and test causal hypotheses.
5. **Fully address response biases:** Focus also on underreporting, careless responding and any combination of response biases.
6. **Diversify publication outlets:** Expand dissemination beyond specialized journals for broader impact.
7. **Embrace diversity:** Explore across cultures and demographics.
8. **Move beyond the honesty preoccupation:** Look into the communicative and therapeutic ramifications.

*Do you have recommendations? Let us know!*



# Culture and Validity Assessment



## Interplay of Ethnoracial and Gender Characteristics of Assessee in Decision-making About Malingering/Feigning

Ali Yunus Emre Akca

Maastricht University, Maastricht, The Netherlands & University of Turin, Turin, Italy

### Abstract

It is unclear whether the gender or race of the person being assessed influences the interpretation of the validity scale results. Participants rated on an 11-point Likert scale the likelihood and confidence that the MMPI validity scale profiles indicated a credible symptom report in a person of a particular race (“white” vs. “black”) and/or gender (“male” vs. female). Each participant was randomly assigned to one of four conditions, using equivalent scenarios matched to each race and gender. The (preliminary) sample consisted of 117 professionals ( $n = 60$  forensic and legal assessors), 56.41% of whom had more than 10 years of experience and 86 of whom reported being moderately to very familiar with symptom validity assessment (SVA). Most participants ( $n = 32$ ) come from the USA, followed by Spain ( $n = 27$ ) and Italy ( $n = 11$ ). The vast majority (81.20%) described themselves as white, European or American-European. Preliminary results indicate a statistically significant difference in the likelihood ratings of credibility from an “ideal unbiased” value of 5 for race or gender ( $p \leq .019$ ). Analysis of variance revealed no statistically significant differences between conditions on likelihood, confidence, and suspicion (likelihood and confidence ratings averaged as a total score; range 0 to 100) ratings,  $F_{(3,113)} < .720$ ,  $p \geq .542$ . The Kruskal-Wallis H-test also do not reveal significant differences in likelihood, confidence, and suspicion ratings between the four conditions ( $p > .05$ ). To summarize, preliminary experiments indicate that professionals with experience in assessment and SVA interpret the MMPI validity scale profiles to a similar extent between races and genders.

## **Cross-cultural applicability of the MENT**

**Kenneth Morel**

National Institutes of Mental Health, Neuropsychiatry Branch (Ret.), USA & U.S. Army Medical Command (Ret.)

### **Abstract**

Assessing response validity across diverse cultural groups presents a significant challenge. Effective methods for evaluating response validity are crucial as they provide insights into the credibility of an examinee's self-reported symptoms or test performance. High error variance can obscure genuine differences, leading to incorrect conclusions about a test's applicability across cultures. The MENT was meticulously designed to minimize error variance while preserving its cross-cultural applicability and clinical utility. By leveraging statistical principles of binomial distribution and neuropsychological insights, the MENT ensures that the measurement tool consistently captures the intended construct across different cultural groups. This enhances the validity and reliability of test results, allowing for clearer interpretation. Throughout the MENT's design and development, sources of error variance were systematically identified and controlled, resulting in an accurate, reproducible, and absolute performance-based measure of response validity. Statistically, the standard error of measurement (SEM) quantifies the precision of an individual's observed score by accounting for error variance. Pooled data from over 1,000 credible examinees from various countries demonstrate a remarkably low SEM of 0.06 for the MENT.

## **Cross-cultural applicability of IOP-29 and SIMS**

**Esteban Puente-López**

Universidad de Valladolid, Valladolid, Spain

### **Abstract**

Assessing the validity of reported symptoms is a fundamental task in clinical and forensic settings. Professionals can use for this purpose, among other resources, a psychometric battery adapted to the characteristics of the population and the context of interest. Such adaptation cannot be taken for granted (i.e., directly using tests that have been created in other countries), but must go through an empirical process in which the possible variation of psychometric properties between populations is analyzed. The objective of this session is to analyze the possible variation between countries in the classification accuracy (sensitivity and specificity) of the Inventory of Problems - 29 (IOP-29) and the Structured Inventory of Malingered Symptomatology (SIMS) in the forensic context. For this purpose, a bivariate meta-analysis will be performed in which the literature findings of both tests since their year of publication will be synthesized.

## Help! The Rest of the World is on My Doorstep!

**Robbi Brockhaus**

Institute for Psychological Diagnostics and Assessment, Duisburg, Germany

### **Abstract**

Because of complex worldwide political, economic, and climatic developments, societies are increasingly becoming the new home for settling immigrants – a phenomenon that offers both challenges and advantages to each single homeland. For psychologists charged with the evaluation of individuals from other countries and minorities, there can be a myriad of technical and methodological difficulties entailed in the assessment procedure. The paucity of answers for unsolved diagnostic issues is often bothersome. Finding a common language for communication and test materials is only one of the problems involved. Frequently, culturally-compatible psychologists are not available in our own homeland. Assessing a subject without having basic insight into the person's cultural background can be a major source of misunderstanding, frustration, and distortion when working with the subject. Assuming others have an implicit Eurocentric world view and reality construction similar to ours can be calamitous. This lecture offers some practical suggestions for dealing with situations in which the assessor has limited understanding of the divergent cultural matters involved.

Secondly, it is well known that validating test results is a necessary, standard element of all psychological assessments. When dealing with culturally divergent individuals, symptom validation SVT and performance validation PVT become even more important as a source of extended information. On the basis of our large data bank including subjects in varying numbers from West European, East European, Turkish, Balkan and African countries (Brockhaus & Dohrenbusch), some results will be shown to illustrate possible cultural trends.

An additional difficulty in the discussion of symptom validity and culture involves the issue of political correctness of statements. Literature suggestions will be offered.

# Validity

## Assessment and Biases: Concealment, Practitioners' Biases, and Other Issues





## Detecting Deception at Single Item Level in Questionnaires: the TF-IDF-Based Approach

Giulia Melis<sup>1</sup>, Graziella Orrù<sup>2</sup>, & Giuseppe Sartori<sup>1</sup>

<sup>1</sup> Università di Padova, Italy

<sup>2</sup> Università di Pisa, Italy

### Abstract

Deception in forensic contexts poses a significant challenge due to the high percentage (from 25% to 45%) of behaviours that are deceptive. Consequently, various efforts have been made to unmask dishonest responding (e.g. L, F, and K scales of the Minnesota Multiphasic Personality Inventory-3 or the X, Y, and Z scales of the Millon Clinical Multiaxial Inventory-III.) Abnormal scores on these scales indicate an overall tendency of the subjects to modulate their responses in the direction of social desirability (faking good) or hyperbolic psychopathology (faking bad). Despite these measures, existing methods fall short in evaluating the intensity of specific symptoms or identify simulation focused on a particular item in a questionnaire. In this work, we explored the potential of the Term Frequency – Inverse Document Frequency (TF-IDF) model as a tool for detecting dissimulation in self-report questionnaires. This approach identifies the infrequency of a particular response and the degree of deviation from other subjects' responses, highlighting deception at the single-item level. We validated the proposed model using the 10-item Big Five Test, a widely used short version of the Big Five personality test. We administered this questionnaire to a group of volunteers ( $n = 694$ ) instructed to answer twice (using a 5-point Likert scale): the first time dishonestly according to specific instructions and the second time honestly. We provided the following three faking contexts: (a) a child custody case ( $n = 221$ ), (b) a job interview for a sales manager position ( $n = 243$ ), and (c) a job interview for a position in a humanitarian organization ( $n = 230$ ). The proposed TF-IDF model has proven to be very effective in distinguishing between authentic responses and those that have been artificially manipulated, enhancing the reliability and validity of test outcomes. One of the main contributions of this approach is that two identical responses (from two participants) to the same item may yield different evaluations. This variation is based on other participants' responses to the same items and the subject's responses to the questionnaire's other items. In other words, TF-IDF aggregates a unique measure that considers the distribution of group responses to a target item and the subject's response style. The TF-IDF index, as presented here, can be regarded as a novelty detector, where a high TF-IDF score of a participant's response to an item would indicate that their response deviates significantly from those of honest respondents.

