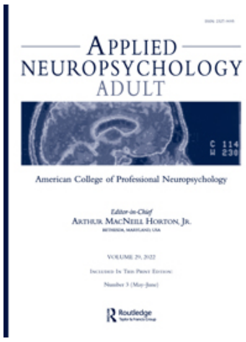

Applied Neuropsychology: Adult

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/hapn21

Using the Woodcock-Johnson IV tests of cognitive abilities to detect feigned ADHD

Laura M. Spenceley, Whitney L. M. Wood & Benjamin J. Lovett

APPLIED NEUROPSYCHOLOGY: ADULT
2022, VOL. 29, NO. 3, 324–332
https://doi.org/10.1080/23279095.2020.1748631

Using the Woodcock-Johnson IV tests of cognitive abilities to detect feigned ADHD

Laura M. Spenceley[a], Whitney L. M. Wood[b], and Benjamin J. Lovett[c]

[a]State University of New York at Oswego, Oswego, NY, USA; [b]Le Moyne College, Syracuse, NY, USA; [c]School Psychology, Teachers College of Columbia University, New York, NY, USA

ABSTRACT
Research has suggested that many young adults can successfully feign ADHD, reporting clinically significant symptom levels and displaying deficits on cognitive tasks when asked to do so. Standalone performance validity tests (PVTs) have shown some success in identifying feigned ADHD, but these tests are rarely used in typical ADHD evaluation batteries. The present study attempted to develop embedded PVT indices from the Woodcock-Johnson IV Tests of Cognitive Abilities (WJ-IV). College students ($N = 150$) completed a battery including tasks from the WJ-IV, as well as an established standalone PVT and a rating scale measuring ADHD and related symptoms. Thirty of the students had been professionally diagnosed with ADHD; of the remaining 120 students, half were asked to perform honestly and to the best of their ability on the battery, whereas the other half were asked to try to simulate ADHD. Several processing speed and working memory scores from the WJ-IV effectively identified students feigning ADHD, detecting at least 50% of those students at score cutoffs that also maintained specificity of 90% or more, close to the efficiency of the standalone PVT. In addition, students with ADHD diagnoses generally did not show deficits on the WJ-IV. Implications for practice and future research are discussed.

KEYWORDS
ADHD; performance validity; WJ-IV

It is now widely acknowledged that valid results from cognitive assessment tools require clients to be motivated to perform well, and that responsible clinicians should formally measure effort during a diagnostic evaluation (e.g., Boone, 2009; Suhr, 2015). Best practice guidelines from professional associations of neuropsychologists endorse such a practice (e.g., Heilbronner et al., 2009), and surveys have found that many neuropsychologists report including measures of effort in their assessment batteries (Brooks et al., 2016; Martin et al., 2015). The supportive evidence for this routine practice includes research suggesting that a substantial minority of clients do exaggerate symptoms or otherwise generate noncredible assessment data (e.g., Mittenberg et al., 2002).

The assessment of attention-deficit/hyperactivity disorder (ADHD) in young adults requires especially careful consideration of issues of client motivation and effort during evaluations (Harrison, 2006; Suhr & Berry, 2017). Not only must clinicians address the usual possibility of insufficient interest in working hard during a lengthy evaluation battery, but clients often also have an incentive to demonstrate impairment. First, an ADHD diagnosis often leads to treatment with stimulant medications—substances that are widely sought and used by individuals without ADHD, either for recreational purposes or to enhance motivation and productivity when studying or working (Drazdowski, 2016; Weyandt et al., 2016). Second, college students with ADHD diagnoses often obtain academic accommodations that may include additional time on tests and quizzes, a classmate paid to provide them with class notes, and even early registration for classes (Weyandt & DuPaul, 2013). These kinds of accommodations are perceived as advantageous by most students, with or without diagnoses (e.g., Lewandowski et al., 2014).

Prior research on feigned ADHD

Much of the early research on feigned ADHD focused on symptom rating scales; college students were asked to simulate ADHD, and their completed rating scales were scored to determine whether the students were able to report clinically significant symptom levels. The vast majority of students could do so (Fisher & Watkins, 2008; Jachimowicz & Geiselman, 2004). Other studies have asked college students to simulate ADHD symptoms on cognitive performance tasks, and nondisabled students can generally perform similarly to students with ADHD successfully (e.g., Harrison et al., 2007).

In parallel with the research on whether young adults can effectively feign ADHD, a second research line developed, this one attempting to differentiate feigners from individuals with genuine ADHD. Most readers will be familiar with symptom validity tests (SVTs), which measure honesty in reported symptoms, and performance validity tests (PVTs),

CONTACT Benjamin J. Lovett bl2799@tc.columbia.edu School Psychology, Teachers College of Columbia University, Teachers Coll., Box 120, 525 W. 120th St, New York, NY 10027, USA.

which measure effort and motivation to perform well on cognitive tasks (see Larrabee, 2012, for a discussion of these terms). Together, SVTs and PVTs have been described as assessing *response validity* (e.g., Nelson et al., 2007). Response validity measures have been applied by researchers to ADHD for over 15 years, and reviews of this literature (Musso & Gouvier, 2014; Tucha et al., 2015) have noted promising practices in detecting feigned or exaggerated ADHD, while also emphasizing the difficult challenges that evaluators face in this task. A recent meta-analysis (Wallace et al., 2019) focusing solely on simulation studies (rather than studies using archival patient records) and on college student samples used the effect size of the difference between genuine ADHD and feigning groups to index the efficacy of response validity measures. That paper found generally medium sized effects, except for standalone PVTs, which had a very large effect size of $g = 0.98$ (by comparison, PVTs embedded in other measures of cognitive functioning had an average effect size of $g = 0.66$).

