ADHD Attention Deficit and Hyperactivity Disorders (2019) 11:73–81
https://doi.org/10.1007/s12402-018-0278-5

ORIGINAL ARTICLE

Apparent lack of practice effects in the Test of Variables of Attention (TOVA) in adult ADHD

Ann Rotem[1] · Yaron Danieli[2] · Joseph Ben-Sheetrit[1] · Amit Bashari[4] · Pavel Golubchik[1,3] · Rachel Ben-Hayun[6] · Abraham Weizman[1,3,5] · Iris Manor[1,3]

Received: 24 May 2018 / Accepted: 23 October 2018
© Springer-Verlag GmbH Austria, part of Springer Nature 2018

Abstract

The test of variables of attention (TOVA) is a continuous performance test commonly used as an aid for diagnosis of ADHD and assessment of treatment response. It has been studied and standardized in both children and adults. As a repetitive measurement of treatment efficacy, used both in research and in the clinic, it's important to disprove a practice effect. A retrospective cohort analysis was done, using only the placebo-arm participants from two different randomized, multicenter, double-blind clinical trials on the efficacy of a non-stimulant (metadoxine-XR). In order to reveal the practice effects, only the participants that showed no placebo effect (< 25% improvement), in the Conners' Adult ADHD Rating Scale–investigator rated (CAARS-Inv), the gold standard, were included. Demographic data, CAARS-Inv baseline and TOVA results during each visit were recorded and analyzed. Ninety-one participants from two studies were pooled (2014 $n = 24$, 2016 $n = 67$). They did not differ significantly in any demographic parameter, most side effect frequencies, and CAARS-Inv baseline scores. The baseline TOVA performances demonstrated similarity in the degree of inattention, variability, impulsivity, and response time. The TOVA scores were not altered significantly between visits, as assessed by repeated-measures analysis of variance. No significant differences were detected between the TOVA baseline-to-endpoint scores as assessed by paired *t* test. No practice effects were detected, in both clinical trials, suggesting that the results of the TOVA are likely to represent genuine changes in attentional performance. Further studies are needed to replicate these findings.

Keywords ADHD · Adults · TOVA · Practice effect · CPT

Introduction

Test of variables of attention (TOVA)

Continuous performance tests (CPTs) are frequently used as assessment tools of attention and impulsivity. The TOVA is among the most commonly used CPTs and has been studied and normed in both children and adults (Greenberg and Waldman 1993; Greenberg et al. 1994). The TOVA is used as an aid for ADHD diagnosis (Forbes 1998) as well as for the assessment of treatment response. TOVA performance correlates with other ADHD measurements, yet provides a different perspective of the ADHD presentation (Ben-Sheetrit et al. 2018; Forbes 1998). The TOVA has been extensively normed and found to have sensitive reaction time (RT) measures and high sensitivity to alterations in performance associated with therapeutic intervention (Greenberg and Waldman 1993).

It is a submission for the Special Issue "65th birthday Jan Buitelaar."

✉ Iris Manor
dr.iris.manor@gmail.com

1 Geha Mental Health Center, 1 Helsinki St, P.O. Box 102, 4910002 Petah Tikva, Israel

2 Yissum Research Development Company, The Hebrew University of Jerusalem, Jerusalem, Israel

3 Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel

4 Bar Ilan University, Ramat Gan, Israel

5 Felsenstein Medical Research Center, Petah Tikva, Israel

6 Cognitive Neurology Institute, Rambam Health Care Campus, Haifa, Israel

The main indices of the TOVA include the omission standard score (O-SS), a measure of inattention; the commission standard score (C-SS), considered as a measure of response inhibition or impulsivity; the response time standard score (RT-SS), considered as a measure of information processing and motor response speed; the RT variability standard score (RTV-SS), calculated as the standard deviation of RT, considered as a measure of consistency or variability of attention; the *d′* prime standard score (*d′* prime—SS) or response sensitivity, considered as a measure of attentional performance decrement, or the rate of deterioration of attentional performance over time; and the attention performance index (API) or Attention Comparison Score (ACS) according to the TOVA version, a measure of the subject's overall performance on the TOVA compared to other individuals diagnosed with ADHD (Greenberg and Waldman 1993).

