



What can we learn about performance validity from TOVA response profiles?

Beth Pollock, Allyson G. Harrison & Irene T. Armstrong

To cite this article: Beth Pollock, Allyson G. Harrison & Irene T. Armstrong (2021) What can we learn about performance validity from TOVA response profiles?, Journal of Clinical and Experimental Neuropsychology, 43:4, 412-425, DOI: [10.1080/13803395.2021.1932762](https://doi.org/10.1080/13803395.2021.1932762)

To link to this article: <https://doi.org/10.1080/13803395.2021.1932762>



Published online: 04 Jun 2021.



Submit your article to this journal [↗](#)



Article views: 113



View related articles [↗](#)



View Crossmark data [↗](#)



What can we learn about performance validity from TOVA response profiles?

Beth Pollock, Allyson G. Harrison  and Irene T. Armstrong

Regional Assessment & Resource Centre, Queens University, Kingston, Ontario, Canada

ABSTRACT

Given the functional impairments associated with Attention Deficit/Hyperactivity Disorder (AD/HD), a valid diagnosis is important. However, particularly when carried out in adulthood, the diagnostic process can be challenging and is complicated by conclusive evidence that a proportion of individuals referred for evaluation of AD/HD exaggerate or feign their symptoms. Relatively few methods, however, exist to identify such feigning. While continuous performance tests (CPTs) may provide useful information regarding performance validity, the question remains as to whether there are consistent patterns of exaggeration demonstrated by those feigning AD/HD. Thus, this study used cluster analysis to determine whether valid and reliable performance clusters would emerge based on CPT performance. Using archival data from a university-based AD/HD screening clinic, we investigated the performance of 305 adults on the Test of Variables of Attention (TOVA). Three profiles emerged, including one cluster who demonstrated exceptionally low performance on the TOVA, exceptionally high reports of AD/HD symptomology, and higher rates of failure on symptom and performance validity tests. The implication from our analysis is that this group most likely represents individuals who were exaggerating or magnifying their difficulties. The results reaffirm previous research showing that performance profiles on a continuous performance test can be used as an indicator of credible performance.

ARTICLE HISTORY

Received 18 January 2021
Accepted 17 May 2021

KEYWORDS

AD/HD; adults; cluster analysis; diagnosis; validity; feigning; TOVA

Attention Deficit/Hyperactivity Disorder (AD/HD) is a neurodevelopmental condition, appearing in childhood and frequently persisting into adulthood (Barkley et al., 2002), characterized by inattention and/or hyperactivity-impulsivity (American Psychiatric Association [APA], 2013). Adults with AD/HD experience significant educational (e.g., poor grades, underachievement), marital (e.g., increased divorce or separation rates, marital dissatisfaction), social (e.g., immaturity, social maladjustment), and occupational (e.g., frequent changes in employment) impairments (e.g., Barkley, 2002; DuPaul et al., 2009; Murphy & Barkley, 1996; Norwalk et al., 2008; Riccio et al., 2005). AD/HD is also often associated with co-morbid disorders, such as mood and anxiety disorders, substance abuse, and personality disorders (Able et al., 2007; Barkley et al., 2008; Miller et al., 2007; Sobanski et al., 2007).

An increasing number of college and university students are presenting to specialists complaining of symptoms of AD/HD and seeking to qualify for disability supports and services (Harrison, 2004; Nugent & Smart, 2014; Weyandt & DuPaul, 2012). Given the wide range of functional impairments and co-morbidities associated with AD/HD, a valid diagnosis

is important (Able et al., 2007; Goodman, 2009). However, particularly when assessed in adulthood, the diagnostic process can be challenging, as symptoms of AD/HD are more heterogeneous than in childhood and can overlap with symptoms of co-morbid disorders (Barkley, 2008; Wasserstein, 2005). To address these problems, specific diagnostic guidelines for adults have been developed in recent years (e.g., Goodman, 2009; Manos, 2010; National Institute for Health and Care Excellence, 2018) that recommend the administration of a complex evaluation including a clinical interview, completion of rating scales, and neuropsychological tests, as well as obtaining multiple sources of information from informants.

Complicating this assessment process is conclusive research showing that a significant proportion of individuals referred for clinical evaluation of AD/HD exaggerate or feign their symptoms (Harrison et al., 2007; Pella et al., 2012; J. Suhr et al., 2008; Sullivan et al., 2007). AD/HD is a condition susceptible to malingering because of a complex etiology (Millichap, 2008; Thome et al., 2012), a lack of distinct and decisive symptoms, vague diagnostic criteria, and the importance of patients' self-reports for diagnosis (Fuermaier et al., 2012). Noncredible representation of current or past

symptoms of AD/HD is also made easier by the face validity of most AD/HD symptom self-report measures, their lack of validity subscales (Quinn, 2003), and the fact that the symptoms associated with AD/HD are well known (Conti, 2004). Base rates for feigned or exaggerated AD/HD symptoms in the post-secondary setting are estimated to range from 14.6 to 47.6% (Harrison & Edwards, 2010; J. Suhr et al., 2008; Sullivan et al., 2007). Individuals may be motivated to be diagnosed with AD/HD by external incentives, such as access to academic accommodations and stimulant medication (Garnier-Dykstra et al., 2012; Jasinski & Ranseen, 2011; Young & Gross, 2011). There are strong societal interests in preventing false-positive diagnoses of AD/HD including: the substantial costs of unnecessary assessments and treatments; unjustified use of limited medical resources; unjustified access to disability-related funding or grants; passive support of drug use; disadvantages for individuals who do not feign AD/HD (and thus are not granted access to academic accommodations); and damage to public confidence in clinical diagnostic practices (Frazier et al., 2008; Harrison, 2006; Harrison et al., 2007, 2012; J. Suhr et al., 2008; Sullivan et al., 2007; Tucha et al., 2014).

Thus, researchers have searched for assessment strategies and/or measures to distinguish between feigned and genuine AD/HD. Research has demonstrated that patients suspected of feigning AD/HD could easily be diagnosed with AD/HD based on their responses to several important clinical interview questions (P. S. Marshall et al., 2016), responses on self-report questionnaires of AD/HD (Fisher & Watkins, 2008; Harrison et al., 2007; Jasinski et al., 2011; Quinn, 2003; P. S. Marshall et al., 2016; Sollman et al., 2010; Tucha et al., 2009; Williamson et al., 2014), self-reported impairments (Bryant et al., 2018; Fuermaier et al., 2018), or performance on commonly used tests of executive functioning, processing speed, and/or academic achievement (Booksh et al., 2010; Frazier et al., 2008; Harrison et al., 2007; Musso & Gouvier, 2012; Sollman et al., 2010).

In an attempt to obtain more reliable, objective, and standardized evidence of AD/HD symptomology, continuous performance tests (CPTs) are often used by clinicians to aid with the diagnosis of AD/HD. Studies have found that adults with AD/HD perform differently on CPTs when compared to controls (e.g., Advokat et al., 2007; Fasmer et al., 2016), though findings have been mixed (see Baggio et al., 2020). However, research has also demonstrated that post-secondary students can easily feign symptoms of AD/HD on these tests. Specifically, J. A. Suhr et al. (2010) demonstrated that students who responded in a non-credible manner on

a performance validity test were difficult to distinguish from those diagnosed with AD/HD in terms of general CPT performance, but performed worse than the AD/HD group on reaction time (RT) variability and RT change. More recently, Nicholls et al. (2020) reported that children who failed a performance validity test (PVT) committed significantly more errors of omission on the TOVA than those who passed a PVT. Harrison and Armstrong (2020) continued this line of research showing that post-secondary students in a suspect effort group returned lower standard scores than honest but symptomatic participants in all TOVA measures except commission errors in the second half of the test. Other research studies comparing analog AD/HD malingerers to a clinical AD/HD group and controls found that those “faking” AD/HD demonstrated the poorest CPT performance (Leark et al., 2002; Quinn, 2003; Robinson & Rogers, 2018). Consistent with these findings, P. Marshall et al. (2010) examined the effectiveness of various performance and symptom validity measures to detect suspect effort. They found that TOVA omission errors (63% sensitivity), the Conners CPT-II omission errors (56%), and TOVA RT variability (54%) measures had modest sensitivity but at least 90% specificity in identifying suspect test-taking effort in adults undergoing AD/HD evaluation.

