Psychiatry Research 291 (2020) 113056

Contents lists available at ScienceDirect

Psychiatry Research

journal homepage: www.elsevier.com/locate/psychres

The interactive effects of test-retest and methylphenidate administration on cognitive performance in youth with ADHD: A double-blind placebo-controlled crossover study

Itai Horowitz[a,b,*], Keren Avirame[a,b], Jodie Naim-Feil[c,d], Mica Rubinson[c], Elisha Moses[c], Doron Gothelf[b,e,f,†], Nava Levit-Binnun[d,†]

[a] Ruth Rappaport Children's Hospital, Rambam Medical Center, Haifa, Israel
[b] Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel
[c] Department of Physics of Complex Systems, The Weizmann Institute of Science, Rehovot, Israel
[d] Sagol Center for Brain and Mind, Baruch Ivcher School of Psychology, Interdisciplinary Center (IDC), Herzliya, Israel
[e] Child and Adolescent Psychiatry Division, The Edmond and Lily Safra Children's Hospital, Sheba Medical Center, Israel
[f] Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel

ARTICLE INFO

Keywords:
ADHD
Methylphenidate
Inhibition
Cognitive tests
Repetition effect

ABSTRACT

Studies have shown that Methylphenidate (MPH) affects cognitive performance on the neuropsychological tests and clinical symptoms of individuals diagnosed with attention deficit/hyperactivity disorder (ADHD). This study investigated the acute effects of MPH on neuropsychological tests to explore the interaction between MPH and test-retest effects. Twenty youths with ADHD were tested before and after MPH intake in a double-blind placebo-controlled crossover design and compared to twenty matched controls. Participants were tested on a range of standardized tasks including sustained attention to response, N-Back, and Word/Color Stroop. Identical tasks were administered twice each testing day, before and 1 hour after MPH/Placebo administration. Healthy controls were tested similarly with no intervention. Decreases in response time (RT) variability across tasks and in commission errors were found in ADHD after MPH. Conversely, a significant increase in RT variability and increase in omission errors were observed after the placebo. In the control group, RT variability and omission errors increased whereas commission errors decreased, suggesting fatigue and practice effects, respectively. Test-retest reliability was higher in controls than ADHD. It is suggested that cognitive tests are sensitive objective measures for the assessment of responses to MPH in ADHD but are also affected by repetition and fatigue.

1. Introduction

The clinical diagnosis of ADHD is often based on subjective experience and/or observer reports of behavioral symptoms. Batteries of cognitive tasks are also administered to measure the performance of individuals with ADHD from preschool to adulthood (Seidman, 2006). These tasks differ not only in difficulty and demands, but also in terms of the attention-related component they measure (e.g. sustained attention, inhibitory control, working memory). For example, continuous performance tests (CPTs) assess inhibitory processing, whereas the Stroop task assesses interference control (Epstein et al., 2003).

Methylphenidate (MPH) is the most commonly prescribed stimulant to treat ADHD (Van der Oord et al., 2008). Its action is relatively well-documented and is thought to correct the underlying brain dysfunction of ADHD (Wood et al., 2014). For example, neuroimaging methods have shown that MPH treatment leads to increased regional cerebral blood flow (rCBF) in the fronto-striato-thalamic circuit, thus palliating what is considered the core pathophysiology of ADHD (Li, Sham, Owen, & He, 2006). Moreover, a recent study has shown that measures of working memory and inhibitory control mediated MPH-induced improvements in classroom performance (Hawk et al. 2018).

Considered to be the first line of treatment, MPH treatment for ADHD has well-documented behavioral and cognitive effects (Van der Oord et al., 2008). However, the responses to MPH treatment are known to vary considerably (Li et al., 2006) and to date, no consistent predictors of responses to ADHD medication have been identified (Lowe et al., 2006). Repeated administrations of cognitive tests of patients on and off MPH have been conducted by researchers, and have

* Corresponding author at: Ruth Rappaport Children's Hospital, 19 Efron road, Haifa, Israel 3109601
E-mail address: itaizen@gmail.com (I. Horowitz).
† These authors contributed equally to the paper

https://doi.org/10.1016/j.psychres.2020.113056
Received 27 August 2018; Received in revised form 14 January 2020; Accepted 30 April 2020
Available online 15 May 2020
0165-1781/

been implemented in some clinical settings alongside standard clinical measures. This procedure is nevertheless problematic since performance on cognitive tests can be affected by repetitive testing or by high intra-individual variability instead of reflecting cognitive improvement induced by treatment. For example, a study on the T.O.V.A® test-retest in children with ADHD off medication revealed a temporal instability pointing to limited test-retest reliability (Llorente et al., 2001).

To the best of our knowledge, no study has explored the interaction between the effects of MPH and the effects of test-retest on the same day in a double-blind placebo-controlled design. This is of particular relevance in the context of ADHD, since inconsistency is a fundamental feature of this disorder (Epstein et al., 2011; Rubia, Smith, and Taylor, 2007). In addition previous studies also differ with respect to the cognitive tasks that were tested (DeVito et al., 2009; Vaurio, Simmonds, and Mostofsky, 2009, Epstein et al., 2011). Thus overall, there is considerable variability in designs, tasks and the observed effects of MPH on neuropsychological measures in ADHD.

To address these disparities, we examined the acute effects of MPH on cognitive performance of adolescents with ADHD using a test-retest protocol. Specifically, the study implemented a double-blind placebo-controlled crossover design with a same-day test-retest of performance on different cognitive tasks. Thus, each participant underwent the same cognitive battery before and after the administration of MPH and before and after a placebo (for a total of four times, consisting of two sessions without treatment and two sessions with treatment). As an additional control condition, a group of matched healthy control participants also underwent the same test-retest procedure on the same day at similar time intervals. We hypothesized that the cognitive changes induced by MPH would be most pronounced in the standard deviations of the measures of reaction times (RTSD) across tasks (Epstein et al., 2011) and that tasks involving inhibitory control would be the most sensitive to the effects of MPH (Huang et al., 2007). We also tested whether there would be clear effects of the test-retest same-day procedure between the participants with ADHD and the control participants. This crossover design enabled us to examine intra-subject variability across different days (repetition effects) by comparing the first session of the first day with the first session of the second day. In addition, we explored possible effects associated with the test-retest procedure such as fatigue, by comparing the session before the placebo with the session after the placebo, and by comparing the findings for the two sessions completed by the control participants.

