The Groundskeeper Gaming Platform as a Diagnostic Tool for Attention-Deficit/Hyperactivity Disorder: Sensitivity, Specificity, and Relation to Other Measures

Stephen V. Faraone, PhD,^{1,2} Jeffrey H. Newcorn, MD,³ Kevin M. Antshel, PhD,⁴
Lenard Adler, MD,⁵ Kurt Roots, MS,⁶ and Monika Heller, MD⁶

Abstract

Objective: The purpose of this study was to assess the relative accuracies of the Conners' Brief Rating Scale, Parent Version, the Conners' Continuous Performance Test II (CPT II), and a novel interactive game called "Groundskeeper" to discriminate child psychiatric patients with and without attention-deficit/hyperactivity disorder (ADHD).

Methods: We administered the three assessments to 113 clinically referred ADHD and non-ADHD patients who had been diagnosed with the Kiddie-Schedule of Affective Disorders and Schizophrenia- Present and Lifetime (K-SADS-PL), Version 19.

Results: As measured by the area under the curve (AUC) statistic from receiver operating characteristic (ROC) analysis, the diagnostic accuracy of Groundskeeper (0.79) was as high as the accuracy of the Conners' parent rating of inattention (0.76) and better than the CPT II percent correct (0.62). Combining the three tests produced an AUC of 0.87. Correlations among the three measures were small and, mostly, not significant.

Conclusions: Our finding of similar diagnostic accuracies between Groundskeeper and the Conners' inattention scale is especially remarkable given that the Conners' inattention scale shares method variance with the diagnostic process. Although our work is preliminary, it suggests that computer games may be useful in the diagnostic process. This provides an important direction for research, given the objectivity of such measures and the fact that computer games are well tolerated by youth.

Keywords: ADHD, gaming, decision-making, diagnosis, biomarker

Introduction

ATTENTION-DEFICIT/HYPERACTIVITY DISORDER (ADHD) is diagnosed by evaluating symptoms of hyperactivity, impulsivity, and inattention, and impaired functioning across settings. The diagnosis of ADHD shows considerable levels of concurrent and predictive validity in its clinical features, course, neurobiology, and treatment response (Faraone et al. 2000; Faraone 2005). The diagnosis has high diagnostic reliability, one of the highest in the *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed. (DSM-5) (American Psychiatric Association 2013; Regier et al. 2013). Nevertheless, concerns about diagnostic accuracy persist. Some suggest that the use of subjective diagnostic procedures may lead to the overdiagnosis of ADHD in the community (Bruchmuller et al. 2012; Visser et al. 2014) (although see Sciotto and Eisenberg [2007] for a contrary view). The diagnosis has been called "sub-

jective" because it relies on clinician evaluation of responses from patients, parents, and/or informants. Other studies have raised concerns about the underdiagnosis of ADHD (Express Scripts Lab 2014; Ginsberg et al. 2014).

In response to such concerns, researchers have sought to develop objective measures to diagnose ADHD or to monitor the course of ADHD symptoms during treatment. The first approach to objectifying ADHD medication response, and eventually diagnosis, was via parent and teacher rating scales. Although ratings scales rely on parent, teacher, or self-reports of symptoms, they evaluate these reports in the context of large, normative databases, which allows for a more data-driven approach to interpreting information obtained in assessment. Other research has examined peripheral biochemical markers as objective measures. Meta-analyses of these studies indicate that five measures differentiated ADHD and control patients (norepinephrine [NE], 3-Methoxy-4-hydroxyphenyl

¹Departments of Psychiatry and of Neuroscience and Physiology, SUNY Upstate Medical University, Syracuse, New York.

²K.G. Jebsen Center for Research on Neuropsychiatric Disorders, Department of Biomedicine, University of Bergen, Bergen, Norway.

³Icahn School of Medicine at Mount Sinai, New York, New York.

⁴Department of Psychology, Syracuse University, Syracuse, New York.

⁵Department of Psychiatry and Child and Adolescent Psychiatry, New York University Langone School of Medicine, New York, New York.

⁶CogCubed, Minneapolis, Minnesota.

Funding: Dr. Faraone is supported by the K.G. Jebsen Center for Research on Neuropsychiatric Disorders, University of Bergen, Bergen, Norway, the European Union's Seventh Framework Program for research, technological development, and demonstration under grant agreement no 602805.

ethylene glycol [MHPG], monoamine oxidase [MAO], zinc, and cortisol) (Faraone, et al. 2014, Scassellati, et al. 2012). Moreover, NE, MHPG, MAO, β -phenylethylamine, and cortisol were responsive to ADHD medications. Meta-analysis also shows that peripheral measures of oxidative stress differ between ADHD and control participants (Joseph et al. 2015). Other approaches to biomarker development for ADHD have used neuropsychological (Ritsner 2009), electroencephalographic (EEG) (Snyder et al. 2015), structural imaging (Silk et al. 2009), and functional imaging (Bush et al. 2005) methods, often with the application of machine learning approaches to optimize diagnostic accuracy (Mueller et al. 2010).

Of particular relevance are continuous performance tests (CPTs), of which many are available (e.g., Corkum and Siegel 1993; Riccio and Reynolds 2001; Homack and Riccio 2006), the Quotient ADHD system (Sumner et al. 2010), which pairs a type of CPT with recording of motor activity, and the Neuropsychiatric EEG-Based Assessment Aid (NEBA) quantitative EEG assessment (Snyder et al. 2015). CPTs are frequently used in neuropsychological testing, but have less than optimal sensitivity and specificity for diagnosing ADHD. Quotient and NEBA are cleared by the United States Food and Drug Administration for augmenting clinical assessments (Dolgin 2014; Snyder et al. 2015); they are not cleared for diagnosing ADHD in the absence of a full clinical assessment. Therefore, these tests can be used to collect adjunct information but do not provide objective diagnoses of ADHD. For example, Snyder et al. (2015) showed that NEBA is useful for clarifying diagnoses when clinicians are uncertain if the patient truly has ADHD. However, because of these limitations, none of the abovementioned tests should be used to diagnose the disorder in the absence of a systematic and comprehensive clinical assessment.

Thome et al. (2012) presented the results of the task force on biological markers of the World Federation of ADHD. They used the following criteria to define a “useful” biomarker: Sensitivity >80%; specificity >80%; and that the putative biomarker is reliable, reproducible, inexpensive, noninvasive, easy to use, and has been confirmed by at least two independent studies. Putative biomarkers examined by Thome et al. were: EEG-based event-related potentials; neuropsychological measures of executive functioning, attention, memory, spatial abilities, and language; olfactory functioning, structural, and functional magnetic resonance imaging; transcranial sonography; genetics; peripheral metabolites; and proteomics. None of these markers met their criteria for utility.

To develop an objective diagnosis for ADHD, we are studying a novel interactive game called Groundskeeper, which captures the ADHD symptoms of impulsivity and inattention, as well as associated features: Motor coordination (Fliers et al. 2008, 2009, 2010), reaction time variability (Perry et al. 2010; Frazier-Wood et al. 2012), impaired temporal processing (Toplak et al. 2006), and impaired decision making (Drechsler et al. 2008). Building diagnostics into game playing has the advantage of offering patients a rewarding, engaging procedure that avoids confounds associated with boredom and random responding (Van der Oord et al. 2012; Dosis et al. 2015). In a preliminary study of 52 outpatients 6–17 years of age, Groundskeeper data predicted inattentive ADHD with a sensitivity of 76.9% and a specificity of 80.7% (Heller 1993). It predicted combined type ADHD with a sensitivity of 58.8% and a specificity of 82.8%. Groundskeeper was a more accurate predictor of gold standard clinician diagnoses than the parent report version of the Conners’ Brief Rating Scale. The present study sought to further characterize the diagnostic accuracy of Groundskeeper and to examine its relationship and comparative predictive accuracy

with parent ratings of ADHD symptoms and patient performance on a CPT.