Thus, while CPTs may provide a data point in the evaluation of AD/HD symptomology, they may also prove useful in detecting malingering, as the CPT performance of non-credible performers appears to differ consistently from those with genuine AD/HD. Specifically, as Harrison et al. (2007) suggested, feigners tend to “over-shoot the mark” and produce much more impaired performance than those with genuine AD/HD. The question remains, however, as to whether there are consistent patterns of exaggeration demonstrated by those feigning AD/HD.

Performance validity studies to date have taken either a simulation or a known-group approach. In the simulation approach, groups are asked to portray specific characteristics, such as AD/HD symptomology. As the preparation and motivation of individuals in a laboratory environment may be quite different from those actually feigning AD/HD in clinics, the ecological validity of these studies may be limited. In the known-group approach, a researcher examines two groups of real-world patients: one that is “known” or strongly suspected of performing invalidly based on an external criterion and one that is “known” or strongly suspected of performing validly based on the same criterion (Larrabee, 2007; Mossman et al., 2012). This approach has clear ecological validity as examinees are being evaluated for realistic purposes. However, as

a participant's group assignment depends on the investigator's application of an independent criterion, misclassifications and misestimates of accuracy are always possible (Mossman et al., 2012).

The current study takes an alternate approach to explore performance validity based on profiles of performance on a CPT, a measure that directly assesses attention. Specifically, we used cluster analysis to determine whether valid and reliable performance clusters would emerge based on TOVA indices. Cluster analysis is a multivariate classification technique that allows for a statistical grouping of like cases in homogeneous clusters based on their similarity across one or more characteristics. To examine the external validity of the cluster solutions, the derived groups were then compared on a measure of performance validity as well as self-reported symptoms of AD/HD, impairment, and psychological symptomology. This study utilized data collected in an AD/HD screening clinic rather than a laboratory setting, which maximizes the ecological validity of the findings. As this study is exploratory in nature and clusters were not specified in advance, no strict a priori hypotheses are provided.

Method

Participants

Archival data from adults ($n = 305$) who attended a university-based AD/HD screening clinic between 2011 and 2019 (48.9% female) were included in this study. Referral sources at colleges and universities within the Eastern Ontario catchment area were made aware of the AD/HD Screening Clinic and encouraged to refer students with no prior diagnosis of AD/HD who, nevertheless, now wondered if they suffered from this disorder. Referrals to the AD/HD screening clinic were initiated by one or more of the following: university-based counseling services (50.1%), physician (27.9%), parent (23.6%), and/or friends (23.3%). Participants were excluded if they did not provide consent to have their data used in archival research, did not complete the TOVA, and/or were not administered a performance validity test. The average age was 22.3 y ($SD = 5.48$, range = 17 to 53 y). Most were attending a 4-year university (69.2%) vs. a 2-year community college (27.6%). The majority were in their first three years of post-secondary education (72.5%), but a sizable minority were pursuing graduate studies (13.3%). Information on socioeconomic status and ethnicity was not collected. None of the individuals had previously been diagnosed with an Attention Deficit/Hyperactivity

Disorder. A small minority had already been prescribed medication to improve their attention (1.8% Concerta; 1.2% Adderall; 1.2% Vyvanse; 0.6% Ritalin; 0.9% Strattera). Learning strategies to address their symptoms, obtaining medication, and accessing additional time for tests/examinations were identified as the top three reasons for attending the AD/HD screening clinic.

The screening appointment usually lasted between two to three hours and involved the student undergoing a comprehensive clinical interview to evaluate whether all five of the DSM-IV or DSM-5 criteria for ADHD had been met in childhood and currently. Students then completed a standard test battery, including a background questionnaire, Conners' Adult AD/HD Rating Scales (CAARS; Conners et al., 1998), Weiss Functional Impairment Rating Scale (WFIRS; Weiss, 2000), Test of Variables of Attention (TOVA; L. M. Greenberg et al., 1994), Symptom Checklist-90-Revised (SCL-90-R; Derogatis, 1994), and Medical Symptom Validity Test (MSVT; Green, 2004). Students with more complex presentations were referred for a full Psychoeducational Assessment to help with differential diagnosis and determination of functional impairments. The Psychoeducational Assessment included additional data collection via a lengthier background questionnaire as well as four to six hours of testing using a flexible battery that included measures of cognitive, academic, and psychological functioning. Based on the screening and/or subsequent testing administered by clinic staff, 11.5% of the students were ultimately diagnosed with AD/HD. Thus, most students received no diagnosis and were referred for learning strategies support or counseling services. A small minority ($n = 3$) admitted to malingering in an attempt to obtain disability-related services and/or medication.

Measures

The measures used in this investigation were administered and scored by trained examiners employed by a university-based assessment center under the supervision of a clinical neuropsychologist. Scores obtained from the core indices of the TOVA were used to derive representative clusters. Once the clusters were identified, cluster performance on an external performance validity measure (Green's Medical Symptom Validity Test: MSVT, 2004) and self-report ratings of AD/HD symptomology (Conners Adult Attention Rating Scales: Self-Report: CAARS, 1998), functional impairment (Weiss Functional Impairment Rating Scale: WFIRS, 2000), and psychiatric symptomology (Symptom

Checklist-90 Revised: SLC-90-R- Global Severity Index, 1994) were compared.

Test of Variables of Attention

The TOVA is a computer-based continuous performance test, measuring attention and impulse control (Leark et al., 2008). Measures of performance on the TOVA include the following: (1) errors of omission: failure to respond to the target, a measure of inattention; (2) errors of commission: inappropriate response to the nontarget stimulus, a measure of impulsivity or failure to inhibit response; (3) mean correct RT: the average latency of correct response, a measure of processing speed and RT; and (4) standard deviation of RT (variability), a measure of variability or consistency. The TOVA standard scores have a mean of 100 and SD of 15. It should be noted that scores below 40 are truncated and were given as 0 in older versions of the test and provided as < 40 (which were represented as 40 in our database) in the newer versions of the test. The TOVA includes an overall score that compares an individual's performance to an age/gender specific AD/HD group (Attention Comparison Score: ACS) and a Symptom Exaggeration Index (SEI: Hughes et al., 2008). The SEI was developed to identify exaggeration in those who deliberately fake bad; a score of 0 or 1 provides "no evidence," a score of 2 indicates "some evidence," a score of 3 indicates "strong evidence," and a score of 4 indicates "very strong evidence" of symptom exaggeration. The TOVA is among the most commonly used CPTs, has been studied and normed in both children and adult populations (L. M. Greenberg & Waldmant, 1993; L. M. Greenberg et al., 1994), and is one of the few CPTs to include a measure of performance validity (L. Greenberg et al., 1996; P. Marshall et al., 2010).

Medical Symptom Validity Test

The Medical Symptom Validity Test (MSVT: Green, 2004) is a computerized forced-choice stand-alone performance validity test. The MSVT measures Immediate Recognition and Delayed Recognition of a word list, as well as the Consistency of answers between the two subtests. According to the test manual, scores at or below 85 on any of these three subtests indicates non-credible performance. Numerous studies have demonstrated the utility of the MSVT in the discrimination between those with genuine memory impairment and those simulating impairment in a range of patient samples (see Carone, 2009 for review).

The Conners' Adult AD/HD Rating Scale

The Conners' Adult AD/HD Rating Scales (CAARS; Conners et al., 1998) 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 Inattention, Hyperactivity/Impulsivity, and Total DSM symptoms; and c) an overall AD/HD Index that is said to measure the "overall level of AD/HD symptoms" (Conners et al., 1998, p. 23). While the manual suggests that individuals obtaining *T*-scores on the AD/HD Index of over 70 are likely to meet the diagnostic criteria for AD/HD, it also cautions that *T*-scores above 80 on any of the subscales should be considered as possible indicators of symptom exaggeration.

While not utilized in the original assessment, the current study calculated retrospectively two newly described methods of evaluating self-report credibility on the CAARS described by J. A. Suhr et al. (2011) and Harrison and Armstrong (2016).

The CAARS Infrequency Index (CII; J. A. Suhr et al., 2011) is composed of 12-items rarely endorsed by either typically developing adults or those diagnosed with AD/HD. J. A. Suhr et al. (2011) identified a cut score of > 21 as producing few false positive identifications for those with AD/HD.