## Symptom concealment and its association with defensive functioning

Louis De Page

Vrije Universiteit Brussel (VUB), Brussels, Belgium. Centre Hospitalier Jean Titeca, Brussels, Belgium. Mediter, Halle, Belgium

### Abstract

Symptom concealment (SC) mostly exist within positive impression management (PIM) schemes, but can also occur within negative IM schemes. Reasons for SC range from obvious (to obtain evident material outcomes) to less obvious reasons (for psychological gains) to no apparent reason at all. In a series of articles, we found that psychotherapy patients engaged far more in PIM/SC than previously thought with no obvious reasons. We posit that SC is inextricably bound to defensive processes whenever we leave the strictly conscious or aware zone of psychological functioning. Paulhus (1997) suggested four dimensions for defensive operations: 1) their cognitive complexity and level of development, 2) their internal-external orientation, 3) their maturity-immaturity, and 4) their level of conscious awareness. We think that instances of PIM/SC also vary along these dimensions, which allows for a finer understanding of the phenomena. We will present data from psychotherapy patients regarding the association between PIM/SC and defensive processes. Preliminary results suggest that PIM/SC is positively correlated with mature defensive functioning, and negatively correlated with immature defensive functioning. Self-deceptive and conscious PIM/SC might reflect partially overlapping but distinct processes at different maturational levels: less conscious, self-deceptive enhancements, naïve forms of PIM/SC seem to be mediated by affect regulating defenses whilst aware positive self-presentations seemed to be associated with higher maturational defenses. We will end this presentation with concluding thoughts on the phenotypic similarity of these concepts: “Is this patient making a good impression, is he/she coping very well, or is he simply healthy?”.

## Uncovering Social Desirability in Responses: Novel Approaches to Psychological Assessment

Francesca Bosco<sup>1</sup>, Cristina Mazza<sup>2</sup>, & Paolo Roma<sup>1</sup>

<sup>1</sup> Department of Human Neuroscience, Sapienza University of Rome, Italy

<sup>2</sup> Department of Dynamic and Clinical Psychology, and Health Studies, Sapienza University of Rome, Italy

### Abstract

Self-report questionnaires, extensively employed in psychological assessment, operate under the assumption that individuals provide accurate and honest responses. However, in specific contexts like child custody evaluations, weapon eligibility, and personnel selection, there exists a possibility that subjects may, whether consciously or not, present themselves in an overly favorable manner. This phenomenon, known as socially desirable responding (SDR), represents a significant source of bias within psychological assessments, leading to considerable social and economic implications, particularly in critical evaluative situations. In this simulation study we aimed to: a) evaluate the potential utility of a new variation of the Implicit Association Test (IAT), termed the SDR-IAT, in distinguishing between genuine responses and those that are distorted; and b) enhance our understanding of the efficacy of self-report questionnaires in measuring SDR by assessing the accuracy of various scales and items. Participants were randomly allocated into two groups: one group completed personality questionnaires honestly (i.e., control group) and then took the SDR-IAT, while the other group answered a personality questionnaire in a socially desirable manner before undergoing the SDR-IAT (i.e., simulator group). The data are not available in this moment, but we expect to find that SDR-IAT aids the distinction between the members of control and simulator groups. The findings from this exploratory research shed light on the effectiveness of novel approaches to measure SDR, proving particularly beneficial in high-stakes scenarios.

## Derivation of a novel PAI discriminant function for the detection of concealed psychopathology

John E. Kurtz

Department of Psychological and Brain Sciences, Villanova University, Villanova, PA, USA

### Abstract

The Personality Assessment Inventory (PAI) provides several scores designed to assess positive response distortion (PRD), including the Positive Impression Management (PIM) scale, the Defensiveness Index (DEF), and the Cashel Discriminant Function (CDF). Morey and Hopwood (2007) propose that CDF scores are specific to effortful PRD and concealed psychopathology. However, CDF has several unusual properties that complicate its interpretation, including sensitivity to both fake-good and fake-bad response set instructions. This presentation describes the derivation of a novel discriminant function from the PAI full scales and subscales designed to identify response patterns observed during fake-good role play that are associated with elevated clinical scores during standard instructions. A sample of 334 undergraduate students responded to the PAI in two testing sessions separated by a 10-21 day interval. Students followed standard instructions to respond honestly in one session, and they read a role-play scenario and followed instructions to engage in PRD to obtain a desirable job in the other sessions. Order of instruction conditions varied in roughly equal proportions. Using the PAI score profiles obtained during standard instructions, two study groups were constructed. Clean respondents ( $n = 98$ ) obtained T-scores  $< 60$  on all PAI full scales in the standard (i.e., honest) instruction condition. Concealer respondents ( $n = 136$ ) obtained T-scores  $\geq 70$  on one or more clinical full scales in the instruction condition. A discriminant function was calculated that classified students into Clean versus Concealer groups based on the PAI scores obtained during role-play (SN = .80, SP = .75; AUC = .87), and this formula performed better than PIM, DEF, or CDF. Efforts to cross-validate this function are also presented. This novel formula may prove useful for identifying deliberate efforts to conceal known psychopathology in contexts with strong incentives for PRD.

## Does Social Desirability Vary Depending on the Purpose of the Assessment? A Comparative Study

Cristina Mazza<sup>2</sup>, Francesca Bosco<sup>1</sup>, Paolo Roma<sup>1</sup>

<sup>1</sup> Department of Human Neuroscience, Sapienza University of Rome, Italy

<sup>2</sup> Department of Dynamic and Clinical Psychology, and Health Studies, Sapienza University of Rome, Italy

### Abstract

The Minnesota Multiphasic Personality Inventory (MMPI-2) is extensively utilized for personality assessments, particularly to evaluate an individual's suitability in various contexts, such as parenting, firearm handling, and complex occupations like airline piloting and driving. The MMPI-2 Lie (L) scale is specifically designed to detect socially desirable self-presentation and the tendency to deny the most common human mistakes. This study aims to explore how social desirability manifests differently across diverse groups and affects L-scale scores. We analyze groups including parents evaluated for parenting skills, individuals assessed for child adoption suitability, and both civilian and military pilots and drivers undergoing license evaluations. Although current data are pending, we anticipate significant variance in L-scale T-scores across these groups. Our findings are expected to clarify the effectiveness of the L-scale across various populations and recommend additional strategies to enhance the accuracy of social desirability assessments. This improvement should lead to more nuanced and sensitive assessments of the motivations behind the assessment, including how individuals may, intentionally or unintentionally, wish to appear more virtuous and better adjusted.