2. Method

2.1. Subjects

Twenty teens and pre-teens (10 males and 10 females, mean age ± SD = 12.3 ± 2.7, range 9 to 17 years) with ADHD diagnosed according to the Diagnostic and Statistical Manual of Mental Disorders version 5 (DSM5) criteria and 20 age and sex-matched controls participated in this study. The ADHD participants were recruited from an ADHD clinic at a tertiary medical center, and the control participants were recruited through ads in the community. Of the ADHD participants, 8 were naive to ADHD medication, 12 were regular users of MPH, and all were diagnosed with at least moderate ADHD symptoms (Clinical Global Impression ≥ 4). Exclusion criteria included a history of psychotic disorders, autism, bipolar disorder, drug abuse, and chronic neurological or cardiovascular diseases. This study was part of a larger project for which ethical approval was obtained from the Medical Ethics Committee of Sheba Medical Center. Both parents signed an informed consent form and agreement to participate in the study was obtained from all the participants.

2.2. Assessment

2.2.1. Clinical Assessment

To establish a diagnosis of ADHD, as well as other possible comorbidities, a psychiatric clinical assessment involving a semi-structured interview was conducted by a child and adolescent psychiatrist. This assessment consisted of the Schedule for Affective Disorders and Schizophrenia for School-aged Children (Kaufman, 1997). In addition, parents and teachers filled out the ADHD rating scale (DuPaul, 1991) and the general severity of the diagnosis was rated using the Clinical Global Impression (Guy, 1976). All the participants also underwent a standardized intelligence test using the Raven's Progressive Matrices scale (Raven, 1973) to exclude other developmental factors. The participants reported on their emotional status using the Positive Affect Negative Affect Scale (PANAS; Bloch et al., 2010; Watson et al., 1988) before and after treatment with MPH and placebo. The PANAS is a 20-item self-report measure where participants rate the extent to which they feel a particular emotion on a five-point scale. In addition, the ADHD participants were asked to report any adverse effects on the Stimulant Drug Side Effects rating scale (Barkley et al., 1990).

2.2.2. Cognitive Assessment

The cognitive assessment included the following tests:

1 **The Sustained Attention to Response Task (SART).** The SART is a Go-No-Go continuous performance task developed to measure lapses in sustained conscious attention and to explore inhibitory control (Robertson et al., 1997). In the current study, participants were presented with a series of single digits (1–9), one on each trial, and asked to press a designated button in response to the presentation of a digit (Go condition) but to refrain from pressing the button if the digit presented was 3 (No-Go condition). Participants were asked to emphasize response speed without sacrificing accuracy. The task comprised 297 trials (264 Go and 33 No-Go) and lasted 12 minutes. The key measures on the SART are the number of omission errors, the number of commission errors, reaction times (RT), and the variability of reaction times (RTSD).

2 The N-Back Task. The N-Back task examines a number of key processes involved in working memory (Owen et al., 2005). Participants were presented with single letters and were instructed to press a designated button for the target stimulus and a different designated button for the non-target stimulus. In the 0-Back condition, "X" was the target letter. During the 1-Back condition, a letter was considered the target if that letter was identical to the one preceding it. In the 2-Back condition, a letter was considered the target if that letter was identical to the one presented two trials back. The 0-2 N-Back tasks thus differs in the relative demand they place on working memory and attention, and thus can evaluate the contribution of attentional deficits to working memory deficits. Importantly, since the participants were required to press a button on each trial, inhibitory control was less necessary. Each of the three tasks consisted of 90 trials (60 target and 30 non-targets) and lasted for 3 minutes. The key measures on the N-Back tasks are the number of errors, the reaction times, and the variability of the reaction times.

3 The Stroop Color and Word Task. In this task participants are required to inhibit the more automatic tendency to read a written word while performing the less automatic task of choosing the color of ink in which the word is written (Stroop, 1935). The colors were represented on the keyboard (red, yellow, blue and green). Presentation of word stimuli were divided into four blocks, each consisting of 16 congruent (the word red written in red ink for example) and16 incongruent colored word stimuli (red written in green, for example), presented randomly (Peterson et al., 2009). Overall, the task lasted for 3 minutes. The key measures on the Stroop task are the number of errors, the RTs, and the variability of the RTs (RTSD).

