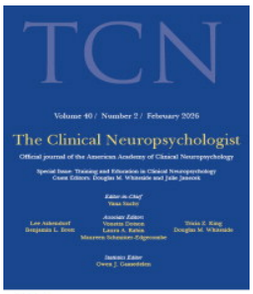

The Clinical Neuropsychologist

ISSN: 1385-4046 (Print) 1744-4144 (Online) Journal homepage: www.tandfonline.com/journals/ntcn20

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

Allyson G. Harrison, Nathaniel Davin & Emma Jamieson

To cite this article: Allyson G. Harrison, Nathaniel Davin & Emma Jamieson (21 Feb 2026): Changes in validity test scores after COVID: Reflection of general distress or something else?, The Clinical Neuropsychologist, DOI: 10.1080/13854046.2026.2630691

To link to this article: https://doi.org/10.1080/13854046.2026.2630691

THE CLINICAL NEUROPSYCHOLOGIST
https://doi.org/10.1080/13854046.2026.2630691

Routledge
Taylor & Francis Group

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

Allyson G. Harrison, Nathaniel Davin and Emma Jamieson

Regional Assessment and Resource Centre, Queen's University, Kingston, 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). **Method:** Archival data were analyzed from 1076 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 1, 2020), during COVID (March 2, 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.

ARTICLE HISTORY
Received 24 October 2025
Accepted 6 February 2026
Published online 22 February 2026

KEYWORDS
COVID-19 pandemic; postsecondary students; test validity; general psychological distress; ADHD assessment

Introduction

The adage "garbage in, garbage out," originally coined in computer science, reflects the principle that even sound analytic logic produces misleading results when input data are flawed (Rouse, 2017). The same principle applies to psychological

CONTACT Allyson G Harrison harrisna@queensu.ca Regional Assessment and Resource Centre, Queen's University, Kingston, Canada.

assessment: diagnostic conclusions are unreliable when symptom reports are inaccurate or performance data are non-credible.

Diagnostic evaluations for many types of non-visible psychological conditions rely heavily on the self-report information provided by patients and scores obtained from diagnostic performance tests. Unfortunately, there are many factors that can undermine the validity of information obtained in the context of a disability-related evaluation (Merckelbach et al., 2019; Schroeder & Martin, 2022a, 2022b). Although malingering represents the most obvious threat to validity (Bender, 2015; Schroeder & Martin, 2022b), intent is seldom assessed directly during psychoeducational assessments and so malingering cannot be inferred from PVT failure alone. While elevated rates of performance validity test (PVT) failure in adult psychoeducational assessments have been well documented, empirical evidence has failed to demonstrate that such failure is clearly due to deliberate or premeditated feigning (Harrison et al., 2021; Martin et al., 2015; Rosenfeld et al., 2000). There are, however, many other factors that can affect the validity of data obtained during a psychological assessment (Merckelbach et al., 2019; Schroeder & Martin, 2022a), including: high anxiety (Merckelbach et al., 2019), depression (Merckelbach et al., 2019), neuroticism (e.g. Denovan et al., 2019), general psychological distress (e.g. Shwartz et al., 2020), somatoform disorders (e.g. Aronson et al., 2006; Lockhart & Satya-Murti, 2015), factitious disorders (e.g. Chafetz et al., 2020), illness identity/diagnosis threat (e.g. Merckelbach et al., 2019; Suhr & Gunstad, 2002; Suhr & Wei, 2017), sociogenic illness (e.g. Fremer et al., 2024; Frey et al., 2022), cluster B personality disorders (e.g. Hong et al., 2019; Renner et al., 2008), and psychotic or delusional disorders (e.g. van der Heide et al., 2020a). Martin and Schroeder (2020) found that the base rate of assessment invalidity ranges from 5% to 50%, depending on the setting and whether external incentives are evident, with a median estimated base rate of invalidity across clinical non-forensic settings of 15%.

Given the misinterpretation that can occur if clinicians rely on non-credible data when formulating a diagnosis and making treatment recommendations (e.g. Roor et al., 2016; Schwarz, 2016; van der Heide et al., 2020b), and research indicating that clinical judgment alone cannot help clinicians reliably identify when clients have produced non-credible data (Dandachi-FitzGerald et al., 2017; Guilmette, 2013), a number of objective self-report and performance validity measures have been developed to help identify when data obtained in any assessment is not credible. These measures cannot determine *why* data are invalid, only that symptom reports or performance scores are unlikely to be interpretable. Importantly, validity test failure does not rule out the presence of a genuine disorder, but it does preclude reliable inference regarding the nature or cause of the current results based on the available (invalid) data.

Since the onset of the COVID-19 pandemic, the mental health of adolescents and young adults has deteriorated globally, with substantial increases in internalizing symptoms, psychological distress, and functional impairment reported across countries (Benton et al., 2021; Haidt, 2024; Jo et al., 2023; Jones et al., 2022; Montero-Marin et al., 2023; Racine et al., 2021). Concurrently, the proliferation of health-related misinformation on social media platforms has been linked to outbreaks of mass sociogenic illness, particularly among adolescents and younger adults (Giedinghagen, 2023;

Shmerling, 2022). During this period, clinicians have documented sharp increases in presentations involving non-credible or atypical symptom patterns resembling Tourette's syndrome (e.g. Frey et al., 2022; Müller-Vahl et al., 2022), dissociative identity disorder (e.g. Porter et al., 2024), autism spectrum disorder (e.g. Harness & Getzen, 2022), and ADHD (e.g. Hartnett & Cummings, 2024; Martin et al., 2023).

Emerging evidence suggests that such sociogenic symptom presentations may be more likely to occur among individuals (disproportionately females) with elevated internalizing psychopathology, heightened suggestibility, or broader psychological distress (e.g. Haltigan et al., 2023). Importantly, although symptom and performance validity tests (SVTs and PVTs) are designed to be largely insensitive to bona fide psychiatric and neurologic conditions, prior research indicates that high levels of psychological distress, maladaptive illness beliefs, exposure to misinformation, and contextual pressures can increase the likelihood of exaggerated or distorted symptom reporting (Larrabee, 2012; Merckelbach et al., 2019). Under these conditions, elevations on symptom validity indicators, particularly those embedded within self-report measures, may occur more frequently, even in the absence of deliberate feigning. As such, given the marked rise in mental health symptoms among young adults concurrent with a "crisis" in health misinformation dissemination on social media platforms both during and post-COVID pandemic (e.g. West et al., 2021), the stage is potentially set for significant symptom overreporting.

Within this context, there is a need to empirically distinguish between changes in performance validity, symptom validity, and psychological distress among assessment-seeking postsecondary students during and following the COVID-19 pandemic. Determining whether such increases are occurring has important implications for diagnostic accuracy and clinical decision-making.

The primary aim of the present study was to examine whether rates of performance validity test (PVT) and symptom validity test (SVT) failure among assessment-seeking postsecondary students differed across pre-, mid-, and post-COVID-19 cohorts. Given prior literature demonstrating the relative robustness of validity tests across psychiatric and contextual factors, we examined whether performance validity outcomes would remain stable across cohorts despite substantial increases in psychological distress observed during and after the pandemic.

A second aim was to evaluate whether rates of non-credible symptom reporting, particularly on ADHD-specific self-report measures, differed across COVID-era cohorts. Based on evidence of increased psychological distress, increased non-credible symptom information being circulated on social media, and prior findings of ADHD symptom over-endorsement in high-distress contexts, we hypothesized that SVT elevations, especially on disorder-specific self-report measures, would be more frequent in the mid- and post-COVID cohorts.

Finally, we examined whether changes in symptom validity outcomes co-occurred with changes in general psychological distress. We assessed general psychological distress using a higher-order distress dimension derived from the PAI. This approach allows symptom reporting patterns to be interpreted within the broader context of global internalizing distress, rather than relying solely on individual symptom scale elevations. Rather than testing a causal model, this aim was intended to contextualize

symptom reporting patterns within broader shifts in distress observed among post-secondary students following the onset of the COVID-19 pandemic.

