Applied Neuropsychology: Child

ISSN: 2162-2965 (Print) 2162-2973 (Online) Journal homepage: www.tandfonline.com/journals/hapc20

Differences in performance on the test of variables of attention between credible vs. noncredible individuals being screened for attention deficit hyperactivity disorder

Allyson G. Harrison & Irene T. Armstrong

To cite this article: Allyson G. Harrison & Irene T. Armstrong (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, DOI: 10.1080/21622965.2020.1750115

To link to this article: https://doi.org/10.1080/21622965.2020.1750115

Published online: 17 Apr 2020.

APPLIED NEUROPSYCHOLOGY: CHILD
2020, VOL. 9, NO. 4, 314–322
https://doi.org/10.1080/21622965.2020.1750115

Routledge
Taylor & Francis Group

Differences in performance on the test of variables of attention between credible vs. noncredible individuals being screened for attention deficit hyperactivity disorder

Allyson G. Harrison and Irene T. Armstrong

Regional Assessment and Resource Centre, Queens University, Kingston, Canada

ABSTRACT
Measuring performance validity in Attention Deficit Hyperactivity Disorder (ADHD) assessments is essential, with multiple studies identifying how easily young adults can feign symptoms on self-report measures. Few methods, however, exist to identify such feigning when it occurs. While some clinicians include computerized tests of attention (e.g., Test of Variables of Attention [TOVA]) when assessing for possible ADHD, it is unclear how symptom exaggerators perform, and whether the TOVA Symptom Exaggeration Index (SEI) adequately identifies performance-based exaggeration when it occurs. Using archival data from a university-based ADHD screening clinic we investigated the performance of 245 late adolescents/emerging adults. Three groups were created: (1) Good effort but not ADHD ($n = 183$); (2) Good effort and diagnosed ADHD ($n = 13$); and (3) suspect effort ($n = 49$), based on final diagnosis and performance on an existing validity measure. Results showed clearly that those with suspect effort performed more poorly than the other two groups on all but second-half commission errors on the TOVA. Similar to Nicholls et al., the suspect effort group showed significantly subaverage (i.e., greater than two standard deviations below the mean) scores in Omission errors; in this replication, however, this was true for both the first and second half of the test. Response time variability was similarly exaggerated, with the suspect effort group again returning extreme scores in both halves of the test. Suspect effort students were indistinguishable from those with genuine ADHD when looking solely at self-reported symptoms; however, embedded symptom validity measures on an ADHD rating scale discriminated well between groups. Overall, results support the use of the TOVA as an embedded performance validity measure in the assessment of late adolescents/emerging adults and support previous findings that symptom report alone cannot distinguish credible from noncredible ADHD presentation.

KEYWORDS
ADHD; adolescent; diagnosis; validity; feigning; TOVA

Attention Deficit Hyperactivity Disorder (ADHD) is a common neurodevelopmental disorder that affects up to 9.4% of school-aged children (American Psychiatric Association (APA), 2013; Danielson et al., 2018) and up to 4.4% of adults in North America (Kessler et al., 2006). The inattention, hyperactivity and/or impulsivity symptoms associated with this disorder significantly impair individuals in two or more major life areas (APA, 2013).

The rate of children and adolescents being diagnosed with ADHD has skyrocketed in the last 20 years (Xu et al., 2018), more so in North America than in other developed countries (Smith, 2017). Specific diagnostic guidelines exist regarding how the diagnosis must be made (APA, 2013); however, questions have been raised about clinicians' lack of adherence to such criteria when making this diagnosis in young adults (Joy et al., 2010; Nelson et al., 2019; Weis et al., 2019), with studies finding that the majority of clinicians and physicians focus mainly on symptom report alone rather than ensuring that all five of the DSM criteria for diagnosis are met (Nelson et al., 2019; Weis et al., 2019). This lack of adherence to these published criteria is said to contribute to overidentification of ADHD (Gathje et al., 2008; Weis et al., 2019).

In addition to issues regarding comprehensiveness of assessment, research has also suggested that adolescents and young adults might feign or exaggerate these symptoms for reasons of secondary gain (Musso & Gouvier, 2014). Reasons for such behavior include access to academic accommodations such as extra test-taking time (Harrison, 2004, 2006); and access to stimulant medication for recreational purposes or as a study aid (Advokat et al., 2008; Hanson et al., 2013; McCabe et al., 2006; White et al., 2006), with many of these studies documenting the increase in illegal use, resale, or diversion of these medications by young adults. Because symptoms of ADHD are feigned easily (Harrison et al., 2007), and because most clinicians focus solely on current symptom report when diagnosing ADHD, it appears that it would be easy for adolescents or young

CONTACT Allyson G. Harrison harrisna@queensu.ca Regional Assessment and Resource Centre, Queen's University, 68 University Ave, Kingston, Ontario K7L 3N6, Canada.

adults to fool a clinician and obtain access to such accommodations and medications.

While clinical neuropsychologists are becoming more aware of and sensitive to the need to include formal validity tests in their assessment battery (Martin et al., 2015), this awareness has not translated into the assessment of children and adolescents (Brooks et al., 2016). In fact, the use of objective measures by school psychologists to identify noncredible presentations is "virtually non-existent" (DeRight & Carone, 2015, p. 3). Additionally, many studies find that almost no ADHD assessment reports submitted to colleges in order to verify the need for academic accommodations include any formal investigation of symptom or performance validity (e.g., Nelson et al., 2019; Weis et al., 2019). This suggests that many children or adolescents who exaggerate or feign symptoms of ADHD during an assessment could go undetected, potentially leading to inflated and inaccurate prevalence rates.