In some studies, investigators have made a more direct estimation of the efficacy of response validity measures, by calculating their sensitivity and specificity at various cut points. Specificity is sometimes considered the more important of the two features, since clinicians would not want to make false accusations of feigning; specificity of 0.90 (90%) has therefore been suggested as a minimum value (Larrabee, 2012). Unfortunately, at the stringent cut points required for high specificity, many response validity measures have shown only low sensitivity—or occasionally moderate sensitivity—when used in ADHD contexts (for a review, see Tucha et al., 2015). Here again, standalone PVTs appear to have shown the most promise in identifying noncredible clinical presentations (see e.g., Booksh et al., 2010; Edmundson et al., 2017).

Response validity measures have also been used to try to estimate the prevalence of exaggeration or feigning during diagnostic evaluations in ecologically valid settings. The estimates have varied from one study to another, but rates of response validity test failure are often disturbingly high. For instance, Sullivan et al. (2007) found that 48% of college students presenting specifically for ADHD (and only ADHD) assessment failed at least one index on a standalone PVT. Similarly, Nelson and Lovett (2019) found that 53% of a large sample of college students undergoing a comprehensive psychoeducational evaluation (for concerns that included ADHD) triggered at least one PVT or SVT index. Marshall et al. (2016) found that 27% of clients referred for adult ADHD assessment triggered *two or more* PVT/SVT indices. Not all validity test failure represents poor effort, certainly, but findings such as these mean that motivation is at least questionable for many adult clients in the ADHD context.

Given that standalone PVTs have shown some efficacy in identifying feigned ADHD, and considering the significant incentives for obtaining an ADHD diagnosis, it is surprising to see that these measures are very rarely included in psychoeducational evaluations. For instance, Nelson et al. (2019) reviewed documentation from a sample of 100 postsecondary students seeking disability accommodations for ADHD, and found data from response validity measures in only 3 cases. It is possible that standalone measures are particularly unattractive to clinicians, given that they take time and money beyond that required for a standard evaluation; a new test must be purchased and then administered for the sole purpose of assessing response validity. If so, embedded measures may prove more acceptable to clinicians, although Wallace et al. (2019) found them to be less effective than standalone PVTs.

The present study

In the present study, we were interested in identifying new potential embedded PVT indices, hoping to find indices approximating the diagnostic efficiency (sensitivity and specificity) of a standalone PVT while being located within a test that clinicians might already be using as part of a diagnostic evaluation. We chose the Woodcock-Johnson IV Tests of Cognitive Abilities (WJ-IV; Schrank et al., 2014) as a potential source of embedded PVTs, since it is commonly administered as part of comprehensive psychoeducational evaluations, including those that lead to ADHD diagnoses (see e.g., Mather & Jaffe, 2010; Nelson et al., 2019). To be clear, our choice was *not* due to the WJ-IV's diagnostic utility in ADHD (indeed, it is not preferred for this purpose; see Fuermaier et al. 2019); however, psychologists often include an intelligence/cognitive ability measure as part of all psychoeducational evaluations.

The particular WJ-IV tests were chosen for multiple reasons. We wished to include measures of cognitive processing speed (*Gs*) and short-term/working memory (STM/WM; *Gwm*), since (a) some prior studies identified embedded PVTs that have been developed from such measures and (b) they relate logically and empirically to deficits that are traditionally associated with ADHD. We also wished to include measures that are relatively *free* from the influence of *Gwm* and *Gs*, to determine if individuals simulating ADHD restricted any feigning to the most relevant areas of ability, leaving their performance on measures of general fluid and particularly crystallized ability intact. We chose the Medical Symptom Validity Test (MSVT; Green, 2004), a well-established measure, as the standalone PVT against which to compare the WJ-IV. Finally, since a student's specific motive for exaggerating or feigning problems can affect responses made during an evaluation (Bryant et al., 2018), we provided one: testing accommodations to improve performance in college classes.

The principal research question in the present study was: Do any measures within the WJ-IV Tests of Cognitive Abilities effectively distinguish between simulated ADHD and independently diagnosed cases of ADHD? One secondary question was: Do college students with independent diagnoses of ADHD perform differently from nondisabled peers, when both groups are asked to respond honestly and perform with sufficient effort? Another secondary question was: Do simulators respond to self-report questionnaires of

ADHD and related symptoms differently than students with independent diagnoses of ADHD?

Method

Participants

The participants included 150 college students with ($n = 30$) and without ($n = 120$) diagnoses of ADHD, with each student attending either a midsized public university or a midsized private college in the northeastern United States. The sample had a mean age of 19.62 years ($SD = 1.69$), with the majority identifying as female ($n = 105$, 70.0%), and the largest proportion identifying as freshmen ($n = 62$, 41.3%). Participants were recruited from undergraduate classes via campus announcements, fliers, and in-person and email recruitment.

We did not perform independent ADHD evaluations on participants. Those participants who reported having received a professional ADHD diagnosis (i.e., from a physician, psychologist, or other health professional) confirmed their diagnosis at multiple time points (i.e., upon registering for the study, after reading the informed consent, and on a demographic form). Additionally, participants reporting an ADHD diagnosis were screened with an ADHD symptom checklist (described below) to ensure that they were still reporting significant levels of symptoms—at least 4 inattentive or 4 hyperactive/impulsive symptoms experienced often; this would place them at or above the 95th percentile according to the screener manual. The data from five participants were not used due to these participants reporting an ADHD diagnosis but failing to report sufficient current symptoms; this still left 30 participants in the Diagnosed ADHD group. Participants had no incentive to exaggerate symptoms or lie about having received a professional diagnosis, as they would have access to compensation for participating regardless of diagnostic status, and participation would not lead to a diagnosis that could be used for medication access or accommodations. Of the 30 students in the Diagnosed ADHD group, 13 (43.3%) reported having an active prescription for medication for ADHD, and seven (23.3%) reported having taken medication for ADHD on the day of the research study.