The E-CAARS includes 18 additional symptom validity items embedded among the regular items and has been described in detail by Harrison and Armstrong (2016). The sum of the symptom validity items produces a Dissimulation Index (DI) score, while a formula that combines items from the DI with extreme scores from existing CAARS indices produces the Exaggeration Index (EI). According to Harrison and Armstrong (2016), the EI has acceptable classification accuracy when discriminating between those feigning AD/HD and other clinical groups (including those with AD/HD) who were presumed to be reporting symptoms accurately, with sensitivity to feigning ranging from .24 to .69 and specificity ranging from .74 to .97, depending on cut score used. A cutoff set at 2 had relatively good classification accuracy, with a score of 3 or more being almost exclusive to those feigning AD/HD (97%).

The Weiss Functional Impairment Rating Scale

The Weiss Functional Impairment Rating Scale (WFIRS: Weiss, 2000) is a self-report measure for impairments that commonly occur in patients with AD/HD. The WFIRS comprises 70 items that are divided into seven domains: Family (8 items), Work (11 items), School (11 items), Life Skills (12 items), Self-

Concept (5 items), Social (9 items), and Risk (14 items). Each item employs a four-point Likert scale scored from 0 to 3 (0 = *never, not at all*; 1 = *sometimes, somewhat*; 2 = *often, much*; 3 = *very often, very much*). A scale score per domain is calculated by summing the response values to all items per domain and dividing this sum by the number of endorsed items. The WFIRS has been found to have excellent internal consistency for the total score (.96), and good to excellent (.85 to .94) for the subscales (Canu et al., 2020). The Canadian AD/HD Resource Alliance (CADDRA) recommends that clinicians consider any domain with a mean score > 1.5 as suggesting impairment (CADDRA, 2011).

The Symptom Checklist-90-Revised

The Symptom Checklist-90-Revised (SCL-90-R; Derogatis, 1994) is a 90-item multidimensional questionnaire developed to screen for a range of psychological symptoms and psychopathological features. In psychotherapy outcome research, the SCL-90-R has been best conceptualized as a one-dimensional measure of symptom load and the Global Severity Index (GSI) has become a widely used measurement of psychological distress (e.g., Osterberg et al., 2002; Skydsbjerg et al., 2001).

Procedure

Participants underwent a one-hour semi-structured interview, were asked to provide report cards from childhood, and supplied rating scales completed by collateral informants to provide information regarding past and present symptoms of AD/HD. During the session, participants also completed the TOVA and MSVT, as well as the self-report version of the CAARS, the SCL-90-R, and the WFIRS. This allowed a determination as to whether they met the five criteria for diagnosis as outlined in DSM-IV-TR (APA, 2000) or DSM-5 (APA, 2013), depending on year tested.

Results

Initial cluster analyses

Means and SDs for the TOVA indices and standard scores used in the cluster analyses are presented in Table 1. As can be seen, the mean ACS ($M = -2.67$) for the entire sample was not within normative limits. According to the TOVA manual, scores below 0 suggest performance more similar to that of individuals with AD/HD. The mean SEI score did not suggest symptom exaggeration. For the remaining standard scores, the TOVA manual suggests interpretation of

Table 1. Mean TOVA scores for the entire sample (N = 305).

TOVA Index and Standard Scores	<i>M</i>	<i>SD</i>	<i>Range</i>
Attention Comparison Score	-2.67	5.82	-27.1–7.61
Symptom Exaggeration Index	0.72	0.97	0–3
Response Time Standard Score			
Half 1	92.54	28.22	0–143
Half 2	102.66	23.03	0–141
Response Time Variability Standard Score			
Half 1	66.22	34.53	0–118
Half 2	72.77	32.43	0–131
Omission Errors Standard Score			
Half 1	72.07	31.74	0–113
Half 2	74.49	31.82	0–112
Commission Errors Standard Score			
Half 1	96.45	21.68	0–117
Half 2	90.75	20.7	0–120

scores above 85 as normal, 80–85 as borderline normal, and below 80 as not within normative limits. Thus, mean scores for the entire sample were within normative limits for RT and Commission Errors, but not within normative limits for RT Variability and Omission Errors. As may be seen in Table 1, the range of scores within the entire sample was quite large, indicating that the mean scores failed to capture the significant variability in the overall sample.

TOVA standard scores from the first and second halves of the test were then subjected to cluster analyses. A hierarchical cluster analysis using Ward's method with Squared Euclidean Distance was initially conducted to estimate the number of clusters present in the data. Examination of the agglomeration coefficients, dendrogram, and individual cluster profiles generated strongly suggested that either a three- or five- cluster solution would provide the best description of the data. To correct for fusion errors, a k-means relocation pass was then applied to the first stage cluster centroids from each solution. Following K-means analysis, it was determined that both the three- and five-cluster solutions were interpretable and were subjected to reliability analyses.

Reliability analyses

Hierarchical and K-means reliability

Morris et al. (1981) contended that the number of cases changing cluster membership from the first stage (hierarchical) solution to the second stage (K-means) solution reflects an index of cluster stability (i.e., reliability). Thus, the reliability of the three- and five-cluster solutions was evaluated by comparing the profiles derived from the initial Ward's analysis (stage 1) to those derived from the K-means analysis (stage 2). Cohen's Kappa and one-way random effects intraclass correlation coefficients (ICCs) were used to compare the solutions in terms of membership

agreement and profile similarity, respectively. Significant ICCs were found for both solutions ($p < .001$). Kappa values were also significant ($p < .001$), with substantial agreement (based on a Kappa interpretive system suggested by Landis & Koch, 1977). Therefore, both solutions were deemed reasonably stable and were retained for multiple-method reliability assessment.

Multiple-method reliability

The prospective solutions were then subjected to three additional hierarchical clustering algorithms (Complete Linkage, Average Linkage Between Groups, and Average Linkage Within Groups), followed by a K-means pass through the data. The level of agreement between cluster solutions generated using the various methods was then calculated. Kappa values for both solutions were significant ($p < .001$) and suggested Fair to Almost Perfect agreement in the three-cluster solution and Moderate to Almost Perfect for the five-cluster solution. All ICCs in the three- and five-cluster solutions were significant ($p < .001$). Thus, both the three- and five-cluster solutions were deemed adequately replicable and were subjected to split-half reliability analyses.

Split-half reliability

To determine the extent to which the derived cluster solutions could be replicated in different samples, the initial sample was randomly split in half and each subsample was subjected to a two-stage Ward's analysis specifying the number of clusters to be recovered. The split-half profiles associated with the three-cluster solutions had good visual agreement, and all ICCs were significant ($p < .01$). Conversely, although the ICCs were all significant ($p < .01$), the split-half profiles from the five-cluster solutions were difficult to match. Based on these findings, the three-cluster solution was considered representative of the data and was selected as the final cluster solution.

Description of clusters

The three clusters generated on the basis of the initial two-stage Ward's analysis were assigned descriptive labels reflecting the most salient features of each mean TOVA profile. Mean TOVA Index scores for each cluster are presented in Table 2. There were significant differences in gender distribution, $\chi^2_{(2)} = 6.87$, $p = .032$, and clinical classification of cases, $\chi^2_{(10)} = 27.10$, $p = .002$, by cluster membership. There

Table 2. Mean TOVA Index for each cluster.

N % of sample	Cluster					
	High 149 49.8		Low 73 23.9		Mixed 83 27.2	
	M	SD	M	SD	M	SD
Response Time Half 1	112.26	13.97	61.90	25.61	83.73	19.14
Response Time Half 2	116.35	11.61	76.34	24.99	101.24	14.70
Response Time	92.54	14.10	28.78	24.84	51.92	28.38
Variability Half 1						
Response Time	96.04	16.50	31.92	24.32	66.93	19.59
Variability Half 2						
Omission Errors Half 1	96.13	12.80	36.10	23.59	60.51	24.69
Omission Errors Half 2	96.99	13.70	37.14	24.33	66.94	26.00
Commission Errors Half 1	101.83	17.50	81.71	28.10	99.77	15.47
Commission Errors Half 2	96.60	16.25	82.99	24.22	87.10	21.64

was no significant difference in age, $F_{(2, 302)} = .743$, $p = .477$.

The first cluster, characterizing 48.9% of the participants ($n = 149$; 87 males), demonstrated TOVA Index scores within normative limits across metrics and time intervals. Not surprisingly, they had a mean ACS above zero (1.71), indicating that their profile was more similar to individuals without AD/HD, and their mean SEI was low (0.18), which was not indicative of possible performance exaggeration. In fact, only two individuals in this cluster had an SEI score of 2 (suggesting some evidence of exaggeration) and no individuals in this cluster had an SEI score of 3 or above. 11.4% of those in this cluster were ultimately diagnosed with AD/HD. None of the individuals classified to this group admitted to malingering. Given their globally strong (or high) performance on the TOVA indices, this group was named "High" for short.

