Neuropsychology Review (2019) 29:338–356
https://doi.org/10.1007/s11065-019-09407-6

REVIEW

A Comparison of Virtual Reality Classroom Continuous Performance Tests to Traditional Continuous Performance Tests in Delineating ADHD: a Meta-Analysis

Thomas D. Parsons [1,2,3] · Tyler Duffield [4] · Justin Asbee [1]

Received: 16 August 2018 / Accepted: 15 April 2019 / Published online: 3 June 2019

Abstract

Computerized continuous performance tests (CPTs) are commonly used to characterize attention in attention deficit-hyperactivity disorder (ADHD). Virtual classroom CPTs, designed to enhance ecological validity, are increasingly being utilized. Lacking is a quantitative meta-analysis of clinical comparisons of attention performance in children with ADHD using virtual classroom CPTs. The objective of the present systematic PRISMA review was to address this empirical void and compare three-dimensional (3D) virtual classroom CPTs to traditional two-dimensional (2D) CPTs. The peer-reviewed literature on comparisons of virtual classroom performance between children with ADHD and typically developing children was explored in six databases (e.g., Medline). Published studies using a virtual classroom to compare attentional performance between children with ADHD and typically developing children were included. Given the high heterogeneity with modality comparisons (i.e., computerized CPTs vs. virtual classroom CPTs for ADHD), both main comparisons included only population comparisons (i.e., control vs. ADHD) using each CPT modality. Meta-analytic findings were generally consistent with previous meta-analyses of computerized CPTs regarding the commonly used omission, commission, and hit reaction time variables. Results suggest that the virtual classroom CPTs reliably differentiate attention performance in persons with ADHD. Ecological validity implications are discussed pertaining to subtle meta-analytic outcome differences compared to computerized 2D CPTs. Further, due to an inability to conduct moderator analyses, it remains unclear if modality differences are due to other factors. Suggestions for future research using the virtual classroom CPTs are provided.

Keywords Attention-deficit/hyperactivity disorder · Executive function · Attention · Continuous performance test · Virtual reality

Introduction

If one extrapolates from National Center for Education Statistics (NCES) data, youth attending public school in the United States spend approximately 1200 h in the classroom annually, and approximately 14,000 h in the classroom by their high school graduation (U.S. Department of Education, 2007–2008). Thus, the classroom represents an environment where youth will spend a considerable amount of their formative years. Additionally, the classroom represents one of the most cognitively and socially demanding environments for youth. Importantly, learners are diverse and various neurodevelopmental and neurologic conditions, such as ADHD, can disrupt a variety of processes relevant to optimal academic functioning (e.g., Bruce, 2011). Further, Hale and colleagues (Hale et al., 2016) discuss the increasingly diverse student population within classrooms, as well as increasing class sizes, potentially stretching the resources and competencies of every teacher. Thus, being able to accurately predict classroom attentional capacity, which is foundational for academic performance and attainment, is important. For example, a meta-analytic review of behavioral ratings demonstrated that children with ADHD show deficient time on task in the classroom compared to peers (75% compared to 88% after

✉ Thomas D. Parsons
Thomas.Parsons@unt.edu

1 Computational Neuropsychology and Simulation (CNS) Laboratory, University of North Texas, 3940 N. Elm, G150, Denton, TX 76207, USA

2 NetDragon Digital Research Center, University of North Texas, 3940 N. Elm, G150, Denton, TX 76207, USA

3 Learning Technologies, College of Information, University of North Texas, 3940 N. Elm, G150, Denton, TX 76207, USA

4 Oregon Health and Science University, Portland, OR, USA

accounting for moderators), and more variable visual attending to required learning stimuli in the classroom (Kofler, Rapport, & Matt Alderson, 2008).

A widely used neuropsychological approach to assessing attentional deficits is the Continuous Performance Test (CPT; Fasmer et al., 2016). Briefly, the CPT is a task-oriented computerized assessment of attention that is often understood via signal-detection theory, either implicitly or explicitly. In alignment with basic signal-detection methods, the CPT requires participants to respond to a target when it is present and ignore it when it is not. This is often accomplished by performing simple motoric responses (e.g., button presses). Correct responses occur when participants respond (e.g., button press) when the target appears (i.e., hit) or inhibit a response when the target is not present (correct rejection). During the CPT, a commonly accepted metric of inattentiveness is failure to respond to a target (i.e., omission error) when the target is present. Likewise, a participant's failure to inhibit their response to a non-target (i.e., commission errors) has been thought to reflect impulsivity. Moreover, sustained attention is believed by many to be reflected in the participant's reaction time and reaction time variability.

Several previous meta-analytic reviews of CPT performance in ADHD have been conducted (i.e., Corkum & Siegel, 1993; Losier, McGrath, & Klein, 1996; Nigg, 2005; Sonuga-Barke, Sergeant, Nigg, & Willcutt, 2008; Willcutt, Doyle, Nigg, Faraone, & Pennington, 2005). Generally, commission and omission errors have demonstrated small to moderate effect sizes, and researchers had been unable to examine reaction time in the aggregate (Huang-Pollock, Karalunas, Tam, & Moore, 2012). Huang-Pollock et al. (2012) posit that the previously reported effect sizes are attenuated due to not using contemporary recommendations for conducting meta-analyses (i.e., using a random effects model and correcting for both sampling and measurement error, not just sampling error), which likely resulted in measurement error in previous reviews. In a more recent meta-analysis, Huang-Pollock et al. (2012) replicated previous findings correcting for just sampling error. However, these authors demonstrated large effect sizes for commission and omission errors between subjects with ADHD and typical controls when controlling for both sampling and measurement error. The reaction time effect size was moderate, but the credibility interval suggested that an effect size of 0 was present within the distribution. Further, after correcting for publication bias, the effect size for reaction time decreased to 0.29. Please refer to this previous meta-analytic literature for in depth conceptual discussion of the commonly reported omission error, commission error, and hit reaction time metrics.

A further difficulty with many traditional assessment methods, such as continuous performance tests (CPTs), is that results generally do not predict everyday functioning in real-world environments for clinical populations (Chaytor, Schmitter-Edgecombe et al., 2006; Spooner & Pachana 2006), and for ADHD specifically (Barkley & Murphy, 2011; Rapport, Chung, Shore, Denney, & Isaacs, 2000). Some have suggested that the psychometric inconsistencies of the CPT may be attributed to its limited capacity for simulating the difficulties persons with ADHD experience in everyday life (Pelham et al., 2011; Rapport et al., 2000). The majority of CPTs in common use are relatively free from the external distractions theorized to significantly impair the attentional performance of children with ADHD. As a result, several authors have called for enhanced ecological validity in assessments of attentional processes (Barkley, 1991; Berger, Slobodin, & Cassuto, 2017; Neguț, Matu, Sava, & David, 2016).

In this respect, virtual classrooms offer attentional assessments in a real-world dynamic simulation with distractors that mimic the conditions found in a youth's classroom. This active testing environment may have contemporaneous relevance for differentiating ADHD from typically developing individuals. Although current gold standard procedures for ADHD diagnosis are behavioral observation and ratings by a clinician, parent, teacher, and so on, evidence has emerged that the increase in academic demands at young ages has coincided with increased prevalence of ADHD predicated upon reporter expectations (Brosco & Bona, 2016). In a similar vein, concerning environmental demands relevant to the expression ADHD type behaviors, a recent meta-analysis revealed that hyperactivity was ubiquitous across ADHD subtypes and best predicted by situations with high executive function demands or low stimulation environments (Kofler, Raiker, Sarver, Wells, & Soto, 2016). If it is the case that normative-based cognitive assessment may have incremental value in the diagnosis of ADHD, we might question whether traditional CPTs (an often used testing adjunct for ADHD evaluation) provide environmental demands that are necessary and sufficient to elicit ADHD behaviors for diagnosis?

In a recent meta-analysis, Neguț et al. (2016) examined several virtual reality (VR) based neuropsychological assessments, which included some virtual classroom studies. Results revealed large effects for virtual reality-based assessments of cognitive impairments. Regarding virtual classroom studies available in the literature, most have utilized a continuous performance test (CPT; see Fig. 1 and Table 2). More specifically, empirical data from research assessing the efficacy of various virtual classroom CPTs for differentiating persons with ADHD from typically developing controls have emerged over the last 10 years. This is likely because VR systems have become less costly, more available, and generally more usable. A number of qualitative reviews of initial research findings have concluded that virtual classroom CPTs have potential as an assessment of attentional processing (Díaz-Orueta, 2017; Parsons & Rizzo, 2018; Rizzo et al., 2006). A potential problem in interpreting and reconciling findings about the nature and extent that attention can be assessed with virtual classroom CPTs is that the vast majority of virtual classroom studies of persons with neurodevelopmental disorders have reported on small sample sizes and made use of inadequate

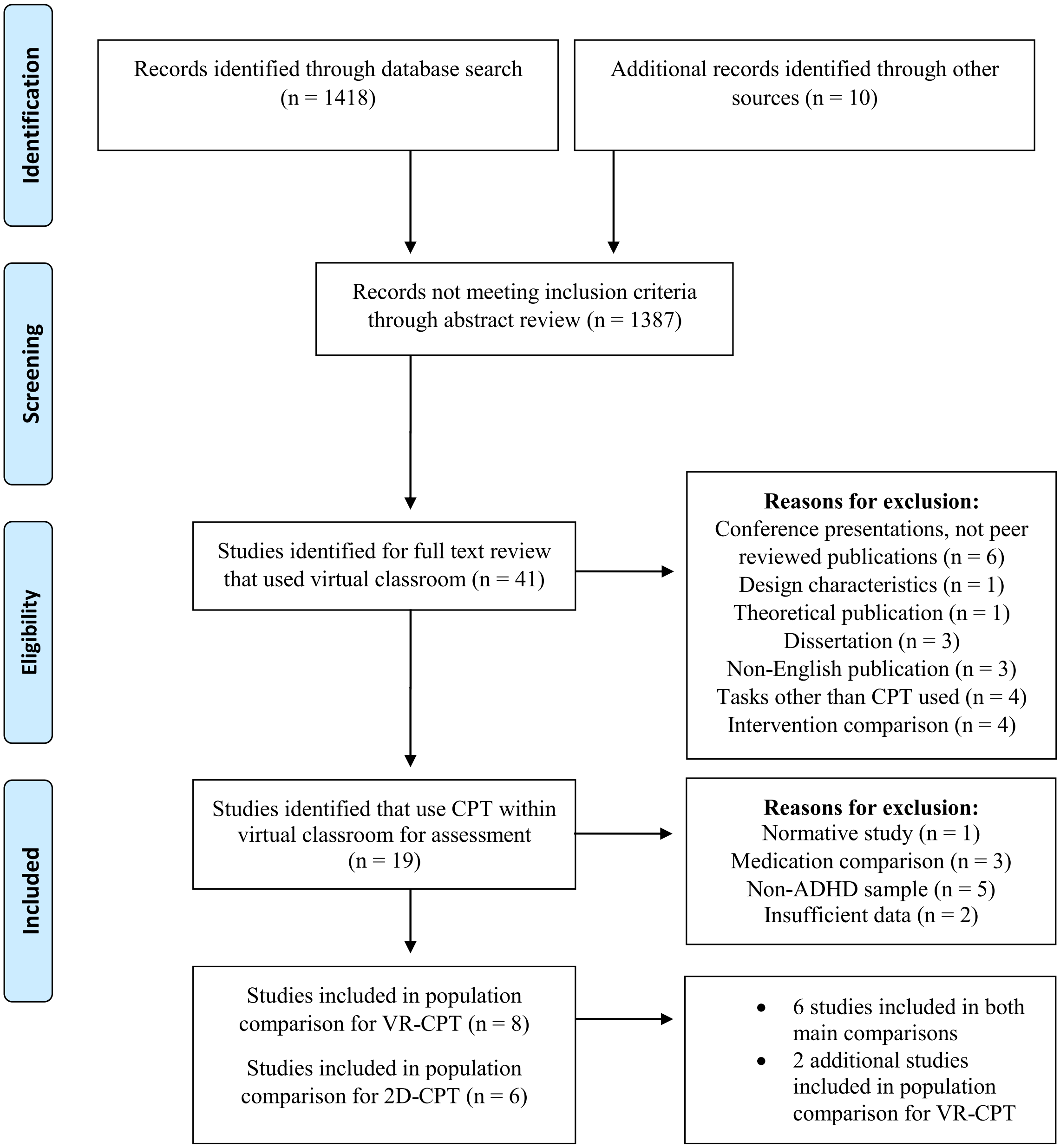

Fig. 1 PRISMA flow diagram

null hypothesis significance testing (Duffield, Parsons, Karam, Otero, & Hall, 2018).