Methods

Participants

Participants in this archival study were consecutively assessed postsecondary students who completed psycho-educational assessments at a regional assessment center between 2018 and 2024 and agreed to allow their data to be used for research purposes. Students from various community colleges and universities across Ontario, Canada were referred to this assessment center for a comprehensive assessment to determine (or verify) whether they had an attention-related disorder that would require academic accommodations and/or medication to treat their symptoms.

In total, 1076 students (45.2% male; mean age = 22.16 years, SD = 6.70, range 17–59) were identified who met these criteria. There was no significant difference found in gender distribution across the time periods (X^2 (4) = 9.184, p = .057). There was, however, a significant difference in age F = (2, 1073) = 4.446, p = .012. Post-hoc testing showed that individuals assessed post-COVID were significantly older than individuals assessed pre-COVID. While specific information about ethnic groups was not included for each client in the database, a review of all referrals made to the center during this time period showed that the majority were White/Caucasian (72%), with the remainder of the sample identifying as Asian (12%), Black (6%), Middle Eastern (3%), or Other/not specified (7%). We therefore assume that the current sample is likely of similar composition.

Materials

Since these assessments were all undertaken to determine not only diagnosis but also to evaluate the need for possible academic accommodations, the assessment protocols employed a variety of measures, including: measures of cognitive and academic ability, self-and observer ADHD symptom reports, at least one stand-alone performance validity test (PVT), at least one embedded PVT, at least one symptom validity test (SVT), and a self-report measure of personality. The measures of interest in the present study are described below. Note that due to clinician preference and referral reasons, not every PVT and SVT listed was given to each student.

Clinical measures of interest

Personality assessment inventory (PAI)

The PAI (Morey, 1991) is a widely used, 344-item self-report inventory that measures a number of personality traits. Test-takers read and rate each statement on a four-point Likert scale (1-"Not true at all, False", 2-"Slightly true", 3-"Mainly true", and 4-"Very true"). The test provides the clinician with data from 11 clinical scales, 5 treatment scales, and 2 interpersonal scales, along with a number of validity indices. Non-content

validity indices are included to determine whether the client responded consistently, paid sufficient attention to item content, and provided a sufficient number of responses. Inconsistency (ICN) scores greater than $T = 72$ and/or those with Infrequency (INF) scores greater than $T = 74$ are deemed invalid due to random or careless responding.

To evaluate content-related validity, the PAI includes two basic negative response bias scales of relevance to the present study: Negative Impression Management (NIM; Morey, 1991), and Malingering (MAL; Morey, 1993, 1996). NIM represents high reporting of rarely endorsed items and is interpreted as a measure of item over-endorsement; MAL is associated with exaggerated or feigned psychopathology and, in applied neuropsychological settings, may co-occur with other indicators of non-credible responding, although it is not designed to assess cognitive effort directly. Musso et al. (2016) report that NIM and MAL showed promise in identifying college students feigning ADHD when using the following cutoffs (NIM >= 77, Mal >= 3). In addition, the PAI also includes indicators of positive distortion. For example, high scores (T >= 68) on the Positive Impression Management scale (PIM) reflect overtly defensive responding or lack of insight. However, low scores on the PIM (i.e. denial of any positive qualities and endorsement of many minor flaws relative to the average individual) may instead reflect deliberate exaggeration of psychological distress (Morey & Lanier, 1998). For example, Musso et al. (2016) found that students who were able to successfully simulate symptoms of ADHD on a self-report rating scale obtained higher scores on the NIM and lower scores on the PIM scale compared with actual ADHD students. A cutoff score of <25 on the PIM scale resulted in 97.3% sensitivity and 91.1% specificity in identifying students feigning ADHD. Similarly, in their study of feigned ADHD, Smith et al. (2017) noted that clinicians should be highly suspicious of symptom exaggeration when PAI profiles include mild to moderate elevations on the NIM as well as markedly below average scores on the PIM. Finally, subtracting PIM from NIM provides a single score that is said to reflect the overall direction and magnitude of impression management (Hopwood et al., 2008). A positive score of 40 or more suggests a greater tendency toward negative impression management, while a negative score (below zero) suggests a greater tendency toward positive impression management.

General psychological distress (GPD) was estimated using the first higher-order component identified in the invariant component structure of the Personality Assessment Inventory (PAI) full scales described by Hoelzle and Meyer (2009). Using multiple extraction and retention methods across six independent clinical and non-clinical samples, Hoelzle and Meyer demonstrated that this component is highly stable and replicable, accounting for the largest proportion of shared variance among PAI scales. Component 1 is said to reflect a broad internalizing distress dimension, with strong loadings from scales assessing depression, anxiety, anxiety-related disorders, somatic complaints, stress, suicidality, and related symptomatology. Hoelzle and Meyer (2009) assert that this first factor of the PAI appears very similar to the second factor they derived from the MMPI-2 –Restructured Clinical Scales, which embodies depressive withdrawal, helplessness, emotional discomfort, and demoralization.

Importantly, although Component 1 includes contributions from Negative Impression Management, it is not conceptualized as a validity or malingering indicator. Rather, it represents a higher-order dimension of general psychological distress that remains interpretable after excluding profiles that meet standard PAI invalidity criteria. In the present study, Component 1 was used as a dimensional index of general psychological distress (GPD) to contextualize symptom reporting patterns rather than as an indicator of response validity. GPD analyses were restricted to participants with valid PAI profiles (i.e. INC, INF, NIM, PIM below interpretive thresholds). Of the participants with available GPD data, 46 cases were excluded due to non–content-related PAI invalidity, defined as Inconsistency (ICN) scores greater than $T = 72$ and/or Infrequency (INF) scores greater than $T = 74$. The proportion of cases excluded for PAI invalidity did not differ significantly across pre-, during-, and post-COVID cohorts, $\chi^2(2) = 0.144$, $p = .938$.

For the present study, we employed the scoring program provided by Hoelzle and Meyer (2009) to calculate a T-score for Component 1 of the PAI (GPD). Similar to other clinical scales of the PAI, T-scores of 70 or more were taken to indicate a significant elevation.

Conners' adult ADHD rating scale (CAARS)

The CAARS itself is a 66-item self-report scale that allows for the calculation of eight different indices, with some items contributing to more than one scale. CAARS items are rated on a 4-point scale (0 = *not at all*, 1 = *just a little*, 2 = *pretty much*, 3 = *very much*). The CAARS provides the following scores: a) four factor-derived subscales: b) three scales that correspond to the DSM-IV symptoms of Hyperactivity/Impulsivity, Inattention, and Total DSM symptoms; and c) an overall ADHD Index that is said to measure the "overall level of ADHD symptoms" (Conners et al., 1998, p. 23). Unfortunately, no symptom validity scales were included in this original version of the CAARS; the test includes only an Inconsistency Index designed to identify inconsistent responses to items measuring similar content rather than overreporting of symptoms. While the CAARS does not include any published SVT, three methods of evaluating self-report credibility on the CAARS do exist. The first is the CAARS Infrequency Index (CII; Suhr et al., 2011). The CII is composed of 12 items rarely endorsed by typically developing adults as well as those diagnosed with ADHD. Suhr et al. (2011) identified a cut score of >21 as producing few false positive identifications for those with ADHD. The index was found to have modest sensitivity (approximately 30%) and high specificity (approximately 95%). Cook et al. (2016) found that the CII had 52% sensitivity to feigning and 97% specificity for ADHD based on extreme elevations of the three CAARS clinical scales derived from DSM-IV ADHD criteria. Because item-level responses were not recorded for all clients, data from only $n = 570$ clients could be used to calculate the CII in the present investigation.