Participants without a diagnosis of ADHD were randomly assigned to one of two groups. Half of the participants comprised the Simulated group ($n = 60$) and were prompted to feign ADHD symptoms across tasks. The other half of participants served as controls ($n = 60$) and were instructed to complete the tasks using honest responding and typical effort. All participants with ADHD diagnoses ($n = 30$) were provided similar instructions to complete the tasks honestly with typical effort. The three groups were found to be comparable across demographic factors, including gender, $\chi^2(4, N = 150) = 8.76$, $p = 0.07$; year in school, $\chi^2(8, N = 150) = 3.41$, $p = .91$; ethnicity, $\chi^2(10, N = 150) = 11.69$, $p = .31$; age, $F(2, 147) = 0.76$, $p = .47$; and self-reported grade point average (GPA), $F(2, 127) = 1.48$, $p = .23$. Comprehensive demographic information of the sample can be found in Table 1.

Table 1. Demographic characteristics of participants.

	Control ($n = 60$)		Simulated ADHD ($n = 60$)		Diagnosed ADHD ($n = 30$)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Age	19.46	1.45	19.47	1.48	19.93	2.32
GPA	3.03	0.60	3.13	0.60	2.87	0.75
	n	%	*n*	%	*n*	%
Sex						
Female	46	76.7	44	73.3	15	50.0
Male	13	21.7	15	25.0	15	50.0
Other	1	1.7	1	1.7	0	0
Year in School						
Freshman	27	45.0	24	40.0	11	36.7
Sophomore	9	15.0	12	20.0	6	20.0
Junior	9	15.0	10	16.7	8	26.7
Senior	14	23.3	13	21.7	5	16.7
Graduate	1	1.7	1	1.7	0	0
Ethnicity						
Asian	3	5.0	3	5.0	0	0
African American	13	21.7	8	13.3	1	3.3
Hispanic/Latino	2	3.3	5	8.3	1	3.3
Hawaiian/Pacific	0	0	1	1.7	0	0
White	35	58.3	38	63.3	23	76.7
Multiracial	7	11.7	5	8.3	5	16.7

Measures

Demographic questionnaire

A demographic questionnaire was used to gather information from the participants regarding age, gender, ethnicity, GPA, primary language, year in college, current and previous psychological or medical diagnoses, and the use of medication prescribed for ADHD.

Woodcock-Johnson tests of cognitive abilities–fourth edition

The Woodcock-Johnson Tests of Cognitive Abilities–Fourth Edition (WJ-IV COG; Schrank et al., 2014) is a battery designed to assess a wide variety of Cattell-Horn-Carroll (CHC) cognitive abilities. For the purposes of this study, we administered the following tests: Letter-Pattern Matching, Verbal Attention, Numbers Reversed, Pair Cancellation, Oral Vocabulary, Number Series, General Information, and Concept Formation. We calculated scores for the individual tests as well as the Gf-Gc cluster, the Short-term/Working Memory (STM/WM; *Gwm*) cluster, the Cognitive Processing Speed (*Gs*) cluster, and the Cognitive Efficiency cluster.

The WJ-IV COG was normed on a large, nationally representative sample of 7,416 individuals, ranging in age from 2 to 90 years (McGrew et al., 2014). Across the clusters of interest, all demonstrate excellent reliability evidence (coefficients of 0.92 or higher). Individual test reliability estimates were always above 0.80, where available.[1] Furthermore, the WJ-IV COG has a wealth of supportive validity evidence, including strong correlations with scores from other cognitive ability measures as well as factor analytic evidence of a coherent internal structure (see McGrew et al., 2014, for a comprehensive review).

[1]The reliability of scores from time-pressured WJ-IV COG tests was estimated through test-retest studies on selected individuals in the normative sample, and the test-retest reliability coefficients were always above 0.80, although these studies were done on groups older and younger than the individuals in our sample.

Medical symptom validity test

Participants also completed the Medical Symptom Validity Test (MSVT; Green, 2004), a computerized task designed to identify suboptimal effort. The MSVT is a forced choice task designed to appear to measure memory, but in fact can be easily passed by individuals with a variety of cognitive impairments. The MSVT presents individuals with a set of ten word pairs twice, and then asks them to identify words from the original list (Immediate Recognition; IR). After a delay of ten minutes, a recognition task is completed using different foils (Delayed Recognition; DR). A Consistency (CS) score is also calculated based on how well the person performed across IR and DR conditions.

In the current study, we utilized the initial trials of the MSVT (yielding the IR, DR, and CS scores) to identify poor effort. Participants' effort was categorized as a pass or fail by utilizing a cutoff score of 85% on these three MSVT indices, per the recommended guidelines. The MSVT has been shown to demonstrate excellent diagnostic efficiency (Green et al., 2009; Singhal et al., 2009).

Barkley adult ADHD rating scale-IV

The Barkley Adult ADHD Rating Scale-IV (BAARS-IV; Barkley, 2011) is a self-report questionnaire of ADHD symptoms in adults. The BAARS-IV contains 18 ADHD items aligned with the DSM symptoms for ADHD diagnosis. These items include 9 inattentive symptoms, 6 hyperactive symptoms, and 3 impulsive symptoms. Nine items measuring symptoms of Sluggish Cognitive Tempo (SCT) are also included in the form. All participants responded to each item according to a 4-point scale (1 = "not at all," 2 = "sometimes," 3 = "often," and 4 = "very often"). Responses to these items yield separate total scores for inattention, hyperactivity, and impulsivity, as well as "symptom count" scores for the number of (a) inattentive and (b) hyperactive and impulsive items rated as occurring often or very often. The BAARS-IV was normed on a sample of 1,249 adults who were representative of the United States adult population. The scale's scores have good reliability (e.g., Cronbach's alpha coefficients: Inattention = 0.90; Hyperactive-Impulsive = 0.80; SCT = 0.90). Moreover, the BAARS-IV has validity evidence including correlations with other measures of relevant psychiatric symptoms, executive function problems, and a number of life outcomes (Barkley, 2011).