## Addressing Bias in DSM Criteria for Malingering Suspicion

Ali Yunus Emre Akca

Maastricht University, The Netherlands & University of Turin, Italy

### Abstract

The Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria for malingering, which include medicolegal context, inconsistencies, uncooperativeness, and antisocial personality disorder, have been criticized for potentially leading to biased judgments. This study aims to investigate whether these criteria are biased for certain sociodemographic characteristics. Participants rated the likelihood and confidence (Likelihood-Confidence Composite Score [LCCS]; range 0 to 1) that each DSM criterion for suspected malingering characterized a person of a particular race ("white" vs. "nonwhite"), gender (male vs. female), or age ("40 years or younger" vs. "40 years or older") using neutral scenarios matched on each criterion. The (preliminary) sample consists of 84 students (95.2% pursuing bachelor's degrees) with a mean age of 21 years ( $SD = 3.85$ ) who self-reported minimal prior knowledge of malingering ( $M = 1.96$ ,  $SD = .93$ ). Most participants (47.6%) are from Germany, 23.8% are from the Netherlands, and 77.4% report being White, European, or American-European. Preliminary results indicate no statistically significant difference in LCCS from an "ideal unbiased" value of .5 for race, gender, or age,  $t(83) \leq 1.863$ ,  $p > .05$ , Hedges'  $g \leq 0.20$ . One-sample  $t$ -tests for each criterion reveal no significant difference ( $p > .05$ ) to an LCCS of 0.5. The Kruskal-Wallis  $H$ -test also do not reveal significant differences in LCCS between the four criteria ( $p \geq .676$ ). In summary, the preliminary results suggest that relatively uninformed college students do not perceive the DSM criteria for suspected malingering as biased toward race, gender, or age.

# Memory and Validity Assessment



## Exploring the Application of the Verifiability Approach in Detecting Malingered COVID-19 Symptoms

Raffaella Maria Ribatti<sup>1</sup>, Tiziana Lanciano<sup>1</sup>, Luciano Fadiga<sup>2,3</sup>, & Antonietta Curci<sup>1</sup>

<sup>1</sup> Department of Education, Psychology, Communication Sciences, University of Bari Aldo Moro, Italy

<sup>2</sup> Department of Neuroscience and Rehabilitation, University of Ferrara, Italy

<sup>3</sup> Italian Institute of Technology, Italy

### Abstract

The present study focuses on the application of the Verifiability Approach (VA), a credibility assessment tool derived from deception detection, to malingering of COVID-19 infection. The VA relies on deceivers' difficulty in balancing appearing truthful while avoiding verifiable specifics. While the VA doesn't directly assess memory, verifiable details may stem from genuine memories of the illness and can be compared with malingerers' accounts. Individuals who had COVID-19 can likely recall specific details like doctor's names, symptom timelines, or medications taken. In contrast, malingerers struggle to fabricate such details due to the lack of a true memory base. Drawing from previous research, we compared reports by honest and feigner respondents, exploring the impact of informing participants about the VA features. Results indicate that being informed influenced participants' disclosure of verifiable details, and honest participants provided more verifiable details than malingerers. Furthermore, informed honest participants provided more verifiable details compared to uninformed honest participants and both informed and uninformed malingerers, while malingerers did not show significant differences between the informed and uninformed conditions. These findings support previous studies indicating that verifiable details signal truthfulness, particularly if an "Information Protocol" is applied. Additionally, malingerers struggle to provide detailed elaborations of their fabricated memory of symptoms, resulting in few perceived success scores. Our results underline the utility of the VA in detecting inconsistencies and assessing the credibility of symptom reports in the realm of malingering.



## Uncovering Lies During Investigative Interviews: Analysis of Response Latencies and Error Rates to Unexpected Questions

Giulia Melis<sup>1</sup>, Andrea Zangrossi<sup>1</sup>, Graziella Orrù<sup>2</sup>, & Giuseppe Sartori<sup>1</sup>

<sup>1</sup> Università di Padova, Italy

<sup>2</sup> Università di Pisa, Italy

### Abstract

In this study, we aim to detect identity deception during investigative interviews by integrating the analysis of response latency and error rate with the unexpected questions technique. Sixty participants were assigned to an honest (n = 30) or deceptive group (n = 30), with the latter instructed to memorize false biographical details of a fictitious identity. Both groups underwent a face-to-face investigative interview comprising a randomized sequence of control, expected, and unexpected open-ended questions regarding their identity. The honest group was required to answer truthfully, whereas the deceptive group was asked to respond using the pre-memorized fictitious identity. The responses were audio-recorded for detailed analysis.

Our findings indicate that deceptive participants exhibited longer latencies and higher error rates for expected (requiring deception) and unexpected questions (for which premeditated deception was not possible). Longer response latencies were also observed in deceptive participants when answering control questions (which required truthful answers). Moreover, a within-subject analysis highlighted that the unexpected questions significantly impaired individuals' performance compared to control and expected questions. Leveraging machine learning algorithms, our approach achieved a classification accuracy of 98% in distinguishing deceptive and honest participants. Additionally, a classification analysis on single responses was conducted.

Our findings underscore the effectiveness of merging response latency metrics and error rates with unexpected questioning as a robust method for detecting identity deception in investigative interviews. This approach addresses a significant gap in the scientific literature by offering a method that allows for lie detection without the individual's awareness that the technique is focused on the credibility assessment. Moreover, it can be applied retrospectively to interviews that have been conducted, allowing for the exploitation of data that are routinely acquired. Lastly, it doesn't impact standard judicial or police procedures, as no ad hoc tests are administered, thus preserving the continuity of the investigative interview.

## Is the Flashbulb Memory Checklist a Valid Tool to Distinguish True from False Autobiographical Memories?

Fabiana Battista, Tiziana Lanciano, Ivan Mangiulli, & Antonietta Curci

University of Bari Aldo Moro, Department of Education, Psychology, and Communication, Italy

### Abstract

The correct identification of memory accuracy is one of the most important issues in legal psychological studies and, in general, for the legal context. In proceedings in which is not possible to have objective evidence, witnesses' memories are fundamental to try to reconstruct the crime. Data show that oftentimes wrongful convictions are due to the acceptance of inaccurate testimonies (i.e., memories) as true. This occurs because of two problems. First, legal practitioners do not always know scientific literature and, consequently, draw from their own experience the criteria to evaluate if a testimony is reliable. Second, despite the numerous studies conducted on false memories (i.e., spontaneous and induced) and on individual differences and true vs false memories, it is still debated whether and how specific memory characteristics can help in differentiating true memories from false memories. The literature on flashbulb memories provides a significant contribution to the debate, as it focuses on the qualitative and phenomenological evaluation of emotional and/or traumatic autobiographical memories, such as testimonies. The Flashbulb Memory Checklist is a validated tool for measuring the flashbulb characteristics of autobiographical emotional events, such as peripheral, idiosyncratic, and sensorial details. All these details are typical of vivid and detailed real memories of emotionally impactful experiences significant for oneself and others, contrary to false memories. The aim of the present contribution is to present a study investigating the possibility of applying the Flashbulb Memory Checklist for discriminating true from false accounts, with the hypothesis that true memories exhibit more peripheral, idiosyncratic, additional, and irrelevant details as compared to false memories. Practical implications for the legal field will be discussed.

