INTRODUCTION

Attention Deficit Hyperactivity Disorder (ADHD) is one of the most common psychiatric disorders of childhood (NIH, 1998). Teacher and parent rating scales are often used to assess ADHD symptoms (Barkley, 1991). Yet there is a trend in the increased use of cognitive measures as an adjunct to subjective rating scales to enhance diagnostic decision-making (Barkley, 1991; Berlin, Bohlin, Nyberg, & Janoels, 2004). Research on ADHD suggests that assessment should be accomplished through a multi-method procedure (Barkley, 1998; Guay, Parent, & Lageix, in press).

The Continuous Performance Test (CPT) is one of the most frequently used tasks in the clinical assessment of ADHD (Rapport, Chung, Shore, Denney et al., 2000). There are several versions of the test. Generally, the child is asked to sustain his attention and react to the presence of targets while ignoring distracters. A large multi-site study compared the performance of 498 children presenting with ADHD according to gender and type of comorbidity on a CPT and ratings scales (Newcorn, Halperin, Jensen, Abikoff et al., 2001). It was found that inattention and impulsivity errors on the CPT were high in all ADHD subgroups, but dominant error type on the CPT and ratings differed with respect to comorbidity and gender. Children with ADHD and conduct disorder were more impulsive on both types of measures. Children with ADHD and anxiety disorders appeared more inattentive on ratings only. Girls' performance was less impaired than boys' performance on most ratings and on several CPT indices, particularly impulsivity. Girls with ADHD and anxiety made fewer impulsivity errors than girls with ADHD only. It was concluded that the CPT is a sensitive and valid measure for the assessment of ADHD with or without the presence of comorbidity, but lacks specificity. Börger & van der Meere (2000) noted that, during CPT performance, children with ADHD tend to look away from the monitor; this relevant behaviour is typically lost in the assessment process using flatscreen stimulus delivery.

It was demonstrated that most current laboratory methods for assessing ADHD symptoms have a low to moderate degree of ecological validity, with some proving to be clearly unsatisfactory (Barkley, 1991). Ecological validity refers to the degree to which measurement results represent the actual target behaviours as they occur in real life settings (Barkley, 1991). Direct observations of behaviour in its natural setting would represent a highly ecologically valid measure. In contrast, weak ecological validity is represented by a measure of behavior that is unlikely to be encountered in a real life setting, as is exemplified by traditional CPTs (Barkley, 1991). The closer the measure is to direct observation in a natural setting, the more ecologically valid its results should be.

Since direct observation of behaviour is time consuming, expensive, and prone to the influence of subjective judgment, an alternate means of assessing behaviour is desirable. Virtual reality (VR) offers an elegant solution. Attention abilities have been addressed using VR (Wann, Rushton, Smyth & Jones, 1997; Rizzo, Buckwalter, Neumann, Chua, et al., 1999), and has shown promising results in the assessment of ADHD symptoms in children (Rizzo, Bowerly, Buckwalter, Klimchuk, et al., 2006), in the assessment of driving abilities of teenagers and
adults presenting with ADHD (Barkley, 2004), and in the treatment of ADHD symptoms (Cho, Ku, Jang, Kim, et al., 2002). VR offers the clear advantage of placing the participant in a realistic environment. Yet, it remains a test and, to some extent, carries the limitation, in terms of ecological validity, of being more attractive and playful than corresponding real-life situations for many children.

In using VR for the assessment of ADHD symptoms, research results (Rizzo et al., 2006) indicate that children with ADHD, compared with normal controls, have slower correct hit reaction times (RT), higher RT variability, and more omission and commission errors. While effect sizes (d) for variables derived from traditional CPT and other psychological tests seldom exceed 1.0 (Frazier, Demaree & Youngstrom, 2004), these effect sizes from the VR CPT ranged from 1.05 to 2.07, the highest one being obtained on the number of omissions. The task used by these researchers consisted of a CPT presented on a chalkboard within a virtual classroom environment. Two conditions were used, with and without distracters, and results were equally or more significant in the distraction condition. Such impressive effect sizes certainly justify further studies. Currently, large effect sizes are only obtained from questionnaires, with d values ranging from 1.3 to 3.7, depending on the questionnaire (Green, Wong, Atkins, Taylor et al., 1999). This, however, is partly tautological since the diagnosis of ADHD is explicitly based on testimonies from parents and teachers concerning specific behaviours also investigated by the questionnaires.