Methods

Recruitment of participants

This was a cohort study in which participants were recruited as consecutive referrals. We did not attempt to enrich the participant pool with ADHD patients. Instead, we invited for participation consecutive patients referred to a child psychiatrist. From this cohort of patients, we formed two groups: Patients who had received an ADHD diagnosis and those who had not. Participants were children and adolescents between 6 and 17 years of age ($n=113$, see Table 1), recruited from two outpatient psychiatric clinics.

Study participants were administered the Kiddie-Schedule of Affective Disorders and Schizophrenia- Present and Lifetime (K-SADS-PL), Version 19, a semistructured diagnostic interview conducted by a psychiatric nurse trained by a child/adolescent psychiatrist at a community based clinic ($\kappa=1.0$) and reviewed by two independent child and adolescent psychiatrists. Patients were eligible to participate if they met criteria for ADHD, major depressive disorder, dysthymia, generalized anxiety disorder, anxiety disorder not otherwise specified (NOS), social phobia, oppositional defiant disorder, panic disorder, or eating disorders on the K-SADS-PL. We excluded participants with a history of psychosis or neurological disorder, low intellectual functioning (i.e., child was not in a mainstream academic class), substance use disorders, conduct disorder, tic disorders, or physical impairments precluding game play.

Fourteen participants were taking stimulant medications: Dextroamphetamine ($n=1$), methylphenidate extended release (ER) ($n=1$), lisdexamfetamine ($n=3$), transdermal methylphenidate ($n=3$), mixed amphetamine salts ($n=2$), and osmotic controlled-release oral delivery system (OROS) methylphenidate ($n=4$). These were withheld on the days of testing for Groundskeeper and Conners’ Continuous Performance Test II (CPT II) but were not otherwise washed out. Eighty participants were on at least one nonstimulant, psychiatric medication: Citalopram ($n=12$), bupropion ($n=8$), aripiprazole ($n=16$), quetiapine ($n=12$), buspirone ($n=6$), mirtazapine ($n=4$), trazodone ($n=11$), clonidine ($n=2$), duloxetine ($n=1$), venlafaxine ($n=3$), guanfacine immediate release (IR) ($n=2$), guanfacine extended release (XR) ($n=8$), escitalopram ($n=2$), fluvoxamine ($n=1$), paroxetine ($n=3$), atomoxetine HCl ($n=1$), sertraline ($n=7$), fluoxetine ($n=8$), risperidone ($n=4$), propranolol ($n=1$), lamotrigine ($n=1$), olanzapine ($n=2$), and topiramate ($n=2$). Twenty-nine percent ($n=33$) of patients were medication free at study initiation. Among the 66

TABLE 1. PSYCHIATRIC DISORDERS IN THE ADHD AND CONTROL GROUPS

Disorder	n in control group	n in ADHD group
Anxiety disorders	45	33
Mood disorder	36	27
Disruptive behavior disorders	8	11
Autism spectrum disorder	2	6
Reading disability	10	23

The total number of disorders is more than the total number of participants because each participant can have more than one disorder.

ADHD, attention-deficit/hyperactivity disorder.

patients diagnosed with ADHD, 25 (38%) were on a medication not typically used to treat ADHD (i.e., medications other than stimulants, bupropion, clonidine, guanfacine, and atomoxetine). Four subjects were on both stimulant and nonstimulant medications.

Institutional review board approval was obtained from the University of Minnesota. Written permission and assent were obtained from one parent and each participant, respectively. Procedures were in accordance with the ethical standards of the responsible committee on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. For eligible families, a parent completed the Conners' Brief Rating Scale, Parent Version using the past month as the time frame for reporting. The scale was completed once, and we extracted five subscales for analysis: Hyperactivity/impulsivity, inattention, learning problems, aggression, and executive functions. The child played the Groundskeeper game and was administered the CPT on separate occasions within 1 week of each other.

Continuous Performance Task

The CPT II was administered by a technician who remained in the room while the test was completed. Following the standard protocol, after a practice session, the actual testing session began. CPT II respondents were asked to press the space bar whenever any letter except the letter "X" appeared on the computer screen. The inter-stimulus intervals (ISIs) were 1, 2, and 4 seconds, with a display time of 250 milliseconds. The CPT II comprises six blocks and three sub-blocks, each containing 20 trials. The presentation order of ISIs varied among blocks. For our analyses, we used the percent certainty that the CPT II results were in the clinical range.

Groundskeeper game

Groundskeeper (Fig. 1) is played using four cubes and a placement board. One cube is used as a "mallet" to hit targets that appear on other cubes. The mallet must be moved by the player. The other cubes are placed in a straight vertical line. These three cubes have an image of green grass and blue sky as a backdrop. Images of a rabbit, a "groundskeeper" (man with a lawn mower), a gopher, or a few small birds appear on these screens for 1, 1.5, or 3 seconds at random. The object of the game is to touch another cube when the gopher image appears. Successful hits are associated with a "bonk" noise. The other images are distractors to be avoided. Each of the 17 game sessions is 90 seconds long, with a 20 second interval in between each session. Game playing in-

structions are administered via a script read to each participant (see Appendix A). Summaries of each game session are in Appendix B. Variables derived from Groundskeeper are in Appendix C.

The Groundskeeper protocol consists of 17 game sessions, numbered 0 through 16, each with different levels and types of distractions: Visual, auditory, and spatial. Low visual distraction consists of a bird appearing on the cube screens. High visual distraction adds large rabbits. Low auditory distraction consists of occasional tweeting noises; high auditory distraction increases tweet frequency. When there is no spatial distraction, the image cubes are in a vertical line. In low spatial distraction, they are set diagonally 5.08 cm apart (Fig. 1). High spatial distraction consists of each cube put 7.62 cm apart. Sessions 0, 1, and 16 have no distraction. Each session is 90 seconds long, consisting of a randomized number of trials and frequency of target stimulus presentations.

The game was designed to measure attentional capabilities on a go/no go task, with the addition of visual, auditory, and visuospatial distractions at various frequencies. Given that we were implementing multiple levels to the game, each level was restricted to 90 seconds in order to keep the full game time at a reasonable time length, while giving time between levels to arrange the cubes in a specific configuration if necessary. The first and the last levels were meant to be a baseline for comparison and to measure any effect of "learning" on performance. Next, we implemented visual distractors of birds and a groundskeeper to test the ability of a patient to alter their go/no go response when additional figures were presented. First, the visual distractors were presented at "low frequency," meaning only the Groundskeeper was introduced as a distractor, in addition to the gopher, and then at a high frequency when a bird, Groundskeeper, and rabbit were presented. A rabbit was chosen because it most closely resembled the gopher pictorially.

Next we added auditory distractors of a bird tweeting at a low frequency (one tweet at various intervals) and then at a high frequency (multiple tweets at various intervals). Lastly, we spread the cubes out diagonally in an attempt to introduce a visuospatial element, thus adding a new distractor and trying to eliminate any habituation a person may have had to the cubes being at a vertical configuration for various levels. We then added low frequency distractors (visual and auditory), to see if the effect of these distractors while we had a visuospatial element had the same effect.

Statistical analyses

For Groundskeeper, we used logistic regression to assess the significance of factors derived from principal factors factor analysis as predictors of the ADHD diagnosis. We retained factors that significantly predicted ADHD at the Bonferroni-corrected α level of 0.0023 (i.e., 0.05/22). We also used logistic regression to assess the ability of the parent-rated Conners' subscales to predict ADHD diagnoses. For this model, the Bonferroni-corrected α level for selecting significant predictors was 0.01.