Until large-scale studies on the efficacy of virtual classroom CPTs for assessment of attentional difficulties in neurodevelopmental disorders (e.g., ADHD) are published, statistical meta-analyses represent an interim remedy. Such analyses provide estimates of a population effect size across independent studies. They increase statistical power to detect true nonzero population effects by lowering the standard error, and consequently narrowing the confidence intervals associated with the population effect size estimate (Cohn & Becker, 2003). Hence, a quantitative meta-analysis, as opposed to a qualitative review, might facilitate a better understanding of the variability and clinical significance of attentional assessment in ADHD using virtual classroom CPTs. In view of this need, the present study sought to examine the efficacy of virtual classroom CPTs for differentiating between persons with ADHD and typically developing controls.

Methods

Given disparate research designs (see Fig. 1) and inconsistency in reported data, there was a paucity of data available for analyses. Therefore this review was limited to two research questions using the commonly reported omission error, commission error, and hit reaction time metrics of the CPT, 1) can virtual classroom CPTs discriminate between persons with ADHD and typically developing controls, and 2) do virtual classroom CPTs offer greater differentiation in performance than traditional computerized CPTs.

Study Selection

The overall objective of study selection was to collect published journal articles that compared 2D CPT versus 3D virtual classroom CPT performance of persons with ADHD and those that were typically developing. A literature search without date restrictions was conducted on December 1, 2018 using MedLine, PsycLIT, EMBASE, Cochrane Library, Google Scholar, and ISI Web of Science electronic databases. Standard searches were performed, which used keywords containing references to a virtual reality classroom, including "virtual classroom," "ClinicaVR," and "AULA." Reference lists of collected articles were visually inspected to locate any cited journal articles. See Fig. 1 for the flow diagram.

Study Eligibility Criteria

Eligibility criteria for study inclusion consisted of studies that utilized a virtual reality classroom. Exclusion criteria consisted of (1) no report of interval or ratio data, (2) no attention-symptom data reported between 2D CPTs and 3D Virtual Classroom CPT or between controls and an ADHD population using the 3D Virtual Classroom CPT (thus excluding non-ADHD populations), (3) intervention studies, (4) conference presentations, (5) dissertations, (6) non-English language studies, (7) insufficient report of study results (e.g. no means and standard deviations) to allow for effect size computation. Two authors independently evaluated abstracts of each article to determine whether they met criteria for inclusion, followed by full text review to assess if criteria were met for exclusion. An interrater reliability analysis using the Kappa statistic was performed to determine consistency between both authors who reviewed abstracts.

Concerning insufficient report of study results, corresponding authors were contacted and if no response was received, studies meeting this criterion were excluded. In more simplistic terms, the current authors sought studies that examined quantitative comparisons of the virtual classroom CPT utilizing an ADHD population. It is important to note that some studies were both between subject designs (ADHD and typically developing), as well as comparisons examining ADHD population performances (2D CPT versus 3D Virtual Classroom CPT). Table 1 provides a summary of studies included in the meta-analysis.

Data Coding

Two authors independently extracted the following information from the published articles and coded (1) number of subjects, (2) exclusion criteria, (3) diagnostic groups, (4) demographics, (5) assessment measures, and (6) summary statistics required for computation of effect sizes. Inconsistencies between raters were resolved by means of discourse. Discourse primarily related to creation of two tables to report study information as opposed to a single table (i.e, Tables 1 and 2), and a means to report statistical data that used a CPT for assessment purposes, but used a clinical population other than ADHD (i.e., Tables 3 and 4) and thus was not included in the two main comparisons.

Data Analytic Considerations

We used the random-effects meta-analytic model (Shaddish & Haddock, 1994). Analysis of continuous outcomes involved comparing standardized differences between assessment modalities (Hedges & Olkin, 1985). Standardization allowed the study results to be transformed to a common scale (standard deviation units), which assisted pooling (Hedges, 1984; Hedges & Olkin, 1985). Adjustments were made to correct for upward bias of effect size estimation in small sample sizes. An unbiased estimation (Cohen's *d*) was calculated for each study in which the effect size is weighted by a sample-size based constant (Hedges, 1984; Hedges & Olkin, 1985). Given the small sample sizes, effect sizes were also calculated (and reported) as Hedges' *g* (Hedges, 1981), a more conservative measure of effect size than the frequently used Cohen's *d*. Instead of using a maximum likelihood estimation to calculate variance (like Cohen's *d*, which generates a biased estimation for n), Hedges' *g* uses the Bessel's correction to reduce overestimation of effect sizes for small studies by calculating the pooled standard deviation using degrees of freedom.

Standardized mean differences were calculated and analyzed for each study. In particular, we started with $d = (M_1–M_2)/SD^*_{pooled}$, where M_1 and M_2 are the mean scores between groups, respectively, and SD^*_{pooled} is the standard deviation for the pooled sample (Shaddish & Haddock, 1994). Given the small sample sizes and the fact that *d* tends to overestimate the absolute value of *d* in small samples, Hedges' *g* was calculated (Hedges, 1981). This statistic results in a weighted average composite unbiased effect-size estimate for each measure. Following general convention (Cohen, 1988) for both Cohen's *d* and Hedges *g*, an effect size of 0.20 was considered a small effect, 0.50 a moderate effect, and 0.80 a large effect.

Prior to combining studies in the meta-analysis, we assessed the homogeneity of the effect size (Hedges & Olkin, 1985; Higgins, Thompson, Deeks, & Altman, 2003). Heterogeneity between studies was assessed by the Higgins' I^2 test ($P > 0.1$ and I2 < 50% indicate acceptable heterogeneity) and a standard chi-square test. The Higgins' I^2 statistic was calculated by dividing the difference between the *Q*-statistic (sum of squared deviations of each study estimate from the overall meta-analytic estimate) and degrees of freedom by the *Q*-statistic itself. This resulted in an estimated percentage

Table 1 Summary of participant information, design, and standardized assessments in virtual classroom studies

	Clinical							Control				
	N	Age	Male %	Dx	Meds	How Dx	Design (between vs. Within subject)	N	Age	Male %	Assessed Cybersickness	Standardized Assessments
Moreau et al. (2006)	15	9–13	100	ADHD	Prior to taking their daily medication	Patients recruited from health agencies	Between	7	9–13	100	Administered Cybersickness Questionnaire; No subjects reported sickness	SDQ (ADHD and total problems subscales), ADHD-RS-IV (total problem subscale and Achenbach System of Empirically Based Assessment), CBCL (ADHD and total problems subscales)
Parsons et al. (2007)	9	10.6	100	ADHD	No	Clinician & Swan & Npsyc tests	Between	10	10.2	100	Administered SSQ; No subjects reported sickness	BNT; Stroop; NEPSY (Visual Attention, Design Fluency, Verbal Fluency); WISC-III (Digit Span, Coding, Arithmetic, Vocabulary); Trail Making Test; JLO; SWAN Behavior Checklist
Nolin et al. (2009)	8	8–12	Not reported	TBI	Not reported	Not reported	Between				Not reported	None
Gutiérrez Maldonado et al. (2009)	10	Not reported	70	Not reported	Not reported	Diagnosis at hospital	Between	10	Not reported	60	Not reported	None
Pollak et al. (2009)	20	12.6	100	ADHD	No	Clinician DSM-IV & interview	Between	17	12.6	100	Not reported	DRS, SFQ
Adams et al. (2009)	19	10.1	100	ADHD	10 of 19	Licensed mental health provider or pediatric physician	Between	16	10.5	100	Administered SSQ; No subjects reported sickness	BASC
Pollak et al. (2010)	27	13.7	59	ADHD	Part of study	Child neurologist	Between No controls				Not reported	ADHD-RS, SFQ
Gilboa et al. (2011)	29	12.2	31	Neurofibromatosis type 1	No	NIH criteria	Between	25	12.2	28	Not reported	CPRS-R:L
Bioulac et al. (2012)	20	8.4	100	ADHD	No	Clinician & CPRS	Between	16	8.21	100	Administered Cybersickness Questionnaire; No subjects reported sickness	STAI
Nolin et al. (2012)	25	13.6	60	Concussion	No	Grade 1 concussion &SCAT2	Between	25	13.8	60	Most subjects reported cybersickness	Presence Questionnaire, developed by Witmer and Singer (1998), Post-Exposure Symptom Checklist
Díaz-Orueta et al. (2014)	57	11	73.7	ADHD	29 of 57	Neuropediatrician	Between No controls				Not reported	WISC-IV
Gilboa et al. (2015)	41	12.8	58.5	ABI	No	ABI requiring medial and neuropsychological follow-up	Between	35	11.8	93	Not reported	TEA-Ch, Sky Search, Sky Search dual task (DT), Score!, CPRS-R:S, WASI

Table 1 (continued)

	Clinical							Control				
	N	Age	Male %	Dx	Meds	How Dx	Design (between vs. Within subject)	N	Age	Male %	Assessed Cybersickness	Standardized Assessments
Mühlberger et al. (2016)	94	11.6	76.5	ADHD	30 of 107	Clinician	Between	54	12.2	52.9	Not reported	One of the following IQ tests: K-ABC, CFT-1, CFT20-R, or Hamburg-Wechsler-Intelligence-Test für Kinder [for children]
Areces et al. (2016)	86	10.7	Not reported	ADHD	No	Identified accordingto DSM-5	Within	27	12.7	Not reported	Not reported	WISC-IV, Scale for the Assessment of ADHD
Iriarte et al. (2016)							Within Normative sample $N = 1272$	1272	10.25	51.8	Not reported	None
Nolin et al. (2016)								102	7 to 16	52.0	TheSimulator Sickness Questionnaire	Therealistic subscale of the Presence Questionnaire
Neguţ et al. (2016)	33	10.2	44	ADHD	Yes	Parent report and medical records	Between	42	8.9	44	Administered SSQ; Two subjects reported cybersickness	Raven IQ, d2, WISC-IV, CAS
Areces et al. (2018)	237	10.67	71.3	ADHD	No	Diagnosed at clinical center and verified by researchers	Between	101	11.14	71.3	Not reported	None
Areces et al. (2018)	50	10.2	75	ADHD	Not reported	Neuro-psychiatrists	Between	38	10.2	75	Not reported	WISC-IV, EDAH