Defining practice effects

Repeated neuropsychological tests are frequently used as a tool to measure change in performance across at least two test administrations in both clinical and research settings (Loring and Meador 2004). A difference between two scores of the same test over time may reflect a true change in performance, a genuine psychological or physiological effect (e.g., placebo effect) or it can be a result of previous exposure to the test (e.g., practice effects) (Heilbronner et al. 2010).

According to Goldberg et al. (2015), practice effects may result from task familiarity or practice-related effects. The first component, task familiarity, is consisted of task comprehension, sequence knowledge and stimulus response mapping. The second component, a practice-related effect, is ascribed to direct familiarity with a specific item that leads to an improvement in performance across test experiences (Goldberg et al. 2015).

Evidence suggests that practice effect sizes vary not only as a function of the test itself, but also as a function of population heterogeneity (e.g., age, clinical diagnosis) and testing factors (e.g., test–retest interval, the use of alternate forms) (Bird et al. 2003). For the purpose of quantifying the magnitude of practice effects on commonly used neuropsychological tests, Calamia et al. (2012) conducted a meta-analysis of nearly 1600 individual effect sizes and revealed several factors that were associated with practice effect sizes. They found that age is negatively associated with practice effect magnitude. With respect to clinical studies, the results indicated a general pattern of smaller practice effect sizes in clinical groups, compared to healthy controls. However, this association is moderated by the type of test and the type of the clinical group (Calamia et al. 2012). These findings emphasize the importance of awareness to the influence of practice effects on measurement tools that assess response to medications. However, to our knowledge, there is no established consensus on what should count as practice effects for CPTs. Moreover, there is a lack of data on the existence or absence of a practice effect of the TOVA. Hence, we chose a rather strict criterion by which a change of ≥ 1 standard deviation in any index of the TOVA was considered as an indicator of clinically significant practice effect for that index. Since we could not preclude the possibility that a practice effect may be limited to some, but not all, indices of the TOVA, we addressed each index, as well as the ACS, separately with regard to the absence or presence of a practice effect.

The present study was aimed to elucidate the possible existence of practice effects in TOVA. The first hypothesis of the study was that the TOVA is devoid of practice effects due to its unsophisticated, repetitive nature. The second hypothesis assumed that a putative practice effect would vary among the TOVA indices due to the different domains that they measure.

Method

Population

A retrospective cohort analysis was done, using only the participants of the placebo arm, who had no response according to the established measure CAARS-Inv (Conners et al. 1999). The participants were pooled from two different studies: a randomized, multicenter, double-blind study of a non-stimulant, metadoxine-XR, 1400 mg/day study, conducted in 2014, and a randomized, multicenter, double-blind study of metadoxine-XR 1400 mg/day study conducted in 2016. Both studies were performed in adults. The studies differed in several ways (see details below), including the number of TOVA tests per patient, and hence, the results of each study were analyzed separately. There was no population overlap in the two studies.

The 2014 study was a 6-week randomized, multicenter, double-blind study of metadoxine-XR 1400 mg/day in adults with ADHD (Weisler et al. 2014). Twenty-four adults, of the 140 completers in the 2014 study, were included in the present study (12 females) aged 18–55 (mean 32.5 ± 9.8) years. They completed the entire 6-week study period in the placebo arm and did not respond to metadoxine-XR. The 2016 study was a 10-week randomized, multicenter, double-blind, parallel, fixed-dose study of metadoxine-XR 1400 mg/day in adults with ADHD (unpublished study). Of the 139 study completers, 67 adults were included in the current study. *M*/*F* ratio 36:31, aged 18–55 (mean 37.3 ± 10.9) years, who completed the entire 10-week placebo arm of the study and did not respond to metadoxine-XR.

The two studies were approved by the Institutional Review Board (IRB) of all the centers involved, including the Geha Mental Health Center (Petah Tikva, Israel), and all participants provided informed consent to participate in the study. The main inclusion criterion was a DSM-IV-TR and DSM-5 diagnosis of ADHD as assessed by the Adult ADHD Clinician Diagnostic Scale (ACDS version 1.2) modified for DSM-IV and DSM-5, with at least moderate clinical severity (Clinical Global Impression-Severity score ≥ 4 and CAARS-Inv ≥ 22 in the 2014 study and CAARS-Inv ≥ 24 in the 2016 study). Main exclusion criteria were reported previously (Weisler et al. 2014).

