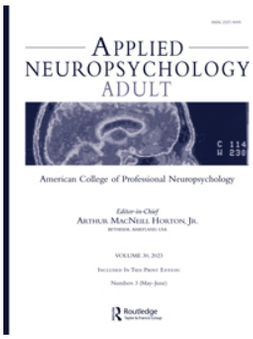

Applied Neuropsychology: Adult

ISSN: 2327-9095 (Print) 2327-9109 (Online) Journal homepage: www.tandfonline.com/journals/hapn21

Predictive value of performance validity testing and symptom validity testing in psychoeducational assessment

Allyson G. Harrison, A. Lynne Beal & Irene T. Armstrong

To cite this article: Allyson G. Harrison, A. Lynne Beal & Irene T. Armstrong (2023) Predictive value of performance validity testing and symptom validity testing in psychoeducational assessment, Applied Neuropsychology: Adult, 30:3, 315-329, DOI: 10.1080/23279095.2021.1943396

To link to this article: https://doi.org/10.1080/23279095.2021.1943396

Published online: 14 Jul 2021.

Submit your article to this journal

Article views: 724

View related articles

View Crossmark data

Citing articles: 9 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=hapn21

APPLIED NEUROPSYCHOLOGY: ADULT
2023, VOL. 30, NO. 3, 315–329
https://doi.org/10.1080/23279095.2021.1943396

Routledge
Taylor & Francis Group

Check for updates

Predictive value of performance validity testing and symptom validity testing in psychoeducational assessment

Allyson G. Harrison[a], A. Lynne Beal[b], and Irene T. Armstrong[a]

[a]Regional Assessment and Resource Centre, Queen's University, Kingston, Canada; [b]Private Practice, Toronto, Canada

ABSTRACT

Using archival data from 2463 psychoeducational assessments of postsecondary students we investigated whether failure on either symptom or performance validity tests (SVTs or PVTs) was associated with score differences on various cognitive, achievement, or executive functioning performance measures or on symptom report measures related to mental health or attention complaints. In total, 14.6% of students failed one or more PVT, 33.6% failed one or more SVT, and 41.6% failed at least one validity test. Individuals who failed SVTs tended to have the highest levels of self-reported symptoms relative to other groups but did not score worse on performance-based psychological tests. Those who failed PVTs scored worse on performance-based tests relative to other groups. Failure on at least one PVT and one SVT resulted in both performance and self-reported symptoms suggestive of greater impairment compared with those who passed all validity measures. Findings also highlight the need for domain-specific SVTs; failing ADHD SVTs was associated only with extreme reports of ADHD and executive functioning symptoms while failing mental health SVTs related only to extreme reports of mental health complaints. Results support using at least one PVT and one SVT in psychoeducational assessments to aid in diagnostic certainty, given the frequency of non-credible presentation in this population of postsecondary students.

KEYWORDS

Assessment/diagnosis; performance validity testing; symptom validity testing

Over the past decade, there has been an increased recognition that, regardless of the motivation driving the behavior, non-credible performance or non-credible symptom reporting by clients interferes with an accurate diagnosis of psychological or neurological conditions (Lippa, 2018; Martin & Schroeder, 2020; Sherman et al., 2020). Indeed, multiple studies have demonstrated how easily adults can falsely report symptoms of conditions such as attention deficit hyperactivity disorder (ADHD; Harrison et al., 2007; Jachimowicz & Geiselman, 2004; Jasinski et al., 2011; Quinn, 2003), traumatic brain injury (Lange et al., 2010) or post-traumatic stress disorder (Henry & Gornbein, 2020). Individuals can also feign functional impairments on performance measures with such accuracy that their performance profiles are indistinguishable from those with true neurological conditions such as a reading disability (RD) or ADHD (e.g., Frazier et al., 2008; Harrison et al., 2007, 2008; Lindstrom et al., 2011; Sollman & Berry, 2011).

As a result, a variety of stand-alone and embedded methods to identify non-credible performance or exaggerated symptom reporting have been investigated. The 2021 AACN consensus statement on validity assessment (Sweet et al., 2021) defines symptom validity tests (SVTs) as indicators of symptom exaggeration, and performance validity tests (PVTs) as measures used to identify diminished performance. The statement endorses the inclusion of multiple empirically-based PVTs and/or SVTs in any neuropsychological or psychoeducational evaluation in which there is the potential for secondary gain.

Given that clinical judgment alone is not an effective means to determine whether clients performed assessment tasks in a credible manner or were investing sufficient effort (e.g., Dandachi-Fitzgerald et al., 2017; Faust, Hart, & Guilmette, 1988; Faust, Hart, Guilmette, et al., 1988; Guilmette, 2013), such validity measures serve as objective warning signs to alert clinicians that obtained scores may not be entirely accurate due to negative response bias (Sweet et al., 2021).

While the current opinion is that clinicians should include more than one objective validity measure in any assessment (Larrabee, 2014; Odland et al., 2015; Sherman et al., 2020; Sweet et al., 2021), it is unclear whether clinicians should rely mainly on measures of symptom exaggeration, performance exaggeration, or both when conducting evaluations in which secondary gain is possible. Also unclear is whether specific validity measures may be more effective in detecting non-credible performance in a specific population of postsecondary students.

At the postsecondary education level and on high-stakes examinations, learning disabilities (LD) and ADHD are the disabilities for which most academic accommodations are sought (Harrison & Wolforth, 2012). Accommodation for

CONTACT Allyson G. Harrison harrisna@queensu.ca Regional Assessment and Resource Centre, Queen's University, B100, Macintosh Corry Hall, 68 University Avenue, Kingston K7L 3K1, Canada.

© 2021 Taylor & Francis Group, LLC

genuine neurological impairment is necessary in order to allow individuals with disabilities equal opportunity to participate. However, the many secondary gain advantages available to postsecondary students in Canada (see Harrison, 2017 for a complete list) offer possible incentives to feign educational disabilities. Extra test-taking time in particular is an accommodation that can provide significant academic advantages if given to students who are not truly disabled (e.g., Lewandowski et al., 2013; Mandinach et al., 2005). Previous studies suggest that anywhere from 15–47% of postsecondary students undergoing formal assessments for LD and/or ADHD may be feigning or exaggerating their impairments (Harrison & Edwards, 2010; Sullivan et al., 2007). Thus, it is essential that clinicians identify non-credible performance when evaluating students for these conditions.

PVTs were developed to identify those feigning symptoms of neurological disorders in medical-legal contexts (e.g., memory or attention), whereas SVTs were developed to evaluate exaggeration of either psychiatric or neurological symptoms. While position papers (e.g., Sherman et al., 2020; Sweet et al., 2021) provide general guidance regarding the need for validity testing in neuropsychological evaluations, no consensus exists currently regarding which validity measures, if any, should be included in psychoeducational assessments (Harrison, 2017). Furthermore, there may be limited convergence between the behaviors of those who feign on performance measures and those who overreport symptoms.

At present, there is no gold standard method to identify feigned ADHD or LD. Diagnosis of ADHD is often based on self-report alone (Nelson et al., 2019). However, in the ADHD literature in general there has been little cross-method convergence between neuropsychological performance and self-reported difficulties (Barkley, 2019), so PVTs may not capture exaggeration of ADHD symptoms. Conversely, while a diagnosis of LD is mainly performance-based (American Psychiatric Association, 2013), it is not known whether those feigning LD would also exaggerate on self-report symptom measures. Finally, it is unclear whether all PVTs and SVTs are equally sensitive to feigned impairment of LD or ADHD, or whether, as suggested by Osmon et al. (2006), validity tests specific to the type of impairment being claimed are required in order to increase sensitivity to different types of invalid test performance. While a few experimental SVTs specific to feigned ADHD have been investigated, (e.g., Becke et al., 2021; Harrison & Armstrong, 2016; Suhr et al., 2011), it is unclear whether such measures are more sensitive to self-report exaggeration than are other existing SVTs or if failure on these ADHD-specific measures is associated with extreme scores in any other assessment domains.

At present, little research exists to help clinicians determine which PVTs or SVTs are most helpful in the diagnostic evaluation of ADHD and LD. Research support for a higher-order dimension of generalized feigning has been reportedly weak (Merten & Merckelbach, 2013), suggesting that the overlap between classification using PVTs and SVTs may be limited. For instance, Copeland et al. (2016) examined the differential relationship of PVTs and SVTs with cognitive performance and self-reported symptom measures. While not directly evaluating ADHD or LD, they found that individuals who passed both a PVT and an SVT, as well as those who passed only a PVT, had better cognitive performance and self-reported fewer symptoms relative to those who failed both tests. They suggest that this finding supports the differential utility of PVTs and SVTs when assessing cognitive as opposed to self-reported symptoms. Similarly, Haggerty et al. (2007) found only modest or slight correlations between scores on the Victoria Symptom Validity Test (VSVT; Slick et al., 1997), a PVT, and feigning measures derived from self-report SVTs on the Personality Assessment Inventory (PAI; Morey, 1991), implying that, in general, individuals feigning psychiatric impairment do not always underperform on cognitive measures of exaggeration. Dandachi-FitzGerald and Merckelbach (2013) found that an SVT measuring general malingered symptoms outperformed a PVT in correctly identifying simulated malingerers, suggesting that memory-oriented PVTs may be useful in detecting malingered memory problems, but not other types of feigned psychopathology. Finally, in a sample of adults assessed for ADHD, White et al. (2020) found that 18% failed PVTs or SVTs or both. Failed PVTs were related to worse scores on WAIS-III working memory subtests, while failed SVTs were related to higher scores on self-reported ADHD symptoms.

By contrast, some research (e.g., Green et al., 2001; Green & Flaro, 2021) suggests that failure on a PVT presumably assessing one cognitive domain is associated with lower scores on neuropsychological tests in both the same and different areas of functioning. In other words, poor performance on a PVT predicted generalized poor performance on all other neuropsychological tests administered. These findings would suggest that individuals feigning cognitive deficits do so across multiple cognitive domains in a somewhat indiscriminate manner. Supporting the view of indiscriminate broad-based feigning, Lange et al. (2010) found that individuals evaluated for mild traumatic brain injury who failed the Test of Memory Malingering (TOMM; Tombaugh, 1996) performed more poorly on cognitive measures than did those who passed the TOMM. The failing individuals were also more likely to fail two other SVTs. Likewise, Lippa et al. (2014) found that individuals who failed the Word Memory Test (WMT, Green, 2005) were more likely to also fail an SVT such as the Fake Bad Scale (FBS; Lees-Haley et al., 1991).