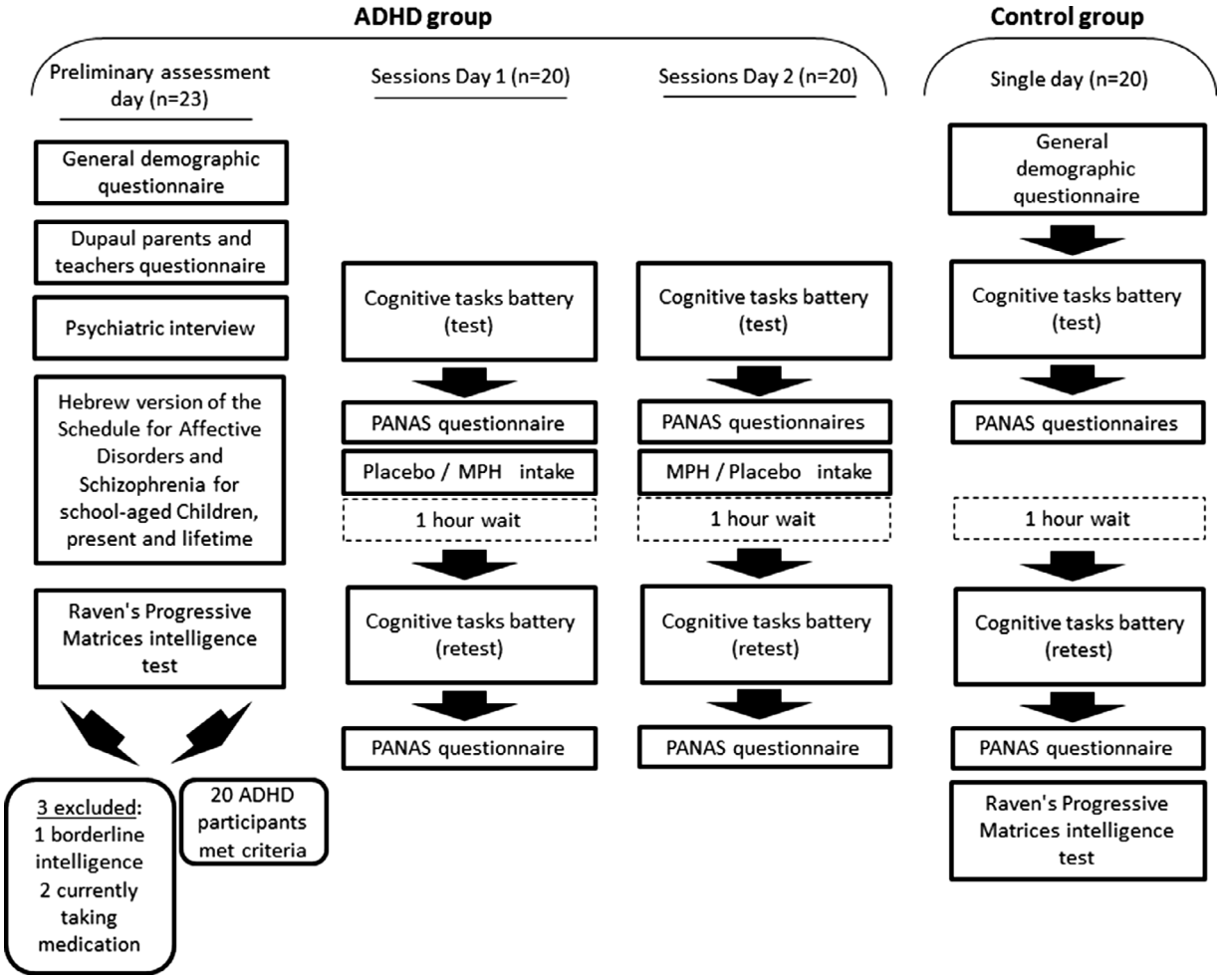

Fig. 1. Experimental procedure: double-blind placebo-controlled cross-over study design. Half of the participants received MPH immediate release on Day 1 and the placebo on Day 2 and half received the placebo on Day 1 and MPH immediate release on Day 2. The control group consisted of healthy age and sex matched participants. PANAS - Positive Affect Negative Affect Scale. SDQ – Strengths and Difficulties Questionnaire.

2.3. Procedure

The experimental procedure is depicted in Fig. 1. In this double-blind placebo-controlled cross-over study design, half of the participants received MPH immediate release (IR) first and then the placebo and the other half received the placebo first and then MPH-IR in a randomized manner. Since the CDC usually restricts maximum dose levels (with occasional exceptions) to a 20 mg capsule regardless of body weight, either 0.5 mg/kg or 20 mg MPH-IR were administered, whichever was lower. Participants already using methylphenidate (MPH) had at least a 7-day medication washout prior to the evaluation day to ensure a complete washout of any stimulant. All participants performed the three tasks in the same order, starting with SART, continuing with the N-Back and ending with the Stroop. For all tasks, the stimulus presentation and response recording were controlled by a computer using the E-Prime software package (Psychology Software Tools Inc., Pittsburgh, PA). After being given the instructions, each task started with 16 practice trials with feedback. The PANAS was administered at the end of each testing session, overall four times, to control for possible effects of mood on cognitive performance and to measure changes in the participants' subjective experience. At the end of each session day, the ADHD participants were also assessed for adverse effects using the Barkley rating scale. The control group were only administered two sessions (T1 and T2) in a single day. Due to the crossover of MPH and placebo, the ADHD group completed four sessions, two sessions per day on two different days (T1, T2, T3, T4). Each session lasted for about 90 minutes with a 60 minute interval between same-day sessions (T1-T2 and T3-T4) in order to permit the participants to rest and for the stimulant to become active (Denney and Rapport, 1999).

2.4. Statistical Analysis

The mean number of errors, the mean response times for correct responses, and the mean response time variance for correct responses were calculated separately for each task (SART, N-Back, Stroop). Only response times within 2 standard deviations of the mean were included in the analysis. A Kolmogorov–Smirnov test was used to assess normality (set to $p < 0.05$) prior to the data analysis. A repeated measure ANOVA was used to test for stimulant effects in the ADHD participants, with Intervention (MPH, placebo) and Time (Before, After) as the within-subject factor, and Order (MPH first, MPH second) as the between-subjects factor. Post-hoc analyses with paired-sample *t*-tests were conducted to explore the direction of the significant differences. The differences between ADHD participants (before and after Placebo) and the controls were tested using a repeated measure ANOVA with Time (T1, T2) as the within-subject factor and Group (HC, ADHD) as the between-subjects factor. Independent-sample and paired-sample *t*-tests were used to further examine possible differences between groups and between sessions (T1, T2).

For each outcome measure, we calculated the difference between the results before MPH intake and after MPH intake (ΔMPH) regardless

Table 1
Characteristics of Study Sample.

	ADHD	Controls
N	20	20
Gender (M:F)	1:1	1:1
Age (Mean ± SD)	12.4 ± 2.7	12.4 ± 2.7
Raven's progressive matrices standardized mark (Mean ± SD)	134.7 ± 15.5	135.5 ± 13.6
Socioeconomic Status (high: medium: low)	17:3:0	18:2:0
Parents education (years; Mean ± SD)	16.6 ± 3.1	17.5 ± 2.2
ADHD-RS Parents (Mean ± SD)	49.7 ± 9	N/A
ADHD-RS Teachers (Mean ± SD)	39.9 ± 12	N/A
Inattentive (*n*)	8	N/A
Hyperactive-impulsive (*n*)	0	N/A
Combined (*n*)	12	N/A
CGI (Mean ± SD)	4.65 ± 1.3	N/A

of day. A Pearson correlation was then used to test whether there was an association between performance at baseline and cognitive changes induced by MPH. To estimate test-retest reliability for the different measures on each task, correlational analyses were conducted separately for the healthy controls (T1-T2), for ADHD on the same day (pre- and post-placebo) and for ADHD on different days (T1-T3). A repeated measure ANOVA was also used to test for changes in subjective experience (positive and negative PANAS scales), with Intervention (MPH, placebo) and Time (Before, After) as the within-subject factor and Order (MPH first, MPH second) as the between-subjects factor. Statistical analyses were carried out using SPSS, Version 21 (IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp). All statistical tests were considered significant at a level of $p < 0.05$, unless specified otherwise.

3. Results

The ADHD subjects and the healthy controls were matched for age, sex, socioeconomic status and fluid intelligence (see Table 1 for demographic details). One healthy control participant was excluded from the analyses because of extremely poor performance (error rates exceeding 2 SD) on the initial examination (T1). One participant from the ADHD group was excluded from the analysis of the N-Back task due to technical problems which occurred at T2. The mean time difference between the first and second day of the examination was 17.63 days (SD = 9.3) with a minimum difference of 7 days. No serious adverse effects were reported following drug administration. At the end of the study, participants were asked to state the order in which they received the medication to confirm the blinding procedure; 80% of the ADHD participants reported receiving MPH on the first session (37.5% were wrong), 20% reported receiving the Placebo (50% were wrong) with no significant correlation between reported and actual pill content ($r = 0.05$, $p = 0.83$).Comparison of neuropsychological scores between groups and conditions is shown in Table 2. Task reptition differences were most prominent in RTSD. (See Figure 2).