Studies also varied on number of conditions, length of conditions, stimulus parameters (e.g., stimulus exposure time), distraction parameters, and language of CPT (e.g., Bioulac VR-CPT adapted to French)

SDQ Strength Difficulties Questionnaire, *CBLC* Achenbach System of Empirically Based Assessment, *AULA* Nesplora CPT, *DMW* Digital Media Works, *BNT* Boston Naming Test, *Dx* diagnosed, *Body Mvmt* Body movement assessed, *JLO* Judgment of Line Orientation, *SSQ* Simulator Sickness Questionnaire, *SWAN* SWAN Behavior Checklist, *WISC-III* Wechsler Intelligence Scale for Children-Third Edition, *CAS* Cognitive Absorption Scale, *DRS* Diagnostic Rating Scale, *SFQ* Subjective feedback questionnaire, *STAI* State Trait Inventory Anxiety, *BASC* Behavior Assessment System for Children, *K-ABC* Kaufmann Assessment Battery for Children, *CFT-1&CFT20-R* Culture Fair Intelligence Test, *ADHD-RS* ADHD rating scale, *CPRS-R:L* Conners' Parent Rating Scales—Revised: Long, *TEA-Ch* Test of Everyday Attention for Children, *CPRS-R:S* Conners' Parent Rating Scales-Revised: Short, *WASI* Wechsler Abbreviated Scale of Intelligence, *EDAH* The Scale for the assessment of Attention Deficit Hyperactivity Disorder

Table 2 VR and 2DCPT Type, stimulus parameters, and VR-CPT hardware configuration

	VR-CPT Type	VR-CPT Stimulus Parameters	Distractor Parameters	VR-CPT Hardware Configuration	2D-CPT Type	2D-CPT Stimulus Parameters
Moreau et al. (2006)	DMW: AK Version	• 300 stimuli • 150 ms stimulus duration • 30% of stimuli were incorrect hit stimuli • Other non-target letters occurred with equal probability	• Distractors consisted of pure auditory, pure visual, and a mix of auditory and visual	• Respond using left mouse button	CPT-II	• Total time = 15 min • Inhibit response to letter X
Parsons, Bowerly, Buckwalter, and Rizzo (2007)	DMW: AX version	• Three 10 min conditions (AX with and w/o distractors, and BNT match) • 400 stimuli per condition • 150 ms stimulus duration • ISI = 1350 ms	• 200 s presentation block • 5 s stimulus duration • Presented in randomly assigned equally appearing intervals of 10, 15, or 25 s • 36 distraction intervals (12 of each) and 36 distracters (9 of each) included in condition	• VR V8 HMD • Ascension tracking device fitted to non-dominant hand and opposite knee • Remote mouse used for responding	Conners'	• Presumably manufacturer settings, not reported
Nolin et al. (2009)	DMW: AX version	• Stimulus parameters not reported	• Stimulus parameters not reported	• Hardware configuration not reported	Vigil	• Presumably manufacturer settings, not reported
Pollak et al. (2009)	DMW: 3–7 Version	• Total time = 10 min • 400 total stimuli • 100 target stimuli • 150 ms stimulus duration • ISI = 1350 ms	• 20 total distractors • 5 s stimulus duration • Distracters presented in randomly assigned intervals of 10, 15 or 25 s	• HMD type not reported • No use of movement tracker • Mouse used for responding	TOVA & VR-CPT on computer screen	• Total time = 21.6 min • ISI = 2 s • Stimuli presented in consistent 3.5:1 ratio for a target infrequent and target frequent conditions
Adams, Finn, Moes, Flannery, and Rizzo (2009)	DMW: AX version	• Total time = 6 min • 400 total stimuli • 150 ms stimulus duration • "The letter X and the letter A followed by the letter X each appeared with a 10% probability, A and H appeared with a 20% probability, and all other letters appeared with a 5% probability."	• "Distracters were presented throughout the duration of the task."	• 3-D virtual reality dome by Elumens • Movement tracker placed on bike helmet worn by subjects (derived via manuscript figure) • Method or responding reportedly entailed "pressing a button"	Vigil	• The cued condition was used, which requires subjects to strike a computer key when the letter K appears immediately after the letter A. Targets appear for a duration of 85 msec; there are 25 targets in the cued condition."
Gutiérrez Maldonado, Letosa Porta, RusCalafell, and Peñaloza Salazar (2009)	Research lab built semantic CPT	• Total time = 10 min • 4 conditions: auditory CPT w/o distracters; auditory CPT w/ distracters; visual CPT w/o distracters; and visual CPT w/ distracters. • 6 blocks containing 100 stimuli each, with 20 target stimuli • 300 ms stimulus duration • ISI = 1 s	• "Distracters were identical in shape, duration (4 s) and presentation for both environments." • "Distracters could be auditory, visual and combined (auditory and visual) stimuli and were randomly distributed along the test administration in their respective conditions."	• No HMD was used • Mouse used for head turning in the environment • VR-CPT displayed on desktop screen	No 2D CPT was used	N/A
Pollak, Shomaly, Weiss, Rizzo, and Gross-Tsur (2010)	DMW: 3–7 Version	• Total time = 10 min • 400 total stimuli • 100 target stimuli • 150 ms stimulus duration • ISI = 1350 ms	• 20 total distractors • 5 s stimulus duration • Distracters presented in randomly assigned intervals of 10, 15 or 25 s	• eMagin Z800 3DVisor • No use of movement tracker • Mouse used for responding	TOVA	• Total time = 21.6 min • ISI = 2 s • Stimuli presented in consistent 3.5:1 ratio for a target infrequent and target frequent conditions
Gilboa, Rosenblum, Fattal-Valevski, Toledano-Alhadef, and Josman (2011)	DMW: 3–7 Version	• Total time = 10 min • 5 identical blocks of 2 min • 400 total stimuli • 100 target stimuli • 150 ms stimulus duration • ISI = 1350 ms	• 20 total distractors • 5 s stimulus duration "and presented identically for the entire sample"	• HMD type not reported • Movement tracker embedded in HMD • Mouse used for responding	No 2D CPT was used	N/A
Bioulac et al. (2012)	DMW: AK version	• 5 blocks (each 100 s in duration) with 20 targets (AK). • 500 total stimuli presented during whole task (500 s).	• Stimulus parameters not reported	• HMD type not reported • No use of movement tracker • Mouse used for responding	Conners'	• Total time = 14 min • 6 blocks (each 140 s in duration) • 54 targets per block (except block 1: 53 targets) and 6 non-targets • 250 ms stimulus duration

Table 2 (continued)

	VR-CPT Type	VR-CPT Stimulus Parameters	Distractor Parameters	VR-CPT Hardware Configuration	2D-CPT Type	2D-CPT Stimulus Parameters
Nolin, Stipanicic, Henry, Joyal, and Allain (2012)	DMW: AX version	• Stimulus parameters not reported, but did note, "As in the traditional VIGIL-CPT, it was possible to obtain performance scores for the 6-min test in three blocks of 2 min, thereby tracking performance over the course of the test."	• Stimulus parameters not reported	• eMagin Z800 HMD • Movement tracker embedded in HMD • Method of responding not reported	Vigil	• ISIs of 1, 2, and 4 s • Total time = 6 min • 300 total stimuli • 60 target stimuli
Díaz-Orueta et al. (2014)	AULA Nesplora	• No-X condition followed by X condition for total of 20 min • 180 stimuli per condition • 250 ms visual stimulus presentation • 470 ms – 891 ms auditory stimulus presentation • ISI = 1250 ms	• No-X task: 9 total distractors, with sequence that alternates 2 visual, 3 auditory, and 4 combined distractors; • X-task: 7 total distractors, with 2 visual, 3auditory and 1 combined distractor	• HMD type not reported • Movement tracker embedded in HMD • Connected one button mouse used for responding	Conners'	• 360 total stimuli • 250 ms stimulus duration
Gilboa et al. (2015)	DMW: 3–7 Version	• Total time = 10 min • 5 identical blocks of 2 min • 400 total stimuli • 100 target stimuli • 150 ms stimulus duration • ISI = 1350 ms	• 20 total distractors • 5 s stimulus duration "and presented identically for the entire sample"	• HMD type not reported • Movement tracker embedded in HMD • Mouse used for responding	No 2D CPT was used	N/A
Neguţ et al. (2016)	DMW: AK version	• Distractor and non-distractor conditions • 374 total stimuli • 55 target stimuli • 200 ms stimulus duration • ISI = 1000 ms	• Only information provided was that used only auditory distractors	• HMD type not reported • No use of movement tracker • Mouse used for responding	Built AX-type 2D-CPT to replicate VR-CPT	• 2D- CPT designed using Inquisit v3 • Identical stimulus parameters to VR-CPT • "...in the condition with distractors we used the audio recording from the VC and the children heard the noises from the classroom through headphones."
Nolin et al. (2016)	DMW: AX version	• Stimulus parameters not reported	• Stimulus parameters not reported	• eMagin Z800 HMD • Movement tracker embedded in HMD • Method of responding not reported	Vigil	• Total time = 6 min • 300 total stimuli • 60 target stimuli
Mühlberger et al. (2016)	DMW: AK version	• Total time = 10 min 42 s • 4 blocks (each with 80 stimuli), each lasting 2 mins 40s • 10 "pseudorandomized" target sequences • 100 ms stimulus duration • ISI = 1900 ms	• Block 1: 26 distracters (12 auditory, 3 visual, 11 auditory/visual) = medium • Block 2: 51 distracters (21 auditory, 8 visual, 22 auditory/visual) = high • Block 3: 5 distracters (3 auditory, 1 visual, 1 auditory/visual) = low • Block 4: 28 distracters (13 auditory, 4 visual, 11 auditory/visual) = high	• eMagin Z800 HMD • 3DOF magnetic tracking device head position tracker • Method of responding reportedly entailed "clicking [a] device"	No 2D CPT was used	N/A
Areces, Rodríguez, García, Cueli, and González-Castro (2016)	AULA Nesplora	• Stimulus parameters not reported	• Stimulus parameters not reported	• HMD type not reported • Movement tracker embedded in HMD • Mouse used for responding	No 2D CPT was used	N/A
Iriarte et al. (2016)	AULA Nesplora	• Task sequence, but parameters not reported	• Stimulus parameters not reported	• HMD type not reported • Movement tracker embedded in HMD	No 2D CPT was used	N/A

Table 2 (continued)

	VR-CPT Type	VR-CPT Stimulus Parameters	Distractor Parameters	VR-CPT Hardware Configuration	2D-CPT Type	2D-CPT Stimulus Parameters
Areces et al., (2018)	AULA Nesplora	• Stimulus parameters not reported	• 16 total distractors (4 visual, 6 auditory, 5 combined)	• Connected one button mouse used for responding • HMD type not reported • Motion sensors and headphones were used • Answers recorded with button press	TOVA	• Total time = 22.5 min • First half: 22.5% targets, 77.5% non-targets • Second half: 77.5% targets, 22.5% non-targets • 100 ms stimuli presentation • ISI = 2000 ms
Areces et al. (2018)	AULA Nesplora	• Total time = 20 min • 360 items • 180 targets • 250 ms visual stimuli presentation • 650 ms auditory stimuli presentation • 2500 ms max response time	• 16 total distractors (4 visual, 6 auditory, 5 combined)	• HMD type not reported • Motion sensors and headphones were used • Answers recorded with button press	No 2D CPT was used	N/A

Studies also varied on the language of CPT (e.g., Bioulac et al., 2012 VR-CPT adapted to French). Some studies specifically reported use of remote or connected mouse, while many studies did not specify type of mouse used for responding. No studies reported type of headphones used. 2D-CPT stimulus parameters of Pollak et al. (2009) pertain to the TOVA, not the 2D presentation of the virtual classroom

DMW Digital Media Works, *DOF* degrees of freedom, *HMD* head mounted display, *ISI* interstimulus interval, *min* minutes, *ms* milliseconds, *s* seconds, *VR-CPT* virtual reality continuous performance test

of study variance explained by heterogeneity (Huedo-Medina, Sanchez-Meca, Marin-Martinez, & Botella, 2006). Values for I^2 range from 0 to 1. An I^2 of 0% indicates no heterogeneity. I^2s of 25% represent low heterogeneity, 50% represent moderate heterogeneity, and 75% represent high heterogeneity (Higgins et al., 2003). I^2 represents a ratio of variance in the true effect in compared to variance due to sampling error (Borenstein, Higgins, Hedges, & Rothstein, 2017). Therefore, τ^2 was also reported which is an indication of absolute variance (Borenstein et al., 2017). Meta-analyses were performed using the meta-analysis software package Review manager 5.3.5 (RevMan, 2014).