Receiver operating characteristic (ROC) curve analysis determined the accuracy of predictions. ROC analysis assesses the diagnostic efficiency of tests for diagnoses and allows for adjustment of cut-points for clinical or research purposes (McNeil and Hanley 1984); this approach has been widely applied to assess the accuracy of diagnostic tests across multiple disorders (Swets 1982; Swets and Pickett 1982; Swets 1986a,b). For each participant, we computed the predicted values, or logits, from the logistic regression models. For each successive point on the logit scale we computed a sensitivity and specificity of the logit as a predictor of ADHD diagnosis, by predicting those higher than the cut-point to have

FIG. 1. Sifteo cubes used for Groundskeeper.

ADHD and the others not to have ADHD. These data were used to draw the ROC curve. ROC analysis summarizes diagnostic efficiency with the area under the curve (AUC) statistic. The AUC ranges from 0.5 (for a diagnostically useless test) to 1.0 (for a diagnostic test that is a perfect predictor). All analyses used STATA 13.1.

Results

The ADHD ($n=66$) and psychiatric control groups ($n=47$) did not differ significantly in sex (57% vs. 38% male, respectively; $\chi^2[1]=3.6$, $p=0.06$) or ethnicity (88% vs. 82% Caucasian, respectively; $\chi^2(3)=4.0$, $p=0.3$). They differed significantly in age (12.3 vs. 13.6; $t[105]=2.5$, $p=0.01$), which was used as a covariate along with sex (marginally significant) in subsequent analyses. Although no participants were taking ADHD medications at the time of testing, we included medication status (yes/no) as a covariate, because the non-ADHD group was significantly more likely to be on other medications at the time of testing compared with the ADHD group (82% vs. 60%; $\chi^2[1]=6.2$, $p=0.01$). Table 1 gives the distribution of psychiatric disorders.

Principal factors factor analysis with varimax rotation reduced the 106 Groundskeeper variables to 22 principal factors having eigenvalues >1.0 . These factors accounted for 68% of the variance of Groundskeeper scores. We entered these factors into a logistic regression model to predict the “gold standard” K-SADS diagnoses of ADHD. The eighth ($z=3.7$, $p<0.001$) and tenth ($z=3.1$, $p=0.002$) factors remained statistically significant after correcting for age, sex, and medication status. The area under the ROC curve (AUC) was 0.78 (Fig. 2; $\chi^2[2]=28$, $p<0.0001$). Neither factor interacted with age or sex in predicting ADHD diagnoses (p 's >0.10). We used the rotated factor loadings to determine which Groundskeeper scores accounted for the two significant factors. Table 2 shows the scores with loadings >0.25 on one or both of the factors. Boldface highlights the highest factor loadings. When these raw scores were used to predict the K-SADS ADHD diagnosis, the AUC was 0.79 ($\chi^2[12]=29.0$, $p=0.003$).

FIG. 2. Receiver operating characteristic curves.

TABLE 2. GROUNDSKEEPER VARIABLES ACCOUNTING FOR FACTORS 8 AND 10

Variable	Factor 8	Factor 10
Movement, incorrect	-0.3002	0.0229
Omissions	0.2579	-0.0475
Reaction time, correct	0.3403	0.1154
Reaction rate, correct	0.1249	0.2645
Reaction rate, correct, 2-7	0.7368	0.1026
Reaction rate, correct, 8-11	0.0490	0.7961
Reaction rate, Incorrect, 2-7	0.3729	-0.0655
Incorrect, S, V, A distractors	0.2566	-0.0722
Performance	-0.2755	-0.4649
Reaction time, correct, 2-7	0.7544	0.1804
Reaction time, Correct, 8-11	0.2150	0.6856
Reaction time, Incorrect, 8-11	-0.0418	0.3437

Movement incorrect: Average amount of tilt/movement during incorrect response.

Omissions: No response when target stimulus presented.

Reaction time, correct: Average reaction time from point image is shown to response (correct).

Response rate, correct: Ratio of time to response divided by total time image is shown.

Correct reaction rate, 2-7: Average of correct reaction rate levels 2-7.

Correct reaction rate, 8-11: Average of correct reaction rate levels 8-11.

Incorrect reaction rate, 2-7: Average of incorrect reaction rate levels 2-7.

Incorrect, S, V, A distractors: Incorrect, effect of spatial distractor with low frequency visual and auditory distractors.

Performance: Difference in reaction time between first and last levels demonstrating change in performance.

Reaction time, correct, 2-7: Average of correct reaction time levels 2-7.

Reaction time, correct, 8-11: Average of correct reaction time levels 8-11.

Reaction time, incorrect, 8-11: Average of incorrect reaction time levels 8-11.

Boldface indicates the highest factor loadings.

For the logistic regression analysis of the parent-rated Conners' subscales as a predictor of ADHD diagnoses, only the inattention scale was significant ($z=3.25$, $p=0.001$) after controlling for age, sex, and medication status. The AUC was 0.76 (Fig. 2; $\chi^2[1]=8.1$, $p=0.004$). This did not differ significantly from the Groundskeeper AUC ($\chi^2[1]=0.1$, $p=0.8$). The subscales did not interact with age or sex in predicting ADHD diagnoses (p 's >0.10). In another logistic regression model, the CPT percent correct score significantly predicted K-SADS ADHD diagnoses after controlling for age, sex, and medication status. The AUC was 0.62 (Fig. 2; $\chi^2[1]=5.0$, $p=0.03$). This was significantly lower than the Groundskeeper AUC ($\chi^2[1]=4.6$, $p=0.03$) and the CPT AUC ($\chi^2[1]=5.8$, $p=0.02$). When we combined the significant Groundskeeper factors with the Conners' inattention subscale and the CPT percent correct in the same model, all terms remained significant after controlling for age, sex, and medication status (all p 's <0.04) and the AUC was 0.87 (Fig. 2; $\chi^2[5]=49.2$, $p<0.0001$). This AUC was significantly greater than the CPT AUC ($\chi^2[1]=15.0$, $p=0.0001$), but did not differ significantly from either the Groundskeeper ($\chi^2[1]=0.5$, $p=0.5$) or Conners AUCs ($\chi^2[1]=1.7$, $p=0.19$).

The correlations between Groundskeeper factor eight and the Conners' scores ranged from -0.02 to 0.11. None were statistically significant (all p 's >0.05). The correlation between factor eight and the CPT percent correct score was 0.13 ($p=0.18$). The correlations between Groundskeeper factor ten and the Conners' scores ranged from -0.28 to 0.001. Only the correlations with the Executive Functioning score ($r=-0.21$, $p=0.03$) and the Aggression subscale ($r=-0.28$, $p=0.003$) were statistically significant. The correlation between factor ten and the CPT percent correct score was also significant ($r=-0.30$, $p=0.002$).

To clarify the degree to which the three logistic models based on the Groundskeeper, Conners', and CPT identified the same cases, we used each of the corresponding logistic regression models to compute the probability of each participant having K-SADS-diagnosed ADHD. We took a median split of these predicted probabilities and classified participants above the median as ADHD and those below the median as not ADHD. Only 14% of participants were predicted to have ADHD by the three methods. In this group, 79% were diagnosed with ADHD. Only 18% of participants were predicted to not have ADHD by all three models. In this group, 16% were diagnosed with ADHD. The κ coefficients of agreement were 0.15 for Groundskeeper versus Conners' ($z=1.6$, $p=0.06$), 0.18 for Groundskeeper versus CPT ($z=1.9$, $p=0.9$), and 0.3 for Conners versus CPT ($z=3.2$, $p=0.0007$).

Discussion

We found that the Groundskeeper game can significantly discriminate ADHD patients from other psychiatric patients. As measured by the AUC statistic from ROC analysis, the diagnostic accuracy of Groundskeeper (0.79) was as high as the accuracy of the Conners' parent rating scale (0.76), which is used as a screening or adjunct diagnostic tool for the diagnosis of ADHD.