# Poster Abstracts

## Replicating the classification accuracy of the IOP-29-M across administration languages in Romanian-English Bilinguals

Crișan, I.<sup>1</sup>, Bunghez, C.<sup>1</sup>, Avram, L.<sup>1</sup>, & Pirciu, A.<sup>2</sup>

<sup>1</sup> West University of Timișoara, Department of Psychology, Romania

<sup>2</sup> West University of Timișoara, Department of American Studies, Romania

### Abstract

**Objective.** This paper replicates previous findings by Crișan (2023) on the classification accuracies of the IOP-29-M in Romanian-English bilinguals.

**Method.** Onehundred-six undergraduates (76 females;  $M_{\text{Age}} = 22.87$ ,  $SD_{\text{Age}} = 5.64$ ;  $M_{\text{Education}} = 14.35$ ,  $SD_{\text{Education}} = 1.44$ ), randomized into controls ( $n = 76$ ) and experimental malingerers ( $n = 30$ ) were administered the IOP-29-M, in both Romanian and English, as part of a comprehensive neuropsychological battery. IOP-29 FDS and IOP-M accuracy scores were compared across languages within criterion groups, as well as between criterion groups. Classification accuracies for the IOP-29-M were computed against experimental malingering as the criterion, at previously published cutoffs.

**Results.** Results largely replicated previous findings, in that the IOP-29-M produced similar scores across languages within the experimental group compared to the control group, and significant differences between criterion groups, regardless of the administration language. Classification accuracies for both indicators were higher than previously reported, with virtually similar values for both administration languages (IOP-29 FDS<sub>EN</sub>  $\geq 50$ :  $Sn = .73$ ,  $Sp = 1.00$ ; IOP-29 FDS<sub>RO</sub>  $\geq 50$ :  $Sn = .80$ ,  $Sp = .99$ ; IOP-M<sub>EN</sub>  $\leq 31$ :  $Sn = .90$ ,  $Sp = 1.00$ ; IOP-M<sub>RO</sub>  $\leq 31$ :  $Sn = .93$ ,  $Sp = .93$ ).

**Conclusions.** In this bilingual population, the IOP-29-M proves cross-culturally valid and relatively robust to language effects.

## Unveiling crime-related amnesia using a visual version of Forced Choice Test

Stefano Zago<sup>1</sup>, Elisabetta Soldini<sup>1</sup>, Teresa Difonzo<sup>1</sup>, Silvia Cattaneo<sup>2</sup>, & Nadia Bolognini<sup>3,4</sup>

<sup>1</sup> Department of Neurosciences and Mental Health, Fondazione IRCCS Ca' Granda, Ospedale Maggiore Policlinico, Milan, Italy

<sup>2</sup> Department of Medicine and Surgery, Laboratory of Clinical Psychology, Clinical Psychophysiology and Clinical Neuropsychology, University of Parma, Parma, Italy

<sup>3</sup> Department of Psychology, University of Milano-Bicocca, Milan, Italy

<sup>4</sup> Laboratory of Neuropsychology, Department of Neurorehabilitation Sciences, IRCCS Istituto Auxologico Italiano, Milan, Italy

### Abstract

The Forced Choice Test (FCT) represents a memory detection test that can be profitably used in criminal setting. In the FCT, the offender, who claims to be amnesic, is asked to answer a series of dichotomous true-false and equally plausible questions (one correct, the other incorrect) about details related to the crime and the circumstances under which it took place. This application can only be used in cases where police reports contain sufficient information about the homicide to compose the ad-hoc FCT, or when the crime details have been leaked to the offender by police detectives or lawyers. During the FCT, questions can be presented orally or in written form on a computer screen; in addition, if crime scene images obtained during police investigation are available, they can be included as stimuli in the FCT. Here we report the case of J.B., a 38-year-old man, who killed his wife in 2018 by multiple stabbing and claimed not to remember the homicide. We selected forty-two images from the police file, pairing each with a similar and plausible alternative. Firstly, we asked a group to judge the matching of visual representation of paired items. Then, J.B. was presented with 50% images of his personal property at the crime scene (e.g., clock, house) and 50% of similar but unrelated objects at the crime scene. At the FCT, J.B. scored 20 out of 42, below 50%, although higher than the value calculated using binomial statistics (<11.7%). Results from the neuropsychological assessment, traditional malingering and FCT suggested malingered crime-specific amnesia. The court also found his claim of amnesia not credible. We argued that visual presentation of crime-related information appears preferable, as it requires lower working memory load, but enough photos must be acquired to create a reasonable number of questions.

## Once is enough! An analogue study on repeated validity assessment in adults with ADHD

Hui Dong<sup>1</sup>, Yvonne Groen<sup>1</sup>, Gerdina H.M. Pijnenborg<sup>1</sup>, Oliver Tucha<sup>2,3</sup>, Stefan Aschenbrenner<sup>4,5</sup>, Matthias Weisbrod<sup>4,6</sup>, Janneke Koerts<sup>1</sup>, & Anselm B.M. Fuermaier<sup>1</sup>

<sup>1</sup> Department of Clinical and Developmental Neuropsychology, University of Groningen, The Netherlands

<sup>2</sup> Department of Psychiatry and Psychotherapy, University Medical Center Rostock, 18147 Rostock, Germany

<sup>3</sup> Department of Psychology, Maynooth University, National University of Ireland, W23 F2K8 Maynooth, Ireland

<sup>4</sup> Department of Psychiatry and Psychotherapy, SRH Clinic Karlsbad-Langensteinbach, 76307 Karlsbad-Langensteinbach, Germany

<sup>5</sup> Department of Clinical Psychology and Neuropsychology, SRH Clinic Karlsbad-Langensteinbach, 76307 Karlsbad-Langensteinbach, Germany

<sup>6</sup> Department of General Psychiatry, Center of Psychosocial Medicine, University of Heidelberg, 69115 Heidelberg, Germany

### Abstract

Performance validity tests (PVTs) can be seen as gatekeepers for valid neuropsychological assessment, by marking cognitive test scores that may not reflect true ability levels. The present study explored the significance of repeated validity testing in independent assessment sessions of adults with attention-deficit/hyperactivity disorder (ADHD), by exploring the potential value of performance (in)consistencies across assessments to distinguish valid from invalid cognitive performance. Neuropsychological test data of 24 individuals diagnosed with ADHD were complemented by an analogue study involving 69 typically developing individuals who were allocated to either a control group or a simulation group instructed to simulate feigned ADHD. All individuals were assessed with embedded and stand-alone PVTs three times with one-month intervals between each assessment. The rate of positive findings per validity test remained rather stable across assessments for the entire group of participants. Significant differences in neuropsychological performance scores occurred between individuals with ADHD and experimental simulators, however, mostly nonsignificant effects of small size emerged when considering performance (in)consistency as a source for detecting invalid cognitive performance. Our data demonstrate that the (in)consistency of cognitive performance over repeated assessments may be no effective approach to complement validity assessment. Replication is needed in independent research on larger samples.