Procedures

Following IRB approval at each data collection site, each participant was individually assessed by one of the principal investigators or a trained graduate assistant. Following consent procedures, participants without ADHD diagnoses were randomly assigned to the Control or Simulated ADHD group. Both the Control and Diagnosed ADHD groups were given instructions to provide full effort and honest responding across all tasks. Participants assigned to the Simulated ADHD group were asked to pretend that they were undergoing an evaluation for ADHD due to (a) concerns over performance in college and (b) a belief that an ADHD diagnosis would lead to testing accommodations that they wanted to obtain. As in many prior studies of simulated ADHD, this group of participants was also given basic information about the symptoms and impairment associated with the disorder. Given that inattentive symptoms are considered most likely to persist into adulthood, several such symptoms were described, as well as "impulsive responding" and "problems with interpersonal relationships and employment" (see Appendix for full description of ADHD given to simulators).

Following consent and assignment to condition, all participants first completed the BAARS-IV. Then, participants were administered the WJ-IV COG tests in the order presented in the manual using standard administration procedures. Between the administration of the Letter-Pattern Matching and Verbal Attention tests, participants completed the initial trials of the MSVT (allowing for calculation of the IR, DR, and CS scores), and then the remaining WJ-IV COG tests were administered. At several points throughout the testing battery, participants in the Simulated ADHD group were reminded that their goal was to "realistically pretend that you are experiencing the effects of ADHD," whereas participants in the other groups were repeatedly reminded to respond honestly and to try their best on the tasks. At the end of the battery, participants completed the demographic form and were debriefed and compensated. Across conditions, participants had been offered a base pay of $20 for participation, and an additional $5 for "providing full effort" (Control and ADHD conditions) or "if you successfully pretend to have ADHD" (Simulated ADHD condition). In actuality, all participants who completed the study were paid $25 regardless of performance.

Results

Group comparisons

Table 2 shows the performance and responses of the three participant groups on all dependent measures from the study. After calculating descriptive statistics for each group on each measure, independent-groups t-tests were conducted (each with $df = 88$), comparing the group with diagnosed ADHD to (a) the control group and (b) the group with simulated ADHD.[2] The first comparison allowed us to determine the association between an ADHD diagnosis and performance, among students who were apparently responding honestly and putting forth sufficient effort on the tasks. The second comparison allowed us to determine how students simulating ADHD responded, relative to those with independent prior diagnoses of ADHD. We calculated effect sizes (Cohen's d values) for each comparison as well.

[2]Since many t-tests were conducted, the possibility of familywise Type-I error is raised. In all, 44 t-tests were conducted, and so the Bonferroni-corrected p-value for each statistical test would be approximately .001. However, only 1 of our 44 p-values was between .001 and .05, so the t-tests' statistical significance can be interpreted in essentially the same way whether or not a Bonferroni correction is applied.

Table 2. Descriptive statistics and group comparisons for dependent measures.

	Control ($n = 60$) M (SD)	Simulated ADHD ($n = 60$) M (SD)	Diagnosed ADHD($n = 30$) M (SD)	Control vs. Diagnosed ADHD p	d	Simulated vs. Diagnosed ADHD p	d
WJ-IV Oral Vocabulary	99.42 (11.04)	97.98 (12.27)	101.57 (11.38)	.39	−0.19	.19	0.30
WJ-IV Number Series	107.93 (12.83)	97.92 (15.17)	110.07 (13.24)	.46	−0.16	.00	0.85
WJ-IV Verbal Attention	100.08 (11.48)	77.65 (18.30)	102.50 (10.81)	.34	−0.22	.00	1.65
WJ-IV Letter Pattern Matching	100.73 (10.24)	74.73 (21.80)	99.97 (15.19)	.78	0.06	.00	1.34
WJ-IV General Information	92.78 (16.76)	91.65 (13.48)	96.27 (13.24)	.32	−0.23	.13	0.35
WJ-IV Concept Formation	101.08 (14.04)	94.82 (17.93)	104.73 (11.65)	.22	−0.28	.01	0.66
WJ-IV Numbers Reversed	101.27 (11.52)	81.43 (17.19)	103.97 (14.25)	.34	−0.21	.00	1.43
WJ-IV Pair Cancelation	100.85 (12.24)	81.15 (18.68)	103.97 (12.12)	.26	−0.26	.00	1.45
WJ-IV STM/WM	100.97 (10.95)	77.18 (17.47)	103.80 (13.01)	.28	−0.24	.00	1.73
WJ-IV Processing Speed	100.50 (11.09)	74.08 (21.20)	101.93 (13.27)	.59	−0.12	.00	1.57
WJ-IV Cognitive Efficiency	100.90 (10.27)	74.00 (19.21)	101.93 (13.04)	.68	−0.09	.00	1.70
WJ-IV Gf-Gc	101.27 (12.86)	95.15 (13.93)	104.63 (11.20)	.23	−0.28	.00	0.75
MSVT IR	99.33 (1.71)	80.58 (19.42)	99.83 (0.91)	.14	−0.36	.00	1.40
MSVT DR	99.08 (2.35)	76.08 (19.44)	99.33 (1.73)	.61	−0.12	.00	1.68
MSVT Consistency	98.58 (2.62)	75.42 (17.25)	99.17 (2.31)	.30	−0.24	.00	1.90
BAARS Inattention	13.88 (3.69)	26.92 (5.41)	25.23 (4.60)	.00	2.72	.15	−0.34
BAARS Hyperactivity	8.15 (2.82)	15.68 (2.86)	13.00 (3.30)	.00	1.58	.00	−0.87
BAARS Impulsivity	6.37 (2.63)	12.28 (2.70)	9.13 (3.09)	.00	.96	.00	−1.09
BAARS SCT	15.47 (4.96)	24.28 (5.51)	24.00 (5.60)	.00	1.61	.82	−0.05
BAARS Inattentive Count	.68 (1.33)	6.45 (2.51)	5.63 (2.48)	.00	2.48	.15	−0.32
BAARS Hyp/Imp Count	1.22 (1.78)	6.73 (2.15)	4.30 (1.99)	.00	1.64	.00	−1.17
BAARS SCT Count	1.37 (1.82)	5.13 (2.36)	5.13 (2.36)	.00	1.78	1.00	0.00

WJ-IV: Woodcock-Johnson IV Tests of Cognitive Abilities; MSVT: Medical Symptom Validity Test; IR: Immediate Recognition; DR: Delayed Recognition; BAARS: Barkley Adult ADHD Rating Scale IV; SCT: Sluggish Cognitive Tempo.