Placebo profile definition

In an attempt to isolate the practice effects, the studied population included only participants who received placebo (no medication effect) and did not exhibit a clinical placebo effect according to the CAARS, an internationally accepted measure of ADHD evaluation and response monitoring (no placebo effect). In these conditions, the practice effects, if exist, could be identified.

In order to minimize any CAARS-Inv placebo effect, the placebo response was defined as a mild improvement, namely ≥ 25% improvement in the CAARS-Inv scores, the gold standard of monitoring response of ADHD symptoms to treatment, at the endpoint (Ben-Sheetrit et al. 2018). Indeed, it's not guaranteed that a "no placebo" response according to the CAARS-Inv indicates a "no placebo" response according to the TOVA. However, a thorough search of the literature [MEDLINE (http://www.pubmed.com), Google Scholar (http://scholar.google.com), and PsycINFO (http://www.apa.org/pub/databases/psycinfo/index.aspx)] failed to find any studies regarding a possible correlation of the CAARS-Inv-placebo response and a CPT (and more specifically the TOVA)-placebo response. Hence, it was assumed that the same participants will respond to a similar extent in both measures but due to no better optional choice, this was used as the rule, due to the lack of data on the placebo effect in ADHD in the current available literature. Also, this population was selected since it was considered a relatively pure non-placebo population, in an attempt to estimate practice effects that would not be biased by a possible placebo effect of the TOVA.

The methods of these two studies were different in several ways (Weisler et al. 2014). In both, the Conners' Adult ADHD Rating Scale–Investigator rated (CAARS-Inv) was recorded at baseline and its improvement was used as a primary endpoint. However, in the 2014 study the timing of the endpoint was determined 6 weeks after the initiation of medication/placebo, while in the 2016 study—it was determined after 10 weeks. The TOVA was used in both, but in 2014 study it was conducted 4 times: baseline, at week 1, at week 6 (the final visit) and at the follow-up visit, which was 2 weeks later, while in the 2016 study the TOVA was conducted 3 times: at baseline, at week 4, and at week 10 (the final visit).There was no TOVA assessment in the follow-up visit in the 2016 study. Due to these differences, the data from the two studies were analyzed separately.

Another difference is the change in version of the TOVA between these years that enabled in the new version recording of absolute numeric results instead of a threshold cut-off score in the previous TOVA version. There were no other significant differences as to the clinical analysis of the results.

Statistical analysis

The pre-specified endpoints were TOVA performance at the endpoint of each study: week 6 and week 10 for study 2014 and study 2016, respectively. Paired-repeated samples *t* test was used in order to assess the statistical significance of the changes. In addition, one-way ANOVA-RM was performed in order to evaluate changes in TOVA measures between visits. Analysis was performed using SPSS for Windows ver. 22 (IBM Inc., Chicago, IL, USA). For statistical power calculations, we used G*Power ver. 3.1. Results are presented as number or mean accompanied by standard deviations (SD) or rates (%) as appropriate. $p < .05$ was considered significant.

Results

Demographic and clinical characteristics

As can be seen in Table 1, participants from the 2014 study ($n = 24$) and the 2016 study ($n = 67$) did not differ significantly in any demographic parameter (age $p = .063$, gender $p = .753$). Around half of the participants in both studies were treatment naive ($p = .971$) Side effect rates were similar in both studies and did not reach a significant level, with the exception of fatigue ($p = .036$), which was more pronounced (20.8% vs. 6.0%) in the 2014 study.

Interestingly, the change in the CAARS-Inv scores from baseline to endpoint was numerically statistically significant in both studies, especially in the 2016 study, although the magnitude of this response, as was pre-defined, was clinically meaningless (less than 25% improvement) (Table 2).