Additionally, Shura et al. (2017) studied the classification of feigned ADHD by comparing individuals identified as feigning using the TOMM (trial 1) versus those identified using various validity scales from the Minnesota Multiphasic Personality Inventory-2 Restructured Form (MMPI-2 RF; Ben-Porath & Tellegen, 2011). While they found that significantly more of their clinical sample failed a self-report validity measure from the MMPI-2 RF (44.7%) than the TOMM (19.3%), those who overreported symptoms on the MMPI-2 RF did not endorse significantly higher levels of ADHD symptoms on self-report than did credible subjects. By

contrast, those who failed the PVT reported significantly higher ADHD symptoms than did those who passed an SVT. In other words, as has been suggested by Marshall et al. (2010), using PVTs that appear to include measures of attention may work well to also identify individuals feigning ADHD symptoms. More recently, Spenceley et al. (2020) also found that the working memory and processing speed subtests of Woodcock Johnson-IV (Schrank et al., 2014) may be used to detect over half of those potentially simulating ADHD while still maintaining at least 90% specificity.

In a recent meta-analysis, Wallace et al. (2019) found that stand-alone PVTs produced larger effect sizes in discriminating between actual versus feigned ADHD than did SVTs. Hence, despite the fact that ADHD is diagnosed mainly by self-report, PVTs appeared better able to identify feigned symptom behavior compared with existing SVTs. The authors note, however, that SVTs in general measure exaggeration of psychiatric symptoms, whereas PVTs, in general, were developed to identify feigned brain injury symptoms, the latter of which may seem more consistent with expected ADHD symptoms.

Finally, Sherer et al. (2020) suggest that the overlap between PVT failure and symptom overreporting may be even more complex. Conducting a cluster analysis of the scores returned by patients with a documented traumatic brain injury who failed a PVT, they identified three separate groups: one who performed poorly on memory tests but did not report excessive symptoms; one who showed impaired memory and processing speed as well as concerns about cognitive functioning; and a third who had both cognitive complaints and excessive self-reported symptoms. Interestingly, the third group scored worse on the PVT relative to the other two groups. In other words, PVT failure may sometimes, but not always, be associated with both symptom overreporting as well as performance deficits.

Nevertheless, we found no published studies evaluating whether postsecondary students who fail only a PVT, only an SVT, or both differ in their obtained scores in a psychoeducational assessment that includes both self-report questionnaires and performance-based tests. The present study, therefore, examined this question. Using an archival database of scores obtained from psychoeducational assessments conducted on postsecondary students over ten years, we investigated whether differences in performance or symptom ratings existed between those who failed no validity tests, one or more PVTs, one or more SVTs, or a combination of failed SVTs and PVTs.

Additionally, we investigated whether failure on SVTs specific to ADHD overreporting was associated with a different pattern of symptom ratings than was a failure on an SVT specific to exaggeration of mental health (MH) symptoms.

Finally, we sought to establish base rates and demographic characteristics of failed PVTs or SVTs or both, in a large sample of postsecondary students who had completed a psychoeducational assessment to support their request for academic accommodations.

Method

Participants

Participants for the current study were drawn from a large database of community college and four-year university students who had completed a psychoeducational assessment through a university-based regional assessment center between 2009 and 2019. All were referred to investigate reported problems with learning, attention, or post-concussion-related impairments, and all were seeking a diagnosis to allow access to disability supports and services. All students had agreed during the informed consent process to allow their de-identified information to be included in this research database. We included data from 2463 students (58.0% female) who had been given at least one PVT and one SVT as part of their assessment. The average age was 21.8 years (SD = 5.9, range 16–57 years). All students were high school graduates or equivalent, with their college or university program in progress. Demographic information on race, ethnicity, and socioeconomic status was not entered into the database. Since students completed their assessments only if the Psychologist determined they were fit to participate, no students were excluded from the study on the basis of severity of reported symptoms, medications, or other conditions.

Participants were categorized into four groups based on their scores on PVTs and SVTs. The Pass All group had passed all validity tests given. The Fail PVT Only group had failed one or more PVTs but passed all SVTs. The Fail SVT Only had failed one or more SVTs but passed all PVTs. The Fail PVT&SVT group had a combination of at least one PVT failed and at least one SVT failed. A further analysis categorized students based on the type of failed SVT: Fail ADHD SVT Only, Fail Mental Health (MH) SVT Only, or Fail ADHD&MH SVT (with failures on both a SVT for ADHD- and a MH-specific SVT).

Materials

Clinicians administered a flexible battery of individually administered tests widely used in clinics and research, depending on the referral question and clinician's preference. Each battery assessed the core abilities and symptoms (listed below) and was reported as either subtest and/or composite scores.

For most students, the intellectual ability was tested with the Wechsler Adult Intelligence Scale Fourth Edition (WAIS-IV; Wechsler, 2008). All 10 core subtests yielded four index scores (Verbal Comprehension, Perceptual Reasoning, Working Memory, Processing Speed), Full-Scale IQ (FSIQ), and General Ability Index (GAI). For some students, Fluid Reasoning and Processing Speed subtests from the Woodcock-Johnson III (WJ-III; Woodcock et al., 2001) or the Woodcock-Johnson IV (WJ-IV; Schrank et al., 2014) were also employed.

Academic fluency in reading was tested with one of more of the WJ-III or WJ-IV Tests of Achievement, the Wechsler Individual Achievement Test III (WIAT III; Wechsler,

Table 1. Validity tests, cut score for failure, failure rate and reference for performance and symptom validity tests used.

Validity test	*n*	Failure cut score	Failure rate	Reference
Word Memory Test	1810	≤82.5	11.4%	Green, 2005
Medical Symptom Validity Test	648	≤85	8.8%	Green, 2004
Victoria Symptom Validity Test	465	Hard Items ≤ 19	22.4%	Frazier et al., 2008
Symptom Exaggeration Index (TOVA)	328	≥3	3.0%	Greenberg et al., 2007
Test of Memory Malingering	110	≤45	13.6%	Tombaugh, 1996
The b test	44	≥82	18.2%	Roberson et al., 2013
PAI Validity Scales				Morey, 1991
Inconsistency ICN	1944	≥73	2.3%	
Infrequency INF	1942	≥75	2.8%	
Negative Impression Management NIM	1942	≥92	2.5%	
Roger's Discriminant Function	1944	> 0.12367	27.6%	Rogers et al., 1992
Malingering Index	1944	≥3	21.2%	Morey, 1996
Behavior Assessment System for Children – 2 Self-Report – College				Reynolds & Kamphaus, 2004
F Validity Index (Fake Bad)	260	≥5	1.9%	
V Validity Index (Endorse nonsense items)	256	≥5	0.8%	
Consistency	259	≥17	1.9%	
Response Pattern Raw (Attention to content)	259	≥130	1.5%	
Conners' Adult ADHD Rating Scale (CAARS)				
CAARS Infrequency Index	1032	≥21	7.6%	Suhr et al., 2011
Exaggeration Index	980	≥3	7.5%	Harrison & Armstrong, 2016
Dissimulation	966	≥20	17.5%	Harrison & Armstrong, 2016

2009), or the Nelson Denny Reading Test (NDRT; Brown et al., 1993). Fluency in written language and mathematics was tested with the WJ-III, WJ-IV, or the WIAT III. Executive functions were evaluated using the Delis Kaplan Executive Function System (D-KEFS, Delis et al., 2001). Self-reports of executive functions were obtained on the Behavior Rating Inventory of Executive Function—Adult (BRIEF-A; Roth et al., 2005). Sustained and divided attention and concentration were evaluated using the Test of Variables of Attention (TOVA; Greenberg et al., 2007), a computer-based continuous performance test. Functional impairment was measured through self-report on the Weiss Functional Impairment Rating Scale (WFIRS; Canadian ADHD Resource Alliance, 2017). Specific subtests used for all of these measures are reported in the tables of results.

All assessments included at least one free-standing PVT and at least one embedded SVT. The PVTs selected have all been used in studies of learning disabilities and ADHD. The embedded SVTs employed have been used in the research literature to identify either exaggeration of mental health and/or ADHD symptoms. Table 1 shows the PVTs and SVTs used, the cut scores for failure (using cut scores either recommended for use in this population as reported in published research, or employing those specified in the test manuals), and the failure rate for each. Only PVTs given to more than 45 students were included. The PVTs were chosen from: the Word Memory Test (WMT; Green, 2005); the Medical Symptom Validity Test (MSVT; Green, 2004); the Victoria Symptom Validity Test (VSVT; Slick et al., 1997); the Symptom Exaggeration Index from the TOVA (SEI; Greenberg et al., 2007); the b Test (Boone et al., 2000), and the Test of Memory Malingering (TOMM, Tombaugh, 1996). Failure of one or more of the first three subtests of the WMT or MSVT was counted as a failed PVT. Only 3% of the total sample (85 people) were given more than two PVTs. Of those, the majority (75) were given three, with only 18 failing one of the three tests given. Nine were given four, and only one person failed only one out of four PVTs given. The maximum number given was six (one person), and that person failed two.

The SVTs were taken from the Personality Assessment Inventory validity scales (all but the positive impression management scale; PAI; Morey, 1991), the Behavior Assessment System for Children-2 Self-Report—College validity scales (BASC-2; Reynolds & Kamphaus, 2004), and three embedded SVTs created to identify overreporting of ADHD symptoms for the Conners Adult ADHD Rating Scale-Self-Report (CAARS; Conners et al., 1999): the CAARS Infrequency Index (CII; Suhr et al., 2011); and the Exaggeration and Dissimulation Indices (EI and DISS; Harrison & Armstrong, 2016). Failing any of the CII, EI, and/or DISS on the CAARS, or failing any one of the relevant validity scales on the PAI or BASC-2 was counted separately as a failure. As a result, a student could fail more than one SVT within a given self-report battery. This was done because the test manuals for the PAI and BASC-2 identify that the various validity scales measure different types of symptom exaggeration and are often computed based on different response patterns. Likewise, the three embedded CAARS validity scales are calculated based on different item endorsement patterns.