Cluster two was comprised of 23.9% of the participants ($n = 73$; 35 males) who performed below normative limits across metrics and time intervals, with the exception of borderline normal scores on the commission standard scores. As such, they attained a low mean ACS score (-9.99), suggesting that they performed similarly to individuals with AD/HD, but a high SEI score (1.85), suggesting possible performance exaggeration. 62.5% of the students in this cluster attained an SEI score of 2 or above. A small minority of individuals in this cluster were ultimately classified as having AD/HD (4.1%) and two of the students admitted to malingering. Given their globally low performance on the TOVA indices, this group was named "Low" for short.

The third cluster was comprised of 27.2% of the participants ($n = 83$; 34 males) who achieved

Table 3. Mean and SD of performance on CAARS-SR, SCL-90, and WFIRS by cluster.

	Cluster									<i>F</i>	<i>p</i>
	High			Low			Mixed				
	<i>n</i>	Mean	SD	<i>n</i>	Mean	SD	<i>n</i>	Mean	SD		
CAARS Self Report											
Inattention/Memory ^a	148	67.4	10.7	73	73.2	9.7	82	71.5	11.1	8.53	.000
Hyperactivity/Restlessness ^b	148	57.8	10.3	73	63.9	9.3	82	60.5	11.9	8.29	.000
Impulsive/Emotional Lability ^a	148	55.7	11.9	73	62.7	11.3	82	60.6	11.9	10.09	.000
Problems with Self-concept	148	60.3	11.1	73	60.5	9.1	82	58.4	10.7	1.06	.348
DSM Inattention Symptoms ^b	147	76.4	10.7	73	81.8	7.9	82	79.2	10.3	7.42	.001
DSM Hyperactivity Symptoms ^b	148	62.0	12.2	73	70.0	12.5	82	76.8	10.7	9.63	.000
ADHD Total Symptoms ^a	148	73.1	10.8	73	80.0	8.7	82	75.8	10.7	11.74	.000
ADHD Index ^a	148	63.4	8.4	73	68.4	7.4	82	66.3	8.7	9.60	.000
Inconsistency	148	6.0	6.2	73	5.2	2.3	82	5.0	2.0	1.42	.243
EI ^d	110	1.9	1.5	50	2.6	1.8	64	2.0	1.5	3.21	.042
CI ^b	108	14.5	5.7	48	18.2	6.0	63	15.8	5.5	6.88	.001
Symptom Check List 90 Revised											
Global Severity Index	73	69.3	11.7	44	69.1	8.3	44	68.2	9.2	.18	.838
Weiss Functional Impairment Rating Scale											
Home ^c	90	.9	.7	52	1.2	.9	55	.8	.6	4.04	.019
Self-concept	90	1.8	.8	52	1.8	.8	56	1.9	.7	0.36	.699
Work	41	.8	.6	25	1.2	.7	39	.9	.6	2.54	.084
Learning	83	1.5	.6	51	1.7	.5	56	1.5	.6	1.89	.154
Daily Living	89	1.4	.5	52	1.4	.6	56	1.3	.5	1.06	.350
Social	89	.9	.5	52	1.0	.6	56	.8	.5	1.14	.321
Risk Taking	90	.5	.5	52	.6	.4	56	.5	.4	1.17	.311

a High < Low, Mixed; *b* High < Low; *c* Mixed < Low; *d* no post hoc differences.

normal scores on the scales assessing for commission errors, a borderline normal score on the first half and a normal score on the second half of the scale assessing for RT, and scores not within normative limits on measures of RT variability and omissions. Their ACS score was similar to individuals with AD/HD (-3.99) but not as profoundly low as the Low group. Their mean SEI score was low (0.67), with 18% of students in this cluster attaining an SEI score of 2 or above. Of those classified to this cluster, 18.1% were ultimately diagnosed with AD/HD and 1 student admitted to

malingering. This group was labeled Mixed TOVA Performance or “Mixed” for short.

External validity of the derived typology

The external validity of a cluster solution addresses the degree to which empirically derived subgroups can be distinguished on the basis of theoretically important variables not used in the cluster analysis (Fletcher, 1985). External validity was assessed by comparing the derived TOVA subgroups on the basis of mean standard score performances on the MSVT, AD/HD Index scores

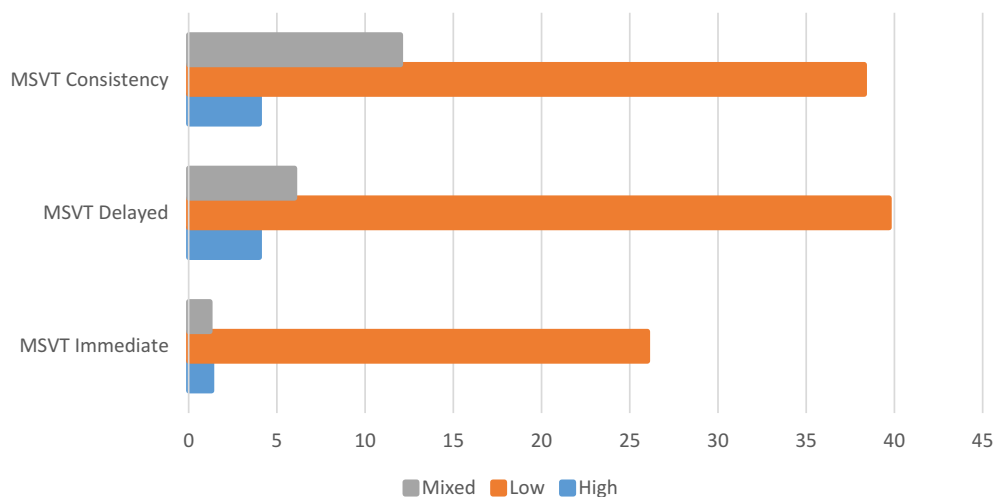


Figure 1. Percentage of MSVT subtests failed as a function of cluster membership.

from the CAARS: Self-Report, WFIRS impairment mean scores, and the SLC-90 Global Severity Index. Analysis of variance (ANOVA) and chi-square tests were conducted to determine if the derived subgroups differed on these external variables. In response to significant ANOVA findings, subsequent post hoc comparisons (Games-Howell procedure) were conducted to determine which TOVA clusters differed (see Table 3).

MSVT performance differences based on TOVA cluster was conducted by translating the MSVT scores for Immediate Recognition, Delayed Recognition, and Consistency to binomial Pass or Fail (score of 85 or below) scores. Figure 1 illustrates the percentage of individuals who failed each Index by TOVA clusters. More of those in the Low group failed the Immediate Recognition subtest (26%) compared to the High (1.3%) and the Mixed (1.2%) groups, $\chi^2_{(4)} = 55.38, p < .001$. Similarly, the Low group failed the Delayed Recognition subtest at a much higher rate (39.7%) than the High (4.0%) or Mixed (6.0%) groups, $\chi^2_{(4)} = 60.58, p < .001$. Furthermore, 38.3% of those in the Low group failed the Consistency Index, which was a significantly higher percentage than the High (4.0%) and the Mixed (12.0%) groups, $\chi^2_{(4)} = 48.10, p < .001$. As a group, 43.8% of the Low group failed at least one MSVT subtest when compared to 4.9% of the High and 13.3% of the Mixed group, $\chi^2_{(2)} = 53.62, p < .001$.

The second group of variables to be compared across the TOVA clusters was participant’s reported AD/HD symptomology on the CAARS: Self-Report. Significant differences were found across all three DSM-IV scales (Inattentive $F_{(2, 299)} = 7.42, p = .001$, Cohen’s $d = .544$;

Hyperactive-Impulsive $F_{(2, 300)} = 9.63, p = .000$, Cohen’s $d = .619$; Total Symptoms $F_{(2, 300)} = 11.74, p = .000$. Cohen’s $d = .687$), the AD/HD Index score ($F_{(2, 300)} = 9.60, p = .001$, Cohen’s $d = .619$), and the CII ($F_{(2, 216)} = 6.88, p = .001$, Cohen’s $d = .619$), but not the EI ($F_{(2, 221)} = 3.21, p = .042$), when a Bonferroni correction is applied for multiple analyses. Figure 2 illustrates the mean CAARS T -scores for each of the TOVA clusters.