3.1. MPH vs. placebo

3.1.1. SART

The analysis of omission errors yielded a significant Intervention × Time interaction ($F(1,18) = 8.238$, $p = 0.01$). Omission errors significantly increased under the placebo ($t(19) = -3.107$, $p = 0.006$) whereas the decrease observed under MPH was not significant. The analysis of commission errors indicated a significant Intervention × Time interaction ($F(1,18) = 7.680$, $p = 0.013$), since participants committed significantly fewer commission errors under MPH ($t(19) = 3.454$, $p = 0.003$). There was also a significant Intervention × Time interaction for the RTSD (reaction time standard deviation) ($F(1,18) = 19.328$, $p < 0.001$). Post- hoc comparisons revealed a significant decrease in the RTSD under MPH ($t(19) = 3.595$, $p = 0.002$) and a significant increase under the placebo ($t(19) = -3.779$, $p = 0.001$).

3.1.2. N-Back Test

There was a significant Intervention × Time interaction for errors on the 1-Back ($F(1,17) = 10.960$, $p = 0.004$) with no significant results on the post hoc tests. The analysis of RTSD yielded a significant Intervention × Time interaction on the 0-Back ($F(1,17) = 11.288$, $p = 0.004$), the 1-Back ($F(1,17) = 7.632$, $p = 0.013$) and the 2-Back ($F(1,17) = 6.848$, $p = 0.019$). A significant decrease in RTSD under MPH was only observed for 0-Back ($t(18) = 4.043$, $p = 0.001$) and 1-Back ($t(18) = 3.538$, $p = 0.002$).

Table 2
Comparison of neuropsychological scores between groups (healthy controls versus ADHD) and conditions (T1 and T2 for healthy controls, before and after placebo or MPH for ADHD, and between testing days).

	HC			ADHD						
TASK	T1	T2	Δ	Before Placebo	After Placebo	Δ Placebo	Before MPH	After MPH	Δ MPH	Δ Day#
SART	Mean (SE)	Mean (SE)	Difference	Mean (SE)	Mean (SE)	Difference	Mean (SE)	Mean (SE)	Difference	Difference
Mean RT	436.3 (17.9)	432.2 (20.4)	-4.1	479.3 (26.3)	465.6 (23.7)	-13.7	441.4 (21.8)	464.6 (29.1)	23.2	-35.8*
RTSD	107.7 (6.7)	112.9 (9.3)	5.2	111.8 (9.1)	127.9 (9.7)	16.11***	118.2 (9.3)	90.4 (7.9)	-17.74***	-7.5
Omission	3.8 (0.7)	8.79 (3.1)	4.9	10.2 (2.4)	20.4 (4.7)	10.15**	10.7 (3.2)	6.4 (2.4)	-4.4	-3.2
Commission	12.2 (1.6)	9.79 (1.50)	-2.5*	10.7 (1.8)	11.6 (1.7)	0.9	15.0 (1.6)	10.5 (1.7)	-4.5***	3.4
0-BACK										
Mean RT	505.7 (26.6)	517.2 (27.1)	11.6	555.6 (30.7)	540.9 (31.5)	-14.7	546.6 (46.6)	578.0 (29.2)	31.4	-3.09
RTSD	109.6 (6.0)	124.4 (7.9)	14.8*	113.7 (9.5)	127.3 (6.2)	13.6	122.7 (6.8)	94.1 (6.4)	-28.60***	16.6
Errors	1.9 (0.7)	3.83 (1.1)	1.9	14.4 (5.4)	8.5 (2.7)	-5.9	6.2 (2.1)	3.1 (1.4)	-3.1	-4.9*
1-BACK										
Mean RT	587.9 (30.1)	594.1 (28.9)	6.1	626.6 (35.2)	633.1 (32.2)	6.5	611.0 (32.8)	584.9 (34.9)	-26.1	-3.4
RTSD	126.8 (6.2)	145.2 (7.9)	18.4***	134.8 (7.9)	142.6 (8.0)	7.8	134.3 (7.2)	114.4 (5.7)	-19.89***	-2.5
Errors	8.2 (2.5)	10.6 (2.7)	2.3	14.3 (3.1)	19.8 (3.9)	5.5	15.3 (3.6)	8.5 (2.6)	-6.8	1.0
2-BACK										
Mean RT	598.2 (29.9)	588.2 (28.6)	-10.1	567.2 (45.7)	552.4 (45.7)	-14.8	572.0 (49.7)	581.0 (44.1)	9.0	6.6
RTSD	146.9 (7.0)	158.2 (9.2)	11.3	143.4 (12.1)	156.5 (12.6)	13.0	146.7 (12.6)	127.8 (9.3)	-18.9	7.7
Errors	17.9 (3.8)	15.5 (3.3)	-2.4	25.3 (5.6)	28.0 (5.4)	2.7	27.1 (5.8)	21.1 (5.4)	-6.0	3.2
STROOP										
Mean RT	758.9 (21.5)	764.7 (20.6)	5.8	829.8 (11.8)	811.0 (11.3)	-18.9	796.6 (16.5)	784.2 (16.5)	-12.3	4.2
RTSD	110.5 (5.2)	117.1 (9.0)	6.6	102.2 (5.9)	108.6 (5.4)	6.4	117.8 (6.0)	103.4 (6.0)	-14.4	3.5
Errors	9.6 (2.3)	9.1 (1.9)	-0.5	13.7 (1.6)	11.4 (1.5)	-2.2	13.3 (2.1)	8.3 (1.6)	-5.0	1.7

SE – Standard error. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ # pre-intervention

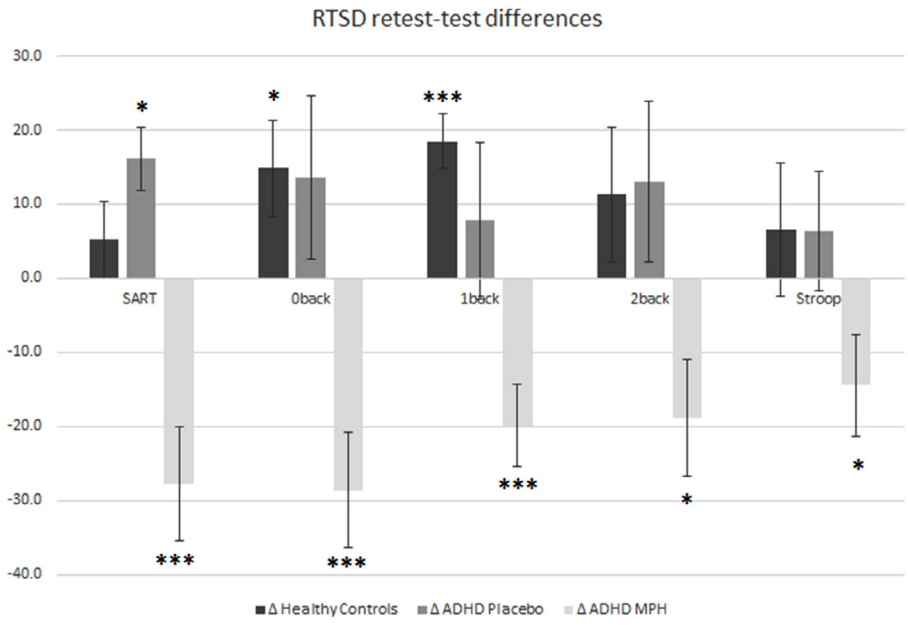

Fig. 2. The effects of test-retest on RTSD. Comparison of variability in differences in response times across tasks and conditions: (Δ Healthy Controls) T2 - T1 for Healthy controls, (Δ ADHD Placebo) after – before placebo intake and (Δ ADHD MPH) before - after MPH intake for ADHD patients; RTSD - reaction time standard deviation; MPH – Methylphenidate; * $P < 0.05$, *** $P < 0.001$

3.1.3. Stroop Test

The Intervention × Time interaction for RTSD was marginally significant ($F(1,17) = 4.289$, $p = 0.053$) with no significant results on the post-hoc comparisons. No other significant interactions were observed.