The other two SVTs that can be calculated for the CAARS are described in detail by Harrison and Armstrong (2016). These authors embedded 18 validity items into the existing CAARS to identify symptoms of over-endorsement. The sum of these items produces a Dissimulation score (DISS), while a formula that combines the DISS with extreme scores from existing CAARS indices produces the Exaggeration Index

(EI). According to Harrison and Armstrong (2016), this Exaggeration Index (EI) is said to have acceptable classification accuracy when discriminating between those feigning ADHD and other clinical groups (including those with ADHD) who were reporting symptoms accurately; a cut score of EI >2 had sensitivity of .34 and specificity of .94, and a cut score of > = 20 on the DISS correctly identified 43% of instructed to feign students. A follow-up study (Harrison et al., 2022) found the DISS to have .76 sensitivity and .91 specificity, and the EI to have .53 sensitivity and .93 specificity when differentiating credible reporters from those suspected of feigning ADHD.

Of note, 117 cases were excluded from analyses involving CAARS-based symptom validity indices due to elevated CAARS inconsistency scores (Consistency Index ≥ 8). The proportion of cases excluded for CAARS inconsistency did not differ across cohorts, $\chi^2(2) = 0.791$, $p = .673$.

Performance validity measures

A variety of PVTs were employed in the archival assessments, depending on clinician preference and usability during COVID. All assessments included at least two PVTs, as this is said to produce excellent specificity when both tests are failed (Odland et al., 2015). Assessments included at least two of the following: the Word Memory Test (WMT; Green, 2003 revised 2005) or the Medical Symptom Validity Test (MSVT; Green, 2004); the Victoria Symptom Validity Test (VSVT; Slick et al., 1997); the Dyslexia Assessment of Simulation or Honesty-Revised (DASH-R; Harrison et al., 2025); the Reliable Digit Span (RDS; Greiffenstein et al., 1994), an embedded measure obtained from the Digit Span subtest of the WAIS-IV; and the Symptom Exaggeration Index (SEI) from the Tests of Variables of Attention (TOVA; Greenberg et al., 2007). Cut scores used for each PVT may be found in Table 1. The present study did not classify individuals as globally credible or non-credible; rather, analyses focused on obtained scores and rates of exceeding established cutoffs on individual PVTs and SVTs across time periods.

Table 1. Validity tests employed and cut scores used.

Validity test	Failure cut score	Reference
Word Memory Test	As per manual	Green, 2003, 2005
Medical Symptom Validity Test	As per manual	Green, 2004
Victoria Symptom Validity Test	Hard Items <= 19	Frazier et al., 2008
Test of Variables of Attention Symptom Exaggeration Index	>=3	Greenberg et al., 2007
DASH-R Feigning Index	>3	Harrison et al., 2025
RDS	<=7	Harrison et al., 2010
PAI Validity Scales		Morey, 1991
Inconsistency ICN	>=73	
Infrequency INF	>=75	
Negative Impression Management NIM	>=77	Musso et al., 2016
Positive Impression Management PIM	<25 (e.g. denial of favorable characteristics)	Musso et al., 2016
NIM - PIM	>40	Hopwood et al., 2008
Malingering Index MAL	>=3	Morey, 1996
Conners' Adult ADHD Rating Scale (CAARS)		
CAARS Infrequency Index	>=21	Suhr et al., 2011
Exaggeration Index	>=3	Harrison & Armstrong, 2016
Dissimulation	>=20	Harrison & Armstrong, 2016

DASH-R: Dyslexia Assessment of Simulation or Honesty-Revised; RDS: Reliable Digit Span; PAI: Personality Assessment Inventory.

Procedure

As part of the informed consent process, clients referred for assessment were asked at the first appointment whether their data could be entered into a database and used, without any identifying information, in subsequent archival research projects. Clients were told that they did not have to agree to allow their de-identified data to be used in any research study, and that they would still be provided with a complete assessment regardless of their decision. All clients who agreed to have their anonymous data used for this research and who agreed to complete the experimental tests (99.9%) were included in this ongoing REB-approved archival data collection study.

Clients underwent a psychoeducational or neuropsychological assessment, including completion of mental health surveys, self and observer-symptom reports, cognitive and achievement tests relevant to the referral question, and tests of performance and symptom validity. To ensure that the current study was evaluating assessments conducted during similar spans of time, we reviewed data obtained from approximately equal periods of time both before and after the start of COVID-19 restrictions. Using the same classification method as that employed by Jamieson et al. (2025), students were grouped based on the date of their testing feedback appointment. If the assessment feedback appointment occurred during the dates of January 1st, 2018 to March 1st, 2020, they were included in the pre-COVID group. If the participants' feedback session occurred between the dates of March 2nd, 2020 to August 31st, 2022, they were placed in the during-COVID group. Given that in-person learning did not resume until the fall of 2022, students were placed in the post-COVID group if their feedback occurred during the dates of September 1st, 2022 to September 3rd, 2024. Note that during the COVID-19 restrictions phase, all in-person testing was completed using personal protective equipment and a plexiglass screen placed in between the client and the assessor, with room for testing materials to be passed underneath the screen.

Results

When examining scores obtained on PVTs during all three time periods, one can see that only the mean scores on the DASH-R Feigning Index (FI) were significantly different across the three time periods (see Table 2). Post-hoc analysis indicates significantly higher scores on the FI both during and after the COVID-pandemic when compared to the pre-pandemic period. Furthermore, the percentage of students deemed to have failed the DASH-R *(i.e. exceeded the published cutoff on that individual measure)* increased significantly both during and after the onset of COVID compared with pre-pandemic levels. On no other PVT, however, were obtained scores significantly different across time periods, nor was the percentage of students deemed to have failed a PVT different during those three time periods.

On the PAI, no significant difference was found across time periods in the number of students with elevations on non-content validity scales (INC and INF; See Table 3). After removing individuals with invalid PAI profiles based on INC and INF elevations, mean scores on the NIM, PIM, and NIM–PIM indices differed significantly across time periods. Post hoc analyses indicated that students assessed during and after COVID reported significantly higher levels of negative impression management (i.e. higher

Table 2. Performance validity tests scores from pre-during- and post COVID.

	Pre	During	Post	F/X^2	p	n^2/Cramer's V
WMT IR *N* = 608						
N	402	66	140			
Mean	94.48	93.16	94.15	0.729	.483	0.002
SD	8.48	8.02	8.13			
%<= 82.5	8.0	16.7	11.4	5.522	.063	0.095
WMT DR *N* = 606						
N	401	65	140			
Mean	94.55	94.35	93.94	0.210	.811	0.001
SD	9.67	8.68	9.70			
% <=82.5	10.0	7.7	11.4	0.695	.706	0.034
WMT CNS *N* = 604						
N	399	65	140			
Mean	92.53	91.09	91.63	0.784	.457	0.003
SD	10.48	10.54	10.05			
% <=82.5	14.8	16.9	17.9	0.815	.665	0.037
MSVT IR *N* = 391						
N	194	81	116			
Mean	97.18	96.54	97.22	0.280	.756	0.001
SD	7.21	7.36	6.30			
% <=85	6.7	8.6	6.9	0.341	.843	0.030
MSVT DR *N* = 390						
N	193	81	116			
Mean	95.48	94.94	95.39	0.097	.908	0.000
SD	9.56	9.60	8.97			
% <=85	11.9	12.3	12.1	0.010	.995	0.005
MSVT CNS *N* = 389						
N	193	81	115			
Mean	94.48	93.58	94.15	0.217	.805	0.001
SD	10.04	12.02	9.65			
% <=85	13.5	16.0	16.5	0.633	.729	0.040
VSVT Hard = 247						
N	168	22	57			
Mean	20.66	19.86	18.98	2.833	.061	0.023
SD	4.41	4.94	5.21			
% <=19	22.6	27.3	35.1	3.470	.176	0.119
TOVA SEI = 270						
N	124	61	85			
Mean	.24	.16	.27	1.097	.336	0.006
SD	.55	.37	.59			
% >=3	0.8	0	1.2	0.682	.711	0.050
RDS = 546						
N	282	109	155			
Mean	8.55	8.50	8.63	0.167	.846	0.001
SD	1.93	1.77	1.80			
% <=7	30.1	28.4	26.5	0.671	.715	0.035
DASH-R FI = 433						
N	257	35	141			
Mean	1.24^a	1.83^b	1.82^b	10.989	<.001	0.054
SD	1.03	1.38	1.39			
% >=3	10.9^a	22.9^b	29.8^b	22.608	<.001	0.229

Note: Means with different superscripts are significantly different from one another based on Tukey HSD Post Hoc analysis; WMT IR: Word Memory Test Immediate Recall; WMT DR: Word Memory Test Delayed Recall; WMT CNS: Word Memory Test Consistency; MSVT IR: Medical Symptom Validity Test Immediate Recall; MSVT DR: Medical Symptom Validity Test Delayed Recall; MSVT CNS: Medical Symptom Validity Test Consistency; VSVT: Victoria Symptom Validity Test; TOVA SEI: Test of Variables of Attention Symptom Exaggeration Index; DASH-R: Dyslexia Assessment of Simulation or Honesty-Revised; RDS: Reliable Digit Span.