The VR Classroom is a computer-based program that uses a head mounted display (HMD) to deliver both visual and auditory stimuli within a simulated classroom virtual environment. Within the VR Classroom, a CPT designed to test attention in school-aged children is administered. The child is immersed in a 360-degree classroom environment and presented with a standard A-K CPT on a chalkboard at the front of the class. This task requires children to hit a response button whenever they see an A-K sequence of letters appear over a six minute period. During the assessment, visual and auditory distracters are presented (i.e. ambient classroom and hall noise, movement of virtual classmates, activity occurring outside the window, etc.). Attention performance in the VR Classroom, like for other CPT measures, is quantified in terms of reaction times and its standard error and of commission and omission errors. While the task is performed, a head tracking device monitors movements, documenting to what extent the child turns away from the stimulus delivery location.

Objectives of this pilot study
1. Compare performance from children diagnosed with ADHD and children in a control group on the VR Classroom test, including head movement measures.
2. Determine if the cognitive profile of children with ADHD outlined by the traditional CPT corresponds to the one outlined by the VR Classroom (ecological CPT).
3. Compare performance on the VR Classroom to a standard neuropsychological battery on variables of commission and omission errors, reaction time (RT) and its standard error and determine ecological validity according to Barkley’s criteria (1991).

RESEARCH METHODOLOGY

Participants: Twenty-two participants (15 boys with ADHD and 7 boys in a comparison group) all aged between 9 and 13. Boys with ADHD were recruited from the Montreal area through various health agencies and Rivière-des-Prairies Hospital. Boys in the comparison group were unaffected siblings of participants with ADHD. Both groups were tested with the VR Classroom, standard neuropsychological tests, and parent ratings on behavioural questionnaires. ADHD-diagnosed participants were tested prior to taking their daily medication and tests were not repeated if recent administration results were already available in medical chart (less than six months).

VR Classroom procedure: Participants sat on a standard “school chair,” wearing the HMD displaying the interior of a classroom. The scenario consisted of a standard rectangular classroom environment containing four rows of desks, a teacher’s desk at the front, a chalkboard across the front wall, a female virtual teacher between the desk and chalkboard, nine virtual children seated at desks around the participant, a large window on the left side wall looking out onto a street with moving vehicles, and a pair of door-
ways, one at each end of the wall facing the window, through which activity occurred. The technician then instructed the participant to spend a minute looking around the room and naming various objects observed. The participant was provided with a one-minute practice of the virtual task before actual testing started. The virtual teacher then warned the participant that the testing proper was about to start and instructed him to view a series of letters appearing on the chalkboard and to hit the left mouse button only after he viewed the letter “K” preceded by an “A” (successive discrimination task) and withhold their response to any other stimulus letter. This A-K version of the CPT consists of the letters A, B, C, F, G, H, I, J, K, L, S, T, U, V, X, Y and Z. The letters are white on a green background (virtual chalkboard) presented at a fixed position directly in front of the child. The stimuli remained on the screen for 150ms, with a fixed 1200ms stimulus onset asynchrony. Three hundred stimuli were presented in the six minutes task. The target letter K (correct hit stimulus) and the letter K without the A (incorrect hit stimulus) each appeared with equal probability of 10%. The letters A and H both appeared with a frequency of 20%. The remaining fourteen letters occurred with equal probability. Stimuli occurred in the presence of mixed 3D immersive audio and visual distracters. Distracters consisted of (a) pure auditory: constant ambient classroom sounds (i.e., whispering, pencils dropping, chairs moving, etc.), (b) pure visual: paper airplane flying directly across the participant’s field of view (occurring three times throughout the 6-minute task), (c) mixed audio and visual: cars and school buses “rumbling by” outside the window on the left (occurring three times each), and a virtual person coming in and out of doors on the right side of the classroom, with sounds of the door “creaking open,” footsteps, and hallway activity (occurring once). Reaction time, response variability, and commission and omission errors were used as performance measures, while the tracking device on the HMD was used to monitor head movement.