Forrest plots offer a synopsis of each study effect and the confidence around the effect sizes. Funnel plots are employed as visual indicators for publication bias, where effect sizes are plotted along standard errors. Studies high on the y axis (low standard error) are more reliable than studies low on the axis (high standard error). Potential publication bias is indicated by a placement of studies from one side of the "funnel" to the other.

Types of Continuous Performance Tests Used

Data related to the typical CPTs and virtual classroom CPTs used in the studies is shown in Table 2. Three of the studies used a Conner's CPT, four used a Vigil CPT, three used a TOVA (Test of Variables of Attention) CPT, one study built an AX-type CPT, one study presented the virtual classroom CPT on a 2D computer screen (in addition to administering the TOVA), and eight studies did not use a traditional CPT. For the virtual classroom CPT, 12 studies used variations of the Digital Media Works virtual classroom CPT (five used the AX version, four used the AK version, four used the 3–7 version), and five used the Aula Nesplora version of the virtual classroom CPT. One study built a semantic based CPT for their study. Regarding display of the virtual classroom, one study used a dome, one study presented the virtual classroom on a 2D computer screen, and only six of the 19 studies that used a head mounted display provided information regarding that hardware. For most of the studies, little to no information was provided regarding headphones (for auditory stimuli) or mouse hardware (used for responding).

Moderator Variables

An attempt was made to evaluate the potential influence on ADHD effect sizes of several potential moderators. Moderators were selected on the basis of prior research identifying these variables as candidate moderators of attentional performance. Personal characteristics such as personality, hypnotizability, and absorption may act as variables that are able to account for the effectiveness of virtual environments (Witmer & Singer, 1998). Additionally, virtual reality system

Table 3 Effect sizes for 2D CPTs between groups

Reference	Clinical Group	2D CPT Omissions g; η2; (OR); [AUC]	Commissions g; η2; (OR); [AUC]	Reaction time g; η2; (OR); [AUC]
Parsons et al. (2007)	ADHD	0.94; 0.19; (5.73); [0.75]	1.06; 0.24; (7.64); [0.79]	0.75; 0.13; (4.00); [0.71]
Adams et al. (2009)	ADHD	0.55; 0.076; (2.83); [0.66]	0.54; 0.074; (2.79); [0.66]	Not Reported
Pollak et al. (2009)	ADHD	1.23; 0.30; (10.78); [0.82]	1.40; 0.35; (14.33); [0.85]	0.20; 0.011; (1.47); [0.56]
Pollak et al. (2010)	ADHD	−0.4; 0.041; (0.47); [0.39]	0.09; 0.002; (1.19); [0.53]	−0.56; 0.074; (0.36); [0.34]
Bioulac et al. (2012)	ADHD	1.29; 0.31; (11.55); [0.83]	0.64; 0.094; (3.21); [0.68]	0.61; 0.092; (3.17); [0.67]
Nolin et al. (2012)	Brain injury	0.28; 0.02; (1.69); [0.58]	−0.45; 0.05; (0.44); [0.37]	0.08; 0.002; (1.16); [0.52]
Neguţ et al. (2016)	ADHD	1.11; 0.24; (7.62); [0.79]	1.31; 0.28; (9.48); [0.81]	0; 0; (1); [0.5]
Areces et al., (2018)	ADHD	0.04; 0.0004; (1.07); [0.51]	0.11; 0.003; (1.21); [0.53]	0.23; 0.01; (1.51); [0.56]

g Hedges's g, *OR* Odds Ratio, *AUC* Area Under the curve. Reaction times not reported for Adams et al. (2009). Gutiérrez Maldonado et al. (2009) did not report commission errors for VR CPT. Control subjects for Pollak et al. (2010) were ADHD patients who received placebo medication, experimental subjects had ADHD and received methylphenidate. Nolin et al. (2009), (2016) were normative studies and did not include experimental/control groups therefore effect sizes could not be calculated. Gilboa, Rosenblum, et al. (2011), Gilboa, Kerrouche, et al. (2015), Areces et al. (2016), Gutiérrez Maldonado et al. (2009), Iriarte et al. (2016), and Mühlberger et al. (2016) did not include 2D CPT. Díaz-Orueta et al. (2014) examined convergent validity for ADHD children and did not include a control group

characteristics may moderate the level of presence felt (Bohil, Alicea, & Biocca, 2011). Furthermore, we aimed to assess for prominent sample characteristics (stratification by subtypes, co-occurring disorders, socioeconomic status, and average full scale IQ). Given that manipulation of CPT task parameters in traditional 2D versions (i.e., Conner's, TOVA, Vigil) can affect behavioral response characteristics, moderator analyses were planned for various procedural variations including increased or decreased target frequency, interstimulus intervals, and overall task length.

Unfortunately, there was inconsistent reporting of study data and the number of studies was very small. It was not possible to calculate correlation coefficients because numerous studies did not report exact values, and for some parameters the number of studies was too small to meaningfully interpret the effect size. The limited number of studies, and

Table 4 Effect sizes for VR CPTs between groups

Reference	Clinical Group	VR CPT Omissions g; η2; (OR); [AUC]	Commissions g; η2; (OR); [AUC]	Reaction time g; η2; (OR); [AUC]
Parsons et al. (2007)	ADHD	1.90; 0.49; (34.99); [0.92]	1.56; 0.39; (17.78); [0.87]	−0.14; 0.005; (0.78); [0.46]
Adams et al. (2009)	ADHD	0.94; 0.19; (5.92); [0.76]	0.5; 0.064; (2.57); [0.64]	Not Reported
Gutiérrez Maldonado et al. (2009)	ADHD	5.69; 0.90; (47,960.14); [1.00]	Not Reported	0; 0; (1); [0.5]
Pollak et al. (2009)	ADHD	1.49; 0.38; (17.36); [0.87]	0.92; 0.19; (5.80); [0.75]	1.09; 0.25; (7.91); [0.79]
Pollak et al. (2010)	ADHD	−0.62; 0.09; (0.32); [0.33]	−0.45; 0.05; (0.44); [0.37]	−0.51; 0.063; (0.39); [0.36]
Gilboa et al. (2011)	NF1	0.74; 0.13; (4.02); [0.71]	0.82; 0.16; (4.77); [0.73]	0.14; 0.005; (1.29); [0.54]
Bioulac et al. (2012)	ADHD	1.70; 0.44; (24.51); [0.89]	0.61; 0.088; (3.08); [0.67]	−0.21; 0.01; (67); [0.44]
Nolin et al. (2012)	Brain injury	0.8; 0.14; (4.41); [0.72]	0.41; 0.042; (2.14); [0.62]	−0.01; 0; (0.98); [0.50]
Gilboa et al. (2015)	Brain injury	0.8; 0.14; (4.41); [0.72]	0.33; 0.028; (1.84); [0.59]	0.08; 0.001; (1.15); [0.52]
Areces et al. (2016)	ADHD	0.91; 0.23; (7.41); [0.78]	0.92; 0.21; (6.62); [0.77]	0.54; 0.083; (2.97); [0.66]
Neguţ et al. (2016)	ADHD	1.58; 0.38; (16.98); [0.87]	1.08; 0.21; (6.50); [0.77]	0.45; 0.049; (2.28); [0.63]
Areces et al. (2018)	ADHD	1.01; 0.22; (6.74); [0.77]	0.22; 0.01; (1.50); [0.56]	0.82; 0.15; (4.59); [0.72]
Areces et al. (2018)	ADHD	0.92; 0.20; (5.99); [0.76]	0.48; 0.06; (2.50); [0.64]	0.31; 0.03; (1.82); [0.59]

g Hedges's g, *OR* Odds Ratio, *AUC* Area Under the curve, *NF1* Neurofibromatosis type 1. Adams et al. (2009), did not report reaction times. Gutiérrez Maldonado et al. (2009) did not report commission errors for VR CPT. Mühlberger et al. (2016) did not report means and standard deviations, therefore effect sizes could not be calculated. Control subjects for Pollak et al. (2010) were ADHD patients who received placebo medication, experimental subjects had ADHD and received methylphenidate. Nolin et al. (2009), (2016) were normative studies and did not include experimental/control groups therefore effect sizes could not be calculated. Díaz-Orueta et al. (2014) examined convergent validity for ADHD children and did not include a control group subsequent small sample size was a major limiting factor, which makes the power to detect the presence of moderators very low and the probability of capitalizing on sampling error, as well as identifying falsely moderators when they are not present, is quite high (Hunter & Schmidt, 2004).

Results

Literature Search

The consort diagram in Fig. 1 displays the various steps in the selection process. A total of 41 studies met the inclusion criteria that used a virtual reality classroom. Of those 41 studies, 19 used a CPT for assessment purposes (see Table 2), and eight studies included a population comparison of interest (ADHD vs. typical control) using a VR-CPT (see Table 5). The interrater reliability for the two authors was found to be kappa = 0.87 ($p < .001$, 95% CI: 0.793, 0.949). This constitutes a substantial level of agreement (Landis & Koch, 1977). The first author found six conference presentations the second author did not. The second author found a pilot study, a French publication, and two dissertations the first author did not. None of the discrepant articles found between the researchers were included in the main comparisons.

Initial results (i.e., Hedges *g*; odds ratio; area under the curve) for between group comparisons considering all clinical populations that used traditional CPT are found in Table 3. Initial results (i.e., Hedges *g*; odds ratio; area under the curve) for between group comparisons considering all clinical populations that used a virtual classroom CPT are found in Table 4. Of the 19 identified studies that utilized a virtual classroom CPT for assessment, eight studies were retained for the two main comparisons that used an ADHD population and consisted of an appropriate research design relevant to the research questions (see Table 5), six studies of which were included in both main comparisons (i.e., both the "Control vs. ADHD in VR CPTs" comparison and "Control vs. ADHD in traditional CPTs" comparison). Two additional studies were included exclusively in the "Control vs. ADHD in VR CPTs" comparison.