NIM scores), were less likely to present themselves in a favorable light (i.e. lower PIM scores), and demonstrated larger NIM–PIM discrepancy scores compared to pre-COVID cohorts (see Table 4).

Table 3. PAI inconsistency and infrequency from pre- during- and post COVID.

	Pre	During	Post	F/X^2	*p*	n^2/Cramer's *V*
PAI INC = 705						
N	383	109	213			
Mean	53.65	54.32	53.49	0.335	.715	0.001
SD	9.06	7.95	8.73			
% >=73	2.3	0.9	2.3	0.907	.635	0.036
PAI INF = 705						
N	383	109	213			
Mean	54.61	54.61	55.36	0.475	.622	0.001
SD	9.47	8.93	9.44			
% >=75	1.3	2.8	1.9	1.118	.572	0.040

PAI INC: Inconsistency score from Personality Assessment Inventory; PAI INF: Infrequency score from Personality Assessment Inventory.

Table 4. PAI symptom validity and general psychological distress scores from pre-during- and post COVID.

	Pre	During	Post	F/X^2	*p*	n^2/Cramer's *V*
PAI NIM = 657						
N	356	101	200			
Mean	57.23^{a}	62.23^{b}	61.34^{b}	7.517	<.001	0.022
SD	13.61	15.08	16.10			
% >=77	7.6	8.9	12.0	3.016	.221	0.068
PAI PIM = 657						
N	356	101	200			
Mean	44.25^{a}	40.91^{b}	40.24^{b}	8.615	<.001	0.026
SD	12.11	10.69	11.36			
%<=24	3.7	5.0	6.5	2.325	.313	0.059
PAI NIM-PIM = 657						
N	356	101	200			
Mean	12.99^{a}	21.31^{b}	21.10^{b}	10.550	<.001	0.031
SD	22.28	22.55	23.63			
%>40	11.0^{a}	19.8^{ab}	19.0^{b}	8.993	.011	0.117
PAI MAL = 543						
N	356	101	86			
Mean	.95	1.01	.98	0.149	.862	0.001
SD	.95	.92	1.03			
% >=3	6.7	6.9	10.5	1.442	.486	0.052
Hoelzle General Psychological Distress = 638						
N	345	96	197			
Mean	60.10^{a}	65.84^{b}	65.02^{b}	13.820	<.001	0.042
SD	12.57	11.90	12.64			
%>=70	23.2^{a}	37.5^{b}	37.6^{b}	15.611	<.001	0.156

Means with different superscripts are significantly different from one another based on Tukey HSD Post Hoc analysis; PAI NIM: Personality Assessment Inventory Negative Impression Management; PAI PIM: Personality Assessment Inventory Positive Impression Management; PAI MAL: Personality Assessment Inventory Malingering Inventory.

Examination of categorical thresholds revealed that, although the proportion of students exceeding interpretive cutoffs on NIM and MAL did not differ significantly across time periods, a significantly greater proportion of students assessed post-COVID exceeded established NIM–PIM discrepancy thresholds indicative of a strong tendency toward negative impression management. Importantly, this pattern reflects increased negative self-presentation rather than elevated rates of invalid or malingered psychological profiles, as elevations on malingering-related indices (e.g. MAL) remained stable across cohorts.

By contrast, the level of General Psychological Distress (GPD) was different during the three time periods. Post-hoc testing indicated not only that scores were significantly higher during COVID, but that they also remained higher than pre-pandemic levels after COVID restrictions were lifted. Furthermore, even though analyses were restricted to participants with valid PAI profiles (see methods), the percentage of students reporting severe levels of GPD (i.e. T-scores of 70 or more) rose significantly from 23.2% pre-COVID to 37.5% during and 37.6% post-COVID (see Table 4).

With respect to non-credible self-reporting of ADHD symptoms, we examined the scores on the three CAARS validity scales. After removing those who answered this measure too inconsistently, we found that scores on the CII, EI, and DISS were all significantly different across the three time periods. Post-hoc testing indicated that these three validity scores were significantly higher only in the post-COVID group relative to the other two earlier time periods (see Table 5). Similarly, the proportion of students obtaining scores on each of these scales suggestive of non-credible reporting increased only in the post-COVID period.

Given the observed post-COVID increases in both general psychological distress (GPD) and non-credible ADHD symptom reporting, additional analyses were conducted to examine the relationship between GPD and elevations on the CAARS symptom validity indicators.

Students who violated at least one CAARS validity scale ($n = 79$) reported substantially higher levels of GPD ($M = 74.22$, SD = 10.55) than those with valid CAARS profiles ($n = 165$; $M = 57.75$, SD = 10.94), a difference that was statistically significant, t(242) = 11.12, $p < .001$, with a large effect size (Cohen's $d = 1.52$).

Receiver operating characteristic (ROC) curve analysis indicated that GPD demonstrated good discriminative ability for CAARS validity scale elevations (AUC = 0.85, 95% CI [0.81, 0.90]). A GPD cut score of 66 yielded a sensitivity of 0.81 and specificity of 0.77 for identifying individuals who violated one or more CAARS validity indicators. Notably, a cut score of $T = 86$ produced no false positives, indicating that extreme elevations in GPD were consistently associated with CAARS validity scale violations.

Table 5. Conners' adult ADHD rating scale (CAARS) symptom validity test scores from pre-during- and post COVID.

	Pre	During	Post	F/X^2	p	n^2/Cramer's V
CAARS CII = 447						
N	236	85	126			
Mean	10.62[a]	11.37[ab]	13.83[b]	7.171	<.001	0.034
SD	7.22	6.92	7.95			
% >=21	8.3[a]	10.8[a]	26.6[b]	23.420	<.001	0.232
CAARS EI = 414						
N	164	91	159			
Mean	.74[a]	.86[ab]	1.26[b]	4.392	.013	0.022
SD	1.38	1.60	1.77			
% >=3	10.4[a]	11.0[ab]	21.41[b]	9.103	.011	0.148
CAARS DISS = 416						
N	163	92	161			
Mean	11.99[a]	14.84[ab]	16.92[b]	11.249	<.001	0.049
SD	7.93	10.28	10.91			
% >=20	16.6[a]	25.0[ab]	33.5[b]	12.449	.002	0.173

Note: Means with different superscripts are significantly different from one another based on Tukey HSD Post Hoc analysis.

To further examine this relationship, a multiple linear regression analysis was conducted with the CAARS CII, EI, and DISS scales entered simultaneously as predictors of GPD in an associative model. The overall model was statistically significant, $F(3, 240) = 102.90$, $p < .001$, accounting for 56.3% of the variance in GPD scores ($R^2 = 0.563$; adjusted $R^2 = 0.557$). All three CAARS validity scales contributed significantly to the model. Consistent with these findings, GPD was strongly and positively correlated with the CAARS CII ($r = 0.72$), EI ($r = 0.52$), and DISS ($r = 0.63$), all ps < .001.