Compounding this problem further, few validity tests exist to help identify exaggeration of attention problems. Almost all existing performance validity tests (PVTs) were developed to identify noncredible memory scores in adults undergoing assessments in medical-legal or neuropsychological settings (Harrison et al., in press). While some of these tests have been shown to also identify noncredible performance in younger populations (Emhoff et al., 2018), it is not clear whether a child or adolescent attempting to exaggerate ADHD would necessarily perform poorly on a recognition memory test.

Although not originally intended for use as PVTs, some research has investigated whether continuous performance tests (CPT) might help differentiate between feigned and genuine ADHD in adults. Given that CPTs were designed to measure sustained attention, it is not surprising that ADHD simulators have been shown to perform worse than genuine patients on various types of CPTs (Booksh et al., 2010; Leark et al., 2002; Marshall et al., 2010; Quinn, 2003; Robinson & Rogers, 2018; Sollman et al., 2010). Marshall and colleagues (2010) found that the Test of Variables of Attention (TOVA; Greenberg et al., 1996) omission errors, TOVA reaction time variability, and Conners' Continuous Performance Test-II (Conners' CPT-II; Conners, 2000) omission errors achieved high specificity (ranging from 87% to 92%), but moderate sensitivity (ranging from 54% to 63%) in identifying suspect effort in ADHD evaluations. Robinson and Rogers (2018) similarly found that the use of cutoff scores for the TOVA response time variability that maintained high sensitivity at 100% only achieved 50% specificity. Overall, extant studies suggest that CPTs as performance measures may assist to some extent in identifying feigned ADHD, at least in adult samples.

Recently, Nicholls et al. (2020) report on the ability of specific scores from the TOVA to differentiate between children given a diagnosis of ADHD who either passed or failed an existing PVT (either the Word Memory Test (WMT: Green, 2003) or the Medical Symptom Validity Test (MSVT: Green, 2004)). They found that those who failed a PVT were not different than those who passed on Response Time (total or on either half of the test), nor did they differ in a total number of commission errors. By contrast, those who failed a PVT committed significantly more errors of omission during the first half, the second half, and overall on the TOVA. Further, although all children made more omission errors during the second half of the test, those who failed a PVT scored almost three standard deviations below the mean during the second half of the test, and about three standard deviations below the mean on the total test Omission score, suggesting that poor effort may be associated with this particular variable.

While a useful addition to the literature, the major difficulties with the Nicholls et al. study are (a) the small sample size, (b) no clinical controls included, and (c) that those who failed an existing PVT were still given a diagnosis of ADHD. The WMT and MSVT have been shown to be insensitive to genuine ADHD in children and adolescents (Harrison et al., 2015), and so one must question the accuracy of data produced by those children who failed one or more of these measures. A more parsimonious method of distinguishing groups would be to first obtain data from those who passed or failed a PVT, including those individuals where a diagnosis of ADHD was withheld due to evidence of poor test-taking effort.

The current study, therefore, sought to validate Nicholls et al.'s findings using a large dataset of late adolescents/emerging adults who were all evaluated for possible ADHD. We used archival data from students who sought an evaluation for ADHD, including those strongly suspected of symptom exaggeration. Using this archival dataset, the present study sought to investigate to what extent those who pass or fail existing PVTs demonstrate significant differences on various TOVA measures. Also of interest was the difference in self-reported ADHD symptoms between honest students compared with those strongly suspected of exaggeration, and whether existing symptom validity measures (SVTs) might assist in identifying symptom overreporting in this sample.

Methods

Participants

Participants in this research-approved study were students who were all referred to an ADHD screening clinic at a Canadian assessment center between 2013 and 2019. All students had never previously been diagnosed with ADHD but now believed that this might be the cause of their current reported difficulties.

In the present study, we identified all clients in an archival database who had agreed to allow their data to be used in research. Exclusion criteria for this investigation included clients who did not consent to have their data used in archival research, did not complete the TOVA, and/or were not administered a PVT during testing.

In total, there were 245 late adolescents/emerging adults (49.4% female) who were included in this study. Average age was 20.4 y (SD = 1.8; range 17–24 years). Most were enrolled in a 4-year university ($n = 182$) vs. a 2-year college

($n = 63$); the majority were in first or second year (57.1%). Information pertaining to ethnicity was not collected directly, however, the majority were Caucasian.

Materials

Medical Symptom Validity Test

The Medical Symptom Validity Test (MSVT; Green, 2004) is a shorter version of the WMT. Like the WMT, the MSVT measures Immediate Recognition (IR) and Delayed Recognition (DR) of a word list, as well as the consistency (CNS) of answers between the two subtests. According to the test manual, scores less than 85 on any of the first three subtests indicate noncredible performance. It has been shown to have high sensitivity and specificity in identifying both good effort as well as simulation of impairment in children and adolescents (Blaskewitz et al., 2008).

Test of Variables of Attention

The TOVA is a computerized CPT used to measure attention, concentration, and inhibitory control and can be used with children as young as 6 years of age (Leark et al., 2008). The visual version was used for the current study. Geometric stimuli are used in the test to minimize the effects of cultural differences or language comprehension problems, and contains two test conditions; target infrequent and target frequent. In the first half of the test (target infrequent), the target to non-target ratio is 1 to 3.5, meaning that the task is boring and the subject must pay close attention to respond correctly to infrequent targets. If the subject fails to respond to a target it is labeled a omission error, whereas if they incorrectly respond to a nontarget item it is a commission error.