TOVA data analyses

In order to evaluate potential practice effects in the TOVA, participants' performance on each index was evaluated. The two groups did not differ significantly in the TOVA baseline performance (data not shown) both were below

Table 1 Demographic and clinical characteristics of the study samples

Demographics	2014 study ($n=24$)	2016 study ($n=67$)	p t test/χ^2
Age[a], M (SD; years)	32.5 (9.8)	37.3 (10.9)	$p=.063$
Gender, n (%)			$p=.753$
Male	12 (50%)	36 (53.7%)	
Female	12 (50%)	31 (46.3%)	
Previous treatment, n (%)	11 (45.8%)	31 (46.3%)	$p=.971$
Side effects, n (%)			
Nasopharyngitis	2 (8.3%)	2 (3.0%)	$p=.273$
Diarrhea	–	2 (3.0%)	$p=.392$
Headache	5 (20.8%)	5 (7.5%)	$p=.072$
Nausea	–	2 (3.0%)	$p=.392$
Fatigue	5 (20.8%)	4 (6.0%)	$p=.036$
CAARS baseline[a]	32.8 (8.0)	35.9 (7.6)	$p=.102$
ASRS baseline[a]	40.8 (9.8)	46.3 (9.7)	$p=.019$

CAARS-Inv, Conners' Adult ADHD Rating Scale–Investigator rating; ASRS, Adult ADHD Rating Scale

[a]Mean with (standard deviation in parentheses)

Table 2 CAARS-Inv scores at baseline and endpoint in the two studies

Measure	2014 study ($n=24$)	2016 study ($n=67$)	Between studies p t *test*
Baseline[a]	32.8 (8.0)	35.9 (7.6)	$p=.102$
End point[a]	30.6 (7.0)	34.0 (8.9)	$p=.092$
p t test[b]	$p=.041$	$p=.001$	

CAARS-Inv, Conners' Adult ADHD Rating Scale–Investigator rating; ASRS, Adult ADHD Rating Scale

[a]Mean with (standard deviation in parentheses)

[b]Paired sample two-tailed t test of the difference between baseline and end point

the normal range, but the magnitude of the deficit was much larger in the 2016 study due to a change in the TOVA scoring method. Previously, large negative standard scores (> 4 SD) were presented by the TOVA simply as "< 40," as it was done in the study of 2014, or in its later version, "< 0," without the specific negative value. It presented a problem for researchers, because this range (< 0) did not allow a calculation of the true mean score, and in the 2016 study the TOVA company was requested by the researchers to calculate the actual standard scores (SS) values using the raw score. However, calculating the actual mean score affected largely the SS, especially in the omission errors index in adults, since in the TOVA norming study, most adults did not exhibit any omission errors at all (http://files.tovatest.com/documentation/9/Professional%20Manual.pdf)). In order to reduce the discrepancy, and have a common ground, and after a consultation with the TOVA company scientists, all the scores below 40 were re-defined as 40, in the same way that it was done in the 2014 study. Thus, the results became comparable. Furthermore, any value that was greater than 4 SD was considered as highly abnormal, and hence the actual number seemed to be irrelevant.

For the sake of accuracy and transparency, the analyses of the original actual values are shown as an appendix.

It appears that the attentional functioning, as captured by the TOVA, reflects a similar population: the O-SS, RTV-SS and D' prime-SS were abnormal all along the study, while the C-SS (which is considered an indicator of impulsivity) was on the borderline verge and the RT-SS was normal, though in the low normal range. The mean API/ACS in both studies was also abnormal (< 0) (Tables 3, 4).

The first analysis compared the TOVA results in each visit. Since the 2014 study included four visits in which the test was conducted, and the 2016 study included only three, their results were separated (Tables 3, 4). They were analyzed using the one-way ANOVA-RM. In both studies, no statistically significant differences were detected in any of the indices throughout the placebo treatment, even when the actual data from 2016 were analyzed (Table 7, "Appendix 1"). It is of note that the scores in the follow-up visit in the 2014 study were very similar to the ones in visit 6 (the endpoint study). It means that neither repeated testing nor becoming aware to the discontinuation of any treatment altered the TOVA performance, indicating a lack of a practice, as well as placebo effect.

The second analysis compared the change from baseline-to-endpoint scores, which is usually the main outcome measure in efficacy studies. It was analyzed by paired-repeated t test. As shown in Table 5, in the 2014 study, all TOVA indices were not altered significantly from baseline to endpoint. Similar nonsignificant alterations were observed in the 2016 study (Tables 5, 8, "Appendix 2").