The Conners' and Groundskeeper models had similar levels of diagnostic accuracy. Both predicted ADHD diagnoses more accurately than the CPT. The similar diagnostic accuracies between Groundskeeper and the Conners is remarkable, given that the questions asked of parents during the diagnostic interview are similar to the questions asked of the parent by the Conners' form. Both require subjective reports of ADHD symptoms and share method variance. In contrast, Groundskeeper is performed by the child with no parent involvement. Therefore, it is encouraging that an objective measure (Groundskeeper) is as accurate as subjective assessments of symptom criteria.

Although the AUCs for Groundskeeper and the Conners' rating scales were of similar magnitude, Figure 2 shows that their tradeoffs between sensitivity and specificity differ in clinically important ways. Groundskeeper maintains a false positive rate of zero for a sensitivity of 37%, a positive predictive power of 100%, and a negative predictive power of 53%. For the Conners' to achieve a sensitivity of 37%, the false positive rate would rise to 9.1%, the positive predictive power would be 85%, and the negative predictive power would be 51%. In Figure 2, this difference is seen as the Groundskeeper ROC being skewed toward the left side of the graph, whereas the Conners' ROC is spread more evenly throughout the graph. This means that the Conners' will be a better test for ADHD if a higher false positive rate can be tolerated. For example, the graphs show that, at a false positive rate of 25%, the Conners' has a sensitivity of 75%, a positive predictive power of 81%, and a negative predictive power of 69%; Groundskeeper has a sensitivity of 65%, a positive predictive power of 75%, and a negative predictive power of 65%. For screening tests, a high false positive rate can be tolerated if the costs of a second stage confirmation test are low, but if the costs are high, then the low false positive rate of Groundskeeper is preferred. A low false positive rate is essential for clinicians seeking to confirm an ADHD diagnosis about which they are uncertain, suggesting that Groundskeeper could be used for diagnostic confirmation. For example, in settings such as college health clinics, where the misuse and diversion of ADHD medications is a major concern, having a means of eliminating false positives would be very important. These examples are only illustrative. It would be premature to suggest specific cut-points for clinical practice.

Groundskeeper should not be used to diagnose ADHD independently from a clinical diagnosis.

The correlations among the Groundskeeper, Conners', and CPT scores were mostly low and not significant. Consistent with this, we found low kappa coefficients of agreement between each model's predictions of ADHD diagnoses. This lack of shared variance is probably because of the unique method variance for each method, the imperfect reliability of each method, and the possibility that each is sensitive to different components of the ADHD syndrome. Consistent with this latter interpretation, each score domain remained significant when all were included in the same model. Future work should address the possibility that multimodal assessments of ADHD are needed to create a highly accurate objective diagnostic tool.

Our work has limitations. We used a medicated, psychiatric control group, which we presumed would be relatively difficult to differentiate from ADHD compared with healthy controls. Use of the latter would likely lead to better diagnostic accuracy statistics. Because this was a preliminary study, we used a wide age range so that we could examine age effects. Although we found no effects of age on diagnostic accuracy, our sample was too small to detect small effects. The sample was also small relative to the number of variables analyzed. This required us to use factor analysis to limit our statistical tests. Because most research participants were taking medications of many different types, we cannot rule out medication effects completely and cannot be certain that our results will generalize to samples with a different profile of medication use. We also excluded youth with conduct disorder and tics, which further reduces the generalizability of our findings. Because the CPT and Groundskeeper tests were not counterbalanced, results could have been biased by sequence effects. Moreover, Groundskeeper benefited from the use of multiple variables derived from the test whereas for the CPT, we only used the percent correct, a commonly used index for the version we used. Our results may not generalize to other CPTs or to more complex analyses of CPT data. For these reasons, our results must be considered tentative until cross validated in an independent sample. We do not know if our results will generalize to adults with ADHD or if Groundskeeper will be useful for studying treatment effects over time. Future work must address these issues.

Conclusions

Despite these limitations, our results are encouraging. We demonstrated good discriminative ability when comparing ADHD patients with other psychiatric patients. The discriminative ability of Groundskeeper will likely be much better for discriminating ADHD patients from subjects not having psychiatric disorders. Future work should examine this possibility and should also determine if a revision to Groundskeeper or the application of different algorithms will be able to improve its diagnostic accuracy.

Clinical Significance

Although Groundskeeper is not ready to be used as a diagnostic tool, this work foreshadows to clinicians a new genre of clinical tool that is in development. For a diagnostic tool to be useful, it must be accepted by patients and must be engaging enough to assure that valid data are obtained. Clinicians also need to be aware that the ability of a tool to be successful depends on the diagnostic context. We have shown here that contrasting ADHD with other disorders is a difficult task, which suggests that future uses of Groundskeeper address a different diagnostic issue.

Disclosures

In the past year, Dr. Faraone received income, potential income, travel expenses, and/or research support from Akili Interactive Labs, Alcobra, Arbor, CogCubed, Impax, Ironshore, NeuroLifeSciences, Neurovance, Pfizer, Shire, and VAYA Pharma. With his institution, he has United States patent US20130217707 A1 for the use of sodium-hydrogen exchange inhibitors in the treatment of ADHD. In previous years, he received income or research support from Alcobra, Eli Lilly, Janssen, McNeil, Novartis, Otsuka, Pfizer, and Shire. Dr. Faraone receives royalties from the following books published by Guilford Press: *Straight Talk about Your Child's Mental Health*; Oxford University Press: *Schizophrenia: The Facts*; and Elsevier: *ADHD: Non-Pharmacologic Interventions*. In the past year, Dr. Newcorn is/has been an advisor/consultant to Alcobra, Biobehavioral Diagnostics, Ironshore, Neos, the National Football League (NFL), Rhodes, and Shire. He receives research support from Enzymotec and Shire, and serves on a data and safety monitoring board (DSMB) for Sunovion. In the previous 2 years he was also an advisor/consultant to GencoSciences, Lupin, and Neurovance. Dr. Heller is the co-founder and chief medical officer for CogCubed. Kurt Roots is the co-founder and chief executive officer of CogCubed. During the past 3 years, Dr. Adler has received grant support from the APSARD/Pound Foundation, Department of Veterans Affairs, Eli Lilly and Company, Enzymotec, Purdue Pharmaceuticals, Shire Pharmaceuticals, Sunovion Pharmaceuticals, and Theravance. He was also a consultant to Alcobra Pharmaceuticals, Enzymotec, Major League Baseball, NFL, Novartis Bioventures, Shire Pharmaceuticals, Sunovion Pharmaceuticals, and Theravance. He has received royalty payments (as inventor) from NYU for the license of adult ADHD scales and training materials since 2004. Dr. Antshel does not have any conflicts of interest.