Procedure

Students were provided with a psychoeducational assessment either to update a previous assessment or to investigate the cause for currently reported academic difficulties. Either a registered psychologist or a supervised graduate student completed the assessments. Students were informed that their effort would be tested throughout the day, were advised to work as quickly and as accurately as possible on all timed tests, as per individual test instructions, and encouraged to put forth a good effort. All research procedures were approved by Queen's University's General Research Ethics Board.

Table 2. The number and percentage of students who failed performance validity tests or symptom validity tests or both.

	Number of SVTs failed					
Number of PVTs failed	0	1	2	3	4 or more	Total
0 … *n*	1438	411	191	51	14	2104
%	58.4	16.7	7.8	2.1	0.5	85.4
1 … *n*	145	61	44	12	4	266
%	5.9	2.5	1.8	0.5	0.2	10.8
2 … *n*	50	19	14	0	3	86
%	2.0	0.8	0.6	0	0.1	3.5
3 or more … *n*	3	2	0	1	1	7
%	0.1	0.1	0	0	0	0.3
Total … *n*	1636	493	249	64	22	2464
%	66.4	20.0	10.1	2.6	0.9	100.0

Statistical analyses

Analysis of variance (ANOVA) was used to assess score differences as a function of PVT/SVT failures, followed by pairwise comparisons using Tukey's b post hoc test, or Hochberg's GT2 when group *n*'s differed. As most of the analyses yielded probability values < 0.001 no correction was made for multiple tests. Effect sizes are reported as Cohen's d for significant findings. Cohen defined an effect size of d = 0.2 as small, of d = 0.5 as medium, and of d = 0.8 as large.

Results

Table 1 shows the failure rates on PVTs and SVTs, as numbers of students, as well as the percentage of the total who took the test. Failure rates for PVTs ranged from 3% to 22.4% and for SVTs ranged from 0.8% to 27.6%.

Table 2 shows the cumulative percentage of failed validity tests by type of validity test, with each percentage reflecting a portion of the grand total, and not of any single column or row. Note that 58.4% of students did not fail any validity tests, while 33.6% of the total sample failed one or more SVT, 14.6% failed one or more PVT, and 41.6% failed one or more validity tests of any type.

There were more women in all groups. Within each gender, the percentage who passed all validity tests was larger for men (61.0%) than for women (56.4%), $\chi^2(3) = 9.65$, $p = 0.022$. Within each of our validity groups, the smallest percentage of women was found in the Pass All group (56.2%) versus the percentage of women in any "fail" group (59.1% Fail PVT Only; 59.4% Fail SVT Only; 68.3% Fail PVT&SVT). The groups did not differ in age.

The remaining results are organized by the type of test, with Tables 3–6 showing results for ADHD questionnaires, mental health questionnaires, executive functioning measures, and finally for scores on tests of academic achievement, cognitive abilities, and processing speed, respectively. For each test, mean scores and SDs are shown for each group, with results of each ANOVA shown as p values and Cohen's d. Superscripts identify significant findings from posthoc comparisons.

Impact of validity scale failure on ratings of ADHD symptoms

Table 3 shows ratings given by all groups on self-report inventories, including both self-reports and observer-reports of ADHD symptoms. Starting with the CAARS Self-Report, the Fail PVT&SVT group gave the highest ratings on all subscales except on Problems with Self Concept (where they were tied with the Fail SVT Only group). The Fail SVT Only group returned higher ratings than the Pass All group on all subscales. In addition, both the Fail PVT Only and the Fail SVT Only groups gave higher ratings than the Pass All group for Inattention/Memory Problems, DSM-Inattentive Symptoms, DSM-Hyperactive-Impulsive Symptoms, and Total ADHD Symptoms. Effect sizes were large for Impulsivity/Emotional Lability and the ADHD Index, and medium for all other subscales. Yet, only on the scales of DSM-Inattentive Symptoms and Total ADHD Symptoms were mean ratings of the Fail PVT&SVT group suggestive of significant difficulty (i.e., above the 98th percentile).

Similarly, on the CAARS observer ratings, the Fail PVT&SVT group provided the most impaired ratings on all subscales except Inattention. The Fail SVT Only group obtained equally high ratings on Impulsive/Emotional Lability, Problems with Self Concept, and the ADHD Index. Effect sizes were medium to small. None of these mean ratings, however, was above the 98th percentile.

Impact of validity scale failure on ratings of mental health symptoms and executive functions

Table 4 presents scores from self-report inventories of mental health symptoms. For all PAI indices except for Dominance, the Fail PVT&SVT group reported the highest mean symptoms, followed by the Fail SVT Only group. Effect sizes were large for Somatic Complaints, Anxiety, Anxiety-Related Disorders, Depression, Paranoia, Schizotypal, Borderline Features, and Suicidal Ideation. Effect sizes were medium for Mania, Anti-social Behavior, Aggression, Drug Problems, Stress, Nonsupport, and Treatment Rejection but small for Alcohol Problems and Warmth. The Fail PVT group was indistinguishable from the Pass All group on all scales except Anxiety and Anxiety-Related Disorders; however, they were still not as high as the Fail SVT groups.

On the BASC-2 Self Report, the Fail PVT&SVT group and the Fail SVT group reported the most symptoms on almost all scales. Large effect sizes were found on the majority of the scales; the exceptions were Alcohol Abuse and Relationship with Parents (medium) and Interpersonal Relationships (small). In addition, the Fail PVT&SVT group also reported significantly higher symptoms than the Pass All on all but two scales (Alcohol Abuse and Self-Esteem). Ratings by the Fail PVT Only group did not differ from the Pass All group on any of these symptom scales, with mean ratings in the average range.

Self-reported executive functioning across the four validity groups may be seen in Table 5. On the BRIEF Self-

Table 3. Means, SD, ANOVA significance and effect size with Tukey post hoc pairwise comparisons of self report and observer report ADHD inventories as a function of group.

	Pass all		FAIL PVT only		FAIL SVT only		FAIL PVT/SVT			Effect
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	*p*	d
Conners' Adult ADHD Rating Scale (CAARS) Self, $n = 1960$										
Inattention	56.6[a]	10.9	60.1[b]	11.5	62.2[b]	12.3	65.5[c]	12.6	<0.001	0.74
Hyperactive	51.1[a]	10.3	53.1[ab]	11.2	55.8[b]	11.2	59.3[c]	12.8	<0.001	0.68
Impulsive/emotional lability	48.7[a]	10.1	51.3[a]	11.9	56.1[b]	12.5	60.2[c]	14.0	<0.001	0.99
Problems with self concept	53.4[a]	10.9	54.8[a]	11.9	59.5[b]	11.5	60.7[b]	11.0	<0.001	0.75
DSM inattentive symptoms	63.9[a]	13.3	67.9[b]	13.0	70.5[b]	14.3	74.3[c]	14.5	<0.001	0.73
DSM hyperactive symptoms	52.3[a]	12.0	56.3[b]	13.9	58.9[b]	13.7	63.5[c]	15.7	<0.001	0.79
Total ADHD symptoms	60.1[a]	13.1	64.4[b]	13.8	67.7[b]	14.6	72.5[c]	15.3	<0.001	0.76
ADHD Index	53.7[a]	9.9	56.3[a]	11.2	61.2[b]	11.4	64.3[c]	12.3	<0.001	1.02
Inconsistency	5.3	2.3	5.4	2.3	5.9	2.6	5.9	2.2	<0.001	0.30
CAARS observer, $n = 1486$										
Inattention	60.6[ab]	12.6	60.0[a]	12.3	62.8[ab]	13.4	64.1[b]	13.5	0.003	0.27
Hyperactive	50.4[a]	11.4	51.7[a]	12.7	53.6[a]	12.3	57.2[b]	13.1	<0.001	0.46
Impulsive/emotional lability	51.1[a]	10.6	51.3[a]	10.5	55.5[b]	12.1	58.7[b]	12.7	<0.001	0.63
Problems with self concept	57.2[a]	11.0	57.3[a]	11.6	60. 7[b]	12.2	61.6[b]	12.3	<0.001	0.43
DSM inattentive symptoms	59.5[a]	11.2	59.7[a]	10.9	62.3[ab]	12.2	64.5[b]	11.8	<0.001	0.38
DSM hyperactive symptoms	50.8[a]	11.4	52.3[ab]	13.1	54.9[b]	12.5	59.3[c]	14.1	<0.001	0.58
Total ADHD symptoms	56.5[a]	11.1	57.4[ab]	11.9	60.1[b]	12.4	63.8[c]	13.3	<0.001	0.52
ADHD Index	56.7[a]	11.1	57.2[a]	11.1	61.2[b]	12.5	63.6[b]	13.4	<0.001	0.58
Inconsistency	5.0	2.5	4.9	2.4	5.2	2.5	5.4	2.5	0.207	

Note. Superscripts with matching letters are not statistically different.
PVT: Performance validity test; SVT: Symptom validity test; DSM: Diagnostic and Statistical Manual of Mental Disorders.

Report Adult, the Fail PVT&SVT group reported significantly more challenges than the Pass All group on all subscales except on Task Monitoring. The Fail SVT Only group gave higher ratings than the Pass All group on Shift, Emotional Control, Self-Monitor, Behavior Regulation Index, and Global Executive Composite (GEC). Effect sizes were large for Emotional Control and otherwise medium to small.

As shown in Table 5, the overall effect size for the BAI was medium, but with no significant pairwise differences among the group. On the BDI-II, the Pass All and Fail PVT Only groups gave the lowest mean ratings of depression, with a large effect size. Reports of functional impairment on the WFIRS showed no significant differences among the groups.