In the second half of the test (target frequent half), the target to nontarget ratio is 3.5 to 1, meaning that the subject expects to respond most of the time but must occasionally inhibit a response to a nontarget item.

Variables measured by the TOVA include Response time variability (consistency); response time (RT); commission errors (impulsivity); omission errors (inattention); and an Attention Comparison Score (ACS), which is a comparison of all obtained scores to an age/gender-specific ADHD group (with scores below zero indicating a greater resemblance to those with an existing ADHD diagnosis). In addition, the test also computes the Symptom Exaggeration Index (SEI: Hughes et al., 2008). This index was developed to identify elevated performance on indicators suggestive of exaggeration in those who deliberately fake bad; a score of 3 or more is said to indicate symptom exaggeration. The professional manual states that this index is only valid for ages 17 and up.

The Conners' Adult ADHD Rating Scale *(CAARS; Conners et al., 1998)*

The CAARS itself is a 66-item self-report scale that allows for the calculation of eight different indices, with some items contributing to more than one scale. CAARS items are rated on a 4-point scale (0 = *not at all*, 1 = *just a little*, 2 = *pretty much*, 3 = *very much*). All students completed the CAARS (self and observer forms), which provide the following scores: (a) four factor-derived subscales: (b) three scales that correspond to the DSM-IV symptoms of Hyperactivity/Impulsivity, Inattention, and Total DSM symptoms; and (c) an overall ADHD Index that is said to measure the "overall level of ADHD symptoms" (Conners et al., 1998, p. 23). The test manual states that the ADHD Index "is the best screen for identifying those '*at risk*' for ADHD" (Conners et al., 1998, p. 23). The manual does not stipulate a specific cutoff score that may be taken to indicate ADHD, but recommends that any score over a T value of 65 might be considered to indicate an area of clinically significant problems, and suggests that T-scores over 70 or 75 be used as a cutoff for inferring clinically significant problems. In addition, while the manual suggests that individuals obtaining T-scores on the ADHD Index of over 70 are likely to meet the diagnostic criteria for ADHD, it also cautions that T-scores above 80 on any of the subscales should be considered as possible indicators of symptom exaggeration. Unfortunately, no symptom validity scales currently exist within the CAARS; it includes only an Inconsistency Index designed to identify inconsistent responses to items measuring similar content rather than overreporting of symptoms. Scores on this index of 8 or greater are said to invalidate the obtained CAARS scores due to highly inconsistent responding by the subject.

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

The CAARS Infrequency Index (CII; Suhr et al., 2011) was retrospectively calculated from participant data where possible. The CII is composed of 12 items rarely endorsed by typically developing adults as well as those diagnosed with ADHD. Suhr and colleagues (2011) identified a cut score of >21 as producing few false positive identifications for those with ADHD. The index was found to have modest sensitivity (approximately 30%) and high specificity (approximately 95%). Cook et al. (2016) found that the CII had 52% sensitivity to feigning and 97% specificity for ADHD based on extreme elevations of the three CAARS clinical scales derived from DSM-IV ADHD criteria. Because item-level responses were not recorded for all clients, data from only $n = 230$ clients could be used in the present investigation.

The E-CAARS includes 18 additional symptom validity items embedded within the regular items, and has been described in detail by Harrison and Armstrong (2016). While allowing for all of the regular CAARS indices to be calculated, the embedded items included in this experimental version also allow clinicians to identify over-endorsement of ADHD-related symptoms. The sum of these symptom validity items produces a Dissimulation (DI) score, while a formula that combines items from the DI with extreme scores from existing CAARS indices produces the

Table 1. Means and SD of TOVA subscales as a function of group with significance level and partial eta square values for each ANOVA.

	Good effort not ADHD		Good effort ADHD		Suspect effort			
	183		13		49			
N	Mean	SD	Mean	SD	Mean	SD	*p* Value	η^2
TOVA subscale								
ACS*	−1.11	4.5	−1.83	3.0	−8.14	6.5	<.001	.246
SEI*	.54	.86	.54	.78	1.54	1.1	<.001	.158
Response time variability								
1st half*	72.8	31.9	69.8	30.9	37.5	28.4	<.001	.170
2nd half*	79.1	30.5	78.2	19.9	49.5	32.0	<.001	.132
Total*	72.8	32.7	71.3	20.2	37.0	30.9	<.001	.168
Response time								
1st half*	98.0	25.6	92.7	23.0	69.4	31.7	<.001	.154
2nd half*	106.0	20.9	109.4	20.4	87.2	30.4	<.001	.099
Total*	104.4	21.9	106.3	21.4	82.5	31.3	<.001	.119
Commission errors								
1st half*	99.8	18.6	101.2	14.6	79.3	28.1	<.001	.137
2nd half+	94.7	17.8	93.0	18.1	83.8	20.3	.002	.052
Total*	93.8	19.8	93.8	16.5	79.7	23.8	<.001	.070
Omission errors								
1st half*	77.9	29.7	69.0	33.2	41.9	21.1	<.001	.203
2nd half*	81.6	29.1	73.5	31.5	53.5	29.5	<.001	.127
Total*	77.2	30.6	69.0	31.0	46.8	25.1	<.001	.143

Note: *Post hoc analyses showed suspect effort group is significantly more extreme that either good effort group. +Post hoc analysis showed no differences between pairwise comparisons.