Table 3 Descriptive statistics by week of the 2014 study ($n=24$)

TOVA Index	Baseline	Week 2	Week 6	Follow-up	ANOVA-RM F	$1-\beta$	Partial η^2	p value
ACS	−3.5 (7.9)	−2.9 (8.2)	−3.4 (9.1)	−3.9 (8.4)	0.76	.49	.032	.483
O-SS	75.1 (28.3)	76.1 (28.9)	74.4 (29.9)	77.8 (27.1)	0.38	.66	.021	.655
C-SS	87.5 (24.4)	93.5 (21.4)	91.7 (24.5)	95.1 (21.5)	2.25	.16	.097	.108
RTV-SS	78.3 (28.7)	81.7 (27.3)	79.3 (29.0)	73.7 (27.8)	0.14	.88	.008	.881
RT-SS	99.3 (23.7)	99.0 (22.3)	98.6 (20.7)	96.8 (20.5)	0.11	.91	.005	.912
D prime-SS	75.1 (23.5)	80.7 (25.4)	78.8 (25.2)	79.0 (27.0)	1.1	.36	.049	.337

ANOVA-RM, one-way analysis of variance with repeated measures; ACS norm > 0; Omission, Commission, RT and RT variability norms > 85; ACS, Attention Comparison Score; O-SS, Omission Errors Standard Score; C-SS, Commission Errors Standard Score; RTV-SS, Response Time Variability Standard Score; RT-SS, Response Time Standard Score

[a]Mean with (standard deviation in parentheses)

Table 4 Descriptive statistics by week of the 2016 study ($n=67$)

TOVA Index	Baseline	Week 4	Week 10	ANOVA-RM F	$1-\beta$	Partial η^2	p
ACS	−2.8 (6.2)	−2.2 (5.8)	−2.8 (5.3)	0.61	.51	.009	.508
O-SS	71.6 (28.2)	71.0 (28.6)	68.2 (28.8)	0.95	.75	.014	.390
C-SS	87.8 (23.3)	91.8 (25.6)	91.3 (25.4)	1.71	.75	.026	.185
RTV-SS	72.9 (26.6)	75.0 (24.9)	69.8 (25.4)	2.03	.74	.030	.140
RT-SS	99.6 (28.3)	98.7 (24.9)	96.5 (25.5)	1.12	.74	.017	.328
D prime-SS	75.3 (25.5)	79.7 (27.7)	75.2 (26.9)	1.27	.74	.019	.283

ANOVA-RM, one-way analysis of variance with repeated measures; ACS norm > 0; Omission, Commission, RT and RT variability norms > 85; ACS, Attention Comparison Score; O-SS, Omission Errors Standard Score; C-SS, Commission Errors Standard Score; RTV-SS, Response Time Variability Standard Score; RT-SS, Response Time Standard Score

[a]Mean with (standard deviation in parentheses)

Table 5 The changes in the various TOVA indices from baseline to endpoint in the two studies

Measure	2014 study ($n=24$)			2016 Study ($n=67$)		
	Mean differences[a]	95% CI*	p value[b]	Mean differences[a]	95% CI*	p value[b]
ACS	−0.04 (3.8)	[−1.7, 1.6]	.957	0.01 (5.4)	[−1.3, 1.3]	.994
O-SS	0.71 (27.2)	[−11.7, 13.1]	.905	3.39 (21.9)	[−2.0, 8.8]	.213
C-SS	−4.18 (15.0)	[−10.8, 2.5]	.206	−3.4 (21.4)	[−8.7, 1.8]	.197
RTV-SS	−1.05 (17.0)	[−8.8, 6.7]	.781	3.15 (24.3)	[−2.8, 9.1]	.293
RT-SS	0.74 (13.1)	[−4.9, 6.4]	.789	3.09 (19.6)	[−1.7, 7.9]	.203
D prime-SS	−3.69 (16.0)	[−10.6, 3.2]	.280	0.07 (25.6)	[−6.2, 6.3]	.983

ACS, Attention Comparison Score; O-SS, Omission Errors Standard Score; C-SS, Commission Errors Standard Score; RTV-SS, Response Time Variability Standard Score; RT-SS, Response Time Standard Score

*95% confidence interval

[a]Mean with (standard deviation in parentheses)

[b]Paired sample two-tailed t test of the difference between baseline and end point performance for each measure

It is of note that no practice effects were detected in any of the TOVA indices. Namely, no significant alterations in the various indices were obtained during the repeated exposures to the TOVA (Tables 3, 4, 5, 7, 8).