Impact of validity scales on achievement and cognitive scores

Table 6 shows that most academic achievement scores did not differ across groups. On the reading tests, mean scores were predominantly in the average range, with variability by group typically one to one and a half standard deviations. Significant group differences were found only on the Woodcock-Johnson IV Sentence Reading Fluency, the WIAT-III Reading Comprehension, and the WIAT-III Reading Comprehension and Fluency test, with small to medium effect sizes. On the WIAT-III Reading Comprehension, and Comprehension and Fluency, the two PVT failure groups scored lowest, whereas on WJ-IV Sentence Reading Fluency it was the Fail PVT Only group with the lowest score. However, this pattern was not seen on other tests of reading comprehension, fluency, or basic word reading.

On the tests of written language, mean scores were all in the normal range. There was a small effect size on WJ IV Sentence Writing Fluency, but with no group scoring significantly worse than the Pass All group. No group differences were seen on the other tests of writing speed or expression.

On tests of mathematics, only WIAT III Math showed differences among groups, with a small effect size. The Fail PVT&SVT group scored lower than the Pass All group.

On tests of cognitive abilities, the lowest mean scores were shared by the Fail PVT Only and the Fail PVT&SVT groups on WAIS IV FSIQ, GAI, Verbal Comprehension, Perceptual Reasoning, and Working Memory, each with a small effect size. On WJ III Processing Speed there was a small effect size with the Fail PVT Only group scoring lower than the Pass All Group. No effect emerged on the WJ IV Cognitive Perceptual Speed.

On measures of simple speed on the D-KEFS, there were no group effects on Color Naming, Word Reading, Letter Fluency, or Category Fluency. There was a small effect size for Inhibition/Switching, with the Fail PVT Only group scoring lower than all other groups including Pass All. On the Trails Test, no group effects were found on Visual Scanning and Letter Sequencing. However, there was a small effect size for Number Sequencing and Number Letter Sequencing, where the Fail PVT Only group scored worse than the Pass All and the Fail SVT groups.

Influence of failure on specialized types of SVTs

To test whether failure on a specific type of SVT resulted in different patterns of symptom reporting, we compared failure on the ADHD-specific SVTs on the CAARS to failure on mental-health-related SVTs on the PAI and BASC-2. Although too complex to show in a quantitative table, qualitative descriptions of these findings are shown in Table A1. Those who failed only an ADHD-specific SVT (Fail ADHD SVT group) shared the highest ratings on all of the CAARS symptom subscales with those who failed both an ADHD-

Table 4. Means, SD, ANOVA significance and effect size with Tukey post hoc pairwise comparisons of self report mental health inventories as a function of group.

	Pass all		FAIL PVT only		FAIL SVT only		FAIL PVT/SVT			Effect
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	*p*	d
Personal Assessment Inventory (PAI), $n = 1940$										
Somatic complaints	51.8[a]	7.5	53.6[a]	8.9	57.2[b]	10.6	61.1[c]	11.8	<0.001	0.96
Anxiety	58.9[a]	10.4	61.9[b]	10.8	65.9[c]	12.7	70.1[d]	12.8	<0.001	0.95
Anxiety related disorder	53.9[a]	8.2	56.3[b]	8.9	59.8[c]	10.5	62.1[c]	10.4	<0.001	0.94
Depression	56.3[a]	9.3	58.7[a]	10.6	64.0[b]	13.1	67.1[c]	12.9	<0.001	1.04
Mania	51.9[a]	8.0	52.3[a]	8.2	55.6[b]	8.9	57.3[b]	9.8	<0.001	0.67
Paranoia	52.1[a]	8.0	54.1[a]	9.1	58.1[b]	10.9	60.8[c]	9.7	<0.001	0.97
Schizotypal	54.2[a]	7.3	56.3[a]	8.7	61.2[b]	10.9	64.4[c]	11.2	<0.001	1.15
Borderline Features	54.7[a]	8.1	56.1[a]	8.9	61.3[b]	10.6	63.6[c]	11.1	<0.001	1.08
Antisocial Behavior	52.1[a]	8.3	51.9[a]	8.7	55.6[b]	9.8	57.6[b]	11.0	<0.001	0.61
Aggression	48.6[a]	8.6	49.6[a]	9.0	52.5[b]	11.0	55.1[c]	12.2	<0.001	0.64
Alcohol, $n = 1180$	47.7[a]	7.9	46.6[a]	6.1	49.0[a]	9.1	51.9[b]	12.0	<0.001	0.42
Drugs $n = 1178$	50.1[ab]	9.3	49.5[a]	7.8	53.0[b]	10.9	56.9[c]	12.7	<0.001	0.59
Suicide Ideation	51.0[a]	9.8	52.9[a]	11.4	58.7[b]	16.3	58.4[b]	14.8	<0.001	0.82
Stress	52.9[a]	9.1	54.4[a]	12.1	57.6[b]	11.1	60.2[c]	10.9	<0.001	0.71
Nonsupport	52.6[a]	10.9	55.4[a]	12.4	58.8[b]	14.0	61.4[b]	13.7	<0.001	0.75
Treatment rejection	48.5[a]	9.9	48.9[a]	10.4	43.3[b]	10.8	41.2[b]	11.0	<0.001	0.77
Dominance	46.1	10.3	44.9	12.3	44.7	11.6	46.2	12.6	0.087	
Warmth	48.6[a]	10.5	47.8[a]	11.0	46.1[a]	11.4	43.2[b]	12.1	<0.001	0.41
Behavior Assessment System for Children 2nd edition (BASC2) Self report, $n = 267$										
Atypical	50.3[a]	9.1	50.3[a]	11.7	63.3[b]	19.3	76.2[c]	14.8	<0.001	1.55
Locus of control	49.8[a]	9.6	49.5[a]	10.7	61.3[b]	12.4	66.5[b]	10.8	<0.001	1.33
Social stress	50.1[a]	10.8	46.9[a]	10.8	60.8[b]	15.0	66.5[b]	14.1	<0.001	1.22
Anxiety	52.8[a]	10.7	54.1[a]	12.2	63.9[b]	12.5	65.6[b]	11.4	<0.001	1.05
Depression	50.2[a]	10.4	49.5[a]	8.9	62.0[b]	14.5	68.8[b]	20.0	<0.001	1.29
Sense of inadequacy	56.3[a]	11.6	59.2[a]	12.3	63.5[a]	11.8	75.6[b]	15.2	<0.001	1.00
Somatic	51.2[a]	10.0	53.3[ab]	9.2	60.5[bc]	14.6	62.1[c]	13.2	<0.001	0.93
Internalizing problem	51.9[a]	10.0	52.5[a]	10.3	65.6[b]	14.2	73.1[b]	12.7	<0.001	1.53
Attention	55.6[a]	11.2	58.4[a]	11.4	62.6[ab]	11.6	70.0[b]	6.6	<0.001	0.97
Hyperactivity	52.5[a]	13.1	53.5[a]	11.4	57.3[a]	13.8	70.3[b]	10.8	<0.001	0.79
Inattention hyperactivity	53.5[a]	12.0	56.6[a]	11.3	61.2[a]	13.5	72.8[b]	7.9	<0.001	1.01
Emotional symptoms	52.9[a]	10.2	53.6[a]	10.9	64.1[b]	13.4	65.2[b]	18.0	<0.001	1.06
Sensation seeking	45.3[a]	9.4	44.2[a]	10.5	46.9[a]	11.4	55.6[b]	11.4	0.009	0.60
Alcohol abuse	46.7[ab]	7.5	46.3[a]	5.5	50.5[ab]	10.4	56.0[b]	16.7	0.001	0.77
School mal-adjustment	49.0[a]	8.8	48.4[a]	8.7	52.8[a]	10.3	67.8[b]	10.3	<0.001	0.88
Relation with parents	49.7	10.6	48.9	10.4	42.5	13.4	43.6	8.2	0.003	0.66
Interpersonal relations	48.1	10.1	49.0	10.7	43.8	12.5	42.6	6.2	0.055	0.48
Self esteem	49.5[b]	10.1	59.0[b]	10.3	39.3[a]	11.7	43.8[ab]	11.9	<0.001	0.95
Self reliance	46.2	9.4	43.8	11.7	43.3	11.8	42.1	13.1	0.228	
Personal adjustment	47.9[b]	9.0	46.8 [ab]	10.2	39.9 [a]	12.4	41.4 [ab]	10.1	<0.001	0.81
Test anxiety	55.8[a]	10.4	58.8[ab]	11.0	63.2[ab]	9.9	67.4[b]	8.4	<0.001	1.41
Anger control	49.1[a]	9.6	45.9[a]	10.0	61.7[b]	13.4	65.3[b]	6.9	<0.001	1.33
Mania	52.8[a]	12.3	56.1[a]	14.3	62.5[a]	13.1	74.0[b]	4.5	<0.001	1.01
Ego strength	48.3[b]	10.0	48.6[b]	10.1	39.9[ab]	12.6	37.6[a]	12.2	<0.001	0.84

Note. Superscripts with matching letters are not statistically different.
PVT: Performance validity test; SVT: Symptom validity test.

specific and a mental health-specific SVT (the Fail ADHD&MH SVTs group), and also with the Fail PVT&SVT group. These results held for both the Self- and Observer-Reports of the CAARS. By contrast, ratings by those who failed only one or more mental health SVTs but passed the ADHD SVTs (Fail MH SVT) were no different from the Pass All group on all the CAARS scales.

Those who failed a mental health SVT reported higher symptoms of depression, suicidal ideation, and dominance on the PAI, and more symptoms of depression on the BDI II compared to those who only failed an ADHD-specific SVT. The scores for the latter group were comparable to the Pass All group on all PAI indices except for anxiety, mania, schizotypal, borderline features, antisocial, and treatment rejection, on which they returned the highest scores. Both SVT groups were high on paranoia. Those who failed both an MH and an ADHD SVT had significantly higher scores than the Pass All group on almost all PAI scales (see Table A1). On the BDI II, only those who failed an MH SVT had significantly higher scores than Pass All group, whereas the Fail ADHD SVT group was equal to the Pass All group. No differences emerged on the WFIRS or BAI. Too few BASC-2 scores were available for use in this subgroup analysis.