Exaggeration Index (EI). According to Harrison and Armstrong (2016), this Exaggeration Index (EI) is said to have acceptable classification accuracy when discriminating between those feigning ADHD and other clinical groups (including those with ADHD) who were reporting symptoms accurately, with sensitivity to feigning ranging from .24 to .69 and specificity ranging from .74 to .97, depending on cut score used.

Procedure

All subjects had been given the TOVA and MSVT as part of an extensive screening battery. Additionally, subjects underwent a one-hour semi-structured interview, were asked to provide report cards from childhood, have their parents/caregiver complete a retrospective rating of DSM-based ADHD symptoms (number and frequency) observed in the client prior to age 12, complete both self- and observer-versions of the CAARS, complete various mental health screening measures, and provide evidence to document substantial impairment in more than one major life activity both prior to age 12 and currently. This allowed a determination as to whether they met all of the five criteria for diagnosis as outlined in DSM-IV (APA, 2000) or DSM-5 (APA, 2013), depending on year tested.

Using either admission of feigning ($n = 3$) or scores at or below 85 on Immediate Recall (IR), Delayed Recall (DR), or Consistency (CNS) to define noncredible performance, and focusing only on subjects between the ages of 17–24, three groups of subjects were created: (1) Good effort (i.e., passed the MSVT) but not diagnosed with ADHD ($n = 183$); (2) Good effort and subsequently diagnosed with ADHD-inattentive or combined type ($n = 13$); and (3) suspect effort ($n = 49$).

Results

A one-way ANOVA was conducted on the variables from the TOVA: ACS, SEI, as well as first half, second half, and total scores for Reaction Time Variability (RTV), Reaction Time (RT), commission errors, and omission errors.

As may be seen in Table 1, significant differences were found between all three groups on all TOVA variables. Post hoc analyses showed that, for all but commission errors in the second half of the test, the suspect effort group (Group 3) returned significantly lower standard scores than honest but symptomatic reporters. Within the honest performers, however, no difference was found between those who were eventually diagnosed with ADHD and those who were not.

Examining the magnitude of impairment demonstrated, it is interesting to note that the ADHD group's mean score on the ACS was just below zero, but with some individuals with ADHD diagnoses who performed above zero as well. With respect to the SEI (the embedded validity measure on the TOVA), the mean score for Group 3 was significantly higher than for honest reporters, but still well below the score of $>=3$ said to indicate definite malingering. While only 5.1% of the "good effort" groups scored in the suspect range on the SEI, 25% of the suspect effort group provided scores in that range. Given that the SEI is meant to measure exaggerated performance we assessed its sensitivity using our groups: Sensitivity = .25; specificity = .949, positive predictive value = .545; negative predictive value = .837. Faking prevalence in our sample is 19.8%. But if actual prevalence is only 10% our positive predictive value is reduced to .351 and the negative predictive value is .919.

Given that Nicholls et al. (2020) found a higher number of omission errors committed by those suspected of exaggeration, we next looked at the magnitude of impairment indicated by those in Group 3 compared with the honest responders. All variables on the TOVA except for ACS and

Table 2. Means and SD for critical CAARS items, experimental measures CII and EI, and p value and partial eta squared for each one way ANOVA.

	Good effort not ADHD		Good effort ADHD		Suspect effort			
	183		13		49			
n	Mean	SD	Mean	SD	Mean	SD	p Value	η^2
CAARS indices								
Inattention	68.8	10.9	70.2	10.3	73.55	9.0	.021*	.032
Hyperactivity	59.7	10.8	59.2	8.7	61.2	12.3	.651	.004
Impulsiveness	57.6	12.1	55.5	13.3	61.5	11.8	.099	.019
Self Concept	59.5	10.6	58.9	9.0	61.1	10.0	.613	.004
DSM Inattention	77.9	10.3	80.0	10.8	82.0	7.9	.034*	.027
DSM hyperactivity	64.0	13.2	63.9	11.1	68.1	77.8	.132	.017
ADHD total	74.9	10.7	77.8	10.8	79.4	8.4	.026*	.030
ADHD Index	64.6	8.5	63.6	9.3	67.4	8.2	.112	.018
Inconsistency	5.2	2.2	5.2	2.4	5.2	2.1	.986	<.001
Experimental validity measures								
N	127		12		37			
CII	15.1	5.9	14.9	5.1	17.2	5.9	.156	.021
EI	1.96	1.5	2.08	1.4	3.05	1.6	.001*	.079

Note: All post hoc pairwise analyses of CAARS subscales and CII showed no pairwise differences between groups. Post hoc analyses of EI showed that the "good effort not ADHD" group scored significantly lower than the "suspect effort group" while the ADHD group differed from neither.
*Statistically significant.

SEI provide standard scores with a mean of 100 and a standard deviation (SD) of 15. As may be seen in Table 1, while those with ADHD demonstrated significantly more omission errors than average (resulting in standard scores that are below average), those in Group 3 had mean omission error scores more than three standard deviations below average for both the first half, second half, and total. Of interest, this same pattern was noted for RT variability; while honest ADHD responders had RT variability that was more than two SDs below average in the first half and 1.4 SDs below average for the second half, those in Group 3 returned, on average, scores that were more than 4 SDs below average in the first half and total, and more than 3 SDs below average in the second half.