References

- American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, 5th ed. Washington, DC: American Psychiatric Association; 2013.
- Bruchmuller K, Margraf J, Schneider S: Is ADHD diagnosed in accord with diagnostic criteria? Overdiagnosis and influence of client gender on diagnosis. *J Consult Clin Psychol* 80:128–138, 2012.
- Bush G, Valera EM, Seidman LJ: Functional neuroimaging of attention-deficit/hyperactivity disorder: a review and suggested future directions. *Biol Psychiatry* 57:1273–1284, 2005.
- Corkum PV, Siegel LS: Is the Continuous Performance Task a valuable research tool for use with children with attention-deficit-hyperactivity disorder? *J Child Psychol Psychiatry* 34:1217–1239, 1993.
- Dolgin E: FDA clearance paves way for computerized ADHD monitoring. *Nat Med* 20:454–455, 2014.
- Dovis S, Van der Oord S, Wiers RW, Prins PJ: Improving executive functioning in children with ADHD: Training multiple executive functions within the context of a computer game. A randomized double-blind placebo controlled trial. *PLoS One*. 10:e0121651, 2015.
- Drechsler R, Rizzo P, Steinhausen HC: Decision-making on an explicit risk-taking task in preadolescents with attention-deficit/hyperactivity disorder. *J Neural Transm* 115:201–209, 2008.
- Express Scripts: Turning Attention to ADHD. An Express Scripts Report. U.S. Medication Trends for Attention Deficit Hyperactivity Disorder. St. Louis, MO, Express Scripts Lab, 2014.
- Faraone SV: The scientific foundation for understanding attention-deficit/hyperactivity disorder as a valid psychiatric disorder. *Eur Child Adolesc Psychiatry* 14:1–10, 2005.
- Faraone SV, Biederman J, Spencer T, Wilens T, Seidman LJ, Mick E, Doyle AE: Attention deficit hyperactivity disorder in adults: An overview. *Biol Psychiatry* 48:9–20, 2000.
- Faraone SV, Bonvicini C, Scassellati C: Biomarkers in the diagnosis of ADHD—promising directions. *Curr Psychiatry Rep* 16:497, 2014.
- Fliers EA, de Hoog ML, Franke B, Faraone SV, Rommelse NN, Buitelaar JK, Nijhuis-van der Sanden MW: Actual motor performance and self-perceived motor competence in children with attention-deficit hyperactivity disorder compared with healthy siblings and peers. *J Dev Behav Pediatr* 31:35–40, 2010.
- Fliers E, Rommelse N, Vermeulen SH, Altink M, Buschgens CJ, Faraone SV, Sergeant JA, Franke B, Buitelaar JK: Motor coordination problems in children and adolescents with ADHD rated by parents and teachers: effects of age and gender. *J Neural Transm* 115:211–220, 2008.
- Fliers E, Vermeulen S, Rijdsdijk F, Altink M, Buschgens C, Rommelse N, Faraone SV, Sergeant J, Buitelaar J, Franke B: ADHD and poor motor performance from a family genetic perspective. *J Am Acad Child Adolesc Psychiatry* 48:25–34, 2009.
- Frazier-Wood AC, Bralten J, Arias-Vasquez A, Luman M, Oosterlaan J, Sergeant J, Faraone SV, Buitelaar J, Franke B, Kuntsi J, Rommelse NN: Neuropsychological intra-individual variability explains unique genetic variance of ADHD and shows suggestive linkage to chromosomes 12, 13, and 17. *Am J Med Genet B Neuropsychiatr Genet* 159B:131–140, 2012.
- Ginsberg Y, Quintero J, Anand E, Casillas M, Upadhyaya HP: Underdiagnosis of attention-deficit/hyperactivity disorder in adult patients: A review of the literature. *Prim Care Companion CNS Disord* 16:pii, 2014.
- Heller W: Neuropsychological mechanisms of individual differences in emotion, personality and arousal. *Neuropsychology* 7:476–489, 1993.
- Homack S, Riccio CA: Conners' Continuous Performance Test (2nd ed.; CCPT-II). *J Atten Disord* 9:556–558, 2006.
- Joseph N, Zhang-James Y, Perl A, Faraone SV: Oxidative stress and ADHD: A meta-analysis. *J Atten Disord* 19:915–924, 2015.
- McNeil BJ, Hanley JA: Statistical approaches to the analysis of receiver operating characteristic (ROC) curves. *Med Decis Making* 4:137–150, 1984.
- Mueller A, Candrian G, Kropotov JD, Ponomarev VA, Baschera GM: Classification of ADHD patients on the basis of independent ERP components using a machine learning system. *Nonlinear Biomed Phys* 4 Suppl 1:S1, 2010.
- Perry GM, Sagvolden T, Faraone SV: Intra-individual variability in genetic and environmental models of attention-deficit/hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet* 153B:1094–1101, 2010.
- Regier DA, Narrow WE, Clarke DE, Kraemer HC, Kuramoto SJ, Kuhl EA, Kupfer DJ: DSM-5 field trials in the United States and Canada, Part II: Test-retest reliability of selected categorical diagnoses. *Am J Psychiatry* 170:59–70, 2013.
- Riccio CA, Reynolds CR: Continuous performance tests are sensitive to ADHD in adults but lack specificity. A review and critique for differential diagnosis. *Ann N Y Acad Sci* 931:113–139, 2001.
- Ritsner MS (ed): *The Handbook of Neuropsychiatric Biomarkers, Endophenotypes and Genes: Volume I: Neuropsychological Endophenotypes and Biomarkers 2009th Edition*. Amsterdam (Netherlands): Springer; 2009.
- Scassellati C, Bonvicini C, Faraone SV, Gennarelli M: Biomarkers and attention-deficit/hyperactivity disorder: A systematic review and meta-analyses. *J Am Acad Child Adolesc Psychiatry* 51:1003–1019 e1020, 2012.
- Sciotto MJ, Eisenberg M: Evaluating the evidence for and against the overdiagnosis of ADHD. *J Atten Disord* 11:106–113, 2007.
- Silk TJ, Vance A, Rinehart N, Bradshaw JL, Cunningham R: White-matter abnormalities in attention deficit hyperactivity disorder: A diffusion tensor imaging study. *Hum Brain Mapp* 30:2757–2765, 2009.
- Snyder SM, Rugino TA, Hornig M, Stein MA: Integration of an EEG biomarker with a clinician's ADHD evaluation. *Brain Behav* 5:e00330, 2015.

- Sumner CR, Haynes VS, Teicher MH, Newcorn JH: Does placebo response differ between objective and subjective measures in children with attention-deficit/hyperactivity disorder? *Postgrad Med* 122:52–61, 2010.
- Swets JA: Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychol Bull* 99:181–198, 1986a.
- Swets JA: Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychol Bull* 99:110–117, 1986b.
- Swets JA: Sensitivities and specificities of diagnostic tests. *JAMA* 248:548–549, 1982.
- Swets JA, Pickett RM. *Evaluation of Diagnostic Systems: Methods From Signal Detection Theory*. New York: Academic Press; 1982.
- Thome J, Ehliis AC, Fallgatter AJ, Krauel K, Lange KW, Riederer P, Romanos M, Taurines R, Tucha O, Uzbekov M, Gerlach M: Biomarkers for attention-deficit/hyperactivity disorder (ADHD). A consensus report of the WFSBP task force on biological markers and the World Federation of ADHD. *World J Biol Psychiatry* 13:379–400, 2012.
- Toplak ME, Dockstader C, Tannock R: Temporal information processing in ADHD: Findings to date and new methods. *J Neurosci Methods* 151:15–29, 2006.
- Van der Oord S, Ponsioen AJ, Geurts HM, Brink EL, Prins PJ: A pilot study of the efficacy of a computerized executive functioning remediation training with game elements for children with ADHD in an outpatient setting: Outcome on parent- and teacher-rated executive functioning and ADHD behavior. *J Atten Disord* 18:699–712, 2014.
- Visser SN, Danielson ML, Bitsko RH, Holbrook JR, Kogan MD, Ghandour RM, Perou R, Blumberg SJ: Trends in the parent-report of health care provider-diagnosed and medicated attention-deficit/hyperactivity disorder: United States, 2003–2011. *J Am Acad Child Adolesc Psychiatry* 53:34–46 e32, 2014.

Address correspondence to:

Stephen V. Faraone, PhD

SUNY Upstate Medical University

750 East Adams St.

Syracuse, NY 13210

E-mail: sfaraone@childpsychresearch.org

Appendix A

This script is read to each participant prior to game administration.