On cognitive and achievement measures, the Fail MH SVT group had significantly lower GAI scores compared to the Fail ADHD SVT group, but comparable scores to both the Fail PVT Only and the Fail PVT&SVT group. On the BRIEF, differences were also evident in the response patterns of these two subgroups: the ADHD-related SVT failure group complained of significantly more problems with inhibition, behavioral regulation, metacognition, and the global executive composite relative to the Fail Only MH SVT group. The latter was no different than the Pass All group. No between-group differences emerged on the D-KEFS.

One group of students failed both an ADHD- and MH-related SVT (Fail ADHD&MH SVTs). Table A1 shows that, on all of the CAARS ADHD symptom scales and for DISS,

Table 5. Means, SD, and ANOVA significance and effect size with Tukey post hoc pairwise comparisons of self report executive function inventories as a function of group.

	Pass all		FAIL PVT only		FAIL SVT only		FAIL PVT/SVT			Effect
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	*p*	d
BRIEF self report, $n = 736$										
Inhibit	58.1[a]	11.4	59.4[a]	12.2	62.3[ab]	11.9	65.6[b]	11.9	<0.001	0.60
Shift	60.2[a]	11.1	61.9[ab]	11.6	65.6[bc]	13.0	66.7[c]	11.8	<0.001	0.64
Emotion control	52.1[a]	11.0	55.8[ab]	12.2	59.3[bc]	12.8	60.9[c]	13.2	<0.001	0.85
Self monitor	52.3[a]	11.1	53.8[a]	12.4	58.6[b]	12.7	59.7[b]	13.3	<0.001	0.73
Initiate	60.6[a]	11.4	63.7[ab]	13.1	65.4[ab]	12.1	67.2[b]	13.6	<0.001	0.58
WM	66.6[a]	12.6	69.9[a]	11.9	71.0[ab]	12.6	74.9[b]	12.3	<0.001	0.61
Planning	62.3[a]	11.9	65.0[ab]	12.3	65.2[ab]	12.5	68.2[b]	12.4	0.001	0.43
Task monitor	62.6	12.3	65.7	13.2	65.6	13.4	67.0	12.8	0.007	0.37
Org. materials	55.0[a]	11.7	59.8[b]	12.4	56.7[ab]	13.3	60.7[b]	12.5	<0.001	0.44
BRI	56.1[a]	10.6	58.8[a]	11.4	63.1[b]	12.0	65.2[b]	12.8	<0.001	0.89
Metacognitive Index	63.5[a]	12.0	66.1[ab]	11.6	66.9[ab]	12.6	70.5[b]	12.3	<0.001	0.51
GEC	61.1[a]	11.1	64.2[ab]	11.6	66.4[bc]	12.3	69.4[c]	12.6	<0.001	0.71
Beck Anxiety Inventory (BAI), $n = 240$										
	18.5	9.1	24.0	10.8	25.2	12.6	21.7	13.0	<0.001	0.79
Beck Depression Inventory (BDI-II), $n = 260$										
	17.0[a]	9.6	22.4[ab]	12.7	24.1[b]	11.6	24.4[b]	11.6	<0.001	0.87
WEISS Functional Impairment Rating Scale, $n = 151$										
Mean total	0.74	0.37	0.85	0.46	0.94	0.52	0.81	0.58	0.106	

Note. Superscripts with matching letters are not statistically different.
PVT: Performance validity test; SVT: Symptom validity test; BRIEF: Behavior Rating Inventory of Executive Function; WM: Working memory; Org. materials: Organization of materials; BRI: Behavior Regulation Index; GEC: Global Executive Composite.

Exaggeration Index, and the Infrequency Index (CII), these students obtained the worst (most extreme) scores, along with the Fail Only ADHD SVT group and those who failed both an SVT and a PVT (Fail PVT&SVT). On the mental health symptom scales, these dual-SVT-failing students returned the worst (most extreme) scores on almost all of the PAI scales, and on the BDI-II. They also shared or had the highest scores on most of the BRIEF scales.

Discussion

This large sample study of postsecondary students in Ontario shows that failure on symptom validity tests or performance validity tests is not uncommon. Only 57.8% of our sample did not fail a single SVT or PVT. This failure rate among postsecondary students seeking academic accommodations and/or medications underlines the importance of including validity tests in assessments undertaken to evaluate such requests. The failure rate is similar to but more specific than, the rates of feigning or exaggerating previously reported (Harrison & Edwards, 2010; Martin & Schroeder, 2020; Sullivan et al., 2007; White et al., 2020). This naturalistic study advances the literature by showing the base rate and functional effects of failing SVTs and PVTs beyond analog studies that ask otherwise non-disabled students to deliberately feign specific difficulties, especially self-report of ADHD or mental health symptoms. Our study also showed a higher incidence of failed PVTs, SVTs, and both for women than for men.

Of interest is that more students failed an SVT than a PVT. The 33.6% SVT failure rate in this large sample underlines the importance of administering SVTs in any disability-related assessment when evaluating self-report of symptoms of ADHD and mental health disorders.

Our study shows clearly that a notable difference exists in the symptom ratings and performance test scores obtained by those who pass all validity tests compared with those who fail one or more. Failure on at least one PVT, either alone or in combination with failing an SVT, was associated primarily with the worst performance on measures of cognitive ability. This finding demonstrates clearly that evidence of non-credible performance on a PVT is associated with significantly reduced scores on these types of cognitive tests.

PVT failure either alone or in combination with failing an SVT was also associated with the lowest scores on some, but not all tests of reading, but not on tests of written language or most tests of mathematics. Failing at least one PVT was also associated with the worst performance on a number of D-KEFS measures of executive functions. These results support the notion that PVT failure is suggestive of withholding best effort on some, but not all, achievement measures. It is not clear if some of these achievement tests are less susceptible to the effects of non-credible performance or whether the effects of withheld effort were more subtle on these measures.

Perhaps, as identified by Lindstrom et al. (2011), many students who attempt to exaggerate academic deficits can produce achievement profiles that are "disturbingly sophisticated" (p. 316), easily meeting commonly-used diagnostic criteria, such as performing below average on psychoeducational tests. Given that many of the Pass All students in our sample were diagnosed with disabilities in reading, writing, and/or mathematics, it would be relatively easy for a motivated student to produce lower scores that nevertheless overlapped with those of genuinely impaired individuals. As such, our findings may suggest that students feigning disabilities such as LD or ADHD can often suppress performance to levels that are equal to those with genuine impairment, rather than significantly overdoing their exaggeration. However, students exaggerating cognitive problems may not realize that individuals with LD and ADHD have otherwise normal levels of intellectual ability. Hence, this

Table 6. Means, SD, and ANOVA significance and effect size with Tukey post hoc pairwise comparisons of achievement, cognitive and processing speed as a function of group.

	Achievement scores									
	Pass all		FAIL PVT only		FAIL SVT only		FAIL PVT/SVT			
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	*p*	Effect d
Reading										
Woodcock Johnson IV (WJ-IV) Sentence Reading fluency, $n = 535$										
	93.4[b]	14.4	88.2[a]	14.7	94.7[b]	14.2	89.4[ab]	14.2	0.017	0.39
Woodcock Johnson III (WJ III) Reading fluency, $n = 1237$										
	93.5	14.5	93.4	13.6	4.2	15.7	93.8	14.8	0.902	
Nelson Denny Reading Test (NDRT) reading comprehension. *Note*. SS Mean = 200, SD = 25										
$n = 886$	190.2	26.4	188.1	21.6	194.2	25.8	189.7	24.6	0.173	
NDRT Reading Rate *Note*. SS Mean = 200, SD = 25										
$n = 460$	184.3	23.4	183.0	20.5	189.1	24.9	184.1	20.0	0.260	
Wechsler Individual Achievement Test 3rd Edition (WIAT 3) reading comprehension, $n = 900$										
	95.6[a]	12.2	90.6[bc]	13.7	93.0[ab]	11.5	88.3[c]	13.7	<0.001	0.51
WIAT 3 Reading Comprehension and Fluency, $n = 460$										
	89.7[b]	12.4	86.4[a]	13.6	89.7[b]	11.6	83.5[a]	14.4	0.011	0.44
WIAT 3 Basic Reading, $n = 536$										
	92.7	14.9	91.4	15.2	94.1	15.7	89.1	15.8	0.240	
Writing										
WJ IV Sentence Writing Fluency, $n = 555$										
	97.7[ab]	13.2	93.2[a]	15.3	99.7[b]	12.7	97.5[ab]	10.2	0.025	0.37
WJ III Writing Fluency, $n = 1080$										
	95.5	13.4	96.2	14.4	94.5	14.0	95.0	14.3	0.727	
WIAT 3 Written Expression, $n = 305$										
	97.6	10.6	92.1	14.1	98.1	13.1	97.7	17.2	0.096	
Kaufman Test of Educational Achievement (KTEA) written expression, $n = 436$										
	90.6	12.5	89.1	11.3	90.9	11.8	87.8	13.8	0.610	
Mathematics										
WJ IV Math Facts Fluency, $n = 276$										
	84.2	13.7	80.6	17.6	87.5	13.3	86.3	15.1	0.128	
WJ III Math Fluency, $n = 631$										
	78.6	15.2	75.4	14.0	78.2	16.0	78.9	13.9	0.628	
WIAT 3 Math, $n = 611$										
	89.7[b]	15.0	85.9[ab]	13.5	87.8[ab]	15.2	82.9[a]	14.6	0.012	0.39
Cognitive Scores										
Wechsler Adult Intelligence Scale 4th Edition (WAIS-IV) FSIQ, $n = 1950$										
	97.8[b]	12.8	92.0[a]	12.4	95.8[b]	13.5	90.5[a]	12.1	<0.001	0.49
WAIS-IV GAI, $n = 1515$										
	101.2[c]	14.7	96.4[ab]	14.5	98.5[bc]	14.5	94.1[a]	14.1	<0.001	0.43
WAIS-IV Verbal Comprehension, $n = 1956$										
	101.9[c]	14.8	98.2[ab]	14.0	99.6[bc]	15.0	95.6[a]	13.7	<0.001	0.35
WAIS-IV Perceptual Reasoning, $n = 1953$										
	101.5[c]	14.2	95.5[ab]	14.6	99.1[bc]	15.0	94.2[a]	15.1	<0.001	0.45
WAIS-IV Working Memory, $n = 1952$										
	90.9[b]	13.0	85.9[a]	11.7	89.9[b]	13.7	85.1[a]	11.9	<0.001	0.39
WAIS-IV Processing Speed, $n = 1950$										
	95.7[b]	12.2	91.2[a]	12.8	95.7[b]	13.2	91.4[a]	12.1	<0.001	0.35
WJ IV Cognitive Perceptual Speed, $n = 80$										
	88.8	16.0	81.6	16.8	90.2	17.0	88.4	13.9	0.565	
WJ III Processing Speed, $n = 1060$										
	92.1[b]	13.4	84.8[a]	14.1	92.6[b]	13.1	88.3[ab]	13.5	<0.001	0.39
Delis-Kaplan Executive Function System (D-KEFS)										
Color naming, $n = 900$	9.0	2.8	8.8	2.9	8.8	2.9	8.9	3.1	0.821	
Word reading, $n = 1345$	9.5	2.9	8.8	3.6	9.3	3.1	9.3	3.3	0.116	
Inhibition switch, $n = 1337$	8.9[b]	3.1	7.6[a]	3.5	8.7[b]	3.2	8.1[ab]	3.4	0.001	0.32
Letter fluency, $n = 768$	10.2	3.2	10.4	3.7	10.4	3.6	10.2	3.4	0.972	
Category fluency, $n = 766$	11.5	3.5	11.4	3.4	11.6	3.4	11.7	3.5	0.944	
Category switch, $n = 767$	10.4	3.4	10.5	3.1	10.2	3.5	10.4	3.5	0.896	
Visual scan-ning, $n = 877$	9.9	2.8	9.2	3.3	10.0	2.7	10.0	3.1	0.331	
Number sequencing, $n = 1237$	9.5[b]	2.9	8.3[a]	3.4	9.5[b]	2.8	9.3[b]	2.8	0.003	0.30
Letter sequen-cing, $n = 1405$	9.3	3.1	8.8	3.3	9.3	3.2	9.0	3.1	0.371	
Number/letter sequencing, $n = 1401$	8.9[b]	2.9	7.9[a]	3.0	9.0[b]	2.9	8.2[ab]	3.1	0.002	0.28