VR performance was also compared with results from standard neuropsychological tests: Color-word interference Test (Stroop; DKEFS, 2001) conditions 3-inhibition and 4-flexibility, on total errors variables; CPT-II (CPT; Conners, 2000; lasts 15 minutes; task is to inhibit response to letter X) on RT, RT standard error, omission and commission errors variables; d2 (Brickenkamp & Zillmer, 1998) on omission and commission errors variables; Strength Difficulties Questionnaire (SDQ; Goodman, 1997) ADHD and total problems subscales, ADHD Rating Scale-IV (DuPaul et al., 1998) total problem subscale and Achenbach System of Empirically Based Assessment (CBCL; Achenbach & Rescorla, 2001) ADHD and total problems subscales. VR-CPT was administered at the end of the 40-minute assessment period.

RESULTS

Side effects
No significant side effects were observed in either group, based on post VR testing using a cybersickness questionnaire (Laboratoire de Cyberpsychologie, 2002).

VR Classroom performance
− Independent samples one-tailed t-tests (with 20 degrees of freedom) were done to compare performance of both groups on various variables of the VR Classroom. It was found that:
  − Participants with ADHD made significantly more omissions than participants in the comparison group (mean raw: 27 omissions versus 8; t=3.426, p=0.0015; after log transform to correct positively skewed distributions: t=2.968, p= 0.011, d= 1.36).
  − RT variability (standard error) was significantly higher for children with ADHD (182ms versus 135; t=1.758, p<0.05; after log transform to correct negatively skewed distributions: t=1.986, p=0.031, d=0.91).
  − The ADHD group had slower RT (568ms versus 544ms) and made more commissions errors than the comparison group but these differences were not significant.

Traditional CPT
Independent samples one-tailed t-tests were done to compare the performance of the ADHD and comparison groups on the Conner’s CPT on equivalent variables reported for the VR Classroom:
− Children with ADHD presented significantly more omission errors than children without ADHD (mean: 41 omissions versus 7; t=3.844, p=0.0005, d=1.76; no transformation required).
− The ADHD group had significantly longer RT (518ms versus 356ms; t=4.406,
The RT had significantly higher standard error of the mean in the ADHD group (26ms versus 8; \( t=5.767, p=0.000006 \); after log transform to correct positively skewed distributions: \( t= 6.164, p=0.0001, d=2.82 \).

Commission errors were exactly the same for both groups (26 commissions).

VR Classroom head movements

Independent samples one-tailed t-tests were done to compare head movement of both groups during the VR Classroom:

- The amplitude of head movement from side to side (Yaw: farthest left to farthest right) of participants from the ADHD group was significantly higher than in controls (Yaw: 154 degrees versus 42; \( t=4.462, p=0.0005, d=2.04 \)).

- Amplitude of head movement up and down (Pitch: farthest position looking up to farthest down) in the ADHD group was significantly higher than in the comparison group (Pitch total absolute amplitude: 74.96 degrees versus 28.5; \( t=3.752, p=0.001; \) after log transform to correct positively skewed distributions: \( t=3.547, p=0.001, d=1.62 \)).

VR Classroom and CPT

Univariate analyses of covariance (with degrees of freedom 1 and 19) were done, after verifying that the slopes were homogeneous across groups, to assess the unique contribution of equivalent variables from both tests to discriminate between the two groups. A significant group difference remaining once the corresponding variable from the other test is taken into account indicates that the test reveals information relevant to group difference that is not already provided by the other test.