Table 2 shows that very few differences exist in self-reported symptoms reported by honest and suspect effort groups. A one-way ANOVA was conducted on the variables from the CAARS, as well as the two experimental embedded SVTs derived from the CAARS. Results show significant group differences in reporting of general inattention symptoms, DSM inattention symptoms, and ADHD total. Post hoc analyses showed no differences among the groups, likely due to the variable n's across groups. When examining the embedded SVTs, the ANOVA showed significant differences between groups only for the EI but not for the CII. Post hoc analyses showed that Group 3 returned significantly higher scores than the honest responders on EI, with the mean score for Group 3 falling just above the cut score said to indicate symptom exaggeration.

The CAARS manual suggests that T scores greater than 80 may be indicative of exaggerated performance. Examining all groups together, 47.3% of these students scores above 80 on the CAARS DSM Inattentive scale. Significantly more (67.9%) of those giving suspect effort had scores above 80 on this scale compared with both those in the ADHD group (52.9%) and those in the Good Effort but not ADHD group (41.6%), $X^2 = 12.6$, $p = .002$. Only 37.4% scored above 80 for the ADHD Total score. Here, it appears that those giving suspect effort scored above 80 at about the same rate as those who were diagnosed with ADHD, 51.8% and 58.8%, respectively. Of those considered in the Good Effort not ADHD category, only 32.1% scored above 80 on this measure, $X^2 = 10.9$, $p = .004$. No other subscale of the CAARS showed significant differences across groups for T scores greater than 80.

Diagnostic efficiency of TOVA and CAARS measures

Clinically relevant performance characteristics were evaluated using sensitivity, specificity, and positive and negative predictive power. These statistics assist clinicians in determining the likelihood that an individual whose score falls above or below a given cut-score actually belongs to the classification group in question. Sensitivity is the proportion of true positive diagnoses or probability of detection, while specificity is the proportion of true negatives (i.e., the percentage of honest subjects who are correctly identified). In order to ensure a low false-positive rate, the accepted convention for most validity research is that specificity should be set at .90 or higher (Boone, 2007; Larrabee, 2012; Vickery et al., 2001); even though this may reduce sensitivity to malingering, it is felt that this is a reasonable tradeoff to ensure that genuine clients are not falsely accused of malingering. Consequently, a validity test with high specificity is unlikely to be failed by someone with a genuine neurologic dysfunction (Larrabee, 2012).

Diagnostic efficiency and optimal cutoff scores for all investigated measures were examined with Receiver Operating Characteristic (ROC) curve analysis (for more on ROC curve analysis, see Mayer, 2004). Table 3 presents data from these analyses. One statistic we compared was the Area Under the Curve (AUC), an integrated measure of the sensitivity and specificity of a test at different possible cut scores. While there is no single, commonly agreed-upon standard for interpreting AUC values, the values, in general, indicate the chance that someone feigning ADHD would obtain a more extreme score than a randomly chosen person undergoing an ADHD assessment who was performing honestly (Lasko et al., 2005). In addition, AUCs of 0.7–0.79 are

Table 3. Receiver operator curve analysis of TOVA and CAARs measures.

Scale	AUC	Cut score for specificity of at least 90% (actual specificity)	Corresponding sensitivity for cut score
TOVA measures			
ACS	.797	−7.775 (.902)	.471
SEI	.730	3 (.947)	.240
RTV 1st half	.788	<40*	*
RTV 2nd half	.747	<40 (.928)	.220
RTV total	.780	<40*	*
RT 1st half	.749	61 (.906)	.340
RT 2nd half	.685	78 (.901)	.280
RT total	.716	77 (.905)	.360
CE 1st half	.731	79 (.919)	.440
CE 2nd half	.642	65 (.901)	.200
CE total	.674	63 (.901)	.280
OE 1st half	.818	<40 (.942)	.120
OE 2nd half	.731	<40 (.937)	.100
OE total	.752	<40 (.933)	.100
CAARS measures			
Inattention	.641	83 (.903)	.214
DSM IV inattention	.639	91*	*
ADHD total	.634	91*	*
EI	.704	5 (.906)	.268

Note: ACS: Attention Comparison Score; SEI: Symptom Exaggeration Index; RTV: Reaction time Variability; RT: Reaction Time; CE: Commission Errors; OE: Omission Errors; EI: Exaggeration Index. *Specificity did not reach .9 before maximum/minimum score thus no sensitivity available.

considered acceptable discrimination between groups, 0.8–0.89 excellent, and 0.9 or above outstanding (Sussman et al., 2019).

Although the AUC itself cannot provide the optimal cut score for each measure examined, we identified the coordinates of the ROC curve for each measure to find the cut score at which the measure had at least 90% specificity (with many having higher scores due to the imprecise nature of the curves and the skewed distribution of many of the TOVA scores). Table 3 shows the performance characteristics of the measures where significant differences were found.

As may be seen, only the AUC value for Omission Errors during the first half of the TOVA feel within the excellent discrimination range; ACS, SEI, RTV, RT and Commission errors in the first (but not second) half of the TOVA, Omission errors in the second half, and the EI score all achieved acceptable discrimination. All other values fell below this range.

For all of these scores, the cut score needed to achieve at least 90% specificity is listed in Table 3, along with the sensitivity values associated with those cut scores. It should be noted that because TOVA scores have a floor value of < 40, some measures' specificity is at floor value, thus no sensitivity score is shown (as it would be 100%). This is also the case for two CAARS items where the specificity of 90% or greater was only the single highest value. In all these cases the resultant sensitivity is not meaningful (as it is 100% but at the ceiling score).