3.1.4. Day 1 vs. Day 2

To examine stability between testing days, the two pre-intervention sessions (i.e., before MPH and before Placebo) were compared using paired-sample *t*-tests. Overall, the results point to a between - days stability in performance since there was only a significant difference for RTs on the SART ($t(19) = -2.188$, $p = 0.042$) and errors on the 0-back ($t(18) = -2.391$, $p = 0.028$).

3.2. ADHD vs. Controls

3.2.1. SART

There was a significant main effect of Time ($F(1,36) = 10.864$, $p = 0.002$) for RTSD on the SART with a significant increase in RTSD from T1 to T2. The analysis also revealed a significant main effect of Time ($F(1,36) = 11.028$, $p = 0.002$) and Group ($F(1,36) = 4.323$, $p = 0.045$) for omission errors on the SART with a higher number of errors at T2 than T1 and in ADHD than HC. The analysis of commission errors in the SART revealed a significant Time x Group interaction ($F(1,36) = 5.224$, $p = 0.028$) with a decrease in commission errors from T1 to T2 solely in the control group ($t(18) = 2.531$, $p = 0.021$).

3.2.2. N-Back Test

There was a significant main effect of Group for errors on the 0-Back ($F(1,36) = 4.242$, $p = 0.047$) with significantly more errors in the ADHD group. The main effect of Time for errors on the 1-Back approached significance ($F(1,36) = 3.863$, $p = 0.057$) with an increase in errors from T1 to T2. The analysis of RTSD yielded a significant main effect of Time on the 0-Back ($F(1,36) = 6.695$, $p = 0.014$) and the 1-Back ($F(1,36) = 7.004$, $p = 0.012$) with a significant increase in RTSD from T1 to T2.

3.2.3. Stroop Test

An analysis of the Stroop task only revealed a significant main effect of Group for RT ($F(1,38) = 5.780$, $p = 0.021$) with shorter RTs for the control group.

3.3. Correlation with baseline performance

Analyses using Pearson's correlations showed that baseline performance (before Placebo) was significantly correlated with changes in performance after MPH intake (Δ = Before MPH-After MPH); specifically, there were significant positive associations for errors on the 0-Back ($r = 0.627$, $p = 0.004$) and for RTSD on the Stroop task ($r = 0.571$, $p = 0.025$) as well as for errors on the Stroop task ($r = 0.659$, $p = 0.002$). These correlations suggest that the effect of MPH increased in relation to baseline difficulty to perform the task, with more robust effects of medication for tasks with high variability in response times and high error rates at baseline.

3.4. Test-retest reliability

Table 3 presents test-retest reliability coefficients for all measures on each task. Correlation coefficients below 0.7 are marked in italics as they might be indicative of questionable (≥ 0.6 < 0.7) or poor

Table 3
Group Test–retest reliability (Pearson correlations, *r*) for Mean RT, RTSD and errors across tasks.

Task	Measure	HC Same day	ADHD Same day retest#	ADHD Different days retest##
SART	*Mean RT*	0.900***	0.778***	0.804***
	RTSD	0.886***	0.908***	*0.606***
	Omission	0.589**	0.746***	*0.378*
	Commission	0.806***	0.773***	*0.495**
0-BACK	*Mean RT*	0.899***	0.889***	*0.354*
	RTSD	0.729***	*0.455**	*0.441*
	Errors	*0.409*	*0.511**	0.932***
1-BACK	*Mean RT*	0.932***	0.941***	0.908***
	RTSD	0.745***	*0.606***	*0.232*
	Errors	0.839***	*0.650***	0.722***
2-BACK	*Mean RT*	0.732***	0.945***	0.915***
	RTSD	*0.549**	0.701***	*0.677***
	Errors	0.768***	0.724***	*0.672***
STROOP	*Mean RT*	0.826***	*0.454**	*0.540**
	RTSD	*0.397*	*0.550**	*0.612***
	Errors	0.770***	0.799***	0.782***

Correlation coefficients (*r*) lower than 0.7 were marked in italic
* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ # before and after Placebo intake ## pre-intervention

reliability ($\geq 0.5 < 0.6$). The results showed that the same measures on the same tasks were the most reliable in healthy controls (3 coefficients < 0.7) and more reliable for same day testing (6 coefficients < 0.7) than on different days when testing the ADHD participants (10 coefficients < 0.7).

3.4. Subjective reports

The analysis of rating scores in the ADHD group on the positive and negative PANAS scales yielded no significant results, suggesting that the participants' subjective reports were not significantly affected by the test-retest procedure or the intervention. This was true for both the directly attention-related PANAS items such as "distracted", "active" or "detached", as for attention-unrelated items such as "upset", "guilty" or "ashamed".

4. Discussion

The present study was designed to examine the acute effects of MPH on a sample of adolescents with ADHD using a well-controlled study design and a wide range of cognitive tasks which were administered repeatedly. The present study mirrors the typical use of neuropsychological tests and the frequent intake of MPH in ADHD. It is also in line with the research tradition in ADHD such as the seminal study by Rapoport and colleagues (1980) that tested the effects of a single dose of Dextroamphetamine on cognitive (e.g., CPT), behavioral (e.g., communication), and subjective measures (e.g., self-report rating scale). To the best of our knowledge, the present study is the first crossover placebo-controlled trial to investigate the interaction between MPH and test-retest effects. Understanding this interaction is important for clinicians and researchers who conduct repeated testing on the same day before and after drug intake to investigate or support clinical decisions concerning the administration of a stimulant (Riccio, 2001). For example, a recent study by Kritchman et al. (2019) administered clinical and cognitive measures on the same day before and after intake of MPH or a placebo in 20 children with ADHD to test the effects of one dose of MPH on patients' anxiety and its correlation with clinical responses.