- When possible shared variance between both variables was removed, the RV omission variable still significantly distinguished the two groups (\( F=10.253, p=0.005 \)), but the CPT omission variable no longer distinguished them (\( F=2.761, p=0.113 \)).

- When possible shared variance was removed, the earlier significant variable on t-test, CPT RT still distinguished the two groups (\( F=18.507, p<0.0005 \)).

- Finally, when possible shared variance was removed, the earlier significant variable on t-test, RV RT standard error, could no longer distinguish the groups (\( F=0.360, p=0.555 \)) but the CPT RT standard error variable still did (\( F=27.895, p<0.0005 \)).

VR Classroom and neuropsychological tests

Univariate analyses of covariance were also done to compare equivalent variables of VR Classroom and neuropsychological tests, after verifying that slopes were homogeneous.

- When possible shared variance between d2 and VR Classroom omissions was removed, the earlier significant variable, the RV omission, still significantly distinguished the two groups (\( F=15.628, p=0.001 \)) but the d2 omission variable no longer distinguished them (\( F=0.009, p=0.927 \)).

- Similarly with the Stroop total errors on the inhibition condition, when possible shared variance was removed, the RV omission variable still significantly distinguished the two groups (\( F=14.282, p=0.001 \)) but not the other variable (\( F=0.154, p=0.699 \)).

- VR omission variable and total errors on flexibility condition of the Stroop test had significantly different slopes in the two groups. Analysis of covariance was therefore not done.

- VR commission compared to d2 commission and total errors of the Stroop test did not significantly distinguish the two groups, once the shared variance was removed.

VR Classroom and parent ratings

Bivariate one-tailed Pearson correlations on various behavioural ratings filled by parent and significant RV variables (as determined by t-tests) were obtained for the ADHD group only (since boys of the comparison group obtained scores near 0 on all three ratings). Significant correlations (based on 13 degrees of freedom) were observed:

- Between VR Classroom omission errors and both SDQ scales (ADHD scale, \( r=0.69, p=0.002 \); Total problems scale, \( r=0.602, p=0.009 \)).

- Total absolute Pitch amplitude of head movement with the SDQ ADHD scale \( (r=0.602, p=0.009) \) and CBCL ADHD (DSM) scale \( (r=0.508, p=0.027) \).

- Total absolute Yaw amplitude of head movement and SDQ ADHD scale \( (r=0.460, p=0.042) \).
DISCUSSION

The first objective of the pilot study was to compare performance from children diagnosed with ADHD and from children in a control group on the VR Classroom test, including head movement measures. The study partly replicates results obtained with a previous form of the VR Classroom (Rizzo et al., 2006) on the omission and RT variability scores (children with ADHD made significantly more omission errors and their RT varied more over the test than children without ADHD). Results for head movements were not reported previously (children with ADHD have wider amplitude of head movement either up and down or from side to side). This study, however, did not replicate the group difference in mean RT and in commission errors. The previous version lasted for a total of 20 minutes compared to 6 minutes for the present version. The 20-minute version is evidently more strenuous in terms of sustained attention compared to the 6-minute version. The difference in duration might explain the generally larger effect sizes obtained with the first version of VR Classroom (Rizzo et al., 2006).

The second objective of the study was to determine if the cognitive profile of children with ADHD outlined by the traditional CPT and VR Classroom (ecological CPT) differed or not. It seems that the VR Classroom is efficient in distinguishing boys with ADHD from those without on a few traditional variables of continuous performance tasks (omission and variability of RT). Traditional CPT appears more efficient in distinguishing both groups if compared on similar variables. Since the standard CPT results, taken from the patient records, contributed to a positive diagnosis of ADHD, the CPT effect sizes are likely biased positively. For that reason, the contribution of one test to discriminate the groups beyond what the other test contributes is more relevant to appreciate the respective merits of the two tests. It turns out that the VR Classroom more often contributes new information than does the traditional CPT.

The final objective of the study was to compare the VR Classroom performance to neuropsychological tests and determine ecological validity. To establish ecological validity, Barkley (1991) recommends four sources of evidence.