## The Utility of the Groningen Effort Test (GET) in Assessing Performance Validity in Early Retirement Claimants

Janina Teßmann<sup>1</sup>, Anselm B. M. Fuermaier<sup>1</sup>, Brechje Dandachi-FitzGerald<sup>2,3</sup> & Johann Lehrner<sup>4</sup>

<sup>1</sup> Department of Clinical and Developmental Neuropsychology, Faculty of Behavioral and Social Sciences, University of Groningen, Groningen, The Netherlands

<sup>2</sup> Department of Clinical Psychological Science, University of Maastricht, Maastricht, The Netherlands

<sup>3</sup> Faculty of Psychology, Open University, Heerlen, The Netherlands

<sup>4</sup> Department of Neurology, Medical University of Vienna, Vienna, Austria

### Abstract

**Introduction.** Feigning attention deficits is a frequently reported simulation strategy in neuropsychological testing which might be overlooked when only memory-based performance validity tests (PVTs) are administered. Accordingly, the Groningen Effort Test (GET) has been developed to detect implausible attention performance.

**Objectives.** This study aimed to examine the utility of the Groningen Effort Test in assessing performance validity in a civil forensic setting with external incentives. Additionally, agreement rates between the GET, the Word Memory Test (WMT), and the Reliable Digit Span (RDS) were explored.

**Methods.** Archival data from neuropsychological assessments of 132 individuals who claimed early retirement and presented for evaluation at the Department of Neurology of the Medical University of Vienna, Austria, were used. To establish an external criterion for performance validity, the sample was divided into individuals who passed the WMT and the RDS (valid performance,  $n = 55$ ) and individuals who failed both of these PVTs (invalid performance,  $n = 31$ ). Lastly, correlations between all PVTs and two true cognitive ability tests (i.e., the Cognitrone and the Test of Attentional Performance) were computed.

**Results.** The classification agreement between the GET and other PVTs was fair to moderate. Performance on the GET was substantially better for individuals exhibiting valid performance than for individuals exhibiting invalid performance, with effect sizes of moderate and large magnitudes. Similarly, the GET showed satisfactory classification accuracy, slightly better at identifying valid performance than invalid performance. Correlations between the GET and the cognitive ability tests were in the medium-to-large range.

**Conclusions.** Both GET measures demonstrated their usefulness in determining the plausibility of test performance in a civil forensic setting with early retirement claimants. Nevertheless, relatively high correlations between the PVTs and cognitive ability tests suggest some overlap in the measured concepts for the current sample.

## Advancing Adult ADHD Assessment? The Utility of Attention-Based Performance Validity Testing

Johanna Kneidinger<sup>1</sup>, Anselm B. M. Fuermaier<sup>2</sup>, Emily Raasch<sup>1,2</sup>, Björn Albrecht<sup>1</sup>, & Hanna Christiansen<sup>1</sup>

<sup>1</sup> Department of Psychology, Philipps-University Marburg, Germany

<sup>2</sup> Department of Clinical and Developmental Neuropsychology, University of Groningen, The Netherlands

### Abstract

A substantial number of adults underperform in cognitive testing during their clinical evaluation of attention-deficit/hyperactivity disorder (ADHD) symptoms. Hence, the implementation of performance validity tests (PVTs) is becoming increasingly important in preventing diagnostic errors. The Groningen Effort Test (GET), a PVT specifically designed to assess performance validity in adults undergoing evaluation for ADHD has not been evaluated in large clinical samples, yet. This study examines the utility and classification correspondence of the GET with two other established PVTs (MSVT and AKGT) in a sample of N = 268 adults referred for adult ADHD assessment in our outpatient clinic. All PVTs show slight to fair classification agreement ( $\kappa = .10 - .36$ ) and positive results decrease as more PVTs are considered. More individuals obtained positive results on the GET compared to other PVTs (3.0% - 16.0%), especially when considering the GET-index (26.9%) rather than the GET error score 11.6%). A positive association between PVTs and cognitive ability measures was observed as long as individuals with invalid test performance were included into the analyses; however, this association largely disappeared when excluding those individuals from the analysis. We conclude that the consideration of proven PVTs supports diagnostic accuracy, and appropriate treatment recommendation. Future research should investigate whether the differences in positive results across PVTs arise from a higher sensitivity of the GET in this population, or rather from an increased number of false positives in the GET. Given that cogniphobia and cognitive disengagement both affect invalid performance, future research should take these factors into account when interpreting performance validity assessment results.



## Performance Validity Assessment in a clinical children and adolescent sample

Emily Raasch<sup>1,2</sup>, Anselm B. M. Fuermaier<sup>2</sup>, Johanna Kneidinger<sup>1</sup>, Björn Albrecht<sup>1</sup>, & Hanna Christiansen<sup>1</sup>

<sup>1</sup> Department of Psychology, Philipps-University Marburg, Marburg, Germany

<sup>2</sup> Department of Clinical and Developmental Neuropsychology, University of Groningen, Groningen, The Netherlands

### Abstract

Neuropsychological performance tests are commonly administered in the evaluation of Attention-Deficit/Hyperactivity (ADHD) and related disorders in children and adolescents. However, clinical research revealed that a substantial number of young patients seen for consultation, display evidence of noncredible symptom reporting and/or test performance. The reasons for such distorted symptom reporting and test underperformance are manifold. A proportion of patients may demonstrate certain levels/ manifestations of poor performance to benefit being diagnosed with attention disorders. Examples include academic benefits, access to medication and excuse-making behaviour. To date, scholars have developed and evaluated performance validity tests (PVTs) in adult populations, distinguishing credible from noncredible performance. Literature regarding PVTs in children and adolescent remains scarce, highlighting the need to develop and validate these in detecting noncredible performance. The Groninger Effort Test (GET) is an attention-based PVT, detecting feigned adult ADHD. This study assesses the utility of the GET, alongside two further PVTs, Theory of Memory Malingering and reliable digit span (derived from digit Span Subtest), in detecting noncredible performance in three criterion groups of children and adolescents, aged 6-18: (1) clinical outpatient participants, visiting the child & adolescent outpatient clinic in Marburg, Germany (n= 20) (2) healthy school-aged participants either allocated to a full effort group (n= 57) or a poor effort group (n= 34). Results will be discussed regarding the utility of attention-based performance validity testing in clinical child and adolescent psychology. Classification statistics will be computed, agreement rates between PVTs determined alongside the consideration of age-adjusted cut scores of the GET.

## Using the Forgetfulness Assessment Inventory (FAI) as an embedded measure for memory over-reporting – First results

Bayrhammer-Savel, M., Lehrner, J.