Today you will be playing a game called Groundskeeper. It will be used to measure your attention. The goal is to move the mallet cube to the left or right side of the cube showing an image of a gopher. When you have a correct hit, you will hear a “boink” noise. There will be no noise for an incorrect hit. Keep in mind you only want to hit the gopher! Do not let other images, like birds or rabbits, or noises distract you from this task.

There will be a short pause between each level. The first level that you play is for practice and will not be counted against your final score. After each level, a voice will instruct you to move the cubes to colored dots on the game board. Note that the mallet cube should not be moved to these dots.

Be sure to use two hands and hold the cubes together until you hear a sound. If you do not hear a boink sound, you do not get a point. Move as quickly as possible!

Appendix B

Summary of Groundskeeper Sessions

Each session is 1.5 minutes with 20 seconds between sessions. There will be a 10 second countdown on the screen prior to each session. Program will automatically progress to the next session without intervention by person conducting the test. Total run time is 24 minutes (plus 5 minutes for pauses).

Session 1: Screen shot of gopher, groundskeeper, or grass (neutral) presented for 1, 2, or 3 seconds at a random frequency.

Go/no go task.

Session 2: Gopher, groundskeeper, or grass screen shot with visual disturbance at a low degree (one bird showing up on screen) alternating with screen shot of grass presented for 1, 2, or 3 seconds at random frequency.

Visual disturbance, low frequency.

Session 3: Gopher, groundskeeper, or grass screenshot with visual disturbance at a high degree (bird and rabbit) alternating with screen shot of grass presented for 1, 2, or 3 seconds at random frequency.

Visual disturbance, high frequency.

Session 4: Screen shot of gopher, groundskeeper, or grass screen shot presented for 1, 2, or 3 seconds at a random frequency with

auditory disturbance at a low degree (one bird chirping) occurring at random frequencies for 1, 2, or 3 seconds, not in concert with screen shot frequency.

Auditory disturbance, low frequency.

Session 5: Screen shot of gopher, groundskeeper, or grass screenshot presented for 1, 2, or 3 seconds at a random frequency with auditory disturbance at a high degree (multiple birds chirping) occurring at random frequencies for 1, 2, or 3 seconds, not in concert with screen shot frequency.

Auditory disturbance, high frequency.

Session 6: Gopher, groundskeeper, or grass screen shot with visual and auditory disturbances at a low degree (one bird chirping) occurring at random frequencies for 1, 2, or 3 seconds.

Visual and auditory disturbance, low frequency.

Session 7: Gopher, groundskeeper, or grass screen shot with visual and auditory disturbances at a high degree (bird and rabbit and chirping) occurring at random frequencies for 1, 2, or 3 seconds.

Visual and auditory disturbance, high frequency.

Session 8: Gopher, groundskeeper, or grass screen shot with spatial disturbance at a low degree. Cube set diagonally. Spaced at 5.08 cm apart and occurring at random frequencies for 1, 2, or 3 seconds.

Spatial disturbance, low frequency.

Session 9: Gopher, groundskeeper, or grass screen shot with spatial disturbance at a low degree combined with low frequency visual disturbance. Cube set diagonally. Spaced at 5.08 cm apart and occurring at random frequencies for 1, 2, or 3 seconds.

Spatial disturbance, low frequency, low frequency visual disturbance.

Session 10: Gopher, groundskeeper, or grass screen shot with spatial disturbance at a low degree combined with low frequency auditory disturbance. Cube set diagonally. Spaced at 5.08 cm apart and occurring at random frequencies for 1, 2, or 3 seconds.

Spatial disturbance, low frequency, low frequency auditory disturbance.

Session 11: Gopher, groundskeeper, or grass screen shot with spatial disturbance at a low degree combined with low frequency visual and auditory disturbance. Cube set diagonally. Spaced at 5.08 cm apart and occurring at random frequencies for 1, 2, or 3 seconds.

Spatial disturbance, low frequency, low frequency visual and auditory disturbance.

Session 12: Gopher, groundskeeper, or grass screen shot with spatial disturbance at a low degree. Cube set diagonally. Spaced at 7.62 cm apart and occurring at random frequencies for 1, 2, or 3 seconds.

Spatial disturbance, high frequency.

Session 13: Gopher, groundskeeper, or grass screen shot with spatial disturbance at a high degree combined with low frequency visual disturbance. Cube set diagonally. Spaced at 7.62 cm apart and occurring at random frequencies for 1, 2, or 3 seconds.

Spatial disturbance, high frequency, low frequency visual disturbance.

Session 14: Gopher, groundskeeper, or grass screen shot with spatial disturbance at a high degree combined with low frequency auditory disturbance. Cube set diagonally. Spaced at 7.62 cm apart and occurring at random frequencies for 1, 2, or 3 seconds.

Spatial disturbance, high frequency, low frequency auditory disturbance.

Session 15: Gopher, groundskeeper, or grass screen shot with spatial disturbance at a high degree combined with low frequency visual and auditory disturbance. Cube set diagonally. Spaced at 7.62 cm apart and occurring at random frequencies for 1, 2, or 3 seconds.

Spatial disturbance, high frequency, low frequency visual and auditory disturbance.

Session 16: Screen shot of gopher, groundskeeper, or grass (neutral) presented for 1, 2, or 3 seconds at a random frequency.

Go/no go task with learning curve.

APPENDIX C: VARIABLES DERIVED FROM GROUNDSKEEPER

<i>Variable</i>	<i>Formula</i>	<i>Level</i>	<i>Category</i>	<i>Description</i>
eval_id	NA	All	Identifier	Unique ID for each transformed evaluation
mCRR	$\frac{correct}{incorrect + correct}$	All	Correct	Ratio of correct to incorrect responses
mIRR	$\frac{incorrect}{incorrect + correct}$	All	Incorrect	Ratio of incorrect to correct responses
mEOM	$\frac{error_omission}{incorrect + correct}$	All	Omission	Ratio of omissions to both correct and incorrect responses
TiltX	$TiltX = TiltLeft + TiltRight$	All	Movement	Total movement left/right as denoted by an acceleration state change
TiltY	$TiltY = TiltUp + TiltDown$	All	Movement	Total movement up/down as denoted by an acceleration state change
TiltZ	$\sum_{m < TiltZ < n} f(TiltZ)$	All	Movement	Total movement diagonal as denoted by an acceleration state change
Movement	$m = TiltFlip, n = TiltFlipBack$			
VirtualTicks	$TiltX + TiltY + TiltZ$	All	Movement Counter	Composite variable totaling accelerometer state changes
correct	$\frac{correct}{correctdisplay} \times 100$	All	Correct	Measurable time unit recorded as ticks, which operates within the state machine
incorrect	$\frac{incorrect}{imagedraws - correctdisplay} \times 100$	All	Incorrect	Normalization of correct response, whereby the participant has hit the gopher
error_omission	$\frac{correct - correctdisplay}{correct - correctdisplay} \times 100$	All	Omission	Normalization of incorrect response, whereby the participant has hit something other than the gopher
correct_reaction	$\frac{\sum_{m < VirtualTicks < n} f(VirtualTicks)}{n}$	All	Reaction time	Normalization of omission response, whereby the participant missed the gopher and hit nothing
incorrect_reaction	$n = Response_{correct} \quad m = NewImageld$ $\frac{\sum_{m < VirtualTicks < n} f(VirtualTicks)}{n}$	All	Reaction time	Average reaction time from point image is shown to correct response
offult_reaction	$n = Response_{incorrect} \quad m = NewImageld$ $\frac{\sum_{m < Movement < n} f(Movement)}{n}$ $n = Response_x \quad m = NewImageld$	All	Movement	Average reaction time from point image is shown to incorrect response
				Average movement from point in time cubes are neighbored until mallet cube removed

(continued)