1. Difference between ADHD and control groups: The present study included a group of non-ADHD boys. Larger studies need to include several clinical comparison groups.

2. Correlation with assessments that have previously established ecological validity: Few neuropsychological tests meet the criteria. To do so, performance on such measures ought to be correlated with observation in similar real life settings. An experimental measure of a spelling test in a simulated real life classroom was done with boys of the ADHD group but results are yet to be compiled and analysed. Results obtained on VR Classroom did correlate with traditional neuropsychological tests on equivalent omission variable and added new information regarding the variable, as outlined in the covariance analyses.

3. The assessment shows similar directional changes as that of the ecological criterion when exposed to experimental manipulations known to affect the criterion, such as medication: This condition was not included in the present study.

4. Correlations between the assessment and ecological criterion such as caregiver ratings: Results of the VR Classroom were compared with various parent ratings for which ecological validity have been established in previous studies. High correlations were found between some VR variables (omission errors and pitch and yaw total amplitude of head movement) and some SDQ and CBCL subscales (parent ratings).

This being a pilot study, there are limitations to take into account for future studies. First, the traditional CPT used does not assess the same skills as the VR CPT, as determined by covariance analyses. Conners’ CPT is a measure of the ability to inhibit a response and to adjust in a changing rhythm in answering. On this respect, the VR CPT is closer to the Gordon Diagnostic System (GDS; Gordon, 1983), which measures the ability to react to a specific stimulus when stimuli are presented at a fixed rate. Also, the VR Classroom used here is shorter in duration (6 minutes) than most traditional CPT (around 15 minutes). Performing longer tasks demands further mental effort from the participant. Duration might have an effect on the effect sizes (d values) obtained since the longer version of the VR Classroom (version 1) obtained higher effect sizes. Future versions of the test should take this observation into consideration.
The number of participants in both groups was limited. Only two types of participants were compared, those with ADHD and the comparison group without ADHD. It would have been relevant to measure the specificity of the VR Classroom relative to traditional CPT’s, knowing that they lack the ability to discriminate between various clinical groups (Berlin et al., 2004; McGee, Clark & Symons, 2000; Riccio & Reynolds, 2001). The particular combination of head movement tracking and distraction offered in the VR environment might contribute to enhance the specificity of continuous performance tasks. Proper ways to process this information, i.e. relating movement or performance levels to distraction events, must be devised.

Only boys were included in the pilot study. Girls and boys perform differently on neuropsychological assessments, and these differences need to be addressed in future research on the VR Classroom. Finally, future studies ought to verify the relevance of age and intelligence on VR performance. It has been found that intelligence (Halperin, Newcorn, Sharma, Healey et al., 1990) correlates with performance on CPT whereas age does not (McGee et al., 2000), since norms take age, but not IQ, into account.

CONCLUSION

The VR Classroom contributes information on ADHD status not provided by Conners’ CPT. By automatically monitoring head movements, it provides information that still needs characterisation but contributes positively to distinguish children with and without ADHD. It is still uncertain to what extent this reflects hyperactivity, distractibility, or loss of focus on the task. Even for variables also available in the standard CPT, the VR Classroom yields information not captured in Conners’ CPT. Further work is needed to assess to what extent this reflects the difference in tasks (press for K following A vs. for all letters but X) or the better ecological validity of VR Classroom.

Concerning ecological validity, most of the suggested sources of evidence were addressed in the pilot study, albeit to a limited extent, and positive preliminary results were obtained. The actual research was not able to determine or rule out ecological validity for the VR Classroom, but showed promising results, especially with the inclusion of head movement tracking.

The virtual classroom is in its second version and early in its development. Future changes should consider returning to a longer duration, and possibly embedding other attention tasks. The VR CPT sensitivity, specificity, and ecological validity in discriminating ADHD participants from non-ADHD have yet to be established in larger scale research.

REFERENCES


Diagnosis and Treatment of Attention Deficit Hyperactivity Disorder. NIH Consensus Statement, 1998, Nov 16-18; 16(2), 1-37.