Post-hoc analyses indicated that the High group endorsed fewer AD/HD symptoms than the Low group on all of the CAARS scales. Specifically, the Low cluster’s mean CAARS T -Scores were in the “clinically significant” range on all of the measures and sufficiently high to suggest symptom exaggeration (with mean T -scores above 80) on the DSM-IV Inattentive and ADH/HD-IV Symptom Total indices, while the High cluster’s mean scores were only “clinically significant” for the DSM-IV Inattentive and AD/HD-IV Symptom Total indices. The Low cluster’s mean CII score was also significantly higher than the High cluster, though neither group’s mean CII score exceeded the recommended cutoff (> 21). As their obtained T -scores and CII score fell between the High and Low groups, the Mixed group was not significantly different from either group. While not significantly different, the EI scores followed a similar trend to the CAARS Index and CII scores, with the High cluster attaining the lowest mean score (1.94), followed by the Mixed (2.03), and the Low group (2.60).

The CAARS CII and EI scores were then translated into binomial Pass and Fail scores. There was

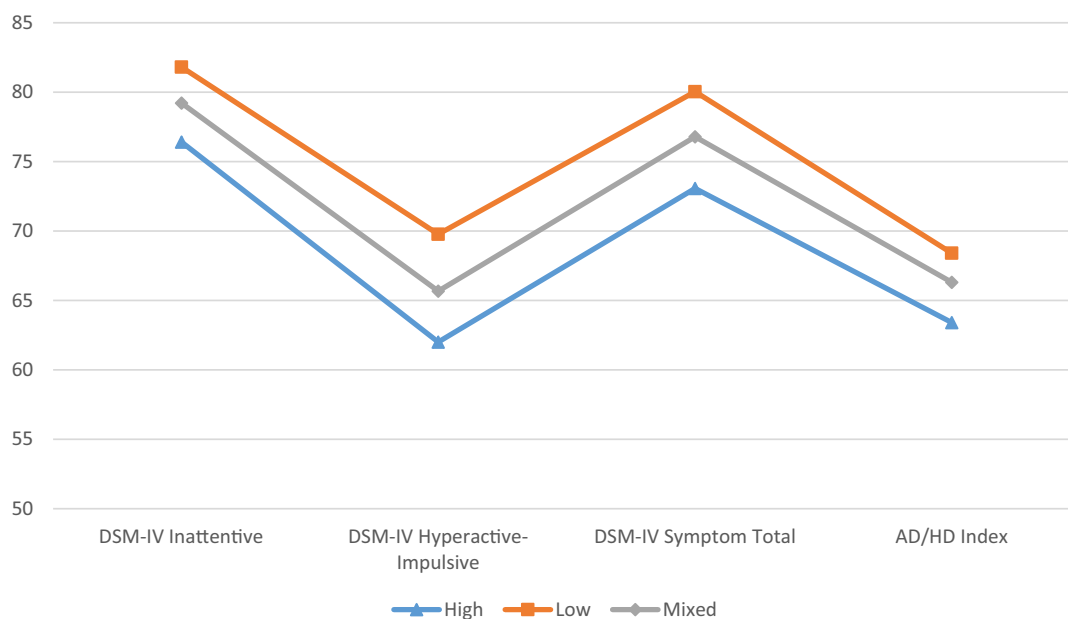


Figure 2. Mean AD/HD symptomology reported as a function of cluster membership.

a significant difference among the groups on the CII, $\chi^2_{(2)} = 7.97, p = .019$, with similar rates of failure (both at 12%) in the High and Mixed groups, but a significantly higher percentage of individuals attaining a score greater than 21 in the Low group (29%). There was no significant difference in failure rates among the groups on the EI, $\chi^2_{(2)} = 3.20, p = .202$, with 27% in the High group, 42% in the Low group, and 29% of the Mixed group attaining scores of 3 or above.

Mean impairment scores from the WFIRS were compared across the groups. There was a significant difference in reported impairment in the Home domain, $F_{(2, 194)} = 4.04, p = .019$, Cohen's $d = 0.5$, with the Low group reporting significantly more impairment than the Mixed group. Significant differences were not found in the remaining areas assessed.

To determine whether the clusters differed in terms of reported psychological symptoms, the SLC-90 Global Severity Index was compared across the three clusters. There was no discrepancy between the groups on reported mental health symptomology, $F_{(2, 158)} = 0.18, p = .838$.

Discussion

The purpose of the present study was to determine whether valid and reliable clusters would emerge based on scores from the TOVA. Results suggest that reliable patterns of TOVA Index scores can be derived using cluster analysis in post-secondary students referred for an AD/HD screening. Specifically, three reliable and clinically meaningful groups emerged.

The first group, labeled High, performed more similarly on the TOVA to those classified as not having AD/HD, with scores within normative limits. It should be noted that a small number ($n = 17$) of individuals in this group were ultimately diagnosed with AD/HD, suggesting that a non-clinical classification on the TOVA does not necessarily preclude a diagnosis of AD/HD. The validity index included in the TOVA (SEI) and their performance on the MSVT suggested credible performance and neither their calculated EI nor CII mean scores exceeded recommended cutoffs for suspected performance or symptom exaggeration. Thus, it seems that students captured in this cluster were generally not exaggerating or feigning difficulties. Their responses on the CAARS indicated some attentional concerns, though their mean AD/HD Index score was not in the clinical range. Their responses on measures of impairment and reported mental health symptomology did not differ from the other cluster groups.

The second cluster, named Low, attained scores on the TOVA indices that fell well below normative

expectations, with performance suggestive of an AD/HD classification. However, they displayed an elevated score on the SEI, raising the potential for non-credible performance, with their mean SEI score exceeding the cut off of 1 and more than half of the individuals classified in this cluster attaining an SEI score of 2 or above. Performance on the MSVT was also suggestive of performance validity issues. The possibility of symptom exaggeration was evident on the CAARS: this group attained the highest scores on the CAARS CII and EI and their mean scores on the indices assessing for Inattention and Symptom Total exceeded 80. Thus, it seems that individuals captured in this group were more likely to be exaggerating or feigning difficulties. This group reported experiencing a higher level of functional impairment in the home than the Mixed group. Their reported level of psychological distress did not differ from the other groups.

Performance on the TOVA indices for the third cluster, Mixed, was variable, with normal scores on the scales assessing commission errors, a borderline normal score on the first half and a normal score on the second half of the scale assessing RT, and scores not within normative limits on measures of RT variability and omission. It should be noted, however, that while the Mixed and Low groups both produced impaired scores on the RT variability and omission measures, the Low group performed at least four SD below the mean on both indices, in both halves of the assessment, while the Mixed group performed three SD below the mean only on the first half of RT variability and just below two SD from the mean on RT variability in the second half and omission errors across time points. The mean ACS score for the Mixed group was similar to individuals with AD/HD (-3.99) but not as profoundly low as the Low group. Their mean SEI score was low (0.67) and their mean MSVT scores suggested credible performance. Neither their calculated EI or CII mean scores exceeded recommended cutoffs for suspected symptom invalidity. Their mean scores across index scores on the CAARS were clinically significant but did not suggest possible symptom exaggeration. This group contained the largest percentage of individuals who were ultimately diagnosed with AD/HD. Thus, it could be concluded that this group contained individuals who were displaying some attentional impairment on the TOVA but were generally performing credibly. It should be noted, however, that one student who ultimately admitted to malingering was classified in this group and, as such, the group was not entirely free of individuals who may have been engaging in some symptom or performance exaggeration. Individuals in the Mixed group reported less impairment in the home when compared to the Low

group. They did not differ from the other two groups in terms of reported level of psychological distress.

Conclusions

In agreement with Robinson and Rogers (2018), the results suggest that good faith assumptions that all AD/HD referrals will put forth their best effort appear unwarranted. By using cluster analysis to identify patterns of performance on the TOVA, a group of participants emerged (about one-quarter of the sample) with globally low TOVA performance, high levels of AD/HD symptoms, and scores suggestive of issues with symptom and performance validity. Consistent with Harrison et al.'s (2007) observation, the results suggest that feigners can produce assessment results that are suggestive of AD/HD, but they tend to overshoot the mark, attaining unrealistically exaggerated scores. In addition, as noted by P. Marshall et al. (2010), individuals in the Low cluster tended to take a sophisticated approach to malingering, producing extreme scores on measures which, from a face validity perspective, appear to be assessing for attentional difficulties, but not doing so on reports of impairment or psychological distress.