As shown in Table 3, ACS and Commission Errors in the first half of the TOVA had the highest sensitivity at .47 and .44, respectively. Response Time Variability in the first half and total also achieve high specificity (well over 90%), identifying 100% of the suspect effort group, but as explained above, this was at cut scores at the extremes of the test metrics. Similarly, the highest scores possible one can achieve on the CAARS ($T > 90$) provide the best specificity for identifying feigning on the DSM-IV Inattentive Symptoms and ADHD Total scales, but relatively few individuals returned such extreme scores.

Discussion

This study sought to replicate the findings of Nicholls et al. (2020) using a larger and more diverse group of students where diagnosis of ADHD was withheld in cases of suspected invalidity. Consistent with the findings of Nicholls et al., late adolescents/emerging adults were found to indeed show differential performance on the TOVA, depending upon whether they passed or failed a standalone performance validity test (PVT). Individuals who failed a PVT committed significantly more errors of omission during the first half, the second half, and overall on the TOVA relative to honest performers, with the suspect effort group returning mean scores that were more than three SD's below average.

Unlike the Nicholls et al. study, however, differences in commission errors during the first half of the TOVA were also identified between suspect effort individuals and honest reporters; post-hoc analyses failed to show significant differences between effort groups for commission errors during the second half of the TOVA. Given that the rate of target presentation in the first half of the test is quite low (minimizing the pull for impulsive commission responses), a large number of impulsive answers during this first half is unexpected and appears to strongly suggest a motivation to feign ADHD symptoms.

Our study also found that both Response Time during the first half of the test and Response Time Variability in both halves appear to demonstrate significant exaggeration effects. In fact, the suspect effort group showed particularly extreme scores on this latter measure, both throughout the test and especially in the first half. Indeed, while the ADHD subjects scored approximately two SDs below standardization norms for RTV in the first half, the feigning group scored more than four SD's below the mean on the first half and overall for Response Time Variability. Given that

Response Time Variability is said to be a hallmark characteristic of those with ADHD (Greenberg et al., 1996), this finding underscores how easily those feigning ADHD can produce credible (yet extreme) performance patterns.

Of note, those in the suspect effort group obtained significantly lower scores on the ACS, the omnibus TOVA measure that is said to combine all TOVA indices to determine if the subject more strongly resembles normal individuals or those with an independent diagnosis of ADHD. Scores returned by the suspect effort group were significantly below zero (the cut score used to determine abnormality on the TOVA), whereas the ADHD group had, on average, scores much closer to zero. The ROC analysis further identified that any score below −7.8 on this variable has a strong likelihood of correctly identifying performance exaggeration, although many of the suspect effort group failed to score at this extreme range. Also of interest is that while the SEI score from the TOVA (the exaggeration index calculated from all the TOVA scores) had only 25% sensitivity to suspect effort, almost no one in the honest responding groups achieved a score of three or more. This should provide clinicians with greater confidence when declaring a score of this magnitude to be indicative of exaggerated responding.

ROC analyses also suggested that extreme scores on Omission Errors and Response Time Variability (below a standard score of 40) very strongly suggest exaggerated performance. Here, too, however, a number of the suspected feigners failed to score at such an extreme range, making the sensitivity of these cut scores relatively weak. A Commission Errors standard score of 79 or worse on the first half of the TOVA, by contrast, had relatively better sensitivity to feigned ADHD while keeping specificity at 90% or higher. This finding mirrors that of Robinson and Rogers (2018), who found an RTV score of less than 68 had high specificity to simulated malingering.

Interestingly, *T* scores of 80 or more on the CAARS DSM-Inattention and ADHD Total indices failed to fully discriminate between those suspected of feigning and those reporting honestly. In fact, the ROC analysis suggested that a cut score of greater than 90 is required on these two indices to identify suspect effort, but due to the large overlap between groups this cut score resulted in low sensitivity to feigning. Similarly, the suggested cut score of three on the EI resulted in specificity below 90%; scores of five or more were required in order to achieve 90% specificity but again with lower sensitivity.

Results from the present study underscore the need for clinicians to include formal validity measures in their assessments. As has been shown in many other studies (see Musso & Gouvier, 2014 for a review of this literature), those suspected of feigning ADHD were easily able to obtain overall scores on the TOVA that fell within the range said to be consistent with ADHD; the majority of these suspected feigners were not flagged by the SEI, the TOVA algorithm meant to catch symptom exaggeration. The suspect effort group also produced profiles on a self-report checklist that mimicked or exceeded the scores found in students with genuine ADHD. On every subscale of the CAARS the mean score of the suspect effort group matched or exceeded that of the ADHD group in terms of reported symptoms. Furthermore, those symptomatic students who believed they had ADHD but did not receive a diagnosis were also often indistinguishable from the true ADHD group, with respect to both self-reported symptoms and also on the ACS and other measures from the TOVA. This underscores the need to base a diagnosis of ADHD on more than just self-reported symptoms or performance on the TOVA, as both measures have high rates of false positive diagnosis (e.g., Harrison et al., 2019; Gualtieri & Johnson, 2005; Musso & Gouvier, 2014; Robinson & Rogers, 2018).