Discussion

To our knowledge, this is the first study to document largely stable rates of exceeding established cutoffs on performance validity test (PVT) across pre-, mid-, and post-COVID cohorts, alongside increased failure on certain disorder-specific symptom validity tests (SVTs) and a single domain-specific PVT among assessment-seeking postsecondary students following the onset of the COVID-19 pandemic. Notably, there was no evidence that students were more prone to non-credible cognitive or memory-related performance problems post-pandemic, nor were most conventional PVTs failed at higher rates relative to pre-COVID assessments.

As noted above, the only increase in PVT failure was on the DASH-R, a dyslexia-specific PVT, on which both FI scores and failure rates were approximately doubled post-COVID relative to pre-pandemic levels. This isolated finding should be interpreted cautiously in the context of otherwise stable PVT performance. Given that the DASH-R stimuli are extremely easy even for individuals with genuine reading disorders (Harrison et al., 2025), these elevations may reflect task-specific non-credible responding on reading-related measures in high-stakes assessments rather than a generalized decline in performance validity.

In contrast, symptom overreporting increased substantially following the onset of COVID, particularly non-credible overreporting of ADHD symptoms. This pattern was not attributable to random responding or elevations on malingering-related indices but rather co-occurred with a marked and sustained increase in general psychological distress (GPD). Levels of GPD were significantly higher during and after COVID, with extreme elevations observed in over one-third of students during these periods. These findings align with national survey data indicating substantial post-pandemic increases in psychological distress among Canadian postsecondary students (American College Health Association, 2022) and suggest that elevated distress has persisted well beyond the resumption of in-person learning.

In addition to increases in general psychological distress, post-COVID cohorts demonstrated a broader shift toward negative self-presentation on personality validity indices. Specifically, a significantly greater proportion of students assessed post-COVID exhibited elevated NIM–PIM discrepancy scores, reflecting a stronger tendency toward negatively biased self-reporting. Importantly, this pattern was not accompanied by increases in malingering-related indices, suggesting that these elevations reflect response style and distress-related self-presentation rather than intentional exaggeration or globally invalid profiles. This broader shift in negative response style provides important context for interpreting disorder-specific symptom validity findings.

Although prior studies did not examine validity indicators directly, Jamieson et al. (2025) reported increased self-reported ADHD symptoms following the COVID-19 pandemic without a corresponding rise in diagnostic rates, speculating that psychological distress or exposure to misinformation may underlie this pattern. Increased social media use during the pandemic coincided with the proliferation of non-credible information about ADHD and other psychological conditions (Karasavva et al., 2025; Verma & Sinha, 2025; Yeung et al., 2022), a phenomenon described as mass sociogenic illness (e.g. Fremer et al., 2024; Giedinghagen, 2023). While causal mechanisms cannot be established, the co-occurrence of elevated general psychological distress (GPD) and increased non-credible ADHD symptom reporting provides an important contextual framework for interpreting the present SVT findings.

An important conceptual clarification is warranted. The present findings do not suggest that psychological distress directly produces failures on SVTs, nor that SVTs are sensitive to distress-related psychopathology. Rather, elevated distress may function as a contextual vulnerability factor that increases reliance on symptom-based explanatory frameworks, such as ADHD, when individuals attempt to interpret nonspecific cognitive, emotional, or functional difficulties. Under these conditions, symptom endorsement may become exaggerated or poorly calibrated relative to objective evidence, increasing the likelihood of elevations on disorder-specific SVTs. This interpretation reflects a population-level association rather than a causal or individual-level linkage between distress and SVT performance.

Consistent with this framework, additional analyses demonstrated a strong association between GPD and elevations on CAARS symptom validity indicators. Students who violated one or more CAARS validity scales reported substantially higher levels of psychological distress than those with valid CAARS profiles, with large effect sizes. Moreover, GPD showed good discriminative ability for CAARS SVT elevations, and CAARS validity scales collectively accounted for a substantial proportion of variance in GPD scores. Importantly, these findings do not indicate that CAARS SVTs measure psychological distress; rather, they suggest that elevated distress is associated with an increased likelihood of non-credible ADHD symptom endorsement.

Taken together, these findings support the interpretation that psychological distress operates as a contextual risk factor for symptom over-endorsement rather than as a direct determinant of validity test performance. This interpretation is consistent with prior work demonstrating associations between internalizing distress, negative affectivity, and exaggerated symptom reporting in the absence of deliberate feigning (Denovan et al., 2019; Merckelbach et al., 2019). Prior research indicates that anxiety, depression, neuroticism, and general psychological distress are associated with symptom over-endorsement, even in the absence of deliberate feigning (Denovan et al., 2019; Merckelbach et al., 2019). Consistent with Harrison et al. (2013), these findings suggest that ADHD symptom elevations may function analogously to a fever in medical contexts, signaling distress or functional difficulty while remaining diagnostically nonspecific. However, when such symptom elevations occur in conjunction with elevations on ADHD-specific symptom validity indicators, they cannot be interpreted as credible evidence of ADHD symptomatology.

Regardless of etiology, these results underscore the importance of incorporating SVTs and PVTs into ADHD and learning disorder assessments. Although performance validity failures remained largely stable across COVID-era cohorts, between one-fifth and one-third of students assessed post-COVID showed evidence of non-credible ADHD symptom reporting. Given that clinicians are poor at detecting symptom exaggeration based on clinical judgment alone, objective validity testing remains essential, particularly when assessments carry implications for diagnosis, treatment, or academic accommodations. Failure to identify invalid data increases the risk of misdiagnosis, inappropriate intervention, and reliance on symptom explanations that do not accurately reflect the primary source of impairment. Alternative explanations, including increased motivation to obtain ADHD diagnoses or academic accommodations in postsecondary settings, cannot be ruled out and may also contribute to elevated rates of non-credible ADHD symptom reporting in post-pandemic cohorts.

Clinical implications for ADHD and LD assessment in postsecondary students

In clinical practice, diagnostic decisions often rely heavily on self-reported symptoms and perceived functional impairment. When such data is invalid, conclusions may be compromised, leading to interventions poorly aligned with students' actual needs. For example, attentional difficulties arising in the context of acute distress may be misattributed to a neurodevelopmental disorder, potentially resulting in inappropriate pharmacologic treatment or delayed access to more suitable mental health interventions.

Invalid assessment data also have implications at the institutional level. Postsecondary accommodations require substantial administrative and financial resources, and when granted on the basis of non-credible data, these resources may be misallocated, reducing access for students with well-documented and persistent disabilities. Failure to detect non-credible responding therefore has both clinical and equity implications.

When clinicians encounter failed PVTs or SVTs in postsecondary ADHD or LD assessments, particularly on measures such as the DASH-R or CAARS, best practice involves neither ignoring these findings nor equating them with intentional deception. Rather, validity test failures should prompt a structured reevaluation of the assessment data and clinical formulation. As outlined by Schroeder and Martin (2025), this process includes examining the pattern and consistency of validity indicators, considering contextual and psychological factors that may influence responding (e.g. acute distress, test-taking expectations, illness beliefs), and integrating collateral information such as academic history, prior documentation, and behavioral observations.

When symptom validity indicators embedded in self-report measures are elevated (e.g. CAARS-EI, PAI validity scales), clinicians should exercise caution in interpreting symptom severity scales and avoid making diagnostic decisions based solely on self-report. Instead, greater weight may be placed on developmental history, objective academic records, performance-based measures, and longitudinal consistency of impairment. Similarly, isolated failures on domain-specific PVTs, such as the DASH-R, should be interpreted in the context of broader performance validity results and task characteristics, rather than as evidence of global non-credible performance.

From a clinical management perspective, the identification of validity concerns may warrant additional steps, such as feedback-focused clarification, deferred diagnostic conclusions, recommendations for follow-up assessment, or referral for treatment targeting distress prior to reevaluation. In this way, validity testing functions as a tool for improving diagnostic accuracy and treatment planning rather than as a gatekeeping or punitive mechanism (Schroeder & Martin, 2025).