In terms of cybersickness (or simulator sickness), of those that reported this variable most studies did not note sickness associated with use of the virtual classroom (e.g., Adams

Table 5 Main comparisons

	Control vs. ADHD in VR CPTs (only distractor conditions)	Control vs. ADHD in traditional CPTs
Includes the following publications:	• Parsons et al. (2007)	• Parsons et al. (2007)
	• Adams et al. (2009)	• Adams et al. (2009)
	• Pollak et al. (2009)	• Pollak et al. (2009)
	• Bioulac et al. (2012)	• Bioulac et al. (2012)
	• Neguţ et al. (2016)	• Neguţ et al. (2016)
	• Areces et al. (2016)	• Areces et al. (2018)
	• Areces et al. (2018)	
	• Areces et al. (2018)	
Not included in analyses due to clinical sample other than ADHD:	• Nolin et al. (2009)	
	• Gilboa et al. (2011)	
	• Nolin et al. (2012)	
	• Gilboa et al. (2015)	
	• Nolin et al. (2016)	
Not included in analyses due to research design or insufficient data reporting:	• Moreau et al. (2006)	
	• Gutiérrez Maldonado et al. (2009)	
	• Pollak et al. (2010)	
	• Díaz-Orueta et al. (2014)	
	• Iriarte et al. (2016)	
	• Mühlberger et al. (2016)	

Adams et al. (2009) did not report hit reaction times and thus this manuscript was not included in forest plot calculations corresponding to Fig. 5 and Fig. 8. Mühlberger et al. (2016) did not report means and standard deviations for measures. Moreau et al. (2006) did not report standard deviations for measures. Areces et al. (2018) randomly assigned subjects two different conditions. In one condition children with and without ADHD ($N = 172$) were assessed with TOVA, while in the other condition, children were assessed with Aula Nesplora ($N = 166$). Thus, for the two main comparisons of the current study can be considered two different samples. Bioulac et al. (2018) did not include control subjects and consisted of a pre-post intervention comparison (virtual cognitive remediation group, methylphenidate group, psychotherapy group) et al., 2009; Bioulac et al., 2012; Mühlberger et al., 2016; Parsons et al., 2007). A single study reported only a small proportion of subjects experienced cybersickness (2 of 75 reported sickness, Neguţ et al., 2016). Two studies reported general mild levels of cybersickness (Nolin et al., 2012; 2016). Cybersickness was not correlated with CPT performance. In terms of presence (or a sense of actually being in a classroom), Nolin et al. (2012) reported a moderate sense of presence with no group differences (i.e., concussion and typical controls), but presence was not correlated with CPT performances. Similarly, Nolin et al. (2016) noted moderate levels of presence in a typical control sample and presence did not correlate with CPT performances. These authors also demonstrated that presence did not differ based upon grade level, gender, or the interaction of these two demographic variables. Overall, few studies to date have examined cybersickness, and even fewer have examined a sense of presence in the virtual classroom. Yet, initial results are generally positive regarding these two important factors when using a virtual reality testing modality.

Tests of Homogeneity of Variance

Regarding the second research question (Do virtual classroom CPTs offer greater differentiation in performance than traditional computerized CPTs?), comparison of performance on traditional CPTs to performance on virtual classroom CPTs in the ADHD participants was examined. Assessment of homogeneity of effects revealed evidence of significant heterogeneity for omission errors ($I^2 = 94\%, Q = 96.86$, df = 6, $p < .001$, $\tau^2 = 1.05$, $\tau = 1.02$), commission errors ($I^2 = 98\%, Q = 100.79$, df = 6, $p < .001$, $\tau^2 = 5.91$, $\tau = 2.43$), and hit reaction time ($I^2 = 98\%, Q = 397.18$, df = 5, $p < .001$, $\tau^2 = 4.52$, $\tau = 2.13$).

As can be seen in the above, our initial assessments revealed a great deal of heterogeneity. As a result, we decided to rerun the analyses to make sure that we could achieve a greater level of homogeneity. Removal of an outlier study regarding commission errors (Parsons et al., 2007) minimally impacted heterogeneity. To increase the dependability of our findings, we completed subsequent meta-analyses using random-effects models stratified by CPT metrics and for all studies combined. The heterogeneity statistics were as follows for the traditional CPT omission ($I^2 = 81\%$, $Q = 26.56$, df = 5, $p < .001$, $\tau^2 = 0.37$, $\tau = 0.61$), commission ($I^2 = 78\%$, $Q = 22.58$, df = 5, $p < .001$, $\tau^2 = 0.30$, $\tau = 0.55$), and hit reaction time ($I^2 = 48\%$, $Q = 7.71$, df = 4, $p < .10$, $\tau^2 = 0.07$, $\tau = 0.26$). The heterogeneity statistics were as follows for the virtual classroom CPT omission ($I^2 = 33\%$, $Q = 10.49$, df = 7, $p < .16$, $\tau^2 = 0.04$, $\tau = 0.2$), commission ($I^2 = 46\%$, $Q = 13.07$, df = 7, $p < .07$, $\tau^2 = 0.06$, $\tau = 0.24$), and hit reaction time ($I^2 = 53\%$, $Q = 12.68$, df = 6, $p < .05$, $\tau^2 = 0.07$, $\tau = 0.26$).

Given the diversity in research designs, stimulus parameters, and hardware configurations for both the 2D CPTs and the virtual classroom CPTs, we do not report comparisons between the virtual classroom CPTs and 2D-CPTs. Figure 2 displays funnel plots for all reviewed CPT metrics and for each CPT modality. The absence of asymmetry would suggest that publication bias in unlikely.

Mean Effects

The average weighted effects were calculated for omission errors, commission errors, and hit reaction times. This involved combining the standardized effect sizes into a composite-mean weighted effect size, and examining each for significance. Forest plots in Figs. 3, 4, and 5 display study effects and the confidence intervals around these estimates for traditional CPTs. Forest plots are in Figs. 6, 7, and 8 display study effects and the confidence intervals around these estimates for virtual classroom CPTs.

Omission errors where the strongest effect sizes for differentiating between children with ADHD and typically developing controls in both the traditional CPTs ($g = 0.81$) and the virtual classroom CPTs ($g = 1.18$). Commission errors were the next largest difference between children with ADHD and typically developing controls in both the traditional CPTs ($g = 0.81$) and the virtual classroom CPTs ($g = 0.70$). Hit reaction times displayed the smallest differences between children with ADHD and typically developing controls in both the traditional CPTs ($g = 0.14$) and the virtual classroom CPTs ($g = 0.45$).

Discussion

This article aimed to quantitatively review results from virtual classroom CPTs for differentiating the attentional performance of persons with ADHD from typically developing controls. Moreover, this study attempted to compare traditional CPTs with virtual classroom CPTs for assessing attention, but given the high heterogeneity (I^2 of >90% for omissions, commissions, and hit reaction time) with modality comparisons (i.e., 2D CPTs vs. virtual classroom-CPTs for ADHD), both main comparisons included population comparisons (i.e., control vs. ADHD) using each CPT modality. Regarding the current inability for direct modality comparisons, as the current meta-analysis demonstrated, the assessment of attention using virtual classroom CPTs has only emerged over roughly the past decade. Research interest appears to be growing based on more recent publications, but in terms of the specificity needed for meta-analyses, a limited number of articles were included to address our specific research questions. Further, a reliable estimation of moderator effects will have to wait for the accumulation of a larger body of research with greater consistency and comprehensiveness of reported results.

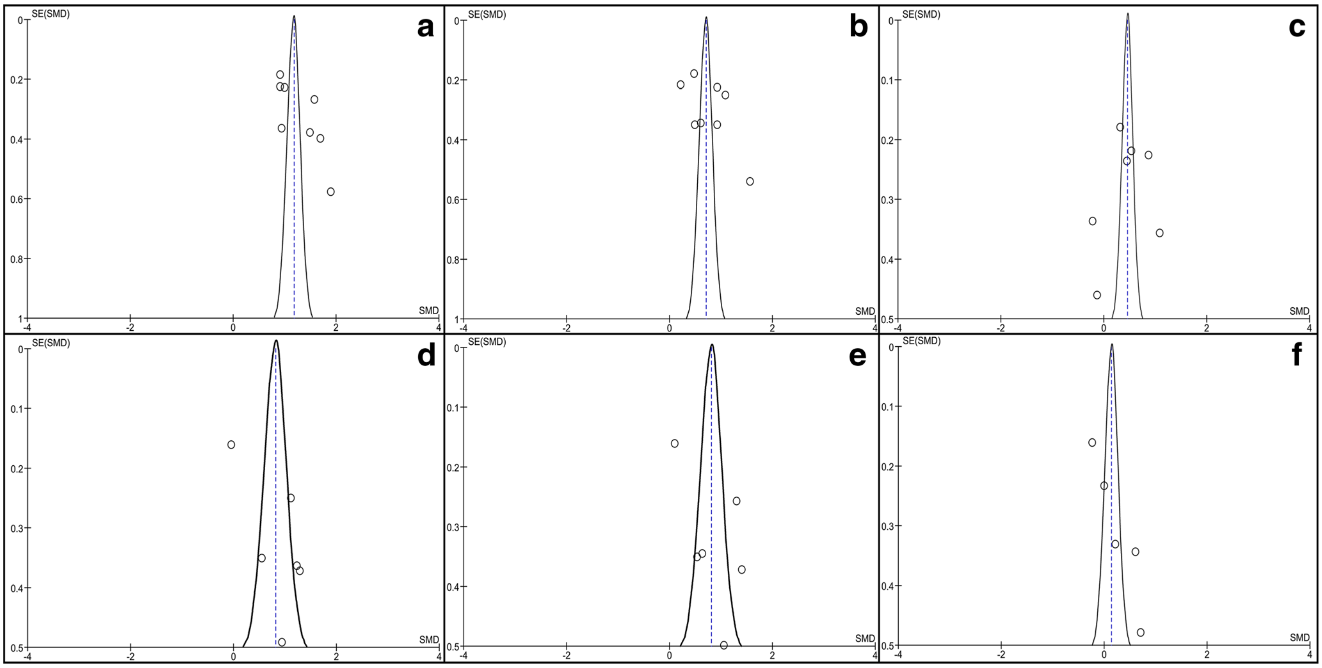

Fig. 2 **a** omission errors on virtual classroom CPTs, **b** commission errors on virtual classroom CPTs, **c** hit reaction times on virtual classroom CPTs, **d** omission errors on traditional CPTs, **e** commission errors on traditional CPTs, and **f** hit reaction times on traditional CPTs

However, as both main comparisons were comprised of primarily the same samples (except Areces et al., (2018)), one can extrapolate modality comparison interpretations from degree of group differences using each modality.

When it comes to attention deficits measured with virtual classroom CPTs, omission errors demonstrated a large effect (Cohen, J., 1992; Cohen, J., 1988), which constitutes one of the most robust and consistent findings in ADHD (Willcutt et al., 2005). Commission errors also demonstrated a large effect (Cohen, J., 1992; Cohen, J. 1988). Hit reaction times were small to trending towards medium at 0.45 (Cohen, J., 1992, Cohen, J., 1988). Thus, virtual classroom CPTs appear to be effective in differentiating individuals with ADHD from a neuropsychological assessment standpoint.