Like most clinical research, this study has a number of minor limitations. First, there were relatively few students who actually received a diagnosis of ADHD. We did, however, have a large number of other students who all complained of ADHD-like symptoms, performed at levels similar to those who were eventually diagnosed, and passed a PVT. As such, our results help demonstrate how best to identify those who are feigning symptoms of ADHD in the context of a screening assessment; further research needs to be done in other contexts.

Second, we identified the suspect effort group based solely on non-credible performance on the MSVT. The difficulty, of course, is that one cannot know for certain whether these suspected individuals were truly malingerers. This, in turn, poses a threat to internal validity; some false positives may be included in this study, therefore diluting the sample. Recent research, however, suggests that the MSVT has a low false positive rate when evaluating children and adolescents for possible ADHD (Harrison et al., 2015), and good sensitivity and specificity for identification of noncredible performance among children with other health conditions (e.g., Carone et al., 2014; Green & Flaro, 2003; Kirkwood & Kirk, 2010). It is also possible, however, that using only the MSVT to identify the suspect effort group failed to accurately identify all individuals who were feigning symptoms of ADHD. This is a problem that plagues all investigations of invisible disabilities that have no objective diagnostic methods available to independently confirm group membership.

In conclusion, our investigation confirms the findings of Nicholls et al. that specific TOVA scores may be valuable as potential embedded effort measures. Those students who failed a stand-alone PVT returned differentially weaker performance on measures of inattention from the TOVA, scoring between three to four standard deviations below the standardization norms on some of the TOVA metrics, and exaggerating self-reported symptoms as well. They also reaffirm previous research showing that neither performance on a continuous performance test alone nor self-reported symptoms alone should be employed to diagnose ADHD, as both types of tests can be feigned easily.

References

Advokat, C. D., Guidry, D., & Martino, L. (2008). Licit and illicit use of medications for attention-deficit hyperactivity disorder in undergraduate college students. *Journal of American College Health*, *56*(6), 601–606. https://doi.org/10.3200/JACH.56.6.601-606

American Psychiatric Association (APA). (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Author.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (4th ed. – Text Revision)* (5th ed.). Washington, DC: Author.

Blaskewitz, N., Merten, T., & Kathmann, N. (2008). Performance of children on symptom validity tests: TOMM, MSVT, and FIT. *Archives of Clinical Neuropsychology*, *23* (4), 379–391. https://doi.org/10.1016/j.acn.2008.01.008

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

Boone, K. B. (2007). *Assessment of feigned cognitive impairment. A neuropsychological perspective*. Guilford.

Brooks, B. L., Ploetz, D. M., & Kirkwood, M. W. (2016). A survey of neuropsychologists' use of validity tests with children and adolescents. *Child Neuropsychology*, *22*(8), 1001–1020. https://doi.org/10.1080/09297049.2015.1075491

Carone, D. A., Green, P., & Drane, D. L. (2014). Word Memory Test profiles in two cases with surgical removal of the left anterior hippocampus and parahippocampal gyrus. *Applied Neuropsychology: Adult*, *21*(2), 155–160. https://doi.org/10.1080/09084282.2012.755533

Cook, C., Bolinger, E., & Suhr, J. A. (2016). Further validation of the CAARS Infrequency Index (CII) for detection of noncredible report of ADHD symptoms. *Archives of Clinical Neuropsychology*, *31*(4), 358–364. https://doi.org/10.1093/arclin/acw015

Conners, C. (2000). *Conners Continuous Performance Test II*. Multi-Health Systems, Inc.

Conners, C. K., Erhardt, D., & Sparrow, E. (1998). *Conners' Adult ADHD Rating Scales (CAARS): ADHD across the life span*. Multi-Health Systems.

Danielson, M. L., Bitsko, R. H., Ghandour, R. M., Holbrook, J. R., Kogan, J. D., & Blumberg, S. J. (2018). Prevalence of parent-reported ADHD diagnosis and associated treatment among U.S. children and adolescents, 2016. *Journal of Clinical Child & Adolescent Psychology*, *47*(2), 199–212. https://doi.org/10.1080/15374416.2017.1417860

DeRight, J., & Carone, D. A. (2015). Assessment of effort in children: A systematic review. *Child Neuropsychology*, *21*(1), 1–24. https://doi.org/10.1080/09297049.2013.864383

Emhoff, S., Lynch, J., & McCaffrey, R. (2018). Performance and symptom validity testing in pediatric assessment: A review of the literature. *Developmental Neuropsychology*, *43*(8), 671–707. https://doi.org/10.1080/87565641.2018.1525612

Gathje, R. A., Lewandowski, L. J., & Gordon, M. (2008). The role of impairment in the diagnosis of ADHD. *Journal of Attention Disorders*, *11*(5), 529–537. https://doi.org/10.1177/1087054707314028

Green, P. (2003). *Word Memory Test for Windows: User's manual and program*. Green's Publishing.

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

Green, P., & Flaro, L. (2003). Word Memory Test performance in children. *Child Neuropsychology*, *9*(3), 189–207. https://doi.org/10.1076/chin.9.3.189.16460

Greenberg, L., Kindschi, C., Dupuy, T., & Corman, C. (1996). *Test of Variables of Attention*. Universal Attention Disorders.

Gualtieri, C. T., & Johnson, L. G. (2005). ADHD: Is objective diagnosis possible? *Psychiatry (Edgmont (Pa.: Township))*, *2*(11), 44–53.