Note: Superscripts with matching letters are not statistically different.
PVT: Performance validity test; SVT: Symptom validity test; GAI: General Ability Index.

may explain why differences were found mainly on the cognitive measures given, rather than the achievement measures, as non-credible performers may have overexaggerated on these scales relative to what is normally found.

The results also highlight that use of only SVTs in a psychoeducational assessment, while identifying symptom overreporting, will fail to identify non-credible performance. As such, at least one well-validated, free-standing PVT should also be administered during a psychoeducational assessment in order to evaluate whether the client was performing credibly. Ideally, more than one of each is needed to confirm the validity of performance tests and symptom

reports (Sherman et al., 2020), and clinicians may wish to consider using embedded PVTs as well to allow for continuous sampling of test-taking engagement.

However, the results showed no overarching dimension of generalized feigning; individuals who failed SVTs tended to have exaggerated scores only on self (or observer) report measures, whereas those who failed PVTs scored lower on performance-based tests in general, although they also obtained low scores on a few self-report subscales. Failure on a combination of at least one PVT and one SVT tended to result in both performance and self-report scores suggestive of greater impairment compared with those who passed all validity measures. This finding underscores the need to administer both types of validity tests in psychoeducational evaluations for postsecondary students.

On measures of anxiety and anxiety-related disorders, the most extreme ratings were by those who failed an SVT or both SVT and PVT. In addition, those who failed a PVT had higher symptom ratings than those in Pass All group. This may speak to symptom magnification in those who fail a PVT or may suggest possible reasons for performance exaggeration. That is, similar to the idea of a "cry for help," those who are more anxious or traumatized may exaggerate performance problems in multiple domains in the hope of receiving services and supports.

While failing only an SVT was not associated with significantly lower performance on any cognitive or achievement measures, SVT failure did predict extreme scores on many self-report measures. Furthermore, failing an SVT, alone or in combination with failing a PVT, was generally related to the highest ratings of difficulty on the ADHD scales and on the mental health questionnaires. The SVT fail group (alone or in combination with PVT failure) returned the highest ratings on all of the CAARS-Self report scales, and typically also had the most extreme ratings on the PAI, the BASC-2, the BRIEF, and the BDI-II. Hence, it does appear that failure on an SVT is associated with extreme ratings on various self-report measures in this student population. White et al. (2020) also obtained this pattern of results in their study of adults with ADHD.

Observer ratings on the CAARS were highest for students failing both a PVT and an SVT. (Fail PVT&SVT group). However, observer ratings did not reach the 98th percentile for any symptom and were equally elevated in the Pass All group for Inattention. This result supports the findings of Sibley et al. (2012), suggesting that observers, in general, may be more reliable symptom reporters. The exception, however, appears to be when the student undergoing the ADHD assessment shows evidence of extreme symptom and performance exaggeration, in which case collateral reports may also be suspect.

Our findings also suggest that there is not a general, higher-order prediction of non-credible self-report, and that grouping results from different types of SVTs together can obscure different patterns of symptom exaggeration. Indeed, it seems clear that domain-specific SVTs are required in order to identify specific types of self-report exaggeration. Specifically, failure on ADHD-specific SVTs was related to the highest ratings of difficulty on both the ADHD-related and executive functioning self-report scales, but not on questionnaires enquiring about mental health problems. Likewise, failure on mental health-specific SVTs was related to the highest ratings of difficulty on the mental health symptom scales, but not on either the ADHD or executive function self-report scales. Similar to the advice of Osmon et al. (2006), these findings certainly support the notion that disorder-specific SVTs are needed to identify specific types of symptom exaggeration.

The clinical implications of these findings are that failed PVTs or SVTs or both are to be expected to occur frequently in psychoeducational assessments of postsecondary students. Failure on such measures should be taken as a serious indication that some or all of the scores obtained in an evaluation may be non-credible. Similar to a carbon monoxide alarm going off even when the home owner cannot smell the toxic fumes, these objective validity measures warn clinicians that the client's reported symptoms and/or performance may not be believable. Results from our study show clearly that failure on such measures is associated with unexpectedly lower performance on various cognitive and academic tests and unexpectedly higher symptom reporting. Failure on one or more of these measures means that clinicians must interpret the remaining data cautiously, since ability and some reading scores are likely suppressed for clients who have failed PVTs, and self-report ratings are likely to be exaggerated for those who fail SVTs. This creates a dilemma for the clinician especially in determining whether diagnostic criteria have been met for a particular disorder. The dilemma is greater in re-assessments when the question is whether a previously diagnosed disorder or disability is still evident.

The varying impact of failed PVTs on the tests we administered raises the question of whether any of these psychological tests are impervious to attempts to feign disability. The impact of failed PVTs on the WAIS-IV subtests that include a speed component, as well as reading fluency tests and simple speed tests on D-KEFS suggests that speed could be the component of these performance tests that is most vulnerable to feigning or exaggeration. This hypothesis is consistent with students' general belief that extra time on tests would help them to perform better (Lewandowski et al., 2014) and findings that extra time provides a significant competitive advantage to non-disabled students when taking timed tests (Lewandowski et al., 2013; Mandinach et al., 2005). However, not all of our tests of reading, written language, and math fluency showed a clear impact secondary to PVT failure.

We emphasize that failure on a validity test does not negate the possibility that the client has a disability. It simply makes it harder to determine the extent to which the client currently has any functional impairments. Therefore, it is important to interpret test results in the context of history, current functioning, and where possible, with collaboration from objective historical documents like school reports and previous assessments, along with reports from those who know the person well.

Limitations and future directions

Like most studies using a naturalistic clinical sample, some limitations exist. Most importantly, our findings may not generalize to populations other than English-speaking postsecondary students seeking psychoeducational assessments. In addition, not all students were tested with a standard test battery. As a result, different numbers of students took the various tests. These numbers, included in Tables 3–6, ranging from1960 to 80, are large samples, often yielding medium to large effect sizes. Future studies should consider giving a standard battery of tests to all students to improve the strength of any conclusions made.

A further limitation is that the test results are not sorted relative to the student's presenting belief about the disorder or disability they might have. Students completed at least one checklist about mental health functioning or ADHD, but the archival data did not specify what was the referring issue that brought them for an assessment. To better understand the effect of symptom or performance exaggeration on self-report and academic functioning, future studies could separate those seeking an LD diagnosis from those suspecting only ADHD, as well as analyzing separately by final diagnosis. In that way, we may better understand whether specific diagnostic expectations influence the pattern of symptom or performance exaggeration in those undergoing different types of assessments.

We cannot be sure that all those who were exaggerating symptoms were identified by the validity measures used. We did not employ continuous sampling of effort using embedded PVTs, which Rickards et al. (2018) have found to be effective in identifying those who were performing non-credibly. While some tests have reportedly shown promise in the detection of feigned ADHD (e.g., Berger et al., 2021; Harrison & Armstrong, 2016) there is currently no "gold standard" for identifying feigned LD or ADHD, and there are no stand-alone SVTs validated for use in ADHD assessments (Harrison et al., in press). As such, it is possible that some of those assigned to the Pass All group were performing noncredibly. This is an unfortunate problem encountered in clinical research; future studies may wish to employ analog malingerers whose motivations can be checked easily, and include embedded PVTs to allow for continuous sampling of effort. We acknowledge that this study did not examine embedded PVTs on tests of cognitive ability, which could increase the incidence of failures on PVTs beyond the 14.6% we found.