In order to eliminate the possibility that the lack of difference was due to sample size, we conducted a third analysis, comparing the change from baseline-to-endpoint scores using combined data from both studies ($n=91$). As can be

Table 6 The changes in the various TOVA indices from baseline to endpoint in a combined sample of both studies ($n=91$)

TOVA Index	Mean differences[a]	95% CI*	t[b]	DF	Power $(1-\beta)$	Cohen's d	p
ACS	−0.01 (5.0)	[−1.0, 1.0]	−0.01	90	.99	−0.001	.988
O-SS	2.8 (23.2)	[−2.2, 7.7]	1.11	86	.51	0.12	.270
C-SS	−3.6 (19.9)	[−7.9, 0.6]	−1.71	87	.50	−0.18	.090
RTV-SS	2.1 (22.8)	[−2.7, 7.0]	0.89	87	.53	0.09	.379
RT-SS	2.5 (18.2)	[−1.3, 6.3]	1.30	89	.51	0.14	.197
D prime-SS	−0.9 (23.5)	[−5.8, 4.0]	−0.36	89	.74	−0.04	.719

ACS, Attention Comparison Score; O-SS, Omission Errors Standard Score; C-SS, Commission Errors Standard Score; RTV-SS, Response Time Variability Standard Score; RT-SS, Response Time Standard Score

*95% confidence interval

[a]Mean with (standard deviation in parentheses)

[b]Paired sample two-tailed t test of the difference between baseline and end point performance for each measure

seen in Table 6, paired sample t test did not yield any significant results. Effect sizes for all indices were small, ranging from −0.18 to 0.14 (Table 6). The effect sizes described in Table 6 were similar to those obtained using the actual values of 2016 (Table 9, "Appendix 3").

Discussion

In the present study, we examined a possible practice effect in repeated administrations of the TOVA in a group of placebo non-responders with ADHD. It is considered important, since the TOVA is used to measure cognitive response to ADHD treatments, and as such, a putative practice effect could weaken its utility. We found that the TOVA lacks a practice effect both in its composite score (the ACS) and in every single index. The literature regarding other CPTs either points to a clear practice effects (e.g., the MOXO CPT, see Shahaf et al. 2018) or is inconsistent in this regard (e.g., the Conners' CPT, see Chen et al. 2009, for some positive results versus Zabel et al. 2009, for negative results). The current study seems to point to a substantial advantage of the TOVA as an aid in diagnosis and monitoring of treatment response in adult ADHD.

It is needed to identify the neuropsychological mechanisms involved in practice effects in order to understand why they are apparently insignificant in the TOVA. Practice effects may stem from memory, learned strategies for solving problems, or general experience with testing (the so-called test sophistication, McCaffrey et al. 2000). Since TOVA stimuli are abstract, repetitive and monotonous squares and the test lasts for 22 min, it seems unlikely that acquired memory or concrete problem-solving factors could contribute to improved performance in this test. Thus, the test itself seems to offer very little room for learning from previous administrations compared to, for example, measures of intelligence, the CPT-II (alphabet letters) and the MOXO (relatively "complex" animations in the visual parts). Moreover, the little that may perhaps be learned from experience with this test (e.g., inner strategies for better inhibitory control) may be ineffectively learned by subjects with ADHD, because of its dependence on working memory, which is very frequently impaired in this disorder. Indeed, a previous meta-analysis of many studies on numerous neuropsychological domains and tests has concluded that practice effects may not be simply a nuisance, but a useful source of information: The lack of practice effects predicts worse cognitive performance in subjects with cognitive decline (Calamia et al. 2012).

As an arbitrary decision, in order to try and certify accuracy, only the participants who received a placebo and were considered "non-responsive to a placebo effect" by the CAARS-Inv were selected to the analysis. This strategy was employed in an attempt to separate between the placebo effect and the practice effects in the TOVA. It was assumed that inclusion of all participants will add a possible placebo effect (about which the literature is very scarce) to a mixture of effects, and no precise discrimination could be done.

The population in both studies was similar. As shown in Table 1, fatigue was a more frequent side effect in the 2014 study. It is of note that the same symptom was reported frequently as a side effect in a previous study, with no significant difference between the medication and the placebo arms (Manor et al. 2012). It seems that fatigue could be a non-specific side effect without a significant clinical meaning.