The p values are from two-tailed independent groups t-tests, each with $df = 88$; p values below .0005 are represented as .00. The WJ-IV values are standard scores, the MSVT values are percent correct scores, and the BAARS values are raw scores. The d values are negative when the Diagnosed ADHD group had lower scores.

On the WJ-IV measures, the comparisons between control participants and those with ADHD diagnoses were never statistically significant, or even trending toward significance (all p values $\geq$ .22). Effect sizes for the comparisons were generally small ($d \approx 0.2$). In contrast, on the same measures, the comparisons between participants simulating ADHD vs. having diagnosed ADHD were generally significant (most p values $\leq$ .001) and large (most d values $>$ 1.00).[3] The only two WJ-IV test scores that did not show significant differences between the Simulated ADHD and Diagnosed ADHD groups were Oral Vocabulary and General Information; all of the cluster scores showed significant differences. In all cases, the Simulated ADHD group scored lower than the Diagnosed ADHD group on the WJ-IV.

On the MSVT measures, findings were similar to the WJ-IV. None of the comparisons between the Control and Diagnosed ADHD groups showed statistically significant differences (all p values $\geq$ 0.14), and although effect sizes were variable, this was likely due to the MSVT variables being highly skewed. As Table 2 shows, means were between 99 and 100 (i.e., percent correct scores) for all MSVT variables. Again, like the WJ, the MSVT showed significant differences between the Simulated and Diagnosed ADHD groups (all p values $\leq$ .0005) and yielded large effect sizes (all d values $>$ 1.00), with the Simulated group obtaining lower scores. We also used the official MSVT criteria to determine the "pass" rates of each group on the initial measures. Almost all of the Control (98%)[4] and Diagnosed ADHD (100%) groups passed the MSVT, suggesting sufficient effort, whereas only 30% of the Simulated ADHD group passed.

Finally, on the BAARS, the Control group obtained significantly lower scores (indicating fewer symptoms) than the Diagnosed ADHD group on all parts of the measure (all p values $<$.0005) and the effect sizes for these comparisons were consistently very large ($0.96 \leq d \leq 2.72$). The Simulated ADHD group reported more symptoms and obtained significantly higher scores than the Diagnosed ADHD group on the parts of the BAARS relating to core ADHD symptoms (inattention, hyperactivity, and impulsivity), with large effect sizes for hyperactivity and impulsivity ($-1.17 \leq d \leq -0.87$) and small-to-medium effect sizes for inattention ($-0.34 \leq d \leq -.32$); however, this was not the case for the BAARS scores relating to sluggish cognitive tempo symptoms, where the comparisons were nonsignificant (p values of .82 and 1.00) and effect sizes were negligible (d values of −0.05 and zero, respectively).

Diagnostic efficiency of PVTs

We used Receiver Operating Characteristic (ROC) curve analysis to examine the diagnostic efficiency of the WJ-IV scores (the proposed embedded PVTs) and the MSVT scores (from an established standalone PVT) in distinguishing between Simulated ADHD and Diagnosed ADHD (for more on ROC curve analysis, see Mayer, 2004). Table 3 presents data from these analyses. One statistic that we computed was the Area Under the Curve (AUC), an integrated measure of the sensitivity and specificity of a test at different

[3]Descriptors for d values were taken from Cohen (1992).

[4]Since the specificity of the MSVT is not always 100%, we chose to leave in the one Control participant who failed the MSVT. However, we also ran our analyses with that participant removed, and found that our results did not substantially change; no significant differences between the Control and Diagnosed ADHD groups became nonsignificant, or vice versa.

Table 3. Receiver Operating Characteristic (ROC) curve analysis of WJ-IV standard scores and MSVT raw scores for detecting simulation of ADHD.

Measure and score	AUC	Cut score for specificity of at least 90% (exact specificity)	Corresponding sensitivity value for cut score
WJ-IV Oral Vocabulary	0.58	88.5 (.90)	0.15
WJ-IV Number Series	0.73	92.0 (.90)	0.37
WJ-IV Verbal Attention	0.88	88.0 (.90)	0.68
WJ-IV Letter Pattern Matching	0.83	87.5 (.90)	0.70
WJ-IV General Information	0.59	79.5 (.90)	0.13
WJ-IV Concept Formation	0.65	89.0 (.97)	0.23
WJ-IV Numbers Reversed	0.85	85.0 (.93)	0.55
WJ-IV Pair Cancelation	0.84	88.5 (.90)	0.67
WJ-IV STM/WM	0.89	88.5 (.93)	0.72
WJ-IV Processing Speed	0.85	89.5 (.90)	0.72
WJ-IV Cognitive Efficiency	0.88	85.0 (.93)	0.70
WJ-IV Gf-Gc	0.71	92.0 (.90)	0.32
MSVT IR	0.85	97.5 (.97)	0.72
MSVT DR	0.87	92.5 (1.00)	0.75
MSVT Consistency	0.89	92.5 (.97)	0.78

WJ-IV: Woodcock-Johnson IV Tests of Cognitive Abilities; MSVT: Medical Symptom Validity Test; IR: Immediate Recognition; DR: Delayed Recognition; BAARS: Barkley Adult ADHD Rating Scale IV; SCT: Sluggish Cognitive Tempo.