Importantly, validity test failures should not be automatically interpreted as evidence of deliberate deception among assessment-seeking students. Rather, they may suggest that elevated distress and contextual pressures, particularly in the post-pandemic period, may compromise the accuracy of symptom reporting. From this perspective, SVTs and PVTs serve not as punitive tools, but as safeguards that promote accurate diagnosis, appropriate intervention planning, and fair allocation of institutional supports.

Limitations

Like most archival studies, ours has a number of limitations which may affect interpretation and generalization of our findings. The primary limitation concerns the classification of students into pre-, during-, and post-COVID groups based on the timing of their feedback sessions. This approach may not have been a completely accurate reflection of the exact time period in which the actual assessment occurred. For instance, the complete lock down did not begin in Canada until mid-March 2020, and pandemic restrictions formally ended in late April 2022. It is therefore possible that some students classified as being assessed during the pandemic may have been less anxious after restrictions had ended, but school had not yet begun. Nevertheless, our data suggest that, in fact, levels of self-reported distress remained high both during and after the pandemic ended, suggesting that our classification approach likely reflects the broad tendencies found during and after the pandemic.

A second limitation is that students tested during COVID were tested in person; however, they were assessed under modified conditions. These conditions included the use of personal protective devices such as masks and a plexiglass barrier placed between the assessor and the client. These procedural differences may have modified both how the students performed and self-reported. However, if this deviation in standardized assessment practice only affected those tested during COVID then one would expect to see specific differences only in that middle group, whereas in fact our findings show that changes in self-reporting either started during COVID and continued afterwards, or became apparent only after COVID restrictions were lifted. It is therefore unlikely that these methodological differences can explain the observed increases in non-credible symptom reporting and self-reported levels of general distress.

Although participants completed varying PVT batteries, this reflects routine clinical practice. Because our analyses focused on performance on individual PVTs rather than case-level validity determinations, heterogeneity in PVT composition was present across all cohorts and would be expected to introduce random measurement noise rather than systematic bias.

Fourth, as we did not have item-level responses from the PAI, we could not calculate other proposed PAI validity indices such as the Negative Distortion Scale (Mogge et al., 2010) or the Cognitive Bias Scale (Gaasedelen et al., 2019). Although scale-level analogues of the CBS (CBS–Scale of Scales; Boress et al., 2022) have been developed and validated for use when item-level data are unavailable, these indices were not part of the original analytic framework for the present study. Future research may benefit from examining whether item-level CBS indices or CBS–SoS measures are sensitive to temporal changes in response style and non-credible self-report across pandemic-related cohorts.

Finally, while the assessment seeking students came from a wide range of post-secondary environments (community college and 4-year universities) in mid-sized communities, the sample was predominantly Caucasian and relatively well educated. This therefore limits the generalizability of the study, and it remains unknown whether these trends would also be observed in extremely large cities or in more ethically and socioeconomically diverse populations. Future research should therefore examine if these patterns of validity test performance and psychological distress differ in settings that deal with more diverse groups of young adults.

Conclusion

The central performance validity finding of this study is the stability of PVT failure rates across pre-, mid-, and post-COVID cohorts, indicating that heightened psychological distress did not translate into widespread non-credible cognitive performance. Although an isolated increase was observed on one reading-specific PVT, this finding should be interpreted cautiously given otherwise consistent PVT results.

By contrast, indicators of non-credible symptom reporting, particularly for ADHD, increased substantially following the onset of COVID and co-occurred temporally with elevated levels of general psychological distress. While causal mechanisms cannot be determined, these findings reinforce the importance of incorporating multiple, independent validity indicators into ADHD and learning disorder assessments. Without such measures, clinicians risk relying on data that may be misleading, increasing the likelihood of misdiagnosis, inappropriate treatment, and misallocation of academic supports.

Overall, the pandemic appears to have amplified psychological distress and vulnerability to non-credible symptom reporting among assessment-seeking postsecondary students. Regardless of etiology, these findings underscore the necessity of routine validity testing to support accurate, equitable, and clinically sound decision-making.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was not preregistered. Data used in this study may be obtained by contacting the first author.

ORCID

Allyson G. Harrison http://orcid.org/0000-0002-0426-2011
Nathaniel Davin http://orcid.org/0000-0002-2687-4617
Emma Jamieson http://orcid.org/0009-0003-5166-9695

References

American College Health Association. (2022). *American College Health Association-National College Health Assessment III: Canadian reference group data report spring 2022*.

Aronson, K. R., Barrett, L. F., & Quigley, K. (2006). Emotional reactivity and the overreport of somatic symptoms: Somatic sensitivity or negative reporting style? *Journal of Psychosomatic Research*, *60*(5), 521–530. https://doi.org/10.1016/j.jpsychores.2005.09.001

Bender, S. D. (2015). Response style and malingering. *In* R. L. Cautin & S. O. Lilienfeld (Eds.), *The encyclopedia of clinical psychology* (pp. 1–6). John Wiley & Sons. https://doi.org/10.1002/9781118625392.WBECP360

Benton, T. D., Boyd, R. C., & Njoroge, W. F. M. (2021). Addressing the global crisis of child and adolescent mental health. *JAMA Pediatrics*, *175*(11), 1108–1110. https://doi.org/10.1001/jamapediatrics.2021.2479

Boress, K., Gaasedelen, O. J., Croghan, A., Johnson, M. K., Caraher, K., Basso, M. R., & Whiteside, D. M. (2022). Replication and cross-validation of the personality assessment inventory (PAI) cognitive bias scale (CBS) in a mixed clinical sample. *The Clinical Neuropsychologist*, *36*(7), 1860–1877. https://doi.org/10.1080/13854046.2021.1889681

Chafetz, M. D., Bauer, R. M., & Haley, P. S. (2020). The other face of illness-deception: Diagnostic criteria for factitious disorder with proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, *34*(3), 454–476. https://doi.org/10.1080/13854046.2019.1663265

Conners, C. K., Erhardt, D., & Sparrow, E. (1998). *Adult ADHD rating scales (CAARS): ADHD across the lifespan*. Multi-Health Systems.

Cook, C. M., Bolinger, E., & Suhr, J. (2016). Further validation of the Conner's adult attention deficit/hyperactivity rating scale infrequency index (CII) for detection of non-credible report of attention deficit/hyperactivity disorder symptoms. *Archives of Clinical Neuropsychology*, *31*(4), 358–364. https://doi.org/10.1093/arclin/acw015

Dandachi-FitzGerald, B., Merckelbach, H., & Ponds, R. W. (2017). Neuropsychologists' ability to predict distorted symptom presentation. *Journal of Clinical and Experimental Neuropsychology*, *39*(3), 257–264. https://doi.org/10.1080/13803395.2016.1223278

Denovan, A., Dagnall, N., & Lofthouse, G. (2019). Neuroticism and somatic complaints: Concomitant effects of rumination and worry. *Behavioural and Cognitive Psychotherapy*, *47*(4), 431–445. https://doi.org/10.1017/S1352465818000619

Frazier, T. W., Frazier, A. R., Busch, R. M., Kerwood, M. A., & Demaree, H. A. (2008). Detection of simulated ADHD and reading disorder using symptom validity measures. *Archives of Clinical Neuropsychology*, *23*(5), 501–509. https://doi.org/10.1016/j.acn.2008.04.001

Fremer, C., Szejko, N., Pisarenko, A., Haas, M., Laudenbach, L., Wegener, C., & Müller-Vahl, K. R. (2024). What distinguishes patients with mass social media-induced illness presenting with Tourette-like behavior from those with Tourette syndrome? Results of a prospective cohort study. *European Archives of Psychiatry and Clinical Neuroscience*, *274*(1), 227–233. https://doi.org/10.1007/s00406-023-01603-z

Frey, J., Black, K. J., & Malaty, I. A. (2022). TikTok Tourette's: Are we witnessing a rise in functional tic-like behavior driven by adolescent social media use? *Psychology Research and Behavior Management*, *15*, 3575–3585. https://doi.org/10.2147/PRBM.S359977

Gaasedelen, O. J., Whiteside, D. M., Altmaier, E., Welch, C., & Basso, M. R. (2019). The construction and the initial validation of the cognitive bias scale for the personality assessment inventory. *The Clinical Neuropsychologist*, *33*(8), 1467–1484. https://doi.org/10.1080/13854046.2019.1612947

Giedinghagen, A. (2023). The tic in TikTok and (where) all systems go: Mass social media induced illness and Munchausen's by internet as explanatory models for social media associated abnormal illness behavior. *Clinical Child Psychology and Psychiatry*, *28*(1), 270–278. https://doi.org/10.1177/13591045221098522

Green, P. (2003). *Green's word memory test for windows: User's manual*. Green's Publishing.