For the TOVA, our findings, similar to Harrison and Armstrong (2020), P. Marshall et al. (2010), and Nicholls et al. (2020), show that the group suspected of noncredible performance (Low group) made significantly more errors of omission and displayed increased RT variability during both halves of the TOVA. While the Mixed group also demonstrated impaired scores in these domains, the magnitude was significantly different, with the Low group consistently attaining scores farther below the mean. As such, the results suggest that clinicians should be wary when individuals attain exceptionally poor scores on TOVA indices (particularly scores more than three SD below the mean).

Consistent with previous findings (e.g., Harrison & Armstrong, 2020; Lark et al., 2002), differences in RT were noted across groups, with the Low group obtaining the most impaired scores. In comparison, the High group performed within normative limits on RT indices and the Mixed group displayed borderline normal results in the first half and a score within normative limits in the second half. A significant difference was also found among groups in terms of commission errors, with the Low group attaining the lowest scores, falling below one SD from the mean. Mean commission error scores were within normative limits in the High and Mixed groups. Lark et al. (2002) too, noted excessive errors of commission in their "fake bad" subjects. It could thus be concluded that consistently poor scores

on measures of RT and a large number of impulsive answers is somewhat unusual in credible responders and might suggest motivation to feign AD/HD symptoms. Further research is needed in this area.

The SEI from the TOVA performed well in detecting performance exaggeration, with 62.5% of the students in the Low cluster attaining an SEI score of 2 or above, in comparison to less than 1% of the Low group and 18% of the Mixed group. The mean scores on the MSVT and CAARS CII did not exceed recommended cutoffs for non-credible performance in the Low group. The mean EI score for the Low and Mixed groups surpassed a cut off of 2 but did not exceed a more stringent cut off of 3. Scores on these indices, however, are suggestive of performance and symptom invalidity in the Low group, providing some clinical utility in detecting non-credible performance. Furthermore, the recommendation that *T*-scores above 80 on CAARS indices suggest possible symptom overreporting appears valid, as these more extreme symptom reports were predominantly found in the Low group (for example, CAARS Total ADHD Index scores over 80 were 52.1% in Low group versus 27.3% and 38.2% in High and Mixed groups). While these measures are certainly performing as expected, more research to fine-tune cutoffs, particularly for more sophisticated malingerers, would be beneficial.

Finally, it is interesting to consider that while there were differences based on cluster membership in terms of reported AD/HD symptomology and measures of symptom and performance validity, there were no differences between the groups based on reported impairment or psychological distress. When considering mean WFIRS index scores, all groups reported experiencing clinically significant challenges with learning and self-concept, which may reflect their student status and possible frustration with self that brought them to the screening initially. With respect to the SCL-90 Global Severity Index, scores were fairly indistinguishable between groups, suggesting that students reported similar levels of psychological distress, regardless of cluster group membership. Thus, possibly due to their appearance of not being directly related to AD/HD symptomology, self-ratings of impairment and psychological distress do not appear to be of utility in detecting differences between those suspected of credible or non-credible performance.

It is important to consider the current investigation within the context of its methodological limitations. One possible limitation relates to sample characteristics. Although the sample used in this investigation was heterogeneous in that it included students presenting with attentional concerns likely

due to a variety of etiologies, the heterogeneity of the sample was limited, as it included only post-secondary students with attentional concerns. In addition, although using retrospectively gathered data enables researchers to carry out studies that may not be possible otherwise, such investigations are constrained by available data. Another limitation relates to changes in score reporting in different versions of the TOVA. Specifically, index scores below 40 were given as 0 in older versions of the test and provided as < 40 (which were represented as 40 in our database) in the newer versions of the test. As the majority of individuals attaining these extreme scores fell in the Low cluster group, mean Index scores on the TOVA in this cluster were likely affected by the changing representation of the lowest score that could be attained. Other limitations of the present investigation relate to the use of cluster analytic methodology. Despite attempts to ensure the reliability and validity of the derived typology, the fact remains that cluster analysis represents a relatively subjective research tool (Lange et al., 2002). Although efforts were made to ensure that selections regarding the similarity coefficient, grouping algorithm and association indexes followed conventional standards and were empirically derived, in the end, a somewhat subjective decision is required by the researcher to determine the metrics to be used. Additionally, with the use of cluster analysis, all participants in a sample are forced into clusters on the basis of relative similarity to other participants without consideration of similarity in an absolute sense (Hair & Black, 2000). Thus, the clusters generated in this investigation likely include some individuals who bear only a minimal similarity to the mean profile derived for that cluster. Furthermore, although Squared Euclidean Distance, the measure of similarity used in the current investigation, is the most commonly used similarity index in taxonomic research, it has been argued that the methodology that maximizes the influence of profile shape and minimizes the influence of profile magnitude may derive clusters that provide more meaningful information (Lange, 2007). Finally, from a clinical standpoint, the final decision about diagnosis of AD/HD was based on the clinical judgment of specific clinicians within the clinic along with results from measures included in the analyses. Future studies may wish to consider using methods to obtain an independent judgment regarding diagnosis. Considering that this investigation represents the first empirical attempt to delineate patterns of performance using the TOVA, it is necessary to

evaluate the reliability and validity of these findings through replication and cross-validation. Nevertheless, this study is the first to demonstrate that cluster analysis may be a useful alternative to simulation and known-groups approaches when conducting research on performance validity.

In conclusion, our investigation confirms that cluster analysis can identify reliable and clinically meaningful groups of young adults seeking initial assessment for possible AD/HD. Three profiles emerged, including one cluster who demonstrated exceptionally low performance on the TOVA and exceptionally high reporting of AD/HD symptomology. The implication from our analysis is that this group likely represents individuals who were exaggerating or magnifying their difficulties to obtain an AD/HD diagnosis. The results also reaffirm previous research showing that neither performance on a continuous performance test alone nor self-reported symptoms alone should be employed to diagnose AD/HD, as both types of tests can be feigned easily. Thus, results underscore the need for clinicians to include formal validity measures in their assessments and to consider symptom exaggeration when interpreting assessment results, particularly when obtaining more extreme scores.

Disclosure statement

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

Funding

The research was supported by the Ministry of Colleges and Universities of Ontario.

ORCID

Allyson G. Harrison  <http://orcid.org/0000-0002-0426-2011>

References

- Able, S. L., Johnston, J. A., Adler, L. A., & Swindle, R. W. (2007). Functional and psychosocial impairment in adults with undiagnosed AD/HD. *Psychological Medicine*, 37(1), 97–107. <https://doi.org/10.1017/S0033291706008713>
- Advokat, C., Martino, L., Hill, B. D., & Gouvier, W. (2007). Continuous Performance Test (CPT) of college students with ADHD, psychiatric disorders, cognitive deficits, or no diagnosis. *Journal of Attention Disorders*, 10(3), 253–256. <https://doi.org/10.1177/1087054706292106>
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., Text Revision).
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.).