Examining the validity and reliability of the use of repeated neuropsychological tests in ADHD is especially important since the behavioral dynamics of individuals with ADHD is inherently characterized by moment-to-moment fluctuations in response accuracy, impulsivity and intra-individual variability, thus leading to behaviors that are more sensitive to task duration and incentives (Castellanos et al., 2005; Uebel et al., 2010; Bubnik et al 2015). In line with previous studies, Bubnik and colleagues (2014) showed that children with ADHD had more difficulty sustaining their attention as time on task increased, resulting in fewer detected targets. Here as well we observed effects related to MPH, as well as intervening factors such as practice and fatigue.

4.1. MPH-related effects

The comparisons between the MPH and placebo enabled us to investigate the direct effects of MPH on performance in ADHD. We found that the increased RT variability in ADHD as compared to controls was suppressed by MPH but not by the placebo. Previous studies have also reported that under MPH, participants with ADHD are more accurate, faster and in better control (Bedard et al., 2003; Epstein et al. 2011). Furthermore, the effect of stimulant medications on reducing RT variability in patients with ADHD was shown to be consistent across multiple cognitive tasks (Epstein, 2011). This finding is in line with a vast corpus of research reporting increased variability in RT across various tasks in the neuropsychological testing of individuals with ADHD (Castellanos and Tannock, 2002). Early studies suggested that increased reaction time variability could reflect reduced stimulus control, thus leading to deficits in self-regulation which are apparent from very early stages in the development of ADHD (Barkley, 1997).

As of its inception, research on the effects of single dose Dextroamphetamine on ADHD has used designs combining a sustained-attention measure (e.g. CPT), tasks involving learning and communication, and a subjective report measure in a cohort of hyperactive and normal boys and young men (Rapoport et al., 1980). The use of various tasks serves to measure different aspects of attention and executive function that capture different cognitive demands and requirements (Coghill et al., 2014). The literature considers that the neuropsychological profile of ADHD appears to include deficits in divided attention, inhibition of response, working memory and RTSD (Pasini et al, 2007). The findings from the three tasks administered in the current study (SART, N-Back, and Stroop) hint that the tasks comprising inhibitory processes and working memory were more sensitive to attentional deficits and behavioral changes following MPH. Finally, RTSD at baseline correlated with the effects of MPH on this measure in that the benefits of MPH were greater for subjects presenting with increased reaction time variability at baseline.

In line with earlier studies, there was no significant effect of a single administration of MPH on performance in the Stroop task (Bedard, Ickowicz, and Tannock, 2002). This is consistent with previous results indicating that the Stroop task is not sensitive to detecting interference control in ADHD (Van Mourik, Oosterlaan, and Sergeant, 2005). Nevertheless, it has been argued that the Stroop task is appropriate for clinically monitoring responses to MPH (Langleben et al., 2006) and that deficits in interference inhibition are considered a reliable predictor of deficits on working memory tests and point to a link between these two functions in the inattentive subtype (Pasini et al., 2007). These contradictory results regarding the Stroop task further emphasize the need to administer several tasks in clinical setups, given the pathophysiological heterogeneity of ADHD (Maoz et al., 2018). Many studies have indeed argued for a complex approach in which a diversity of endophenotypes come into play in different clusters of ADHD (Nigg et al., 2005). For example, some individuals may have executive deficits related to the fronto-striatal system (Nigg, Stavro, et al., 2005) whereas others may suffer from deficient motivational processes associated with the limbic-striatal circuit (Sagvolden et al., 2005). Studies have indicated that in certain clusters of ADHD that did not respond to MPH, the attentional symptoms could partly result from anxiety (Moshe, Karni, and Tirosh, 2012). It is therefore crucial to combine a range of behavioral and clinical measurements to obtain a full profile of the disorder.

In the first session, 80% of the participants in the ADHD group reported receiving MPH (37.5% were wrong) and 20% reported receiving Placebo (50% were wrong) with no significant correlation between the reported and actual pill content ($r = 0.05$, $p = 0.83$), suggesting that true Placebo control was achieved. It has been reported that the average rate of the positive clinical response to a placebo ranges from about 20% to 30% (Waschbusch et al., 2009). However, omission errors and RTSD in the SART increased following the administration of the placebo. Elsewhere, the placebo effect for stimulant treatment was found to enhance subjective arousal but not cognitive performance when tested with continuous performance tests such as the SART (Looby and Earleywine, 2011). Although we found no change in the reported subjective experience as measured by the PANAS, it seems likely that the observed increase in omission errors and RTSD in the SART was due to fatigue and not a nocebo effect.

4.2. Effects related to the test-retest procedure

By examining the changes in performance during the placebo intervention in ADHD and the two testing sessions in the control group we were able to investigate effects related to the test-retest procedure. Again, RTSD was sensitive enough to detect changes in the controls and in the ADHD participants under placebo. It is possible that this effect

was related to mental fatigue, a phenomenon which has been previously documented (Boksem, Meijman, and Lorist, 2005; Wang, Ding, and Kluger, 2014). Normally, mental fatigue is associated with a decline in information processing as a result of 'resource depletion', loss of motivation and fluctuations in executive control. On the behavioral level, mental fatigue is often expressed in increased latency and RTSD (Kato, Endo, and Kizuka, 2009). In the present study, we observed increased RTSD on the N-back tasks in the controls and increased RTSD and omission errors on SART in the ADHD group under placebo. It is plausible that MPH blocked the tendency to commit fatigue-related omission errors during the SART (which occurred under the placebo and in the control group). Additionally, we found that the controls committed fewer commission errors on the SART when repeating the task within a short interval. Previous work based on the CPT has also reported a practice effect on commission errors (Wright et al., 2014). There is an ongoing debate on the practice effects in test-retest constellations although the authors of various commercial ADHD computerized tasks claim that the impact of test-retest practice is negligible (Raz et al., 2014). Our results from the control group raise doubts as to the reliability of repeated measurement of commission errors to accurately reflect impulsivity. This corresponds to data from earlier research using another CPT task, the T.O.V.A®, which pointed to a significant improvement in commission errors in healthy and ADHD participants on retest (Leark, Wallace, and Fitzgerald, 2004). Similarly, Llorente and colleagues reported that omission and commission mean scores increased upon retest in children with ADHD (Llorente et al., 2001). We also found relatively poor reliability on omission errors in the SART for HC and for different days in ADHD. Conversely, the correlational analyses yielded acceptably high reliability for the SART and the 2-Back in ADHD when tested on the same day. The results also indicated that the same measures on the same tasks are more reliable in healthy controls than in ADHD when tested on different days. Although the interpretation of the observed correlations cannot be definitive, the results suggest that reliability appeared to differ between populations and between procedures and therefore might have an important effect on diagnostic accuracy.