These general group differences were similar using the traditional CPT. However, group differences for omission errors and hit reaction times were augmented using the virtual classroom CPT compared to the traditional CPT ($g = 1.18$ vs. 0.81 & 0.45 vs. 0.14), but reduced for commission errors ($g = 0.70$ vs. 0.81). These findings make theoretical sense considering that the virtual classroom CPTs include an ecologically valid testing environment with naturalistic distractors, which traditional CPTs lack. Thus, performance on metrics suggestive of inattention or vigilance should be more negatively impacted, unlike impulsive responding, as indicated by the current meta-analytic findings. However, the current effect size estimates are roughly equivalent with Huang-Pollock et al. (2012) meta-analysis of 2D CPTs, yet this may be due to the current meta-analysis being under powered compared to Huang-Pollock et al. (2012; see Table 6). An interesting trend across meta-analyses (current study; Huang-Pollock et al., 2012; Pievsky & McGrath, 2017) examining the neurocognitive profile of ADHD is the greatest group differences with omission errors, intermediate differences with commission errors, and the smallest group differences regarding hit reaction time (see Table 6).

Further, in terms of ecologic validity, it is likely that the current iteration of the virtual reality classroom has some degree of verisimilitude (i.e., test or testing conditions must resemble demands found in the everyday world; Franzen & Wilhelm, 1996). Yet, some authors argue that including traditional

	ADHD			Control				Std. Mean Difference
Study	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI
Adams et al. (2009)	25.82	24.93	18	14.5	12.48	16	16.3%	0.55 [-0.14, 1.24]
Bioulac et al. (2012)	62.8	12.89	20	47.84	8.94	16	15.8%	1.29 [0.56, 2.02]
Negut, Jurma, and David (2016)	16.68	9.23	33	6.4	9.13	42	18.5%	1.11 [0.62, 1.60]
Parsons et al. (2007)	60.14	21.67	9	45.05	4.73	10	13.2%	0.94 [-0.02, 1.91]
Pollak et al. (2009)	11.75	9.69	20	2.65	1.58	17	16.0%	1.23 [0.52, 1.94]
Rodriguez et al., (2018)	83.18	23.16	113	84.08	22.27	59	20.3%	-0.04 [-0.35, 0.28]
Total (95% CI)			**213**			**160**	**100.0%**	**0.81 [0.26, 1.37]**

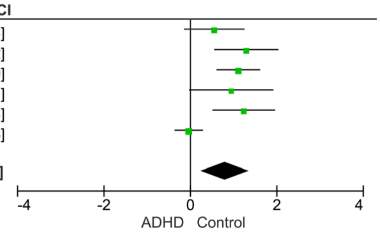

Fig. 3 Results from comparisons between groups for omission errors on traditional CPTs

Study	ADHD Mean	ADHD SD	ADHD Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
Adams et al. (2009)	21.11	25.72	18	10.31	8.06	16	16.1%	0.54 [-0.15, 1.23]
Bioulac et al. (2012)	26.25	5.06	20	22.56	6.35	16	16.3%	0.64 [-0.04, 1.31]
Negut, Jurma, and David (2016)	59.62	55.27	33	10.81	6.53	42	18.6%	1.31 [0.80, 1.81]
Parsons et al. (2007)	52.12	6.42	9	43.64	8.56	10	12.5%	1.06 [0.08, 2.04]
Pollak et al. (2009)	5.58	3.33	20	1.68	1.74	17	15.6%	1.40 [0.67, 2.13]
Rodriguez et al., (2018)	98.91	13.38	113	97.46	14.1	59	20.9%	0.11 [-0.21, 0.42]
Total (95% CI)			**213**			**160**	**100.0%**	**0.81 [0.29, 1.32]**

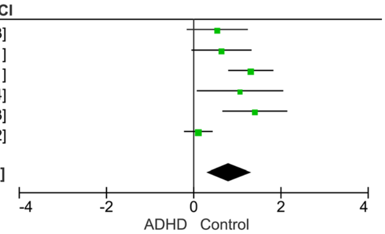

Fig. 4 Results from comparisons between groups for commission errors on traditional CPTs

neuropsychological tasks in an ecologically valid environment, still lacks the capacity to assess functions reflective of real world behaviors (Parsons, Carlew, Magtoto, & Stonecipher, 2017). Adaption of traditional tests is simply assessing antiquated theoretical cognitive constructs in a different environment (albeit a real-world one) that does not improve the ability of test performances to predict some aspect of an individual's functioning on a day-to-day basis, or veridicality (Franzen & Wilhelm, 1996). For example, traditional CPT performances are largely unrelated to executive function rating scales (Barkley & Murphy, 2011). Given the similar metric profile of omission, commissions, and hit reaction times regarding group differences for each modality, it is unlikely that the virtual classroom as is currently designed has changed that relationship between computerized testing and self or observer report of real-world executive control difficulties exhibited by those with ADHD. However, some studies did use head movements to assess inattention or susceptibility to distraction, although not enough studies to include this metric in the meta-analysis. This is an additional step toward a function-led assessment model where directly observable behaviors are captured. Then automatically logged performance attributes are analyzed to examine the ways in which a sequence or hierarchy of actions leads to a given behavior in normal functioning and how this may become disrupted. Veridicality, or the ability to model actual classroom attentional capacity, is possible through ongoing inclusion of ecologically valid attributes to the virtual classroom, such as stimuli or variables to induce more real world impulsivity (e.g., checking a text message while in class), hand or foot motion sensors to model motor hyperactivity during tasks, incorporation of social demands or cues by the classroom teacher, and so on. These are all suggested next steps in the progression towards a function-led neuropsychological testing model that is more ecologically valid.

Limitations of Meta-Analysis

Findings from this meta-analysis must be interpreted with caution given limitations of meta-analysis in general and data available for this analysis in particular. Meta-analysis is limited by the quality of studies included, and we attempted to address this by having fairly strict study inclusion and exclusion criteria. As in any review of studies in a given area, it is possible that studies with nonsignificant results are underreported. The practice of publishing only studies with significant outcomes may create a distortion of the subject under investigation, especially if a meta-analysis is done (Rosenthal, 1979). The random-effect model was utilized in the present analysis because heterogeneity was apparent, the random effects model tends to yield more generalizable parameter estimates.

A further issue for this meta-analysis, as is true of any systematic review, is deciding which trials or studies to include and which to exclude. Many systematic reviews are indeterminate because they include insufficient research designs. This is true in studies of virtual classroom CPTs, a domain where standards for consistent and comprehensive research data is limited. Depending on the study, not all outcome variables were reported (e.g., reported omission errors and commission errors, but not hit reaction time), CPT stimulus parameters were not reported, distracter parameters were not reported, hardware configurations were not reported, and so on. Further, even of the studies that utilized the CPT within the virtual classroom that did report these variables, research designs varied. This diversity constituted a major limitation of the virtual classroom CPT studies and future studies need to improve data reporting, hardware configurations, and consistency in research designs. Another significant limitation is the limited number of studies used in this analysis. This limitation

Study	ADHD Mean	ADHD SD	ADHD Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
Bioulac et al. (2012)	0.46	0.086	20	0.412	0.063	16	16.2%	0.61 [-0.06, 1.29]
Negut, Jurma, and David (2016)	0.3	0.15	33	0.3	0.001	42	24.6%	0.00 [-0.46, 0.46]
Parsons et al. (2007)	0.555	0.144	9	0.477	0.048	10	10.2%	0.71 [-0.22, 1.65]
Pollak et al. (2009)	0.609	0.178	20	0.578	0.089	17	17.0%	0.21 [-0.44, 0.86]
Rodriguez et al., (2018)	0.083	0.018	113	0.087	0.017	59	31.9%	-0.23 [-0.54, 0.09]
Total (95% CI)			**195**			**144**	**100.0%**	**0.14 [-0.21, 0.48]**

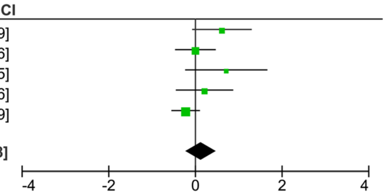

Fig. 5 Results from comparisons between groups for hit reaction times on traditional CPTs. Adams et al. (2009) did not report hit reaction times

Study	ADHD Mean	ADHD SD	ADHD Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
Adams et al. (2009)	13.1	12.6	18	3.7	5	16	9.1%	0.94 [0.22, 1.65]
Areces et al. (2016)	28.76	26.31	89	7.44	7.35	28	17.4%	0.91 [0.47, 1.35]
Areces et al., (2018)	15.38	11.6	50	5.89	5.32	38	17.1%	1.00 [0.55, 1.45]
Bioulac et al. (2012)	32.05	11.54	20	14.19	8.48	16	8.0%	1.70 [0.92, 2.47]
Negut, Jurma, and David (2016)	19.47	7.63	33	9.2	5.32	42	14.1%	1.58 [1.05, 2.10]
Parsons et al. (2007)	22.89	10.06	9	7	5.52	10	4.3%	1.90 [0.77, 3.03]
Pollak et al. (2009)	22.34	14.67	20	5.06	5.1	17	8.7%	1.49 [0.75, 2.23]
Rodriguez et al., (2018)	56.6	7.97	124	49.73	5.78	42	21.2%	0.91 [0.55, 1.28]
Total (95% CI)			**363**			**209**	**100.0%**	**1.18 [0.93, 1.42]**

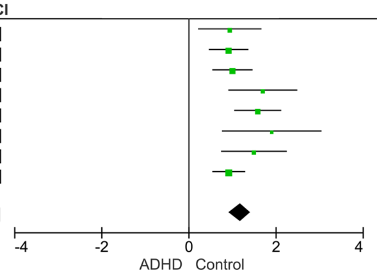

Fig. 6 Results from comparisons between groups for omission errors on virtual classroom CPTs

highlights the lack of consistent and comprehensive data reporting and limitations of many of the research designs found in virtual classroom CPT studies.

Absent or inconsistent reporting of stimulus parameters, or diversity of stimulus parameters used between studies, or different stimulus parameters for each CPT modality used within a single study constitutes a major issue for this meta-analysis. These various parameters (e.g., interstimulus interval) can have major influence on performance. The virtual classroom and traditional 2D CPTs found in studies included in this meta-analysis had numerous procedural variations including increased or decreased target frequency, interstimulus intervals, overall task length, and stimulus type (e.g., letters or numbers). In the same way that manipulation of CPT task parameters in traditional 2D versions (e.g., Conner's; Test of variables of attention; T.O.V.A) can affect behavioral response characteristics (some of which are used as markers of ability to maintain attention), parameters of the virtual classroom CPTs can be impacted. For example, higher target frequencies in traditional 2D CPTs have been found to be related to faster mean reaction times, as well as increases in errors (Beale, Matthew, Oliver, & Corballis, 1987; Silverstein, Weinstein, & Turnbull, 2004). Contrariwise, low target frequency changes result in a slower overall reaction time (Ballard, 2001). Likewise, manipulations of the interstimulus intervals in traditional 2D CPT tasks can also effect response characteristics. Shorter interstimulus intervals (<500 ms) are associated with faster mean reaction times, as well as increases in omission errors (Ballard, 2001). Conversely, longer interstimulus intervals are associated with slower reaction times, and with increased intra-individual variability (Conners, Epstein, Angold, & Klaric, 2003).