- Baggio, S., Hasler, R., Giacomini, V., El-Masri, H., Weibel, S., Perroud, N., & Deiber, M.-P. (2020). Does the Continuous Performance Test predict ADHD symptoms severity and ADHD presentation in adults? *Journal of Attention Disorders*, 24(6), 840–848. <https://doi.org/10.1177/1087054718822060>
- Barkley, R. A. (2002). International consensus statement on ADHD: January 2002. *Clinical Child and Family Psychology Review*, 5(2), 89–111. <https://doi.org/10.1023/A:1017494719205>
- Barkley, R. A. (2008). Challenges in diagnosing adults with ADHD. *Journal of Clinical Psychiatry*, 69(12), e36. <https://doi.org/10.4088/JCP.1208e36>
- Barkley, R. A., Fischer, M., Smallish, L., & Fletcher, K. (2002). The persistence of attention-deficit/hyperactivity disorder into young adulthood as a function of reporting source and definition of disorder. *Journal of Abnormal Psychology*, 111(2), 279–289. <https://doi.org/10.1037/0021-843X.111.2.279>
- Barkley, R. A., Murphy, K. R., & Fischer, M. (2008). *AD/HD in adults: What the science says*. Guilford.
- Booksh, R. L., Pella, R. D., Singh, A. N., & Gouvier, W. D. (2010). Ability of college students to simulate ADHD on objective measures of attention. *Journal of Attention Disorders*, 13(4), 325–338. <https://doi.org/10.1177/1087054708329927>
- Bryant, A. M., Lee, E., Howell, A., Morgan, B., Cook, C. M., Patel, K., Menatti, A., Clark, R., Buelow, M. T., & Suhr, J. A. (2018). The vulnerability of self-reported disability measures to malingering: A simulated AD/HD study. *The Clinical Neuropsychologist*, 32(1), 109–118. <https://doi.org/10.1080/13854046.2017.1346145>
- Canadian Attention Deficit Hyperactivity Disorder Resource Alliance (CADDRA). (2011). *Canadian ADHD practice guidelines* (3rd ed.). Toronto, ON: Author. <http://caddra.ca/pdfs/caddraGuidelines2011.pdf>
- Canu, W. H., Hartung, C. M., Stevens, A. E., & Lefler, E. K. (2020). Psychometric properties of the Weiss Functional Impairment Rating Scale: Evidence for utility in research, assessment, and treatment of ADHD in emerging adults. *Journal of Attention Disorders*, 24(12), 1648–1660. <https://doi.org/10.1177/1087054716661421>
- Carone, D. A. (2009). Test review of the Medical Symptom Validity Test. *Applied Neuropsychology*, 16(4), 309–311. <https://doi.org/10.1080/09084280903297883>
- Conners, C. K., Erhardt, D., & Sparrow, E. (1998). *Conners' Adult ADHD Rating Scales (CAARS): ADHD across the life span*. Multi-Health Systems.
- Conti, R. P. (2004). Malingered ADHD in adolescents diagnosed with Conduct Disorder: A brief note. *Psychological Reports*, 94(3), 987–988. <https://doi.org/10.2466/pr0.94.3.987-988>
- Derogatis, L. R. (1994). *SCL-90-R : Symptom checklist-90-R : Administration, scoring & procedures manual* (3rd ed.). National Computer Systems.
- DuPaul, G. J., Weyandt, L. L., O'Dell, S. M., & Varejao, M. (2009). College students with ADHD: Current status and future directions. *Journal of Attention Disorders*, 13(3), 234–250. <https://doi.org/10.1177/1087054709340650>
- Fasmer, O. B., Mjeldheim, K., Forland, W., Hansen, A. L., Syrstad, V. E. G., Oedegaard, K. J., & Berle, J. Ø. (2016). Linear and non-linear analyses of Conner's Continuous Performance Test-II discriminate adult patients with attention deficit hyperactivity disorder from patients with mood and anxiety disorders. *BMC Psychiatry*, 16(1), Article 284. <https://doi.org/10.1186/s12888-016-0993-4>
- Fisher, A. B., & Watkins, M. W. (2008). ADHD rating scales' susceptibility to faking in a college student sample. *Journal of Postsecondary Education and Disability*, 20(2), 81–92. <http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ825775&site=ehost-live%5Cnhttp://www.ahead.org/publications/jped>
- Fletcher, J. M. (1985). External validation of learning disability typologies. In B. P. Rourke (Ed.), *Neuropsychology of learning disabilities: Essentials of subtype analysis* (pp. 187–211). Guilford Press.
- Frazier, T. W., Frazier, A. R., Busch, R. M., Kerwood, M. A., & Demaree, H. A. (2008). Detection of simulated AD/HD 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>
- Fuermaier, A. B. M., Tucha, L., Koerts, J., Mueller, A. K., Lange, K. W., & Tucha, O. (2012). Measurement of stigmatization towards adults with Attention Deficit Hyperactivity Disorder. *Plos One*, 7(12), e51755. <https://doi.org/10.1371/journal.pone.0051755>
- Fuermaier, A. B. M., Tucha, O., Koerts, J., Butzbach, M., Weisbrod, M., Aschenbrenner, S., & Tucha, L. (2018). Susceptibility of functional impairment scales to non-credible responses in the clinical evaluation of adult ADHD. *Clinical Neuropsychologist*, 32(4), 671–680. <https://doi.org/10.1080/13854046.2017.1406143>
- Garnier-Dykstra, L. M., Caldeira, K. M., Vincent, K. B., O'Grady, K. E., & Arria, A. M. (2012). Nonmedical use of prescription stimulants during college: Four-year trends in exposure opportunity, use, motives, and sources. *Journal of American College Health*, 60(3), 226–234. <https://doi.org/10.1080/07448481.2011.589876>
- Goodman, D. W. (2009). AD/HD in adults: Update for clinicians on diagnosis and assessment. *Primary Psychiatry*, 16(11), 21–30. <https://www.addadult.com/wp-content/uploads/2014/02/Goodman-DW.-2009.-ADHD-in-Adults-Update-for-Clinicians-on-Diagnosis-and-Assessment.pdf>
- Green, P. (2004). *Green's Medical Symptom Validity Test (MSVT) for Microsoft Windows: User's manual*. Green's Publishing.
- Greenberg, L., Kindschi, C., Dupuy, T., & Corman, C. (1996). *Test of Variables of Attention*. Universal Attention Disorders.
- Greenberg, L. M., Kindschi, C. L., Dupuy, T. R., & Hughes, S. J. (1994). *Test of variables of attention continuous performance test*. The TOVATM Company.
- Greenberg, L. M., & Waldmant, I. D. (1993). Developmental normative data on the Test of variables of attention (TOVA™). *Journal of Child Psychology and Psychiatry*, 34(6), 1019–1030. <https://doi.org/10.1111/j.1469-7610.1993.tb01105.x>
- Hair, J. F., & Black, W. C. (2000). Cluster analysis. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 147–205). American Psychological Association.
- Harrison, A. G. (2004). An investigation of reported symptoms of AD/HD in a university population. *AD/HD Report*, 12(6), 8–11. <https://doi.org/10.1521/adhd.12.6.8.55256>