possible cut-scores. The AUC values for the MSVT were between 0.85 and 0.89 (where 1.00 is perfect and 0.50 is chance). Some WJ-IV scores (Verbal Attention, Letter Pattern Matching, Numbers Reversed, Pair Cancelation, the STM/WM Composite, the Processing Speed Composite, and the Cognitive Efficiency Composite) were in or near this range as well (AUC values between 0.83 and 0.89) whereas the other scores were lower (Oral Vocabulary = 0.58, Number Series = 0.73, General Information = 0.59, Concept Formation = 0.65, and the Gf-Gc Composite = 0.71). There is no single, commonly accepted standard for judging AUC values, but the values can be straightforwardly interpreted as the chance that a college student simulating ADHD will obtain a lower score than a randomly chosen student with a prior diagnosis of ADHD (Lasko et al., 2005). In addition, Hosmer and Lemeshow (2000) suggested that AUC values between 0.70 and 0.80 indicate "acceptable" discrimination between two groups, whereas values between 0.80 and 0.90 indicate "excellent" discrimination.

The AUC cannot itself provide the optimal cut point to use in clinical contexts, and so we examined the coordinates of the ROC curve for each measure to find a cut-score at which the measure had at least 90% specificity (exact specificities ranged from 90% to 100% due to curves that were not precisely smooth), and we recorded the sensitivity value that the measure had at that particular cut score (to determine what proportion of the Simulated ADHD group scored below that cut score). For the MSVT scores, the cut score needed to obtain at least 90% specificity ranged from 92.5 to 97.5, and the sensitivity values for those cut scores ranged from 72% (IR) to 78% (CS). For the WJ-IV scores, the cut scores needed to achieve at least 90% specificity ranged from standard scores of 79.5 to 92.0, and sensitivity values for those cut scores varied widely, from 13% to 72%. WJ-IV scores that had shown larger and more statistically significant differences between the Simulated and Diagnosed ADHD groups (in the group comparisons presented earlier) yielded higher sensitivity values at the cut scores needed for 90% or greater specificity.

Discussion

The principal research question in the present study was whether any WJ-IV COG scores could effectively discriminate between (a) individuals deliberately simulating ADHD and (b) individuals with independent prior diagnoses of the disorder. We addressed this question through two analytic strategies. First, we compared the two groups' performance on each of the WJ-IV scores. We found that on measures of working memory and processing speed, the two groups differed significantly (even when correcting for multiple comparisons) and the group differences showed effect sizes in the large range (or greater). However, on the WJ-IV measures of fluid and crystallized ability, effect sizes were smaller and group differences were generally not statistically significant. In a second strategy, we used ROC curve analysis to distinguish between the two groups, finding that the WJ-IV measures of working memory and processing speed showed excellent discrimination between simulators of ADHD and those with independent diagnoses of ADHD, whereas for the measures of fluid and crystallized ability, the ROC curves did not indicate even adequate discrimination. Both in terms of group differences and ROC curve-estimated diagnostic efficiency, the WJ-IV measures of working memory and processing speed performed almost as well as scores from an established standalone PVT, the MSVT.

A secondary research question was whether WJ-IV measures would show differences between participants with and without ADHD diagnoses, when all participants were instructed to respond honestly and put forth adequate effort. None of these comparisons even approached statistical significance, or yielded effect sizes greater than $|d| = 0.28$. Therefore, performance on the selected WJ-IV measures does not appear to relate strongly to independent diagnoses of ADHD in college students under naturalistic conditions.

Another secondary research question was how students simulating ADHD and those with independent ADHD diagnoses would respond to the BAARS. The latter group consistently reported much higher ADHD and SCT symptoms than controls in all areas. Simulators reported even higher ADHD symptom levels than those with independent diagnoses of ADHD but roughly equivalent levels of SCT symptoms.

Clinical implications and integration with prior findings

The results from the present study converge with those from other recent studies suggesting that response validity should be formally measured when assessing ADHD (for review, see Lovett & Davis, 2017; Suhr & Berry, 2017). Simulators were able to report levels of ADHD symptoms even greater than those of participants with prior ADHD diagnoses on the BAARS, with only minimal information about the disorder provided by the investigators. The WJ-IV measures of working memory and processing speed were effective indicators of simulated ADHD—almost as effective as an established standalone PVT. The latter findings relate to a prior study by Erdodi et al. (2017), who found the processing speed portions of the Wechsler Adult Intelligence Scale to be effective embedded PVTs in a mixed, clinically referred sample, although Wechsler processing speed measures have not always proven to be so helpful in this way (for further studies on this point showing mixed results, see Booksh et al., 2010; Harrison & Edwards, 2010; Suhr et al., 2008).

The sensitivity and specificity details in Table 3 provide further clinical guidance; with certain limitations noted below, clinicians can use the working memory and processing speed WJ-IV scores to detect over half of those potentially simulating ADHD while still maintaining at least 90% specificity (making false accusations of low effort or malingering unlikely). Viewing the cutoff scores in Table 3, it appears that poor WJ-IV cognitive scores in the context of ADHD evaluations in college populations are actually far more likely to indicate exaggeration or feigning rather than diagnosed ADHD, since students in the latter case generally do not show substantial deficits in those areas. Certainly, the results from the present paper do not provide any support for using WJ-IV performance to make a diagnosis of ADHD, although this may be different in a sample where no ADHD medication effects are present.