Green, P. (2004). *Green's Medical Symptom Validity Test (MSVT) for Microsoft windows: User's manual*. Green's Publishing.

Green, P. (2005). *Manual for the Word Memory Test (Revised)*. Green's Publishing.

Greenberg, L. M., Kindschi, C. L., Dupuy, T. R., & Hughes, S. J. (2007). *T.O.V.A. clinical manual: Test of variables of attention*. The T.O.V.A. Company.

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesic measures with a large clinical sample. *Psychological Assessment*, *6*(3), 218–224. https://doi.org/10.1037/1040-3590.6.3.218

Guilmette, T. (2013). The role of clinical judgment in symptom validity assessment. *In* D. Carone & S. Bush (Eds.), *Mild traumatic brain injury: Symptom validity assessment and malingering* (pp. 31–44). Springer.

Haidt, J. (2024). *The anxious generation: How the great rewiring of childhood is causing an epidemic of mental illness*. Penguin Random House.

Haltigan, J. D., Pringsheim, T. M., & Rajkumar, G. (2023). Social media as an incubator of personality and behavioral psychopathology: Symptom and disorder authenticity or psychosomatic social contagion? *Comprehensive Psychiatry*, *121*, 152362. https://doi.org/10.1016/j.comppsych.2022.152362

Harness, J., & Getzen, H. (2022). TikTok's sick-role subculture and what to do about it. *Journal of the American Academy of Child & Adolescent Psychiatry*, *61*(3), 351–353. https://doi.org/10.1016/j.jaac.2021.09.312

Harrison, A. G., Alexander, S. J., & Armstrong, I. T. (2013). Higher reported levels of depression, stress, and anxiety are associated with increased endorsement of ADHD symptoms by postsecondary students. *Canadian Journal of School Psychology*, *28*(3), 243–260. https://doi.org/10.1177/0829573513480616

Harrison, A. G., & Armstrong, I. T. (2016). Development of a symptom validity index to assist in identifying ADHD symptom exaggeration or feigning. *The Clinical Neuropsychologist*, *30*(2), 265–283. https://doi.org/10.1080/13854046.2016.1154188

Harrison, A. G., Davin, N., & Pollock, B. (2025). Enhanced detection of suboptimal effort in psychoeducational assessments for dyslexia. *The Clinical Neuropsychologist*, *39*(6), 1559–1580. https://doi.org/10.1080/13854046.2024.2422139

Harrison, A. G., Harrison, K. A., & Armstrong, I. T. (2022). Discriminating malingered attention deficit hyperactivity disorder from genuine symptom reporting using novel Personality Assessment Inventory validity measures. *Applied Neuropsychology. Adult*, *29*(1), 10–22. https://doi.org/10.1080/23279095.2019.1702043

Harrison, A. G., Lee, G., & Suhr, J. (2021). Use of performance validity tests and symptom validity tests in assessment of specific learning disorders and attention deficit hyperactivity disorder. *In* Kyle Boone (Ed.). *Assessment of feigned cognitive impairment* (2nd ed.). Guilford Press.

Harrison, A. G., Rosenblum, Y., & Currie, S. (2010). Examining unusual digit span performance in a population of postsecondary students assessed for academic difficulties. *Assessment*, *17*(3), 283–293. https://doi.org/10.1177/1073191109348590

Hartnett, Y., & Cummings, E. (2024). Social media and ADHD: Implications for clinical assessment and treatment. *Irish Journal of Psychological Medicine*, *41*(1), 132–136. https://doi.org/10.1017/ipm.2023.40

Hoelzle, J. B., & Meyer, G. J. (2009). The invariant component structure of the Personality Assessment Inventory (PAI) full scales. *Journal of Personality Assessment*, *91*(2), 175–186. https://doi.org/10.1080/00223890802634316

Hong, V., Pirnie, L., & Shobassy, A. (2019). Antisocial and borderline personality disorders in the emergency department: Conceptualizing and managing "malingered" or "exaggerated"

symptoms. *Current Behavioral Neuroscience Reports*, *6*(4), 127–132. https://doi.org/10.1007/s40473-019-00183-4

Hopwood, C. J., Talbert, C. A., Morey, L. C., & Rogers, R. (2008). Testing the incremental utility of the negative impression-positive impression differential in detecting simulated personality assessment inventory profiles. *Journal of Clinical Psychology*, *64*(3), 338–343. https://doi.org/10.1002/jclp.20439

Jamieson, E., Pollock, B., Davin, N., & Harrison, A. G. (2025). AD/HD symptoms in assessment seeking post-secondary students: Has the COVID-19 pandemic made a difference? *The Clinical Neuropsychologist*, 1–13. https://doi.org/10.1080/13854046.2025.2533298

Jo, E., Seo, K., Nam, B., Shin, D., Kim, S., Jeong, Y., Kim, A., & Kim, Y. (2023). Deterioration of mental health in children and adolescents during the COVID-19 pandemic. *Journal of the Korean Academy of Child and Adolescent Psychiatry*, *34*(1), 21–29. https://doi.org/10.5765/jkacap.220041

Jones, S. E., Ethier, K. A., Hertz, M., DeGue, S., Le, V. D., Thornton, J., Lim, C., Dittus, P. J., & Geda, S. (2022). Mental health, suicidality, and connectedness among high school students during the COVID-19 pandemic – Adolescent behaviors and experiences survey, United States, January-June 2021. *MMWR Supplements*, *71*(3), 16–21. https://doi.org/10.15585/mmwr.su7103a3

Karasavva, V., Miller, C., Groves, N., Montiel, A., Canu, W., & Mikami, A. (2025). A double-edged hashtag: Evaluation of #ADHD-related TikTok content and its associations with perceptions of ADHD. *PLOS One*, *20*(3), e0319335. https://doi.org/10.1371/journal.pone.0319335

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society*, *18*(4), 625–630. https://doi.org/10.1017/s1355617712000240

Lockhart, J., & Satya-Murti, S. (2015). Symptom exaggeration and symptom validity testing in persons with medically unexplained neurologic presentations. *Neurology. Clinical Practice*, *5*(1), 17–24. https://doi.org/10.1212/CPJ.0000000000000092

Martin, E., Dourish, C. T., & Higgs, S. (2023). Interoceptive accuracy mediates the longitudinal relationship between attention deficit hyperactivity disorder (ADHD) inattentive symptoms and disordered eating in a community sample. *Physiology & Behavior*, *268*, 114220. https://doi.org/10.1016/j.physbeh.2023.114220

Martin, P. K., & Schroeder, R. W. (2020). Base rates of invalid test performance across clinical non-forensic contexts and settings. *Archives of Clinical Neuropsychology*, *35*(6), 717–725. https://doi.org/10.1093/arclin/acaa017

Martin, P. K., Schroeder, R. W., & Odland, A. P. (2015). Neuropsychologists' validity testing beliefs and practices: A survey of North American professionals. *The Clinical Neuropsychologist*, *29*(6), 741–776. https://doi.org/10.1080/13854046.2015.1087597

Merckelbach, H., Dandachi-FitzGerald, B., van Helvoort, D., Jelicic, M., & Otgaar, H. (2019). When patients overreport symptoms: More than just malingering. *Current Directions in Psychological Science*, *28*(3), 321–326. https://doi.org/10.1177/0963721419837681

Mogge, N. L., Lepage, J. S., Bell, T., & Ragatz, L. (2010). The negative distortion scale: A new PAI validity scale. *Journal of Forensic Psychiatry & Psychology*, *21*(1), 77–90. https://doi.org/14789940903174253

Montero-Marin, J., Hinze, V., Mansfield, K., Slaghekke, Y., Blakemore, S.-J., Byford, S., Dalgleish, T., Greenberg, M. T., Viner, R. M., Ukoumunne, O. C., Ford, T., & Kuyken, W. (2023). Young people's mental health changes, risk, and resilience during the COVID-19 pandemic. *JAMA Network Open*, *6*(9), e2335016. https://doi.org/10.1001/jamanetworkopen.2023.35016

Morey, L. C. (1991). *Personality Assessment Inventory professional manual*. Psychological Assessment Resources.