Department of Neurology, Medical University of Vienna, Austria

### Abstract

**Background.** Courts rely on neuropsychological experts to evaluate early retirement claims based on memory loss. Given the potential incentive for applicants to exaggerate their memory symptoms, symptom validity assessment plays a crucial role in this regard.

**Objectives.** This study aims to identify a cut-off value for the FAI to differentiate between symptom overreporting and genuine memory reporting in, firstly, patients with mental disorders undergoing forensic neuropsychological evaluation in private office (PO) and, secondly, individuals undergoing medical examination in an university-based memory clinic (UBMC) subdivided into healthy controls (HC) and patients with subjective cognitive decline (SCD) and mild cognitive impairment (MCI).

**Methods.** Analyses are based on  $n=351$  protocols in PO and  $n=503$  protocols in UBMC. In addition to FAI, participants were subjected to SIMS, SRSI and BDI-II with symptom exaggeration being assumed, if two or more of these measures' respective cutoffs for overreporting were exceeded. Subsequently, ROC analysis and binary logistic modelling were performed.

**Results.** Within PO patients, overreporting testees (35.7%) show a significantly higher FAI-score exhibiting a large effect ( $d=0.94$ ). ROC suggests that a cutoff of 3.91 achieves a specificity of 0.91 and a sensitivity of 0.38. Binary logistic regression (Nagelkerke's  $R^2$  40.8%) shows that FAI (OR 2.82), age (OR 0.97), IQ (OR 0.95) and education (OR 0.89) are significant factors, while sex had no significant effect. Within UBMC, overreporting testees (2.3%) show a significantly higher FAI-score exhibiting a large effect ( $d=1.23$ ). The cutoff of 3.91 demonstrates a specificity of 0.93 and a sensitivity of 0.36. Binary logistic regression (Nagelkerke's  $R^2$  37.7%) indicated a significant effect by FAI (OR 3.49) and IQ (OR 0.90) with sex, age and education showing no significant effect.

**Conclusion.** The FAI was capable of distinguishing between testees who overreport their symptoms and those who state authentic symptoms. Thus, the FAI might be used as an embedded measure for memory overreporting. However, further research is needed to support these results.

## Does Coaching Weaken the Accuracy of Symptom (SVT) and Performance (PVT) Validity Tests?

Sara Pasqualini, Andrea Corgiat Loia, & Luciano Giromini

Department of Psychology, University of Turin, Turin, Italy

### Abstract

A major issue in the efficacy of validity tests is represented by coaching. Coaching has been defined as any attempt to alter the results of psychological or neuropsychological tests in such a way that distorts the true representation of the examinee's cognitive, emotional, or behavioral state. Especially in forensic assessment, it is possible that attorneys would instruct their clients on how to feign their symptoms in order to present a worse or better mental state. Two studies were conducted to investigate: a) whether individuals informed about the possible presence of PVTs and/or SVTs would be able to identify them among a series of administered tests (Study 1); and b) whether identifying an SVT/PVT as a "feigning measure" would decrease its effectiveness (Study 2). The sample consists of 210 volunteer subjects divided as follows: 58 non-clinical adults instructed to respond honestly (i.e., non-clinical control group), and 152 feigners divided into three groups: 56 non-clinical adults instructed to feign schizophrenia (i.e., experimental feigners group) who are only explained what the symptoms of schizophrenia are, 42 non-clinical adults instructed to feign schizophrenia (i.e., experimental coached feigners group) who, in addition to the symptoms, are told that there will be validity tests, and 54 non-clinical adults instructed to feign schizophrenia who, in addition to being given the symptoms and being told there will be validity tests, are instructed on how the validity tests work (i.e., experimental coached feigners group). Contrary to our a-priori hypotheses, the more detailed instructions, designed to facilitate a credible feigning, did not produce significant differences in the performance of the study's target SVTs and PVTs.

## Dutch Insurance Physicians' Practices and Beliefs on Using Validity Tests in Their Insurance Evaluations: A Survey Study

Sophie Giles<sup>1,2</sup>, Brechje Dandachi-FitzGerald<sup>2,3</sup>, Nina Wijnands<sup>1</sup>, & Harald Merckelbach<sup>2</sup>

<sup>1</sup> Dutch National Social Security Agency (Uitvoeringsinstituut Werknemersverzekeringen), The Netherlands

<sup>2</sup> Section Forensic Psychology, Faculty of Psychology and Neuroscience, Maastricht University, The Netherlands

<sup>3</sup> Department of Clinical Psychology, Faculty of Psychology, Open University, The Netherlands

### Abstract

**Introduction.** In the Netherlands performance and symptom validity tests (PVTs/SVTs) are currently implemented through independent medical examinations (IMEs). However, the Dutch national social security agency is considering the use of PVTs/ SVTs by the insurance physician within their own assessments. This study sought to gain insight into insurance physicians' current practices and views on PVTs/SVTs and their opinions regarding using these tests themselves.

**Method.** Participants were recruited via email to anonymously participate. The survey included 4 sections; 1) occupational background, 2) current practice, 3) education and knowledge on validity assessment, and 4) views regarding use of SVT/PVT. The following results are based on preliminary findings from 212 respondents. Data collection is still ongoing.

**Results.** Responses were given from 125 specialized insurance physicians (59%,  $n = 125/212$ ), 68 in specialization training (32%,  $n = 68/212$ ), and 19 medical doctors (9%,  $n = 19/212$ ). Currently, physicians rely on observation of inconsistencies to recognize symptom exaggeration and minimization. Furthermore, 34.8% ( $n = 67/192$ ) did not request an IME in the year prior. The main reasons were: 1) extra time required (27%,  $n = 52/192$ ); 2) the cost (21%,  $n = 41/192$ ); and 3) observing inconsistencies is considered enough to detect symptom distortion (30%,  $n = 58/192$ ). Regarding education; 8% ( $n = 16/192$ ) admitted to not knowing what PVTs/SVTs were and 66% ( $n = 129/196$ ) stated that not enough education is provided. Finally, when asked if they thought if it would be beneficial to be able to perform PVTs/SVTs themselves, 72% answered yes ( $n = 138/192$ ). However, more education within the specialization training (67%,  $n = 93/138$ ) and practical training regarding implementation (97%,  $n = 109/138$ ) is required.

**Conclusion.** Most Dutch insurance physicians appear interested in integrating SVTs/PVTs into their own assessments, provided they receive sufficient training to competently use these tools.