Hanson, C. L., Burton, S. H., Giraud-Carrier, C., West, J. H., Barnes, M. D., & Hansen, B. (2013). Tweaking and tweeting: Exploring twitter for nonmedical use of a psychostimulant drug (Adderall) among college students. *Journal of Medical Internet Research*, *15*(4), e62. https://doi.org/10.2196/jmir.2503

Harrison, A. G. (2004). An investigation of reported symptoms of ADHD in a university population. *The* ADHD *Report*, *12*(6), 8–11. https://doi.org/10.1521/adhd.12.6.8.55256

Harrison, A. G. (2006). Adults faking ADHD: You must be kidding! *ADHD. The* ADHD *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., 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., Flaro, L., & Armstrong, I. (2015). Rates of effort test failure in children with ADHD: An exploratory study. *Applied Neuropsychology: Child*, *4*(3), 197–210. https://doi.org/10.1080/21622965.2013.850581

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.). NY: Guilford Press.

Harrison, A. G., Nay, S., & Armstrong, I. T. (2019). Diagnostic accuracy of the Conners' adult ADHD rating scale in a postsecondary population. *Journal of Attention Disorders*, *23*(14), 1829–1837. https://doi.org/10.1177/1087054715625299

Hughes, S. J., Leark, 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]. Poster paper at the American College of Clinical Neuropsychology Annual Conference, Boston, MA.

Joy, J., Julius, R. J., Akter, R., & Baron, D. (2010). Assessment of Attention Deficit Hyperactivity Disorder (ADHD) documentation from candidates requesting Americans with Disabilities Act (ADA) accommodations for the National Board of Osteopathic Medical Examiners COMLEX exam. *Journal of Attention Disorders*, *14*(2), 104–108. https://doi.org/10.1177/1087054710365056

Kessler, R. C., Adler, L., Barkley, R., Biederman, J., Conners, C. K., Demler, O., Faraone, S. V., Greenhill, L. L., Howes, M. J., Secnik, K., Spencer, T., Bedirhan Ustun, T., Walters, E. E., & Zaslavsky, A. M. (2006). The prevalence and correlates of adult ADHD in the United States: Results from the National Comorbidity Survey replication. *American Journal of Psychiatry*, *163*(4), 716–723. https://doi.org/10.1176/ajp.2006.163.4.716

Kirkwood, M. W., & Kirk, J. W. (2010). The base rate of suboptimal effort in a pediatric mild TBI sample: Performance on the medical symptom validity test. *The Clinical Neuropsychologist*, *24*(5), 860–872. https://doi.org/10.1080/13854040903527287

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society*, *18*(4), 625–630. https://doi.org/10.1017/S1355617712000240

Lasko, T., Bhagwat, J., Zou, K., & Ohno-Machado, L. (2005). The use of receiver operating characteristic curves in biomedical informatics. *Journal of Biomedical Informatics*, *38*(5), 404–415. https://doi.org/10.1016/j.jbi.2005.02.008

Leark, 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.

Leark, 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

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., & Odland, A. P. (2015). Neuropsychologists' validity testing beliefs and practices: A survey of

North American professionals. *The Clinical Neuropsychologist*, *29*(6), 741–776. https://doi.org/10.1080/13854046.2015.1087597

Mayer, D. (2004). *Essential evidence-based medicine*. Cambridge University Press.

McCabe, S. E., Teter, C. J., & Boyd, C. J. (2006). Medical use, illicit use and diversion of prescription stimulant medication. *Journal of Psychoactive Drugs*, *38*(1), 43–56. https://doi.org/10.1080/02791072.2006.10399827

Musso, M. W., & Gouvier, W. D. (2014). 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

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

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)*.

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

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

Smith, M. (2017). Hyperactive around the world? The history of ADHD in global perspective. *Social History of Medicine*, *30*, 767–787. https://doi.org/10.1093/shm/hkw127

Sollman, M. J., Ranseen, 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. 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

Sussman, Z.W., Peterson, R. L., Connery, A. K., Baker, D. A., & Kirkwood, M. W. (2019). Utility of matrix reasoning as an embedded performance validity indicator in pediatric mild traumatic brain injury. *Applied Neuropsychology: Child*, *8*(1), 70–75. https://doi.org/10.1080/21622965.2017.1382359

Vickery, C. D., Berry, D. T. R., Inman, T., Harris, M., & Orey, S. (2001). Detection of inadequate effort on neuropsychological testing: A meta-analytic review of selected procedures. *Archives of Clinical Neuropsychology*, *16*(1), 45–73. https://doi.org/10.1093/arclin/16.1.45

Xu, G., Strathearn, L., Liu, B., Yang, B., & Bao, W. (2018). Twenty-year trends in diagnosed attention-deficit/hyperactivity disorder among US children and adolescents, 1997–2016. *JAMA Network Open*, *1*(4), e181471–9. https://doi.org/10.1001/jamanetworkopen.2018.1471

Weis, R., Till, C. H., & Erickson, C. P. (2019). ADHD assessment in college students: Psychologists' adherence to DSM-5 criteria and multi-method/multi-informant assessment. *Journal of Psychoeducational Assessment*, *37*(2), 209–225. https://doi.org/10.1177/0734282917735152

White, B. P., Becker-Blease, K. A., & Grace-Bishop, K. (2006). Stimulant medication use, misuse, and abuse in an undergraduate and graduate student sample. *Journal of American College Health*, *54*(5), 261–268. https://doi.org/10.3200/JACH.54.5.261-268