Morey, L. C. (1993). Validity and reliability of the Personality Assessment Inventory. *Journal of Personality Assessment*, *60*(1), 89–99. https://doi.org/10.1207/s15327752jpa6001_7

Morey, L. C. (1996). *An interpretive guide to the Personality Assessment Inventory*. Psychological Assessment.

Morey, L. C., & Lanier, V. W. (1998). Operating characteristics of six response distortion indicators for the Personality Assessment Inventory. *Assessment*, *5*(3), 203–214. https://doi.org/107319119800500301

Müller-Vahl, K. R., Pisarenko, A., Jakubovski, E., & Fremer, C. (2022). Stop that! It's not Tourette's but a new type of mass sociogenic illness. *Brain*, *145*(2), 476–480. https://doi.org/10.1093/brain/awab316

Musso, M. W., Hill, B. D., Barker, A. A., Pella, R. D., & Gouvier, W. D. (2016). Utility of the Personality Assessment Inventory for detecting malingered ADHD in college students. *Journal of Attention Disorders*, *20*(9), 763–774. https://doi.org/10.1177/1087054714548031

Odland, A. P., Lammy, A. B., Martin, P. K., Grote, C. L., & Mittenberg, W. (2015). Advanced administration and interpretation of multiple validity tests. *Psychological Injury and Law*, *8*(1), 46–63. https://doi.org/10.1007/s12207-015-9216-4

Porter, C. A., Mayanil, T., Gupta, T., & Horton, L. E. (2024). #DID: The role of social media in the presentation of dissociative symptoms in adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry*, *63*(2), 101–104. https://doi.org/10.1016/j.jaac.2023.03.021

Racine, N., McArthur, B. A., Cooke, J. E., Eirich, R., Zhu, J., & Madigan, S. (2021). Global prevalence of depressive and anxiety symptoms in children and adolescents during COVID-19: A meta-analysis. *JAMA Pediatrics*, *175*(11), 1142–1150. https://doi.org/10.1001/jamapediatrics.2021.2482

Renner, K. H., Enz, S., Friedel, H., Merzbacher, G., & Laux, L. (2008). Doing as if: The histrionic self-presentation style. *Journal of Research in Personality*, *42*(5), 1303–1322. https://doi.org/10.1016/j.jrp.2008.04.005

Roor, J. J., Dandachi-FitzGerald, B., & Ponds, R. W. (2016). A case of misdiagnosis of mild cognitive impairment: The utility of symptom validity testing in an outpatient memory clinic. *Applied Neuropsychology Adult*, *23*(3), 172–178. https://doi.org/10.1080/23279095.2015.1030018

Rosenfeld, B., Sands, S. A., & Van Gorp, W. G. (2000). Have we forgotten the base rate problem? Methodological issues in the detection of distortion. *Archives of Clinical Neuropsychology*, *15*(4), 349–359. https://doi.org/10.1016/S0887-6177(99)00025-6

Rouse, M. (2017). *Garbage in, garbage out*. Techopedia. https://www.techopedia.com/definition/3801/garbage-in-garbage-out-gigo

Schroeder, R. W., & Martin, P. K. (2022a). Validity assessment in clinical settings: How it differs from forensic settings and why it is important. *In* R. W. Schroeder & P. K. Martin (Eds.), *Validity assessment in clinical neuropsychological practice: Evaluating and managing noncredible performance* (pp. 3–10). The Guilford Press.

Schroeder, R. W., & Martin, P. K. (2022b). Explanations of performance validity test failure in clinical settings. *In* R. W. Schroeder & P. K. Martin (Eds.), *Validity assessment in clinical neuropsychological practice; evaluating and managing noncredible performance* (pp. 11–30). The Guilford Press.

Schroeder, R. W., & Martin, P. K. (2025). *Clinician's guide to validity assessment and management in neuropsychology*. Guilford Press.

Schwarz, A. (2016). *ADHD nation: Children, doctors, big pharma, and the making of an American epidemic*. Scribner/Simon & Schuster.

Shwartz, S. K., Roper, B. L., Arentsen, T. J., Crouse, E. M., & Adler, M. C. (2020). The behavior rating inventory of executive function–adult version is related to emotional distress, not executive dysfunction, in a veteran sample. *Archives of Clinical Neuropsychology*, *35*(6), 701–716. https://doi.org/10.1093/arclin/acaa024

Shmerling, R. (2022, January 18). Tics and TikTok: Can social media trigger illness? Harvard Health Publishing. https://www.health.harvard.edu/blog/tics-and-tiktok-can-social-media-trigger-illness-202201182670

Slick, D., Hopp, G., Strauss, E., & Thompson, G. B. (1997). *Victoria symptom validity test, version 1.0 professional manual*. Psychological Assessment Resources.

Smith, S. T., Cox, J., Mowle, E. N., & Edens, J. F. (2017). Intentional inattention: Detecting feigned attention-deficit/hyperactivity disorder on the Personality Assessment Inventory. *Psychological Assessment*, *29*(12), 1447–1457. https://doi.org/10.1037/pas0000435

Suhr, J. A., Buelow, M., & Riddle, T. (2011). Development of an infrequency index for the CAARS. *Journal of Psychoeducational Assessment*, *29*(2), 160–170. https://doi.org/10.1177/0734282910380190

Suhr, J. A., & Gunstad, J. (2002). "Diagnosis threat": The effect of negative expectations on cognitive performance in head injury. *Journal of Clinical and Experimental Neuropsychology*, *24*(4), 448–457. https://doi.org/10.1076/jcen.24.4.448.1039

Suhr, J. A., & Wei, C. (2017). Attention deficit/hyperactivity disorder as an illness identity: Implications for neuropsychological practice. *In* K. B. Boone (Ed.), *Neuropsychological evaluation of somatoform and other functional somatic conditions: Assessment primer* (pp. 251–273). Routledge/Taylor & Francis Group.

van der Heide, D., Boskovic, I., van Harten, P., & Merckelbach, H. (2020a). Psychosis as a confounder of symptom credibility testing in a transcultural sample. *The Journal of Forensic Psychiatry & Psychology*, *31*(1), 64–75. https://doi.org/10.1080/14789949.2019.1663899

van der Heide, D., Boskovic, I., van Harten, P., & Merckelbach, H. (2020b). Overlooking feigning behavior may result in potential harmful treatment interventions: Two case reports of undetected malingering. *Journal of Forensic Sciences*, *65*(4), 1371–1375. https://doi.org/10.1111/1556-4029.14320

Verma, S., & Sinha, S. K. (2025). How evidence-based is the "hashtag ADHD test" (#adhdtest). A cross-sectional content analysis of TikTok videos on attention-deficit/hyperactivity disorder (ADHD) screening. *Australasian Psychiatry*, *33*(1), 82–88. https://doi.org/10.1177/10398562241291956

West, R., Leskovec, J., & Potts, C. (2021). Misinformation in and about science. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(38), 1–8. https://doi.org/10.1073/pnas.1912444117

Yeung, A., Ng, E., & Abi-Jaoude, E. (2022). TikTok and attention-deficit/hyperactivity disorder: A cross-sectional study of social media content quality. *Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie*, *67*(12), 899–906. https://doi.org/10.1177/07067437221082854