## Equivalency of In-Person Versus Remote Assessment of the PAI, IOP-29, and IOP-M

Claudia Pignolo & Luciano Giromini

Department of Psychology, University of Turin, Turin, Italy

### Abstract

Studies on the equivalence of face-to-face and tele-testing should be warranted for all published tests; however, very few studies have addressed the potential differences between remote and face-to-face administration methods, and this gap in the literature becomes even more apparent when considering the validity of computerized and/or remotely administered tests that assess the credibility of reported symptoms and response styles. Therefore, the aim of this study was to examine the comparability and validity of remote and face-to-face administration of the Personality Assessment Inventory (PAI), the Inventory of Problems-29 (IOP-29), and the Inventory of Problems-Memory (IOP-M) by using a simulation study design. Nine hundred Italian, adult volunteers were recruited for this study. We administered the PAI, IOP-29, and IOP-M under three different conditions: (a) online/computerized, (b) face-to-face/computerized, and (c) face-to-face/ paper-and-pencil. In addition, we adopted a simulation study design, so that, for each condition, participants were randomly divided into two groups: the honest responders group, in which participants were asked to take the tests honestly, and the experimental feigners group, in which participants were instructed to feign psychopathology (i.e., depression, psychosis, PTSD, mild traumatic brain injury, and ADHD) while taking the tests. Our results suggest that indicators of negative response bias perform similarly across different modes of administration when individuals are asked to take the PAI, IOP-29, and IOP-M honestly. Our study responds to the call for scientific evidence of equivalence between remote and face-to-face testing and suggests that clinicians should carefully consider the administration in forensic cases.

## Feigning judgments are hampered by a failure to understand the negative predictive power of symptom validity tests

Isabella Niesten<sup>1</sup>, Anselm Fuermaier<sup>2</sup>, Nathalie Faber<sup>2</sup>, & Marie Papenfuss<sup>2</sup>

<sup>1</sup> Open University, The Netherlands

<sup>2</sup> Groningen University, The Netherlands

### Abstract

**Background.** Clinicians' decisions are affected by salient peripheral information about the patient. If such information has limited diagnostic value but is given considerable weight, this may bias decision-making and cause diagnostic error and misclassification. An influential source that provides salient, yet misleading, information on how to differentiate patients who feign symptoms from patients with genuine symptoms is the DSM, which states feigning is closely tied to antisocial personality features. The present study examined whether antisocial features in a case description raise initial suspicion of feigning and whether they serve as an anchor so that subsequent corrective information is not integrated into the diagnostic conclusion.

**Method.** Professionals (N=106) were randomly allocated to an antisocial, hysterical (e.g., strong emotional displays), or a neutral case description. In all case descriptions, potential motivation for feigning was included (i.e., financial compensation after an accident). In rounds, professionals received new information, including corrective information in the form of non-deviant SVT-scores. After the case description and each round, they rated the likelihood of feigning. Finally, they indicated whether they would mention (possible) feigning in a diagnostic report.

**Results and Discussion.** Baseline likelihood estimates of feigning were raised across conditions and were higher for the antisocial case description than the neutral but not the hysterical case description. Furthermore, regardless of case description, professionals did not adjust initial estimates to a clinically meaningful degree in response to non-deviant SVT-scores. Finally, the proportion of professionals endorsing (possible) feigning in their final report was comparable across conditions: 58% of the sample endorsed this option. Taken together, although professionals' initial suspicion of feigning may be slightly raised in cases where the patient fits an antisocial typology, this effect seems subtle. It seems professionals' failure to integrate non-deviant SVT-scores into their judgment may be more so due to insufficient knowledge about the clinical value of these scores, an issue that may be countered with education and training.

## Say What? The Interpreter Effect: Impact of Interpreters on Symptom and Performance Validity Test Scores

Megan Alsop

Jefferson Neurobehavioral Group, Louisiana, USA

### Abstract

**Background.** Close to 300 million international migrants globally need interpreters to access healthcare (WHO). Between 10% and 20% of healthcare providers regularly use such services in the US and European Union. The increasing demand highlights the growing need for research on the impact of interpreters on the outcome of assessments. The National Center for State Courts (NCSC) in the U.S. reports an over 200% increase in the use of an interpreter in the last two decades. Best practice guidelines for using an interpreter (Fletcher, Jansen, Strickland, Reynolds, 2000) have been developed for psychological services. Yet, there is a paucity of empirical research on the effect of reliance on interpreters' on test scores. This is particularly relevant for evaluating symptom and performance validity, given the highly contentious nature of these results – especially in culturally diverse populations.

**Purpose.** The purpose of this study was to determine whether performance or symptom validity test scores differed when an interpreter was involved in the assessment.

**Methods.** 211 adults were evaluated as part of a psychological injury claim [55 (27%) native English speakers, 150 (73%) non-native English speakers assessed through an interpreter. A standard battery of psychological tests was administered to all participants.

**Results.** T-tests were run looking at group means. No significant differences were seen in the overall score between the groups. Correlation matrixes between PVT and SVT tests showed generally favorable correlations between task scores.

**Conclusion.** Implications for using PVTs and SVTs with non-English speaking examinees are discussed.

## Uncovering Drug Use Using Mouse Kinematics: Validity Assessment Through Symptom Sentence Analysis

Giulia Melis, Giuseppe Sartori, & Merylin Monaro

University of Padua, Department of General Psychology, Italy

### Abstract

The consumption of psychoactive substances in the workplace poses a significant challenge for many countries, with a considerable proportion of individuals who engage in the use of illegal drugs being employed across various sectors (European Monitoring Centre for Drugs and Drug Addiction, 2022; Frone, 2013). Recent analyses by the European Monitoring Centre for Drugs and Drug Addiction highlight that, beyond impacting the individual health of employees, the consumption of alcohol and drugs in the workplace leads to increased workplace accidents and injuries, higher rates of absenteeism, and inappropriate behaviors. These issues culminate in a considerable economic burden for employers, governments, and society. The most common workplace drug testing methods include blood testing, breath analysis, and urine analysis (Ones & Viswesvaran, 2001). However, these measures are subject to significant legislative variations across Europe, raising complex moral, ethical, and legal concerns, particularly regarding privacy and discrimination. Blood tests, for example, are often seen as highly invasive and, for this reason, less commonly used (Shahandeh & Caborn, 2003). On the contrary, in instances where drug testing is mandated, employees often attempt to circumvent testing by abstaining from drug use shortly before tests, highlighting the difficulty of effectively addressing workplace drug use (El-Bassel et al., 1997). To address these concerns, the employment sector is increasingly using paper-and-pencil tests (e.g., Drug Abuse Screening Test, DAST; Skinner, 1982) to assess current and prospective employees for drug abuse. These tests aim to predict or confirm substance abuse by evaluating candidates' attitudes and perceptions towards drug use. However, a major drawback is their transparency, which makes them susceptible to deceitful responses (Ramachandran et al., 2019). This study investigates the potential of using mouse kinematics as an innovative lie detection technique to spot drug use. We hypothesize that the cognitive load associated with lying will manifest in distinct mouse movement patterns, differentiating truth-tellers from liars. Two experiments were conducted involving university students (n=81), divided into two groups: those who reported drug use in the past year (liars) and those who did not (truth-tellers). Participants completed a computer-based task, which recorded their mouse trajectories in response to a series of true/false statements about drug use. A significant component of our methodology was the use of symptom sentences designed to increase cognitive load. These sentences were constructed based on general life experiences that were likely to have occurred at



least once in the past year, very rare events (e.g., “I stayed awake for more than 48 hours”), or conditions unlikely to have consistently occurred in the past year (e.g., “I have always been calm”). These statements were intended to confuse users into associating them with potential symptoms of drug use, thereby increasing the cognitive load on those lying about their drug use. Our analysis focused on various kinematic variables, revealing significant differences between the two groups. Furthermore, classification algorithms applied to the kinematic data achieved accuracy rates between 70% and 80% in distinguishing liars from truth-tellers. The use of symptom sentences proved effective in increasing the cognitive load on participants, making it more challenging for liars to maintain consistent responses. These findings suggest that mouse kinematics, combined with carefully crafted symptom sentences, can serve as a supplementary tool for lie detection in forensic investigations. The technique's non-invasiveness and potential to provide real-time analysis make it a promising avenue for further research and application.