Finally, we did not evaluate the efficacy of stand-alone SVTs such as the Structured Interview of Reported Symptoms (SIRS; Rogers et al., 1992), the Structured Inventory of Malingered Symptomatology (SIMS; Rogers et al., 2014), or the Subtle ADHD Malingering Screener (SAMS; Ramachandran et al., 2019). Future research could investigate whether these measures yield the same results as those found in the present study.

In conclusion, it appears that failure on different types of validity tests is associated with differing patterns of test performance. For maximal identification of both non-credible performance and symptom reporting, clinicians should include at least one free-standing PVT and one SVT in their psychoeducational assessment battery for postsecondary students and, wherever possible, use measures shown to be most sensitive to the type of symptoms feigned in the population being evaluated. Otherwise, they run the risk of drawing incorrect diagnostic conclusions based on invalid and non-credible data.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The present work was financially supported by the Ontario Ministry of Colleges and Universities.

References

Barkley, R. (2019). Neuropsychological testing is not useful in the diagnosis of ADHD: Stop it (or prove it)! *The ADHD Report, 27*, 1–7. https://doi.org/10.1521/adhd.2019.27.2.1

Becke, M., Tucha, L., Weisbrod, M., Aschenbrenner, S., Tucha, O. & Fuermaier, A. B. M. (2021). Non-creditle symptom report in the clinical evaluation of adult ADHD: Development and initial validation of a new validity index embedded in the Conners' adult ADHD rating scales. *Journal of Neural Transmission*. Advance online publication. https://doi.org/10.1007/s00702-021-02318-y

Ben-Porath, Y. S., & Tellegen, A. (2011). *MMPI-2-RF (Minnesota Multiphasic Personality Inventory-2-Restructured Form): Manual for administration, scoring, and interpretation*. University of Minnesota Press.

Berger, C., Lev, A., Braw, Y., Elbaum, T., Wagner, M., & Rassovsky, Y. (2021). Detection of feigned ADHD using the MOXO-d-CPT. *Journal of Attention Disorders*, *25*(7), 1016–1032. https://doi.org/10.1177/1087054719864656

Boone, K. B., Lu, P., Sherman, D., Palmer, B., Back, C., Shamieh, E., Warner-Chacon, K., & Berman, N. G. (2000). Validation of a new technique to detect malingering of cognitive symptoms: The b Test. *Archives of Clinical Neuropsychology*, *15*(3), 227–241. https://doi.org/10.1016/S0887-6177(99)00014-1

Brown, J. A., Fishco, V. V., & Hanna, G. (1993). *Nelson–Denny Reading Test: Manual for scoring and interpretation, forms G & H*. Riverside.

Canadian ADHD Resource Alliance. (2017). *Canadian ADHD practice guidelines*. http://www.caddra.ca/

Conners, C. K., Erhardt, D., & Sparrow, E. (1999). *Conners' Adult ADHD Rating Scales (CAARS)*. MHS.

Copeland, C. T., Mahoney, J. J., Block, C. K., Linck, J. F., Pastorek, N. J., Miller, B. I., Romesser, J. M., & Sim, A. H. (2016). Relative utility of performance and symptom validity tests. *Archives of Clinical Neuropsychology*, *31*(1), 18–22. https://doi.org/10.1093/arclin/acv065

Dandachi-FitzGerald, B., & Merckelbach, H. (2013). Feigning ≠ feigning a memory deficit: The Medical Symptom Validity Test as an example. *Journal of Experimental Psychopathology*, *4*(1), 46–63. https://doi.org/10.5127/jep.025511

Dandachi-FitzGerald, B., Merckelbach, H., & Ponds, R. W. H. M. (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

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis Kaplan executive function system D-KEFS*. The Psychological Corporation.

Faust, D., Hart, K., & Guilmette, T. J. (1988). Pediatric Malingering: The capacity of children to fake believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology*, *56*(4), 578–582. https://doi.org/10.1037/0022-006X.56.4.578

Faust, D., Hart, K., Guilmette, T. J., & Arkes, H. R. (1988). Neuropsychologists' capacity to detect adolescent malingerers.

Professional Psychology: Research and Practice, *19*(5), 508–515. https://doi.org/10.1037/0735-7028.19.5.508

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

Green, P., & Flaro, L. (2021). Performance validity test failure predicts suppression of neuropsychological test results in developmentally disabled children. *Applied Neuropsychology. Child*, *10*(1), 65–81. https://doi.org/10.1080/21622965.2019.1604342

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

Green, P. (2005). *Green's Word Memory Test*. Green's Publishing Inc.

Green, P., Rohling, M. L., Lees-Haley, P. R., & Allen, L. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury*, *15*(12), 1045–1060. https://doi.org/10.1080/02699050110088254

Greenberg, L. M., Holder, C., Kindschi, C. L., & Dupuy, T. R. (2007). *Test of variables of attention continuous performance test*. The TOVA Company.

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.

Haggerty, K., Frazier, T., Busch, R., & Naugle, R. (2007). Relationships among Victoria Symptom Validity Test indices and Personality Assessment Inventory validity scales in a large clinical sample. *The Clinical Neuropsychologist*, *21*(6), 917–928. https://doi.org/10.1080/13854040600899724

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., & Edwards, M. J. (2010). Symptom exaggeration in post-secondary students: Preliminary base rates in a Canadian sample. *Applied Neuropsychology*, *17*(2), 135–143. https://doi.org/10.1080/09084281003715642

Harrison, A. G., & Wolforth, J. (2012). Findings from a Pan-Canadian survey of disability services providers in postsecondary education. *International Journal of Disability, Community and Rehabilitation*, *11*(1).

Harrison, A. G. (2017). Clinical, ethical, and forensic implications of a flexible threshold for LD and ADHD in postsecondary settings. *Psychological Injury and Law*, *10*(2), 138–150. https://doi.org/10.1007/s12207-017-9291-9

Harrison, A. G., Edwards, M. J., & Parker, K. C. (2007). Identifying students faking ADHD: Preliminary findings and strategies for detection. *Archives of Clinical Neuropsychology*, *22*(5), 577–588. https://doi.org/10.1016/j.acn.2007.03.008

Harrison, A. G., Edwards, M. J., & Parker, K. C. (2008). Identifying students feigning dyslexia: Preliminary findings and strategies for detection. *Dyslexia (Chichester, England)*, *14*(3), 228–246. https://doi.org/10.1002/dys.366

Harrison, A. G., Lee, G., & Suhr, J. (in press). Use of performance validity tests and symptom validity tests in assessment of specific learning disorders and attention deficit hyperactivity disorder. In K. Boone (Ed.). *Assessment of Feigned Cognitive Impairment* (2nd Ed.). Guilford Press.

Henry, G., & Gornbein, J. (2020). Posttraumatic stress disorder and noncredible symptom reporting on the Trauma Symptom Inventory-2. *Applied Neuropsychology: Adult*. Advance online publication. https://doi.org/10.1080/23279095.2019.1710511

Jachimowicz, G., & Geiselman, R. E. (2004). *Comparison of ease of falsification of attention deficit hyperactivity disorder diagnosis using standard behavioral rating scales*. Cognitive Science Online. http://cogsci-online.ucsd.edu

Jasinski, L. J., Harp, J. P., Berry, D. T. R., Shandera-Ochsner, A., Mason, L. H., & Ranseen, J. D. (2011). Using symptom validity tests to detect malingered ADHD in college students. *The Clinical Neuropsychologist*, *25*(8), 1415–1428. https://doi.org/10.1080/13854046.2011.630024

Lange, R. T., Iverson, G. L., Brooks, B. L., & Rennison, V. L. (2010). Influence of poor effort on self-reported symptoms and neurocognitive test performance following mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology*, *32*(9), 961–972. https://doi.org/10.1080/13803391003645657

Larrabee, G. L. (2014). Minimizing false positive error with multiple performance validity tests: Response to Bilder, Sugar, and Hellemann (2014 this issue). *The Clinical Neuropsychologist*, *28*(8), 1230–1242. https://doi.org/10.1080/13854046.2014.988754

Lees-Haley, P. R., English, L. T., & Glenn, W. J. (1991). A Fake Bad Scale on the MMPI-2 for personal injury claimants. *Psychological Reports*, *68*(1), 203–210. https://doi.org/10.2466/pr0.1991.68.1.203

Lewandowski, L. J., Cohen, J. A., & Lovett, B. J. (2013). Effects of extended time allotments on reading comprehension performance of college student with and without learning disabilities. *Journal of Psychoeducational Assessment*, *31*(3), 326–336. https://doi.org/10.1177/0734282912462693

Lewandowski, L., Lambert, T. L., Lovett, B. J., Panahon, C. J., & Sytsma, M. R. (2014). College students' preferences for test accommodations. *Canadian Journal of School Psychology*, *29*(2), 116–126. https://doi.org/10.1177/0829573514522116

Lindstrom, W., Coleman, C., Thomassin, K., Southall, C., & Lindstrom, J. (2011). Simulated dyslexia in postsecondary students: Description and detection using embedded validity indicators. *The Clinical Neuropsychologist*, *25*(2), 302–322. https://doi.org/10.1080/13854046.2010.537280

Lippa, S. M. (2018). Performance validity testing in neuropsychology: A clinical guide, critical review, and update on a rapidly evolving literature. *The Clinical Neuropsychologist*, *32*(3), 391–421. https://doi.org/10.1080/13854046.2017.1406146

Lippa, S. M., Pastorek, N. J., Romesser, J., Linck, J., Sim, A. H., Wisdom, N. M., & Miller, B. I. (2014). Ecological validity of performance validity testing. *Archives of Clinical Neuropsychology*, *29*(3), 236–244. https://doi.org/10.1093/arclin/acu002

Mandinach, E. B., Bridgeman, B., Cahalan-Laitusis, C., & Trapani, C. (2005). The impact of extended time on SAT test performance. *Ets Research Report Series*, *2005*(2), 1–35. https://doi.org/10.1002/j.2333-8504.2005.tb01997.x