4.3. Conclusion

The current study investigated the validity of the administration of objective measures in a short interval test-retest procedure using a crossover placebo-controlled design. In general, we observed that the effects of task repetition were detectable in this design in both the controls and in the ADHD participants and were distinct from MPH effects, most noticeably in RTSD. Our results underscore the need to further investigate practice and fatigue effects in adolescent ADHD populations, as well as their potential impact on ADHD diagnostic procedures. Future work should aim to develop clinical and behavioral instruments that can better measure ADHD-related symptoms and subjective experiences which are sensitive to the state of the individual, thus complementing the clinical scales commonly used to assess the severity of ADHD symptoms which are mostly focused on individual traits.

4.4. Limitations

The current findings are limited by the small sample size and the heterogeneity of the participants in terms of drug naivety, age, and ADHD specifiers. The small sample size is a key limitation when examining neuropsychological tests given the variability in performance on these tests. The crossover design used here minimized the problems of comparability in terms of age, sex and inter-individual variance since the participants served as their own controls. However, other sources of variability likely remain and further studies with larger groups are needed to confirm the current findings. Moreover, the study followed traditional measurement of variability in Mean RT despite the conceptual and statistical limitations of RTSD due to its use in clinical setup. Note as well that each testing session comprised three consecutive tasks lasting about one hour, which might have constituted a larger cognitive load than usual testing in clinical setups. Although the SART has been considered to be a brief and valid measure of failure in sustained-attention since the end of the 1990's (Robertson et al., 1997), there is an ongoing debate as to whether it is the most suitable test to measure sustained attention. The conclusions that can be drawn from the observed changes are also limited in the sense that since validated clinical assessment tools (e.g. the ADHD rating scale) were not used to assess immediate changes in ADHD symptoms, possible association between cognitive and clinical improvement could not be fully evaluated . Further studies with larger groups are required to confirm the current findings. Finally, the clinical use of computerized test-retest assessment tools in ADHD management varies across countries. Future studies should explore the differences in practices as a function of country to achieve a better picture of the findings reported from the use of these assessment tools.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. This study was supported by the National Institute for Psychobiology in Israel - Founded by the Charles E. Smith Family.

References

Barkley, R.A., 1997. Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. Psychol. Bull. 121 (1), 65.
Barkley, R.A., McMurray, M.B., Edelbrock, C.S., Robbins, K., 1990. Side effects of metlyiphenidate in children with attention deficit hyperactivity disorder: a systemic, placebo-controlled evaluation. Pediatrics 86 (2), 184–192.
Bedard, A.-C., Ickowicz, A., Logan, G.D., Hogg-Johnson, S., Schachar, R., Tannock, R., 2003. Selective inhibition in children with attention-deficit hyperactivity disorder off and on stimulant medication. J. Abnorm. Child Psychol. 31 (3), 315–327.
Bedard, A.-C., Ickowicz, A., Tannock, R., 2002. Methylphenidate improves Stroop naming speed, but not response interference, in children with attention deficit hyperactivity disorder. J. Child Adolesc. Psychopharmacol. 12 (4), 301–309.
Bloch, Y., Harel, E.V., Aviram, S., Govezensky, J., Ratzoni, G., Levkovitz, Y., 2010. Positive effects of repetitive transcranial magnetic stimulation on attention in ADHD Subjects: a randomized controlled pilot study. The World J. Biol. Psychiatry 11 (5), 755–758.
Boksem, M.A., Meijman, T.F., Lorist, M.M., 2005. Effects of mental fatigue on attention: an ERP study. Cognitive Brain Res. 25 (1), 107–116.
Castellanos, F.X., Sonuga-Barke, E.J., Scheres, A., Di Martino, A., Hyde, C., Walters, J.R., 2005. Varieties of attention-deficit/hyperactivity disorder-related intra-individual variability. Biol. Psychiatry 57 (11), 1416–1423.
Castellanos, F.X., Tannock, R., 2002. Neuroscience of attention-deficit/hyperactivity disorder: the search for endophenotypes. Nat. Rev. Neurosci. 3 (8), 617–628.
Coghill, D.R., Seth, S., Pedroso, S., Usala, T., Currie, J., Gagliano, A., 2014. Effects of methylphenidate on cognitive functions in children and adolescents with attention-deficit/hyperactivity disorder: evidence from a systematic review and a meta-analysis. Biol. Psychiatry 76 (8), 603–615.
Denney, C.B., Rapport, M.D., 1999. Predicting methylphenidate response in children with ADHD: theoretical, empirical, and conceptual models. J. the American Academy of Child & Adolescent Psychiatry 38 (4), 393–401.
DeVito, E.E., Blackwell, A.D., Clark, L., Kent, L., Dezsery, A.M., Turner, D.C., Sahakian, B.J., et al., 2009. . Methylphenidate improves response inhibition but not reflection–impulsivity in children with attention deficit hyperactivity disorder (ADHD). Psychopharmacology (Berl.) 202 (1–3), 531.
DuPaul, G.J., 1991. Parent and teacher ratings of ADHD symptoms: psychometric properties in a community-based sample. J. Clin. Child. Adolesc. Psychol. 20 (3), 245–253.
Epstein, J.N., Erkanli, A., Conners, C.K., Klaric, J., Costello, J.E., Angold, A., 2003. Relations between continuous performance test performance measures and ADHD behaviors. J. Abnorm. Child Psychol. 31 (5), 543–554.
Epstein, J.N., Langberg, J.M., Rosen, P.J., Graham, A., Narad, M.E., Antonini, T.N., et al., 2011. . Evidence for higher reaction time variability for children with ADHD on a range of cognitive tasks including reward and event rate manipulations. Neuropsychology 25 (4), 427.
Guy, W., 1976. Clinical global impression scale. The ECDEU Assessment Manual for Psychopharmacology-Revised Volume DHEW Publ No ADM 76 (338), 218–222.
Hawk Jr, L.W., Fosco, W.D., Colder, C.R., Waxmonsky, J.G., Pelham Jr, W.E., Rosch, K.S., 2018. How do stimulant treatments for ADHD work? Evidence for mediation by improved cognition. J. Child Psychol. Psychiatry 59, 1271–1281.
Huang, Y.S., Chao, C.C., Wu, Y.Y., Chen, Y.Y., Chen, C.K., 2007. Acute effects of