- Harrison, A. G. (2006). Adults faking AD/HD: You must be kidding! *AD/HD Report*, 14(4), 1–7. <https://doi.org/10.1521/adhd.2006.14.4.1>
- 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., & Armstrong, I. T. (2020). Differences in performance on the Test of Variables of Attention between credible vs. noncredible individuals being screened for Attention Deficit Hyperactivity Disorder. *Applied Neuropsychology (Child)*, 9(4), 314–322. <https://doi.org/10.1080/21622965.2020.1750115>
- 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., Edwards, M. J., & Parker, K. C. H. (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., Green, P., & Flaro, L. (2012). The importance of symptom validity testing in adolescents and young adults undergoing assessments for learning or attention difficulties. *Canadian Journal of School Psychology*, 27(1), 98–113. <https://doi.org/10.1177/0829573512437024>
- Hughes, S. J., Lark, R. A., Henry, G. K., Robertson, E. L., & Greenberg, L. M. (2008, July). *Using the TOVA to detect deliberate poor performance during assessment of attention* [Poster presentation]. American College of Clinical Neuropsychology Annual Conference, Boston, MA.
- Jasinski, L. J., Harp, J. P., Berry, D. T. R., Shandera-Ochsner, A. L., 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>
- Jasinski, L. J., & Ranseen, J. D. (2011). Malingered ADHD evaluations: A further complication for accommodations reviews. *The Bar Examiner*, 80(4), 6–16. https://thebarexaminer.org/wp-content/uploads/PDFs/800411_Jasinski_Ranseen.pdf
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Lange, R. (2007). WAIS-III index score profiles in the Canadian standardization sample. *Journal of Clinical & Experimental Neuropsychology*, 29(1), 47–58. <https://doi.org/10.1080/13803390500409591>
- Lange, R. T., Iverson, G. L., Senior, G. J., & Chelune, G. J. (2002). A primer on cluster analysis applications to cognitive rehabilitation research. *Journal of Cognitive Rehabilitation*, 20(1), 16–33.
- Larrabee, G. J. (2007). *Assessment of malingered neuropsychological deficits*. Oxford University Press.
- Lark, R. A., Dixon, D., Hoffman, T., & Huynh, D. (2002). Fake bad test response bias effects on the test of variables of attention. *Archives of Clinical Neuropsychology*, 17(4), 335–342. <https://doi.org/10.1093/arclin/17.4.335>
- Lark, R. A., Greenberg, L. M., Kindschi, C. L., Dupuy, T. R., & Hughes, S. J. (2008). *T.O.V.A. professional manual: Test of variables of attention continuous performance test*. T.O.V. A Company.
- Manos, M. J. (2010). Nuances of assessment and treatment of ADHD in adults: A guide for psychologists. *Professional Psychology: Research and Practice*, 41(6), 511–517. <https://doi.org/10.1037/a0021476>
- 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>
- Marshall, P. S., Hoelzle, J. B., Heyerdahl, D., & Nelson, N. W. (2016). The impact of failing to identify suspect effort in patients undergoing adult Attention-Deficit/Hyperactivity Disorder (ADHD) assessment. *Psychological Assessment*, 28(10), 1290–1302. <https://doi.org/10.1037/pas0000247>
- Miller, T. W., Nigg, J. T., & Faraone, S. V. (2007). Axis I and II comorbidity in adults with AD/HD. *Journal of Abnormal Psychology*, 116(3), 519–528. <https://doi.org/10.1037/0021-843X.116.3.519>
- Millichap, J. G. (2008). Etiologic classification of Attention-Deficit/Hyperactivity Disorder. *Pediatrics*, 121(2), e358–e365. <https://doi.org/10.1542/peds.2007-1332>
- Morris, R., Blashfield, R., & Satz, P. (1981). Neuropsychology and cluster analysis: Potentials and problems. *Journal of Clinical Neuropsychology*, 3(1), 79–99. <https://doi.org/10.1080/01688638108403115>
- Mossman, D., Wygant, D. B., & Gervais, R. O. (2012). Estimating the accuracy of neurocognitive effort measures in the absence of a “Gold Standard”. *Psychological Assessment*, 24(4), 815–822. <https://doi.org/10.1037/a0028195>
- Murphy, K., & Barkley, R. A. (1996). Attention deficit hyperactivity disorder adults: Comorbidities and adaptive impairments. *Comprehensive Psychiatry*, 37(6), 393–401. [https://doi.org/10.1016/s0010-440x\(96\)90022-x](https://doi.org/10.1016/s0010-440x(96)90022-x)
- Musso, M. W., & Gouvier, W. D. (2012). Why is this so hard? A review of detection of malingered ADHD in college students. *Journal of Attention Disorders*, 18(3), 186–201. <https://doi.org/10.1177/1087054712441970>
- National Institute for Health and Care Excellence. (2018). *Attention deficit hyperactivity disorder: Diagnosis and management, clinical guideline*. Author. <https://www.nice.org.uk/guidance/NG87>
- Nicholls, C., Winestone, L., Ross, E., & Foley, M. (2020). Test of Variables of Attention among ADHD children with credible vs. noncredible PVT performance. *Applied Neuropsychology: Child*, 9(4), 307–313. <https://doi.org/10.1080/21622965.2020.1751787>
- Norwalk, K., Norvilitis, J. M., & MacLean, M. G. (2008). ADHD symptomatology and its relationship to factors associated with college adjustment. *Journal of Attention Disorders*, 13(3), 251–258. <https://doi.org/10.1177/1087054708320441>
- Nugent, K., & Smart, W. (2014). Attention-deficit/hyperactivity disorder in postsecondary students. *Neuropsychiatric Disease and Treatment*, 10(10), 1781–1791. <https://doi.org/10.2147/NDT.S64136>
- Osterberg, K., Karlson, B., & Orbaek, P. (2002). Personality, mental distress, and risk perception in subjects with multiple chemical sensitivity and toxic encephalopathy.

- Scandinavian Journal of Psychology*, 43(2), 169–175. <https://doi.org/10.1111/1467-9450.00283>
- Pella, R. D., Hill, B. D., Shelton, J. T., Elliott, E., & Gouvier, W. D. (2012). Evaluation of embedded malingering indices in a non-litigating clinical sample using control, clinical, and derived groups. *Archives of Clinical Neuropsychology*, 27(1), 45–57. <https://doi.org/10.1093/arclin/acr090>
- 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>
- Riccio, C. A., Wolfe, M., Davis, B., Romine, C., George, C., & Lee, D. (2005). Attention deficit hyperactivity disorder: Manifestation in adulthood. *Archives of Clinical Neuropsychology*, 20(2), 249–269. <https://doi.org/10.1016/j.acn.2004.07.005>
- Robinson, E. V., & Rogers, R. (2018). Detection of feigned ADHD across two domains: The MMPI-2-RF and CAARS for faked symptoms and TOVA for simulated attention deficits. *Journal of Psychopathology and Behavioral Assessment*, 40(3), 376–385. <https://doi.org/10.1007/s10862-017-9640-8>
- Skydsbjerg, M., Lunn, S., & Hutchings, B. (2001). Psychosocial aspects of human immunodeficiency virus (HIV) infection in a pre-HAART sample. *Scandinavian Journal of Psychology*, 42(4), 327–333. <https://doi.org/10.1111/1467-9450.00244>
- Sobanski, E., Brüggemann, D., Alm, B., Kern, S., Deschner, M., Schubert, T., Philipsen, A., & Rietschel, M. (2007). Psychiatric comorbidity and functional impairment in a clinically referred sample of adults with attention-deficit/hyperactivity disorder (AD/HD). *European Archives of Psychiatry and Clinical Neuroscience*, 257(7), 371–377. <https://doi.org/10.1007/s00406-007-0712-8>
- Sollman, M. J., Ranssen, J. D., & Berry, D. T. (2010). Detection of feigned ADHD in college students. *Psychological Assessment*, 22(2), 325–335. <https://doi.org/10.1037/a0018857>
- Suhr, J., Hammers, D., Dobbins-Buckland, K., Zimak, E., & Hughes, C. (2008). The relationship of malingering test failure to self-reported symptoms and neuropsychological findings in adults referred for AD/HD evaluation. *Archives of Clinical Neuropsychology*, 23(5), 521–530. <https://doi.org/10.1016/j.acn.2008.05.003>
- 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., Sullivan, B. K., & Rodriguez, J. L. (2010). The relationship of noncredible performance to continuous performance test scores in adults referred for Attention-Deficit/Hyperactivity Disorder evaluation. *Archives of Clinical Neuropsychology*, 26(1), 1–7. <https://doi.org/10.1093/arclin/acq094>
- Sullivan, B. K., May, K., & Galbally, L. (2007). Symptom exaggeration by college adults in AD/HD and learning disorder assessments. *Applied Neuropsychology*, 14(3), 189–207. <https://doi.org/10.1080/09084280701509083>
- Thome, J., Ehrlis, A. C., Fallgatter, A. J., Krauel, K., Lange, K. W., Riederer, P., Romanos, M., Taurines, R., Tucha, O., Uzbekov, M., & Gerlach, M. (2012). Biomarkers for Attention-Deficit/hyperactivity Disorder (ADHD). A consensus report of the WFSBP Task Force on biological markers and the world federation of ADHD. *The World Journal of Biological Psychiatry*, 13(5), 379–400. <https://doi.org/10.3109/15622975.2012.690535>
- Tucha, L., Fuermaier, A. B. M., Koerts, J., Groen, Y., & Thome, J. (2014). Detection of feigned Attention Deficit Hyperactivity Disorder. *Journal of Neural Transmission*, 122(S1), 123–134. <https://doi.org/10.1007/s00702-014-1274-3>
- Tucha, L., Sontag, T. A., Walitza, S., & Lange, K. W. (2009). Detection of malingered Attention Deficit Hyperactivity Disorder. *Attention Deficit and Hyperactivity Disorders*, 1(1), 47–53. <https://doi.org/10.1007/s12402-009-0007-1>
- Wasserstein, J. (2005). Diagnostic issues for adolescents and adults with AD/HD. *Journal of Clinical Psychology*, 61(5), 535–547. <https://doi.org/10.1002/jclp.20118>
- Weiss, M. D. (2000). *Weiss Functional Impairment Rating Scale (WFIRS) self-report*. University of British Columbia. <https://www.caddra.ca/wp-content/uploads/WFIRS-S.pdf>
- Weyandt, L. L., & DuPaul, G. J. (2012). Introduction to special series on college students with ADHD: Psychosocial issues, comorbidity, and treatment. *Journal of Attention Disorders*, 16(3), 199–201. <https://doi.org/10.1177/1087054711427300>
- Williamson, K. D., Combs, H. L., Berry, D. T. R., Harp, J. P., Mason, L. H., & Edmundson, M. (2014). Discriminating among ADHD alone, ADHD with a comorbid psychological disorder, and feigned ADHD in a college sample. *Clinical Neuropsychologist*, 28(7), 1182–1196. <https://doi.org/10.1080/13854046.2014.956674>
- Young, J. C., & Gross, A. M. (2011). Detection of response bias and noncredible performance in adult Attention-Deficit/Hyperactivity Disorder. *Archives of Clinical Neuropsychology*, 26(3), 165–175. <https://doi.org/10.1093/arclin/acr013>