The cutoff scores suggested by the ROC curve analyses might be surprising to some readers; for the WJ, they are generally standard scores of 85 or above, suggesting that scores below 85 would be indicators of exaggeration or feigning. It must be kept in mind that the present study was conducted at 4 year colleges, in participants who were not clinically referred for cognitive problems. Therefore, cognitive ability scores of 85 (even in narrow cognitive areas) were unusual among those putting forth their full effort. The cutoffs would likely be lower in a different setting that does not select for higher-ability individuals (relative to the general population). Moreover, triggering embedded PVTs should never, by itself, lead to conclusions of exaggeration or feigning anyway; the data from the evaluation must be viewed as a whole, and in particular external historical records of impairment should be utilized to validate genuine cognitive problems (cf. Lovett & Bizub, 2019). In any case, we view the specific cutoffs in the present paper as preliminary, and even then as only suggesting *potential* low effort in individuals with high education levels. In a clinical setting with young adults presenting for ADHD evaluations, the cutoffs would further be best considered in the case of individuals who are already taking ADHD medication (as many are, both in the case of reevaluations, and when a physician has already made a "soft diagnosis" or initiated a medication trial that will be continued if a psychological evaluation results in a diagnosis).

The question of why a prior professional ADHD diagnosis was not associated with deficits in cognitive abilities—even working memory and processing speed—is an important one, and cannot be answered conclusively based on the present study alone. However, Wood et al. (2019) speculated that students at a 4 year college who report ADHD diagnoses may be less impaired than is typical for the disorder (allowing them to gain entry to more competitive colleges), and may have benefited from effective ADHD medication or other treatments. The ADHD may also have remitted over time if a diagnosis was made early in a student's life (Caye et al., 2016). Misdiagnosis/overdiagnosis is another possible explanation, since we did not perform our own diagnostic evaluation to verify whether the full ADHD diagnostic criteria were met, but even if this is the case, one clinical implication is clear: college students who report an ADHD diagnosis do not, on average, show substantial deficits in key areas of cognitive functioning, as measured by a validated test battery, at least under naturalistic conditions (in which some students with diagnoses are taking medication).

Another important clinical implication stems from our finding that simulated ADHD was not associated with across-the-board low WJ-IV scores. Instead, even though participants were only provided minimal information about typical symptoms and impairment of college students with ADHD, participants in the Simulated ADHD group tailored their performance to do poorly on measures of working memory and processing speed, sparing more general intellectual skills (rather than feigning problems with e.g., vocabulary). Clinicians, then, should not view variability in diagnostic test performance or the presence of *some* average and above-average scores as evidence against exaggeration/feigning.

We would *not* conclude from the present study that use of embedded measures in the WJ-IV is an *optimal* technique for identifying insufficient effort. Although some WJ-IV scores showed diagnostic efficiency close to that of the MSVT in identifying simulators, we would still recommend that a standalone PVT be administered, even if the WJ-IV is already included in an assessment battery. The administration of *multiple* PVTs is always preferred when response bias is a concern (Larrabee, 2014), as it is in college ADHD assessment, due to the incentives for diagnosis. Moreover, standalone PVTs like the MSVT have been validated in many populations and settings, especially with individuals who may not be as high-functioning as our college students were. However, critically, since many clinicians do *not* choose to administer standalone PVTs when conducting ADHD evaluations (see e.g., Nelson et al., 2019), the present study should alert some clinicians to a possible new use of a measure that they are more likely to already be giving.

Limitations and future research directions

One limitation of the present study was noted above; although we screened our Diagnosed ADHD group to ensure that they reported clinically significant symptom levels, we did not perform our own comprehensive diagnostic evaluations to verify the ADHD status of these participants. Such a procedure would lead to greater internal validity, increasing group homogeneity, and so we would encourage future research on that point, particularly including participants who complete a full ADHD evaluation but do not receive a diagnosis. A second, related limitation is that the examiners were not blind to which condition participants were in. This was necessary to allow examiners to remind participants repeatedly of the motivation with which to approach the tasks and to ensure that no participant with ADHD was randomly assigned to simulate ADHD, but it is possible that examiners' knowledge of participants' study conditions could have led to systematic errors in recording participants' responses. A third limitation is that we only examined a single motivation for simulating ADHD: a desire to obtain testing accommodations such as extended testing time. Future studies should examine other motives, such as medication access, since this may affect WJ-IV performance of simulators (particularly on time-pressured tests). While not connected to a limitation, we would also recommend examining features of WJ-IV behavior beyond scores, such as patterns of errors, to determine if these are helpful indicators of feigning as well. Finally, although the present study was focused on performance validity, many of the core assessment measures in ADHD evaluations involve symptom reports, and so we would encourage further work on symptom validity measures for ADHD.

Disclosure statement

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

Funding

This work was supported by the 'Woodcock Institute for the Advancement of Neurocognitive Research and Applied Practice'.

References

Barkley, R. A. (2011). *Barkley adult ADHD rating scale-IV*. 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

Boone, K. B. (2009). The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *The Clinical Neuropsychologist*, *23*(4), 729–741. https://doi.org/10.1080/13854040802427803

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

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 ADHD study. *The Clinical Neuropsychologist*, *32*(1), 109–118. https://doi.org/10.1080/13854046.2017.1346145

Caye, A., Swanson, J., Thapar, A., Sibley, M., Arseneault, L., Hechtman, L., Arnold, L. E., Niclasen, J., Moffitt, T., & Rohde, L. A. (2016). Life span studies of ADHD: Conceptual challenges and predictors of persistence and outcome. *Current Psychiatry Reports*, *18*(12), Article 111. https://doi.org/10.1007/s11920-016-0750-x

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Drazdowski, T. K. (2016). A systematic review of the motivations for the non-medical use of prescription drugs in young adults. *Drug and Alcohol Dependence*, *162*, 3–25. https://doi.org/10.1016/j.drugalcdep.2016.01.011

Edmundson, M., Berry, D. T. R., Combs, H. L., Brothers, S. L., Harp, J. P., Williams, A., Rojas, S. L., Saleh, A. K., & Scott, A. B. (2017). The effects of symptom information coaching on the feigning of adult ADHD. *Psychological Assessment*, *29*(12), 1429–1436. https://doi.org/10.1037/pas0000478

Erdodi, L. A., Abeare, C. A., Lichtenstein, J. D., Tyson, B. T., Kucharski, B., Zuccato, B. G., & Roth, R. M. (2017). Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) processing speed scores as measures of noncredible responding: The third generation of embedded performance validity indicators. *Psychological Assessment*, *29*(2), 148–157. https://doi.org/10.1037/pas0000319

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.