Marshall, P., Schroeder, R., O'Brien, J., Fischer, R., Ries, A., Blesi, B., & Barker, J. (2010). Effectiveness of symptom validity measures in identifying cognitive and behavioral symptom exaggeration in adult attention deficit hyperactivity disorder. *The Clinical Neuropsychologist*, *24*(7), 1204–1237. https://doi.org/10.1080/13854046.2010.514290

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

Merten, T., & Merckelbach, H. (2013). Introduction to malingering research and symptom validity assessment. *Journal of Experimental Psychopathology*, *4*(1), 3–5. https://doi.org/10.1177/2043808713000400102

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

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

Nelson, J., Whipple, B., Lindstrom, W., & Foels, P. (2019). How Is ADHD assessed and documented? Examination of psychological reports submitted to determine eligibility for postsecondary disability. *Journal of Attention Disorders*, *23*(14), 1780–1791. https://doi.org/10.1177/1087054714561860

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

Osmon, D. C., Plambeck, E. A., Klein, L., & Mano, Q. (2006). The word reading test of effort in adult learning disability: A simulation

study. *The Clinical Neuropsychologist*, *20*(2), 315–324. https://doi.org/10.1080/13854040590947434

Quinn, C. A. (2003). Detection of malingering in assessment of adult ADHD. *Archives of Clinical Neuropsychology*, *18*(4), 379–395. https://doi.org/10.1093/arclin/18.4.379

Ramachandran, S., Holmes, E. R., Rosenthal, M., Banahan, B. F., Young, J., & Bentley, J. P. (2019). Development of the subtle ADHD malingering screener. *Assessment*, *26*(3), 524–534. https://doi.org/10.1177/1073191118773881

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior assessment system for children* (2nd Ed.). Pearson.

Rickards, T. A., Cranston, C. C., Touradji, P., & Bechtold, K. T. (2018). Embedded performance validity testing in neuropsychological assessment: Potential clinical tools. *Applied Neuropsychology. Adult*, *25*(3), 219–230. https://doi.org/10.1080/23279095.2017.1278602

Roberson, C. J., Boone, K. B., Goldberg, H., Miora, D., Cottingham, M., Victor, T., Ziegler, E., Zeller, M., & Wright, M. (2013). Cross validation of the b test in a large known groups sample. *The Clinical Neuropsychologist*, *27*(3), 495–508. https://doi.org/10.1080/13854046.2012.737027

Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *Structured interview of reported symptoms (SIRS) and professional manual*. Psychological Assessment Resources.

Rogers, R., Robinson, E., & Gillard, N. (2014). The SIMS screen for feigned mental disorders: The development of detection-based scales. *Behavioral Sciences & the Law*, *32*(4), 455–466. https://doi.org/10.1002/bsl.2131

Roth, R. M., Isquith, P., & Gioia, G. A. (2005). *BRIEF-A Behavior Rating Inventory of Executive Function – Adult Version*. Psychological Assessment Resources Inc.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock-Johnson IV*. Riverside.

Sherer, M., Sander, A. M., Ponsford, J., Vos, L., Poritz, J., Ngan, E., & Leon Novelo, L. (2020). Patterns of cognitive test scores and symptom complaints in persons with TBI who failed performance validity testing. *Journal of the International Neuropsychological Society*, *26*(9), 932–938. https://doi.org/10.1017/S1355617720000351

Sherman, E. M. S., Slick, D. J., & Iverson, G. L. (2020). Multidimensional malingering criteria for neuropsychological assessment: A 20-year update of the malingered neuropsychological dysfunction criteria. *Archives of Clinical Neuropsychology*, *35*(6), 735–764. https://doi.org/10.1093/arclin/acaa019

Shura, R. D., Denning, J. H., Miskey, H. M., & Rowland, J. A. (2017). Symptom and performance validity with veterans assessed for attention-deficit/hyperactivity disorder (ADHD). *Psychological Assessment*, *29*(12), 1458–1465. https://doi.org/10.1037/pas0000436

Sibley, M. H., Pelham, W. E., Molina, B., Gnagy, E. M., Waxmonsky, J. G., Waschbusch, D. A., Derefinko, K. J., Wymbs, B. T., Garefino, A. C., Babinski, D. E., & Kuriyan, A. B. (2012). When diagnosing ADHD in young adults emphasize informant reports, DSM items, and impairment. *Journal of Consulting and Clinical Psychology*, *80*(6), 1052–1061. https://doi.org/10.1037/a0029098

Slick, D., Hopp, G., Strauss, E., & Thompson, G. B. (1997). *Victoria Symptom Validity Test, version 1.0 professional manual*. Psychological Assessment Resources.

Sollman, M. J., & Berry, D. T. (2011). Detection of inadequate effort on neuropsychological testing: A meta-analytic update and extension. *Archives of Clinical Neuropsychology*, *26*(8), 774–789. https://doi.org/10.1093/arclin/acr066

Spenceley, L. M., Wood, W., & Lovett, B. J. (2020). Using the Woodcock-Johnson IV tests of cognitive abilities to detect feigned ADHD. *Applied Neuropsychology: Adult*. Advance online publication. https://doi.org/10.1080/23279095.2020.1748631.

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

Sullivan, B. K., May, K., & Galbally, L. (2007). Symptom exaggeration by college students in attention-deficit hyperactivity disorder and learning disorder assessments. *Applied Neuropsychology*, *14*(3), 189–207. https://doi.org/10.1080/09084280701509083

Sweet, J. J., Heilbronner, R. L., Morgan, J. E., Larrabee, G. J., Rohling, M. L., Boone, K. B., Kirkwood, M. L., Schroeder, R. W., & Suhr, J. A. (2021). American Academy of Clinical Neuropsychology (AACN) 2021 consensus statement on validity assessment: Update of the 2009 AACN consensus conference statement on ueorpyshchological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*. Advance online publication. https://doi.org/10.1080/13854046.2021.1896036

Tombaugh, T. N. (1996). *Test of Memory Malingering TOMM*. Multi-Health Systems Inc.

Wallace, E. R., Garcia-Willingham, N. E., Walls, B. D., Bosch, C. M., Balthrop, K. C., & Berry, D. T. (2019). A meta-analysis of malingering detection measures for attention-deficit/hyperactivity disorder. *Psychological Assessment*, *31*(2), 265–270. https://doi.org/10.1037/pas0000659

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale – 4th ed.* Pearson.

Wechsler, D. (2009). *Wechsler Individual Achievement Test – 3rd ed.* Pearson.

White, D. J., Ovsiew, G. P., Rhoads, T., Resch, Z. J., Lee, M., Oh, A. J., & Soble, J. R. (2020). The divergent roles of symptom and performance validity in assessment of ADHD. *Journal of Attention Disorders*. Advance online publication. https://doi.org/10.1177/1087054720964574

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Riverside.

Appendix A

Table A1. List of variables in each validity failure group with scores that are significantly different from the Pass All group, with emphasis on the SVT failure group subtypes.

Variable	Pass all	fail PVT only	fail SVT only	Fail ADHD SVT	Fail MH SVT	FAIL ADHD&MH SVTs	Fail PVT&SVT
Self report							
CAARS Self			Inattention, Hyperactivity, Impulsivity, Problems with self concept, DSM Inattention, DSM Hyperactivity, DSM Total, ADHD Index	All subscales, Dissimulation, Exaggeration Index, CAARS Infrequency Index		All subscales, Inconsistency, Dissimulation, Exaggeration Index, CAARS Infrequency Index	All subscales, Dissimulation, Exaggeration Index, CAARS Infrequency Index
n				297	323	128	
CAARS observer			Impulsivity, Problems with self concept, DSM Inattention, ADHD Index	All Subscales		All Subscales	All Subscales except Inattention
n				230	244	98	
Personality Assessment Inventory (PAI) Validity Indices			Inconsistency, Infrequency, RDF NIM, PIM, Malingering Index		Infrequency, Negative Impression, RDF, Malingering Index	Inconsistency, RDF, Negative Impression, Malingering Index	Inconsistency, Infrequency, NIM, PIM, RDF, Malingering Index
n				171	410	126	
PAI Scales		Anxiety, Anxiety related disorders	All scales except Dominance and Warmth	Anxiety, Mania, Paranoia, Schizotypal, Borderline features, Antisocial, Treatment rejection	Depression, Paranoia, Suicidal ideation, Dominance	All scales	All scales except Dominance
Behavior Assessment System for Children 2nd edition**							
n				17	13	3	
Beck Depression Inventory II		highest			highest	highest	highest
N				26	28	16	
Beck Anxiety Inventory							
Achievement, cognitive							
Variable	Pass all	fail PVT only	fail SVT only	Fail ADHD SVT	Fail MH SVT	FAIL ADHD&MH SVTs	Fail PVT/SVT
Woodcock Johnson IV		Reading					Reading
n				70	91	36	
WIAT 3 **							Math
Wechsler Adult Intelligence Scale 4th Edition		All IQ scales and indices		FSIQ, General Ability Index, Perceptual Reasoning, Working Memory	General Ability Index		All IQ scales and indices
n				260	343	104	
Woodcock Johnson III		Processing Speed					Processing Speed
n				105	77	36	

(*continued*)

Table A1. Continued.

Variable	Pass all	fail PVT only	fail SVT only	Fail ADHD SVT	Fail MH SVT	FAIL ADHD&MH SVTs	Fail PVT&SVT
BRIEF		Organization of materials	Shift, Emotional control, Self-monitoring, Behavior Regulation Index, Global Executive Composite	Inhibit, Behavior Regulation Index, Metacognitive Index, Global Executive Composite		Inhibit, Shift, Emotional control, Self monitor, Behavior Regulation Index, Global Executive Composite	All scales and subscales except Task monitoring
n				88	134	44	
D-KEFS		Inhibition/ switching, Number sequencing, Number letter sequencing					
n				134	117	30	

Note: **post hoc tests not performed due to small sample size in some cells; *n*'s reported only for SVT subgroups.

PVT: Performance validity test; SVT: Symptom validity test; MH: Mental Health; RDF: Roger's discriminant function; NIM: Negative Impression Management; PIM: Positive Impression Management; BRIEF: Behavior Rating Inventory of Executive Function; DKEFS: Delis–Kaplan Executive Function System.