## The Tangram Task: first results of a newly developed symptom validation test

Riedl, J. & Exner, C.

University Leipzig, Clinical Psychology and Psychotherapy, Neumarkt 9-19, 04109 Leipzig, Germany

### Abstract

**Theoretical background.** Symptom validity tests aim to be as opaque as possible in order to be less susceptible to coaching. Many test procedures are language-based and therefore not suitable for all patients. A newly developed task based on tangram stimuli is intended to make it easier to recognize negative response bias. The developed task is based on the principle of concealed ease. This validation study provides initial information on the suitability of the test paradigm.

**Method.** 32 people with acquired brain damage as well as 32 experimental simulants completed the tangram task and further neuropsychological tests. In the tangram task, four targets from nine stimuli were to be recognized in four rounds of increasing difficulty. Two targets each were meaningful. Simulation were assessed.

**Results.** A mixed ANOVA revealed that recognition performance differed significantly as a function of the group ( $F(1, 62) = 46.42, p < .001, \eta_p^2 = 0.43$ ) and the level of difficulty ( $F(3, 60) = 10.60, p < .001, \eta_p^2 = .35$ ). Meaningful targets were recognized more frequently from patients ( $F(1, 62) = 12.57, p < .001, \eta_p^2 = .17$ ) and in easier trials ( $F(3, 60) = 7.51, p < .001, \eta_p^2 = .27$ ). With both scores, 83% of subjects could be classified correctly ( $\chi^2(2, N = 64) = 43.22, p < .001; R^2_{Nagelkerke} = .66$ ). As expected, memory performance did not significantly influence test performance. Simulants showed more distinct behavior in another, already established performance validity test. Different, partially elaborated simulation strategies could be identified.

**Conclusion.** The stimuli and the task principle are generally suitable for use in the validation of memory complaints. Further research is needed.

# Keynote

## **Malingering in Psychological Injury and Law: New Developments**

**Gerald Young**

Glendon College York University, North York, Canada

### **Abstract**

The area of psychological injury and law continues both to solidify what is known as accepted standards in assessment in the field, and to redefine itself by investigating conceptually and empirically key areas of practice and research. This presentation focuses on the area of malingering and invalid response sets. The latter refers to negative response bias as evaluated by failing a battery of PVTs or SVTs. However, the threshold used depends on cut-offs for both the individual tests and the full battery administered. The attribution of malingering depends on the analysis of the full file and test profile once an invalid response set has been ascertained. Moreover, test and battery interpretation needs to consider multiple contextual and individual factors, in a combined actuarial and clinical judgment determination. There are many open questions, such as cut-offs for SVT batteries, how to combine PVT and SVT data, and even what constitutes an SVT (e.g., does the MMPI-3 include five F scale SVTs, or do they constitute one SVT?).

# Roundtable

## Validity assessment in Europe: Where are we heading?

Thomas Merten<sup>1</sup>, Esteban Puente López<sup>2</sup>, Brechje Dandachi-FitzGerald<sup>3,4</sup>, Harald Merckelbach<sup>3</sup>, & Antonietta Curci<sup>5</sup>

<sup>1</sup> Department of Neurology, Vivantes Klinikum im Friedrichshain, Berlin, Germany

<sup>2</sup> Universidad de Valladolid, Spain

<sup>3</sup> University of Maastricht, Maastricht, The Netherlands

<sup>4</sup> Faculty of Psychology, Open University, Heerlen, The Netherlands

<sup>5</sup> Department of Education, Psychology, Communication Sciences, University of Bari Aldo Moro, Italy

### Abstract

When the first SVA conference took place in 2009, a developmental delay of roughly ten years was a realistic estimate of the state of the art when validity assessment in Europe was compared to that established in North America at that point in time. Matters have considerably changed since then. What has remained is a remarkable heterogeneity among countries and a lack of properly validated instruments for the majority of European languages and for cross-cultural evaluations.

The symposium will try and look ahead to where the journey is going and what the most pressing problems to solve are. With some structured input from some participants, the discussion will be open to all conference participants.

# Workshops

## **The Morel Emotional Numbing Test: A Groundbreaking Performance Validity Tool for Ensuring the Accuracy of PTSD Assessments**

**Kenneth Morel**

National Institutes of Mental Health, Neuropsychiatry Branch (Ret.), USA & U.S. Army Medical Command (Ret.)

### **Abstract**

This workshop explores how neuroscience and statistical analysis can be combined to accurately assess response validity, with a focus on minimizing error variance. It also covers topics such as scoring, applications in clinical and forensic settings, and effective strategies for handling challenges from lawyers during depositions.



## **Assessment of Negative Response Distortion with the Personality Assessment Inventory (PAI)**

**John E. Kurtz**

Department of Psychological and Brain Sciences, Villanova University, Villanova, PA, USA

### **Abstract**

This workshop describes the various scales and index scores used to assess negative response distortion with the PAI. A configural analysis of multiple indicators is proposed to discriminate between exaggerated responding and deliberate falsification of symptoms. The added value of newer indicators from the PAI-plus update will also be addressed. Research findings and case studies will be presented to demonstrate the utility of these measures for clinical and forensic settings.

## **Multivariate Models of Performance and Symptom Validity Tests: Conceptual and Practical Considerations**

**Laszlo Erdodi**

Department of Psychology, Neuropsychology Track, University of Windsor, Windsor, ON, Canada  
Star UBB Institute, Babeş-Bolyai University, Cluj-Napoca, Romania

### **Abstract**

This workshop explores interpreting cumulative PVT failures in neuropsychological profiles. Participants will learn about selecting optimal cutoffs for individual PVTs and using multivariate models to interpret failures from multiple PVTs. It also discusses controversies related to using multiple PVTs to assess cognitive deficits and the appropriateness of generalizing multivariate models developed for PVTs to SVTs.

## **Assessing Symptom and Performance Validity with the Inventory of Problems – 29 (IOP-29) and its Memory Module (IOP-M)**

**Luciano Giromini**

Department of Psychology, University of Turin, Turin, Italy

### **Abstract**

This workshop presents the Inventory of Problems - 29 (IOP-29) and its memory module (IOP-M), a highly cost-effective solution for rapidly assessing both symptom and performance validity.

Participants will learn about key research findings emerging from the more than 30 published IOP articles and gain insights into administration and interpretation guidelines.

## **The Self-Report Symptom Inventory (SRSI): Five Years After its Publication**

**Thomas Merten**

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

### **Abstract**

This workshop gives an introduction into the development of the SRSI, a symptom validity test, which is currently available in eleven languages and was used in several dozens of studies. A main topic will be the proper administration of the questionnaire and the correct interpretation of its results.