<i>Variable</i>	<i>Formula</i>	<i>Level</i>	<i>Category</i>	<i>Description</i>
ontilt_reaction	$\frac{\sum_{m < \text{Movement} < n} f(\text{Movement})}{n}$ $m = \text{Response}_x, n = \text{NewImageld}$	All	Movement	Average movement from point in time gopher image is displayed until neighbored with mallet
neighbor_reaction	$\frac{\sum_{m < \text{VirtualTicks} < n} f(\text{VirtualTicks})}{n}$ $m = \text{NeighborEvent}_{on}, n = \text{NeighborEvent}_{off}$	All	Reaction time	Average amount of time cubes held together
correct_reaction_movement	$\frac{\sum_{m < \text{Movement} < n} f(\text{Movement})}{n}$ $n = \text{Response}_{correct}, m = \text{NewImageld}$	All	Movement	Average amount of tilt during a correct response
incorrect_reaction_movement	$\frac{\sum_{m < \text{Movement} < n} f(\text{Movement})}{n}$ $n = \text{Response}_{incorrect}, m = \text{NewImageld}$	All	Movement	Average amount of tilt during an incorrect response
omission_movement	$\frac{\sum_{m < \text{Movement} < n} f(\text{Movement})}{n}$ $n = \text{Response}_{null}, m = \text{NewImageld}$	All	Movement	Average amount of tilt during an omission
correct_speed_variance	$\frac{\sqrt{((\text{TiltX} \times \text{TiltX}) + (\text{TiltY} \times \text{TiltY}))}}{\max(\text{Response}_{correct})} \div n$	All	Movement	Average amount of speed during a correct response
incorrect_speed_variance	$\frac{\sqrt{((\text{TiltX} \times \text{TiltX}) + (\text{TiltY} \times \text{TiltY}))}}{\max(\text{Response}_{incorrect})} \div n$	All	Movement	Average amount of speed during an incorrect response
mCRRate2_7	$\frac{\text{correct}}{\text{incorrect} + \text{correct}}$	2-7	Correct	Ratio of correct to incorrect responses
mICRRate2_7	$\frac{\text{incorrect}}{\text{incorrect} + \text{correct}}$	2-7	Incorrect	Ratio of incorrect to correct responses
mNR2_7	$\frac{\sum_{m < \text{VirtualTicks} < n} f(\text{VirtualTicks})}{n}$ $m = \text{NeighborEvent}_{on}, n = \text{NeighborEvent}_{off}$	2-7	Reaction time	Average of neighbor reaction
mCT2_7	$\text{TiltX} + \text{TiltY} + \text{TiltZ}$	2-7	Movement	Sum of all movement
mCR2_7	$\frac{\sum_{m < \text{VirtualTicks} < n} f(\text{VirtualTicks})}{n}$ $n = \text{Response}_{correct}, m = \text{NewImageld}$	2-7	Reaction time	Average of correct reaction

(continued)

Variable	Formula	Level	Category	Description
mICR2_7	$\frac{\sum_{m < \text{VirtualTicks} < n} f(\text{VirtualTicks})}{n}$ $n = \text{Response}_{\text{incorrect}} m = \text{NewImageld}$	2-7	Reaction time	Average of incorrect reaction
mOTR2_7	$\frac{\sum_{m < \text{Movement} < n} f(\text{Movement})}{n}$ $n = \text{Response}_x m = \text{NewImageld}$	2-7	Movement	Average of offtilt_reaction
mONTR2_7	$\frac{\sum_{m < \text{Movement} < n} f(\text{Movement})}{n}$ $m = \text{Response}_x n = \text{NewImageld}$	2-7	Movement	Average of ontilt_reaction
mMMO2_7	$\frac{\text{TiltX} + \text{TiltY} + \text{TiltZ}}{\text{correct}}$	2-7	Movement	Average of all other movement
mAOM2_7	$\frac{\text{TiltX} + \text{TiltY} + \text{TiltZ}}{\text{correct} + \text{incorrect}}$	8-11	Correct	Ratio of correct responses to incorrect
mCRRate8_11	$\frac{\text{incorrect}}{\text{incorrect} + \text{correct}}$	8-11	Incorrect	Ratio of incorrect responses to correct
mICRRate8_11	$\frac{\text{incorrect}}{\text{incorrect} + \text{correct}}$	8-11	Incorrect	Ratio of incorrect responses to correct
mNR8_11	$\frac{\sum_{m < \text{VirtualTicks} < n} f(\text{VirtualTicks})}{n}$ $m = \text{NeighborEvent}_{\text{on}} n = \text{NeighborEvent}_{\text{off}}$	8-11	Reaction time	Average of neighbor reaction
mCT8_11	$\frac{\text{TiltX} + \text{TiltY} + \text{TiltZ}}{\text{correct}}$	8-11	Movement	Sum of all movement
mCR8_11	$\frac{\sum_{m < \text{VirtualTicks} < n} f(\text{VirtualTicks})}{n}$ $n = \text{Response}_{\text{correct}} m = \text{NewImageld}$	8-11	Reaction time	Average of correct reaction
mICR8_11	$\frac{\sum_{m < \text{VirtualTicks} < n} f(\text{VirtualTicks})}{n}$ $n = \text{Response}_{\text{incorrect}} m = \text{NewImageld}$	8-11	Reaction time	Average of incorrect reaction
mOTR8_11	$\frac{\sum_{m < \text{Movement} < n} f(\text{Movement})}{n}$ $n = \text{Response}_x m = \text{NewImageld}$	8-11	Movement	Average of offtilt_reaction
mONTR8_11	$\frac{\sum_{m < \text{Movement} < n} f(\text{Movement})}{n}$ $m = \text{Response}_x n = \text{NewImageld}$	8-11	Movement	Average of ontilt_reaction
mMMO8_11	$\frac{\text{TiltX} + \text{TiltY} + \text{TiltZ}}{\text{correct}}$	8-11	Movement	Amount of all other movement
mAOM8_11	$\frac{\text{TiltX} + \text{TiltY} + \text{TiltZ}}{\text{correct} + \text{incorrect}}$	8-11	Movement	Amount of all other movement excluding the mallet

(continued)