Further, researcher consensus regarding distractor type (e.g., social vs. non-social; auditory vs. visual; etc.),sequence (e.g., random or stratified), and relation to presentation of task stimuli or participant responding need ongoing conceptual and quantitative exploration. Given that these CPT parameters impact behavioral outcome measures that may have clinical utility, future studies should comprehensively report the parameters used, use the same parameters for each modality of CPT utilized, and attempt to replicate parameters from previous publications or specifically examine parameter manipulations regarding subsequent group differences in performance.

Even though we had planned, a priori, to identify possible moderators of attention assessment, this was not possible because necessary information was not reported or reported in insufficient detail. This lack of information related to self-reports of presence, levels of immersion, personality, hypnotizability, absorption, stratification by subtypes, co-occurring disorders, socioeconomic status, and average full scale IQ in participants with ADHD may reflect a limited range of values given the selection criteria employed by most studies. Thus, the findings of this meta-analysis may not generalize to patients with attentional deficits in general. Similarly, a host of other factors that could not be directly analyzed might moderate attention assessment, including differences among research centers in terms of beliefs about best practices concerning diagnosis of ADHD, counterbalancing, and types of CPTs used.

Study	ADHD Mean	ADHD SD	ADHD Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
Adams et al. (2009)	12.3	20.2	18	4.7	4.1	16	9.7%	0.50 [-0.19, 1.18]
Areces et al. (2016)	15.67	7.82	89	9	4.55	28	15.8%	0.92 [0.48, 1.36]
Areces et al., (2018)	6.5	5.36	50	5.5	3.3	38	16.4%	0.22 [-0.21, 0.64]
Bioulac et al. (2012)	21.05	9.87	20	14.87	10.05	16	9.9%	0.61 [-0.07, 1.28]
Negut, Jurma, and David (2016)	33.76	30.92	33	10.75	6.24	42	14.4%	1.08 [0.59, 1.57]
Parsons et al. (2007)	12.11	8.81	9	2	1.89	10	5.1%	1.56 [0.50, 2.62]
Pollak et al. (2009)	7.11	5.73	20	2.82	2.52	17	9.7%	0.92 [0.24, 1.61]
Rodriguez et al., (2018)	54.5	9.05	124	50.37	7.15	42	18.9%	0.48 [0.12, 0.83]
Total (95% CI)			**363**			**209**	**100.0%**	**0.70 [0.44, 0.97]**

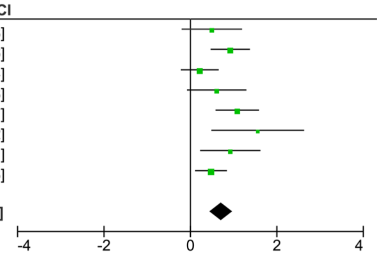

Fig. 7 Results from comparisons between groups for commission errors on virtual classroom CPTs

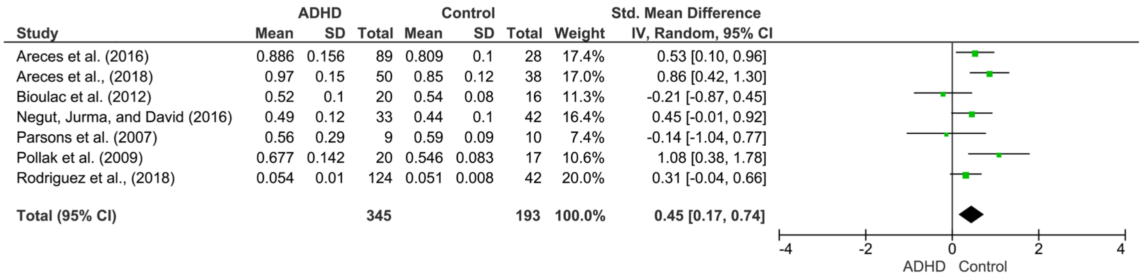

Fig. 8 Results from comparisons between groups for hit reaction times on virtual classroom CPTs. Adams et al. (2009) did not report hit reaction times

Caution is also invited in interpreting the clinical significance of the reported effect sizes. Specifically, effect size classification is somewhat arbitrary in its distinctions between magnitudes (Cohen, 1988). Hence, while a statistical consideration of data may describe 0.80 as a large effect size, statistical and clinical significance are not synonymous (Ogles, Lunnen, & Bonesteel, 2001) and an effect size is not fully informative for clinical interpretation.

A further limitation is that the dearth of sensitivity reporting in the reviewed studies made it impossible to establish a definite metric of sensitivity. Hence, we were unable to estimate completely the potential impact of diagnostic and task reliability on the sensitivity of virtual classroom CPTs. While it is important to know the reliability of clinical diagnosis and sensitivity of the measures used to detect ADHD, few studies reported the diagnostic methods that were used or the reliability of the measures utilized. Of course, even if a diagnostic approach with established reliability was utilized, there is no guarantee that the diagnostic approach was used reliably in a given study. As a result, effect sizes found in this meta-analysis are summary statistics of what is found in the literature.

Methodological Implications for Future Studies

Our study findings have several implications for future research concerning attentional assessment with virtual classroom CPTs. The effect sizes determined in this study suggest that in order for studies to have adequate power (above 0.80) to detect attentional deficits (using between groups design, and two-tailed tests with alpha set at 0.05), they would require a minimum sample size of 32 subjects (16 per group; actual power = 0.95) concerning omissions errors, 68 total subjects (34 per group; actual power = 0.96) concerning commission errors, and 298 total subjects (149 per group; actual power = 0.95) concerning hit reaction time (Faul, Erdfelder, Buchner, & Lang, 2009). Obviously, this is a minimal standard, and adequate evaluation of attentional deficits, at least using instruments applied to ADHD thus far, would ideally involve samples much larger than this. Thus,

Table 6 Effect size comparison of recent meta-analyses using traditional cpt and the current meta-analysis for adhd vs. controls

Reference	CPT Modality	*k*	*N/n*	Omissions	Commissions	Reaction time
Huang-Pollock et al. (2012)	2D CPT	39	3192	1.34	–	–
		33	3165	–	0.98	–
		26	1342	–	–	0.61*
Current Meta-Analysis	2D CPT	6	213	0.81	–	–
		6	213	–	0.81	–
		6	195	–	–	0.14
Current Meta-Analysis	VR-CPT	8	363	1.18	–	–
		8	363	–	0.70	
		8	345	–	–	0.45
Pievsky and McGrath (2017)	2D CPT & other cognitive tasks conceptualized to measure the same domain	–	–	–	–	–
		438	–	–	0.52**	–
		497	–	–	–	0.38

k number of studies, *N/n* total number of participants; The meta-analysis by Huang-Pollock et al. (2012) contained Adams et al. (2009), Parsons et al. (2007), and Pollak et al. (2009), which are articles included in the current meta-analysis. Huang-Pollock et al. (2012) reported population effect sizes and Pievsky and McGrath (2017) reported mean summative SMDs of their review of meta-analyses. Unweighted SMDs are reported

Different neuropsychological measures included to create reaction time summative SMD not reported in Pievsky and McGrath (2017)

*The effect size for reaction time was 0.29 after correcting for publication bias

**Pievsky and McGrath (2017) combined CPT commission error and stop signal task reaction time (SSRT) to create a response inhibition variable while in future small-sample studies detecting significant effects would be of interest, studies with positive findings will probably be of interest only if they are adequately powered.

Another issue is that it may behoove research groups to reach consensus regarding critical variables that should be examined as possible indicators of treatment efficacy in multi-center studies. Attempts to perform moderator analyses to identify factors that may play a role in attentional assessment were unsuccessful because mean values of potential moderator variables (e.g., sense of presence in virtual environment) were too narrow in range to allow meaningful analyses or were not adequately reported. Future studies should seek uniformity in reporting details of various patient, disorder, treatment, and virtual classroom CPT procedural variables. For example, it may be critical to identify the optimal type of virtual environments for treatment success, although this itself is beset by methodological controversy, the number of patients belonging to a diagnostic group (such as ADHD; autism; brain injury), and the relationship of these factors to attentional assessment using virtual classroom CPTs. It is anticipated that such reporting will facilitate identification of factors underlying attentional assessment and sensitivity of virtual classroom CPTs.

Conclusions

Given the currently available data, it appears that virtual classroom CPTs are relatively effective from an assessment standpoint in carefully selected studies. Virtual classroom CPTs can differentiate between persons with ADHD and typically developing controls. Whether the attentional differences are directly related to virtual classroom environments, or some other factor, remains to be specified. The meta-analytic findings parallel qualitative reviews revealing that virtual classroom CPTs have potential for assessing attentional performance in the presence of distractors. There is a need for additional well-designed and adequately powered studies investigating the efficacy of virtual classroom CPTs for assessing attentional performance in neurodevelopmental disorders, as well as more extensive and uniform reporting of data.

References

Adams, R., Finn, P., Moes, E., Flannery, K., & Rizzo, A. S. (2009). Distractibility in attention/deficit/hyperactivity disorder (ADHD): The virtual reality classroom. *Child Neuropsychology, 15*(2), 120–135.

Areces, D., Rodríguez, C., García, T., Cueli, M., & González-Castro, P. (2016). Efficacy of a continuous performance test based on virtual reality in the diagnosis of ADHD and its clinical presentations. *Journal of Attention Disorders*, 1–11. https://doi.org/10.1177/1087054716629711

Areces, D., Rodríguez, C., García, T., Cueli, M., & González-Castro, P. (2018a). Efficacy of a continuous performance test based on virtual reality in the diagnosis of ADHD and its clinical presentations. *Journal of Attention Disorders, 22*(11), 1081–1091.

Ballard, J. C. (2001). Assessing attention: Comparison of response-inhibition and traditional continuous performance tests. *Journal of Clinical and Experimental Neuropsychology, 23*(3), 331–350.

Barkley, R. A. (1991). The ecological validity of laboratory and analogue assessment methods of ADHD symptoms. *Journal of Abnormal Child Psychology, 19*(2), 149–178.

Barkley, R. A., & Murphy, K. R. (2011). The nature of executive function (EF) deficits in daily life activities in adults with ADHD and their relationship to performance on EF tests. *Journal of Psychopathology and Behavioral Assessment, 33*(2), 137–158.

Beale, I. L., Matthew, P. J., Oliver, S., & Corballis, M. C. (1987). Performance of disabled and normal readers on the continuous performance test. *Journal of Abnormal Child Psychology, 15*(2), 229–238.

Berger, I., Slobodin, O., & Cassuto, H. (2017). Usefulness and validity of continuous performance tests in the diagnosis of attention-deficit hyperactivity disorder children. *Archives of Clinical Neuropsychology, 32*(1), 81–93.

Bioulac, S., Lallemand, S., Rizzo, A., Philip, P., Fabrigoule, C., & Bouvard, M. P. (2012). Impact of time on task on ADHD patient's performances in a virtual classroom. *European Journal of Paediatric Neurology, 16*(5), 514–521.

Bioulac, S., Micoulaud-Franchi, J. A., Maire, J., Bouvard, M. P., Rizzo, A. A., Sagaspe, P., & Philip, P. (2018). Virtual remediation versus methylphenidate to improve distractibility in children with ADHD: A controlled randomized clinical trial study. *Journal of Attention Disorders*, 1087054718759751.