Fuermaier, A. B., Fricke, J. A., de Vries, S. M., Tucha, L., & Tucha, O. (2019). Neuropsychological assessment of adults with ADHD: A Delphi consensus study. *Applied Neuropsychology: Adult*, *26*(4), 340–354. https://doi.org/10.1080/23279095.2018.1429441

Green, P. (2004). *Manual for the medical symptom validity test for windows*. Green's Publishing.

Green, P., Flaro, L., & Courtney, J. (2009). Examining false positives on the Word Memory Test in adults with mild traumatic brain injury. *Brain Injury*, *23*(9), 741–750. https://doi.org/10.1080/02699050903133962

Harrison, A. G. (2006). Adults faking ADHD: You must be kidding!. *The Adhd Report*, *14*(4), 1–7. https://doi.org/10.1521/adhd.2006.14.4.1

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. (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

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., & Millis, S. R, & Conference Participants. (2009). American academy of clinical neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, *23*(7), 1093–1129. https://doi.org/10.1080/13854040903155063

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). Wiley.

Jachimowicz, G., & Geiselman, R. E. (2004). Comparison of ease of falsification of attention deficit hyperactivity disorder diagnosis using standard behavioral rating scales. *Cognitive Science Online*, *2*(1), 6–20.

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

Larrabee, G. J. (2014). False-positive rates associated with the use of multiple performance and symptom validity tests. *Archives of Clinical Neuropsychology*, *29*(4), 364–373. https://doi.org/10.1093/arclin/acu019

Lasko, T. A., Bhagwat, J. G., Zou, K. H., & 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

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

Lovett, B. J., & Bizub, A. L. (2019). Pinpointing disability accommodation needs: What evidence is most relevant? *Psychological Injury and Law*, *12*(1), 42–51. https://doi.org/10.1007/s12207-019-09341-1

Lovett, B. J., & Davis, K. M. (2017). Adult ADHD assessment: An integrated clinical-forensic perspective. *Professional Psychology: Research and Practice*, *48*(6), 438–444. https://doi.org/10.1037/pro0000159

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

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

Mather, N., & Jaffe, L. E. (2010). *Comprehensive evaluations*. Wiley.

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

McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Technical manual: Woodcock Johnson IV*. Riverside.

Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, *24*(8), 1094–1102. https://doi.org/10.1076/jcen.24.8.1094.8379

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. M., & Lovett, B. J. (2019). Assessing ADHD in college students: Integrating multiple evidence sources with symptom and performance validity data. *Psychological Assessment*, *31*(6), 793–804. https://doi.org/10.1037/pas0000702

Nelson, N. W., Sweet, J. J., Berry, D. T., Bryant, F. B., & Granacher, R. P. (2007). Response validity in forensic neuropsychology: Exploratory factor analytic evidence of distinct cognitive and psychological constructs. *Journal of the International Neuropsychological Society*, *13*(03), 440–449. https://doi.org/10.1017/S1355617707070373

Nelson, J. M., Whipple, B., Lindstrom, W., & Foels, P. A. (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

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

Singhal, A., Green, P., Ashaye, K., Shankar, K., & Gill, D. (2009). High specificity of the medical symptom validity test in patients with very severe memory impairment. *Archives of Clinical Neuropsychology*, *24*(8), 721–728. https://doi.org/10.1093/arclin/acp074

Suhr, J. A. (2015). *Psychological assessment: A problem-solving approach*. Guilford.

Suhr, J. A., & Berry, D. T. (2017). The importance of assessing for validity of symptom report and performance in attention deficit/hyperactivity disorder (ADHD): Introduction to the special section on noncredible presentation in ADHD. *Psychological Assessment*, *29*(12), 1427–1428. https://doi.org/10.1037/pas0000535

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 ADHD evaluation. *Archives of Clinical Neuropsychology*, *23*(5), 521–530. https://doi.org/10.1016/j.acn.2008.05.003

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

Tucha, L., Fuermaier, A. B., Koerts, J., Groen, Y., & Thome, J. (2015). Detection of feigned attention deficit hyperactivity disorder. *Journal of Neural Transmission*, *122*(S1), 123–134. https://doi.org/10.1007/s00702-014-1274-3

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

Weyandt, L. L., & DuPaul, G. J. (2013). *College students with ADHD*. Springer.

Weyandt, L. L., Oster, D. R., Marraccini, M. E., Gudmundsdottir, B. G., Munro, B. A., Rathkey, E. S., & McCallum, A. (2016). Prescription stimulant medication misuse: Where are we and where do we go from here? *Experimental and Clinical Psychopharmacology*, *24*(5), 400–414. https://doi.org/10.1037/pha0000093

Wood, W. L. M., Lewandowski, L. J., & Lovett, B. J. (2019). Profiles of diagnosed and undiagnosed college students meeting ADHD symptom criteria. *Journal of Attention Disorders*, 108705471882499. https://doi.org/10.1177/1087054718824991

Appendix

Instructions given to assist simulators with feigning ADHD

Today, we would like you to pretend that you are being evaluated for ADHD because you are concerned about your performance in college and you want to receive test accommodations, such as extra time to complete exams and quizzes in your courses. Pretend that your performance on the tests today determines whether you will get these accommodations. Because you think that these accommodations will help you academically, you are eager to convince us that you are experiencing these difficulties. Therefore, you should try to perform on these tests the way you think someone with ADHD would perform.

College students with ADHD often experience symptoms that include making careless mistakes, forgetfulness, distractibility, impulsive responding, and difficulty attending to several things at once. Some students may experience restlessness and trouble engaging with quiet activities. In addition, adults with ADHD often have problems with interpersonal relationships and employment.