<i>Variable</i>	<i>Formula</i>	<i>Level</i>	<i>Category</i>	<i>Description</i>
mCRRate12_15	$\frac{\text{correct}}{\text{incorrect} + \text{correct}}$	12-15	Correct	Ratio of correct to incorrect responses
mICRRate12_15	$\frac{\text{incorrect}}{\text{incorrect} + \text{correct}}$	12-15	Incorrect	Ratio of incorrect to correct responses
mNR12_15	$\frac{\sum_{m < \text{VirtualTicks} < n} f(\text{VirtualTicks})}{n}$	12-15	Reaction time	Average of neighbor reaction
mCT12_15	$m = \text{NeighborEvent}_{on} \quad n = \text{NeighborEvent}_{off}$			
mCR12_15	$\frac{\text{TiltX} + \text{TiltY} + \text{TiltZ}}{\sum_{m < \text{VirtualTicks} < n} f(\text{VirtualTicks})}$	12-15 12-15	Movement Reaction time	Sum of all movement Average of correct reaction
mICR12_15	$n = \text{Response}_{correct} \quad m = \text{NewImageld}$ $\frac{\sum_{m < \text{VirtualTicks} < n} f(\text{VirtualTicks})}{n}$	12-15	Reaction time	Average of incorrect reaction
mOTR12_15	$n = \text{Response}_{incorrect} \quad m = \text{NewImageld}$ $\frac{\sum_{m < \text{Movement} < n} f(\text{Movement})}{n}$	12-15	Movement	Average of offilt_reaction
mONTR12_15	$n = \text{Response}_x \quad m = \text{NewImageld}$ $\frac{\sum_{m < \text{Movement} < n} f(\text{Movement})}{n}$	12-15	Movement	Average of onilt_reaction
mMMO12_15	$m = \text{Response}_n \quad n = \text{NewImageld}$ $\frac{\text{TiltX} + \text{TiltY} + \text{TiltZ}}{\sum_{m < \text{Movement} < n} f(\text{Movement})}$	12-15	Movement	Amount of mallet movement
mAOM12_15	$\frac{\text{TiltX} + \text{TiltY} + \text{TiltZ}}{\sum_{m < \text{Movement} < n} f(\text{Movement})}$	12-15	Movement	Amount of all other movement excluding the mallet
mMRATIO13_15	$\frac{\text{mMMO13_15}}{\text{mAOM13_15}}$	13-15	Movement	Ratio of mallet movement to all other cube movement
mMRATIO12_15	$\frac{\text{mMMO12_15}}{\text{mAOM12_15}}$	12-15	Movement	Ratio of mallet movement to all other cube movement
VSC_VFC	$\frac{\text{mCRR2/mCRR3}}{\text{mIRR2/mIRR3}}$	2-3	Distraction	Low/high visual distraction success/failure comparison
VSC_VFC_O	$\frac{\text{mEOM2/mEOM3}}{\text{mCRR2/mCRR3}}$	2-3	Distraction	Low/high visual distraction omission/correct comparison
ASC_AFC	$\frac{\text{mCRR4/mCRR5}}{\text{mIRR4/mIRR5}}$	4-5	Distraction	Low/high auditory distraction correct/failure comparison
ASC_AFC_O	$\frac{\text{mEOM4/mEOM5}}{\text{mCRR4/mCRR5}}$	4-5	Distraction	Low/high auditory distraction omission/correct comparison

(continued)

APPENDIX C. (CONTINUED)

<i>Variable</i>	<i>Formula</i>	<i>Level</i>	<i>Category</i>	<i>Description</i>
LA2TSC_LA2TFC_HA2TSC_HA2TFC	$\frac{((mCRR2/mCRR6)/(mIRR2/mIRR6))}{((mCRR3/mCRR7)/(mIRR3/mIRR7))}$	2,3,6,7	Distraction	Low/high audio and low/high visual to total success/failure comparison
LA2TSC_LA2TFC_HA2TSC_HA2TFC_O	$\frac{((mCRR2/mCRR6)/(mEOM2/mEOM6))}{((mCRR3/mCRR7)/(mEOM3/mEOM7))}$	2,3,6,7	Distraction	Low/high audio and low/high visual to total correct/omission comparison
CVSC_CVFC	$\frac{((mCRR2/mCRR6)/(mCRR3/mCRR7))}{((mIRR2/mIRR6)/(mIRR3/mIRR7))}$	2,3,6,7	Distraction	Low visual to low/high visual and audio to total success/failure comparison
CVSC_CVFC_O	$\frac{((mEOM2/mEOM6)/(mEOM3/mEOM7))}{((mCRR2/mCRR6)/(mCRR3/mCRR7))}$	2,3,6,7	Distraction	Low visual to low/high visual and audio distraction to total omissions/correct comparison
CASC_CAFc	$\frac{((mCRR4/mCRR6)/(mCRR5/mCRR7))}{((mIRR4/mIRR6)/(mIRR5/mIRR7))}$	4,5,6,7	Distraction	Low audio to low/high visual and audio distraction success/failure comparison
CASC_CAFc_O	$\frac{((mEOM4/mEOM6)/(mEOM5/mEOM7))}{((mCRR4/mCRR6)/(mCRR5/mCRR7))}$	4,5,6,7	Distraction	Low audio to low/high visual and audio distraction omission/correct comparison
SSC_SFC	$\frac{(mCRR6/mCRR11)}{(mIRR6/mIRR11)}$	6,11	Distraction	Low spatial distraction success/failure comparison during low frequency visual and auditory disturbance
SSC_SFC_O	$\frac{(mEOM6/mEOM11)}{(mCRR6/mCRR11)}$	6,11	Distraction	Low spatial distraction omission/success comparison during low frequency visual and auditory disturbance
SSC_SFC_NR	$\frac{mNR6}{mNR11}$	6,11	Distraction	Low spatial distraction of neighbor reaction comparison during low frequency visual and auditory disturbance
SSC_SFC_CR	$\frac{mCR6}{mCR11}$	6,11	Distraction	Low spatial distraction of correct response comparison during low frequency visual and auditory disturbance
SSC_SFC_ICR	$\frac{mICR6}{mICR11}$	6,11	Distraction	Low spatial distraction of incorrect response comparison during low frequency visual and auditory disturbance
CSSC	$\frac{(mCRR11/mCRR15)}{(mIRR11/mIRR15)}$	11,15	Distraction	Low spatial, visual, auditory distraction correct/incorrect comparison to high spatial, visual, auditory distraction
CSSC_O	$\frac{(mEOM11/mEOM15)}{(mCRR11/mCRR15)}$	11,15	Distraction	Low spatial, visual, auditory distraction omission/correct comparison to high spatial, visual, auditory distraction
LearningCurve	$\left(\left(\frac{mCRR16}{mCRR1} \right) - 1 \right) \times 100$	1,16	Learning curve	Measure correct responses from first level compared with last level
LearningCurve_O	$\left(\left(\frac{mEOM16}{mEOM1} \right) - 1 \right) \times 100$	1,16	Learning curve	Measure omissions from first level compared with last level

(continued)

APPENDIX C. (CONTINUED)

<i>Variable</i>	<i>Formula</i>	<i>Level</i>	<i>Category</i>	<i>Description</i>
LearningCurve_CT	$\left(\left(\frac{mCT16}{mCT1}\right) - 1\right) \times 100$	1,16	Learning curve	Measure movement amount from first level compared with last level
LearningCurve_NR	$\left(\left(\frac{mNR16}{mNR1}\right) - 1\right) \times 100$	1,16	Learning curve	Measure neighbor reaction amount from first level to last level
LearningCurve_CR	$\left(\left(\frac{mCR16}{mCR11}\right) - 1\right) \times 100$	1,16	Learning curve	Measure correct reaction time from first level to last level
LearningCurve_ICR	$\left(\left(\frac{mICR16}{mICR1}\right) - 1\right) \times 100$	1,16	Learning curve	Measure incorrect reaction time from first level to last level
mallet_movement_only	<i>TiltX+TiltY+TiltZ</i>	All	Movement	Total amount of mallet movement
all_other_movement	<i>TiltX+TiltY+TiltZ</i>	All	Movement	Total of all other movement excluding the mallet
subsecond_response_total	<i>K</i>	All	Reaction time	Total number of times when time to respond is ≤ 1 tick (1 tick = 10 Hz)
topbottomhit	<i>1 or 0</i>	All	Auxiliary movement	Total for device hits from top or bottom. Measured by response to top or bottom of device.
cubepressed	<i>1 or 0</i>	All	Auxiliary movement	Total number of times player pushes or presses cube screen - no real reason why player should do this
doublehit	<i>K</i>	All	Auxiliary movement	Total number of times cubes are neighbored together after a response has already been made
Gender	<i>NA</i>	NA	Demographic	M or F assignment
Age	<i>NA</i>	NA	Demographic	Discrete assignment of age
Inattention	<i>NA</i>	NA	Condition	Diagnosis made by clinicians
Hyperactivity	<i>NA</i>	NA	Condition	Diagnosis made by clinicians
Combined	<i>NA</i>	NA	Condition	Inattention and hyperactivity
Depression	<i>NA</i>	NA	Condition	Diagnosis made by clinicians
Autism	<i>NA</i>	NA	Condition	Diagnosis made by clinicians
Anxiety	<i>NA</i>	NA	Condition	Diagnosis made by clinicians