methylphenidate on performance during the test of variables of attention in children with attention deficit/hyperactivity disorder. Psychiatry Clin. Neurosci. 61 (3), 219–225.
Kato, Y., Endo, H., Kizuka, T., 2009. Mental fatigue and impaired response processes: event-related brain potentials in a Go/NoGo task. Int. J. Psychophysiol. 72 (2), 204–211.
Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., Williamson, P., Ryan, N., 1997. Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. J. Am. Acad. Child Adolesc. Psych. 36, 980–988.
Kritchman, M., Koubi, M., Mimouni Bloch, A., Bloch, Y., 2019. Effect of Methylphenidate On State Anxiety In Children With ADHD-A Single Dose, Placebo Controlled, Crossover Study. Frontiers in behavioral neuroscience 13, 106.
Langleben, D.D., Monterosso, J., Elman, I., Ash, B., Krikorian, G., Austin, G., 2006. Effect of methylphenidate on Stroop Color–Word task performance in children with attention deficit hyperactivity disorder. Psychiatry Res. 141 (3), 315–320.
Leark, R.A., Wallace, D.R., Fitzgerald, R., 2004. Test-retest reliability and standard error of measurement for the test of variables of attention (TOVA) with healthy school-age children. Assessment 11 (4), 285–289.
Li, D., Sham, P.C., Owen, M.J., He, L., 2006. Meta-analysis shows significant association between dopamine system genes and attention deficit hyperactivity disorder (ADHD). Hum. Mol. Genet. 15 (14), 2276–2284.
Llorente, A.M., Amado, A.J., Voigt, R.G., Berretta, M.C., Fraley, J.K., Jensen, C.L., Heird, WC., 2001. Internal consistency, temporal stability, and reproducibility of individual index scores of the Test of Variables of Attention in children with attention-deficit/hyperactivity disorder. Arch. Clin. Neuropsychol. 16 (6), 535–546.
Lowe, N., Barry, E., Gill, M., Hawi, Z., 2006. An overview of the pharmacogenetics and molecular genetics of ADHD. Curr. Pharmacogenomics 4 (3), 231–243.
Maoz, H., Aviram, S., Nitzan, U., Segev, A., Bloch, Y., 2018. Association between continuous performance and response inhibition tests in adults with ADHD. J. Atten. Disord. 22 (3), 293–299.
Moshe, K., Karni, A., Tirosh, E., 2012. Anxiety and methylphenidate in attention deficit hyperactivity disorder: a double-blind placebo-drug trial. ADHD Attention Deficit and Hyperactivity Disorders 4 (3), 153–158.
Nigg, J.T., Stavro, G., Ettenhofer, M., Hambrick, D.Z., Miller, T., Henderson, J.M., 2005. Executive functions and adhd in adults: Evidence for selective effects on ADHD symptom domains. J. Abnorm. Psychol. 114 (4), 706.
Owen, A.M., McMillan, K.M., Laird, A.R., Bullmore, E., 2005. N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. Hum. Brain Mapp. 25 (1), 46–59.
Pasini, A., Paloscia, C., Alessandrelli, R., Porfirio, M.C., Curatolo, P., 2007. Attention and executive functions profile in drug naive ADHD subtypes. Brain Dev. 29 (7), 400–408.
Peterson, B.S., Potenza, M.N., Wang, Z., Zhu, H., Martin, A., Marsh, R., et al., 2009. . An FMRI study of the effects of psychostimulants on default-mode processing during Stroop task performance in youths with ADHD. Am. J. Psychiatry 166 (11), 1286–1294.
Raven, J.C., Foulds, G., Forbes, A., 1973. Advanced Progressive Matrices: Plan and Use of the Scale with a Report of Experimental Work. Lewis, London.
Raz, S., Bar-Haim, Y., Sadeh, A., Dan, O., 2014. Reliability and validity of the online continuous performance test among young adults. Assessment 21 (1), 108–118.
Riccio, CA, Waldrop, JJ, Reynolds, CR, Lowe, P., 2001. Effects of stimulants on the continuous performance test (CPT) implications for CPT use and interpretation. J. Neuropsychiatry Clin. Neurosci. 13 (3), 326–335.
Robertson, I.H., Manly, T., Andrade, J., Baddeley, B.T., Yiend, J., 1997. Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. Neuropsychologia 35 (6), 747–758.
Rapoport, J.L., Buchsbaum, M.S., Weingartner, H., Zahn, T.P., Ludlow, C., Mikkelsen, E.J., 1980. Dextroamphetamine: its cognitive and behavioral effects in normal and hyperactive boys and normal men. Arch. Gen. Psychiatry 37 (8), 933–943.
Rubia, K., Smith, A., Taylor, E., 2007. Performance of children with attention deficit hyperactivity disorder (ADHD) on a test battery of impulsiveness. Child Neuropsychology 13 (3), 276–304.
Sagvolden, T., Johansen, E.B., Aase, H., Russell, V.A., 2005. A dynamic developmental theory of attention-deficit/hyperactivity disorder (ADHD) predominantly hyperactive/impulsive and combined subtypes. Behav. Brain Sci. 28 (3), 397–418.
Seidman, L.J., 2006. Neuropsychological functioning in people with ADHD across the lifespan. Clin. Psychol. Rev. 26 (4), 466–485.
Stroop, J.R., 1935. Studies of interference in serial verbal reactions. J. Exp. Psychol. 18 (6), 643.
Uebel, H., Albrecht, B., Asherson, P., Börger, N.A., Butler, L., Chen, W., Christiansen, H., et al., 2010. . Performance variability, impulsivity errors and the impact of incentives as gender-independent endophenotypes for ADHD. J. Child Psychol. Psychiatry 51 (2), 210–218.
Van der Oord, S., Prins, P.J., Oosterlaan, J., Emmelkamp, P.M., 2008. Efficacy of methylphenidate, psychosocial treatments and their combination in school-aged children with ADHD: a meta-analysis. Clin. Psychol. Rev. 28 (5), 783–800.
Van Mourik, R., Oosterlaan, J., Sergeant, J.A., 2005. The Stroop revisited: A meta-analysis of interference control in AD/HD. J. Child Psychol. Psychiatry 46 (2), 150–165.
Vaurio, R.G., Simmonds, D.J., Mostofsky, S.H., 2009. Increased intra-individual reaction time variability in attention-deficit/hyperactivity disorder across response inhibition tasks with different cognitive demands. Neuropsychologia 47 (12), 2389–2396.
Wang, C., Ding, M., Kluger, B.M., 2014. Change in intraindividual variability over time as a key metric for defining performance-based cognitive fatigability. Brain Cogn. 85, 251–258.
Watson, D., Clark, L.A., Tellegen, A., 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. J. Pers. Soc. Psychol. 54 (6), 1063.
Wood, S., Sage, J.R., Shuman, T., Anagnostaras, S.G., 2014. Psychostimulants and cognition: a continuum of behavioral and cognitive activation. Pharmacol. Rev. 66 (1), 193–221.
Wright, L., Lipszyc, J., Dupuis, A., Thayapararajah, S.W., Schachar, R., 2014. Response inhibition and psychopathology: A meta-analysis of go/no-go task performance. J. Abnorm. Psychol. 123 (2), 429.