Bohil, C. J., Alicea, B., & Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nature Reviews Neuroscience, 12*(12), 752.

Borenstein, M., Higgins, J. P., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Research Synthesis Methods, 8*(1), 5–18.

Brosco, J. P., & Bona, A. (2016). Changes in academic demands and attention-deficit/hyperactivity disorder in young children. *JAMA Pediatrics, 170*(4), 396–397.

Bruce, K. S. (2011). Academic underachievement: A neurodevelopmental perspective. *Revista Médica Clínica Las Condes, 22*(2), 211–217.

Chaytor, N., Schmitter-Edgecombe, M., & Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Archives of Clinical Neuropsychology, 21*(3), 217–227.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159.

Cohn, L. D., & Becker, B. J. (2003). How meta-analysis increases statistical power. *Psychological Methods, 8*, 243–253.

Conners, C. K., Epstein, J. N., Angold, A., & Klaric, J. (2003). Continuous performance test performance in a normative epidemiological sample. *Journal of Abnormal Child Psychology, 31*(5), 555–562.

Corkum, P. V., & Siegel, L. S. (1993). Is the continuous performance task a valuable research tool for use with children with attention-deficit-hyperactivity disorder? *Journal of Child Psychology and Psychiatry, 34*(7), 1217–1239.

Díaz-Orueta, U. (2017). Advances in neuropsychological assessment of attention. In R. Kane & T. Parsons (Eds.), *The role of Technology in Clinical Neuropsychology* (pp. 103–142). Oxford: Oxford University Press.

Díaz-Orueta, U., Garcia-López, C., Crespo-Eguílaz, N., Sánchez-Carpintero, R., Climent, G., & Narbona, J. (2014). AULA virtual reality test as an attention measure: Convergent validity with Conners' continuous performance test. *Child Neuropsychology, 20*(3), 328–342.

Duffield, T., Parsons, T. D., Karam, S., Otero, T., & Hall, T. (2018). Virtual reality as an assessment modality with pediatric ASD populations: A systematic review. *Child Neuropsychology.*

Fasmer, O. B., Mjeldheim, K., Førland, W., Hansen, A. L., Syrstad, V. E. G., Oedegaard, K. J., & Berle, J. Ø. (2016). Linear and non-linear analyses of Conner's continuous performance test-II discriminate adult patients with attention deficit hyperactivity disorder from patients with mood and anxiety disorders. *BMC Psychiatry, 16*(1), 284.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160.

Franzen, M. D., & Wilhelm, K. L. (1996). Conceptual foundations of ecological validity in neuropsychology. In R. J. Sbordone & C. J. Long (Eds.), *Ecological validity of neuropsychological testing* (pp. 91–112). Delray Beach, FL: GR Press/St. Lucie Press.

Gilboa, Y., Kerrouche, B., Longaud-Vales, A., Kieffer, V., Tiberghien, A., Aligon, D., … Paule Chevignard, M. (2015). Describing the attention profile of children and adolescents with acquired brain injury using the virtual classroom. *Brain Injury, 29*(13–14), 1691–1700.

Gilboa, Y., Rosenblum, S., Fattal-Valevski, A., Toledano-Alhadef, H., & Josman, N. (2011). Using a virtual classroom environment to describe the attention deficits profile of children with neurofibromatosis type 1. *Research in Developmental Disabilities, 32*(6), 2608–2613.

Gutiérrez Maldonado, J., Letosa Porta, A., RusCalafell, M., & Peñaloza Salazar, C. (2009). The assessment of attention deficit hyperactivity disorder in children using continous performance tasks in virtual environments. *Anuario de Psicología, 40*(2), 211–222.

Hale, J. B., Chen, S. A., Tan, S. C., Poon, K., Fitzer, K. R., & Boyd, L. A. (2016). Reconciling individual differences with collective needs: The juxtaposition of sociopolitical and neuroscience perspectives on remediation and compensation of student skill deficits. *Trends in Neuroscience and Education, 5*(2), 41–51.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*(2), 107–128.

Hedges, L. V. (1984). Advances in statistical methods for meta-analysis. *New Directions for Program Evaluation, 24*, 25–42.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ: British Medical Journal, 327*(7414), 557–560.

Huang-Pollock, C. L., Karalunas, S. L., Tam, H., & Moore, A. N. (2012). Evaluating vigilance deficits in ADHD: A meta-analysis of CPT performance. *Journal of Abnormal Psychology, 121*(2), 360.

Huedo-Medina, T. B., Sanchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychological Methods, 11*, 193–206.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.

Iriarte, Y., Diaz-Orueta, U., Cueto, E., Irazustabarrena, P., Banterla, F., & Climent, G. (2016). AULA—Advanced virtual reality tool for the assessment of attention: Normative study in Spain. *Journal of Attention Disorders, 20*(6), 542–568.

Kofler, M. J., Rapport, M. D., & Matt Alderson, R. (2008). Quantifying ADHD classroom inattentiveness, its moderators, and variability: a meta-analytic review. *Journal of Child Psychology and Psychiatry, 49*(1), 59–69.

Kofler, M. J., Raiker, J. S., Sarver, D. E., Wells, E. L., & Soto, E. F. (2016). Is hyperactivity ubiquitous in ADHD or dependent on environmental demands? Evidence from meta-analysis. *Clinical Psychology Review, 46*, 12–24.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.

Losier, B. J., McGrath, P. J., & Klein, R. M. (1996). Error patterns on the continuous performance test in non-medicated and medicated samples of children with and without ADHD: A meta-analytic review. *Journal of Child Psychology and Psychiatry, 37*(8), 971–987.

Moreau, G., Guay, M., Achim, A., Rizzo, A., & Lageix, P. (2006). The virtual classroom: An ecological version of the continuous performance test–A pilot study. *Annual Review of Cybertherapy and Telemedicine, 4*, 59–66.

Mühlberger, A., Jekel, K., Probst, T., Schecklmann, M., Conzelmann, A., Andreatta, M., ... & Romanos, M. (2016). The influence of methylphenidate on hyperactivity and attention deficits in children with ADHD: A virtual classroom test. *Journal of Attention Disorders*, 1–13 https://doi.org/10.1177/1087054716647480.

Neguţ, A., Matu, S. A., Sava, F. A., & David, D. (2016). Virtual reality measures in neuropsychological assessment: A meta-analytic review. *The Clinical Neuropsychologist, 30*(2), 165–184.

Nigg, J. T. (2005). Neuropsychologic theory and findings in attention-deficit/hyperactivity disorder: The state of the field and salient challenges for the coming decade. *Biological Psychiatry, 57*(11), 1424–1435.

Nolin, P., Martin, C., & Bouchard, S. (2009). Assessment of inhibition deficits with the virtual classroom in children with traumatic brain injury: a pilot-study. *Annual Review of Cybertherapy and Telemedicine, 7*, 240–242.

Nolin, P., Stipanicic, A., Henry, M., Joyal, C. C., & Allain, P. (2012). Virtual reality as a screening tool for sports concussion in adolescents. *Brain Injury, 26*(13–14), 1564–1573.

Nolin, P., Stipanicic, A., Henry, M., Lachapelle, Y., Lussier-Desrochers, D., & Allain, P. (2016). ClinicaVR: Classroom-CPT: A virtual reality tool for assessing attention and inhibition in children and adolescents. *Computers in Human Behavior, 59*, 327–333.

Ogles, B. M., Lunnen, K. M., & Bonesteel, K. (2001). Clinical significance: History, application, and current practice. *Clinical Psychology Review, 21*(3), 421–446.

Parsons, T. D., Bowerly, T., Buckwalter, J. G., & Rizzo, A. A. (2007). A controlled clinical comparison of attention performance in children with ADHD in a virtual reality classroom compared to standard neuropsychological methods. *Child Neuropsychology, 13*(4), 363–381.

Parsons, T. D., Carlew, A. R., Magtoto, J., & Stonecipher, K. (2017). The potential of function-led virtual environments for ecologically valid measures of executive function in experimental and clinical neuropsychology. *Neuropsychological Rehabilitation, 37*(5), 777–807.

Parsons, T. D., & Rizzo, A. A. (2018). A virtual classroom for ecologically valid assessment of attention-deficit/hyperactivity disorder. In P. Sharkey (Ed.), *Virtual reality Technologies for Health and Clinical Applications: Psychological and neurocognitive interventions (pp. ##- ##)*. Germany: Springer-Verlag.

Pelham, W. E., Waschbusch, D. A., Hoza, B., Gnagy, E. M., Greiner, A. R., Sams, S. E., … Carter, R. L. (2011). Music and video as distractors for boys with ADHD in the classroom: Comparison with controls, individual differences, and medication effects. *Journal of Abnormal Child Psychology, 39*(8), 1085–1098.

Pievsky, M. A., & McGrath, R. E. (2017). The neurocognitive profile of attention-deficit/hyperactivity disorder: A review of meta-analyses. *Archives of Clinical Neuropsychology, 33*(2), 143–157.

Pollak, Y., Shomaly, H. B., Weiss, P. L., Rizzo, A. A., & Gross-Tsur, V. (2010). Methylphenidate effect in children with ADHD can be measured by an ecologically valid continuous performance test embedded in virtual reality. *CNS Spectrums, 15*(2), 125–130.

Pollak, Y., Weiss, P. L., Rizzo, A. A., Weizer, M., Shriki, L., Shalev, R. S., & Gross-Tsur, V. (2009). The utility of a continuous performance test embedded in virtual reality in measuring ADHD-related deficits. *Journal of Developmental & Behavioral Pediatrics, 30*(1), 2–6.

Rapport, M. D., Chung, K. M., Shore, G., Denney, C. B., & Isaacs, P. (2000). Upgrading the science and technology of assessment and diagnosis: Laboratory and clinic-based assessment of children with ADHD. *Journal of Clinical Child Psychology, 29*(4), 555–568.

Review Manager (RevMan) [Computer program]. (2014) Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration.

Rizzo, A., Bowerly, T., Buckwalter, J., Klimchuk, D., Mitura, R., & Parsons, T. D. (2006). A virtual reality scenario for all seasons: The virtual classroom. *CNS Spectrums, 11*(1), 35–44.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638.

Shaddish, W. R., & Haddock, C. K. (1994). Combining estimates of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 226–285). New York: Russell Sage Foundation.

Silverstein, M. L., Weinstein, M., & Turnbull, A. (2004). Nonpatient CPT performance varying target frequency and interstimulus interval on five response measures. *Archives of Clinical Neuropsychology, 19*(8), 1017–1025.

Sonuga-Barke, E. J., Sergeant, J. A., Nigg, J., & Willcutt, E. (2008). Executive dysfunction and delay aversion in attention deficit hyperactivity disorder: Nosologic and diagnostic implications. *Child and Adolescent Psychiatric Clinics of North America, 17*(2), 367–384.

Spooner, D. M., & Pachana, N. A. (2006). Ecological validity in neuropsychological assessment: A case for greater consideration in research with neurologically intact populations. *Archives of Clinical Neuropsychology, 21*(4), 327–337.

U.S. Department of Education, (2007-2008) National Center for Education Statistics, Schools and Staffing Survey (SASS), "Public School Data File," 2007–08.

Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., & Pennington, B. F. (2005). Validity of the executive function theory of attention-deficit/hyperactivity disorder: A meta-analytic review. *Biological Psychiatry, 57*(11), 1336–1346.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence, 7*(3), 225–240.