The similarity between the populations persisted in the TOVA indices, which demonstrated the same attentional profile in both studies (2014, 2016): abnormal (dysfunctional) sustained attention (O), consistency and vigilance (RTV and *d′* prime) but borderline to normal impulsivity (C) and velocity (RT). It seems that in adults with ADHD, as they are represented by the participants, most of the impairment is derived from inattention and instability, much less from hyperactivity, which is in agreement with the literature (Wilens et al. 2009). This profile may explain the specific deviation in the O-SS, and its huge SD, since this index reflects a highly impaired component of the ADHD presentation in adults.

Clinical perspective

Clinical experience shows that many intelligent, well-compensated individuals are often baffled by their test's results, which tend to unmask much more severe deficits than they previously imagined. It is emphasized by the lack of practice effects. These patients believed that they have learned to adapt their life to the disorder by employing multiple strategies (e.g., meticulously crafted to-do lists) and using many cues (e.g., context awareness in conversations to infer what was previously said) that make life possible despite frequent absent mindedness and impairments in inhibitory control. Faced with a square on a screen with no context and no room for creative solutions, the true extent of their impairment is much more likely to surface, especially as they can't cope with it through learning processes. It is suggested that this is also the reason why clinical symptoms in everyday life are often at odds with CPT results conducted in the laboratory (Ben-Sheetrit et al. 2016). Careful consideration of the compensations and environmental accommodations that have developed in response to the patient's disorder is crucial to comprehensive assessment and treatment.

Research implications

Failing to take into account practice effects can compromise the validity of research findings (Calamia et al. 2012). Improvement in measures that correlate with the desired construct being tested can diminish after repeated testing (Calamia et al. 2012). Using measures and methods that minimize practice effects is therefore of great importance. In ADHD research, the gap between the TOVA, as an objective measure that seems to be devoid of practice effects, and items of symptomatic measures, is of considerable potential. An interesting question is the possibility of practice effects in other measures, such as self-reported questionnaires (like the CAARS, ASRS). It was not the topic of this study, but an important topic to investigate further on, as the significant placebo effect, which is found in these measures (Ben-Sheetrit et al. 2018) may be partly due to the practice effects. The inflated placebo response, which is currently detected in self-reported questionnaires (Ben-Sheetrit et al. 2018), may be related to combined placebo and practice effects. Looking for a placebo effect in the TOVA, or any other CPT with similar characteristics, is another topic of interest to study in adult ADHD research.

Limitations and conclusions

The current study has several limitations. First, the relatively small sample sizes of our samples limit the statistical power of the study and justify further studies in order to replicate the current findings. Second, some variables such as the ADHD presentations and the presence of comorbid disorders were not assessed, and this should be taken into account in future investigations. Moreover, possible selection bias due to inclusion and exclusion criteria in clinical trials is another concern that limits generalizability to real-world patients with ADHD. However, this study is the first to specifically assess possible practice effects of TOVA. The present study's findings suggest that the TOVA is devoid of such effects. If corroborated by future studies, this may have important clinical and research implications and relevance to diagnosis and treatment of adults with ADHD.

Compliance with ethical standards

Conflict of interest All authors declare that they have no conflict of interest regarding the current study.

Informed consent Informed consent was obtained from all individual participants included in the 2014, 2016 studies in all the centers that were involved in them.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

Appendix 1

Analysis of the actual original data that are shown in Tables 4 and 7.

Table 7 Descriptive statistics by week of the 2016 study ($n=67$)

TOVA Index	Baseline	Week 4	Week 10	ANOVA-RM F	$1-\beta$	Partial η^2	p
ACS	−2.8 (6.2)	−2.2 (5.8)	−2.8 (5.3)	0.61	.51	.009	.508
O-SS	−109.9 (591.1)	−103.8 (632.5)	−34.2 (227.5)	0.70	.50	.011	.498
C-SS	83.8 (36.3)	88.0 (37.6)	87.1 (34.1)	1.15	.32	.017	.316
RTV-SS	64.6 (43.4)	67.7 (40.0)	60.1 (39.1)	1.40	.26	.021	.251
RT-SS	99.3 (29.0)	98.2 (26.1)	94.8 (30.1)	1.71	.20	.025	.189
D prime-SS	68.6 (41.4)	72.6 (46.0)	68.3 (40.9)	0.76	.45	.011	.450

ANOVA-RM, one-way analysis of variance with repeated measures; ACS norm > 0; Omission, Commission, RT and RT variability norms > 85; ACS, Attention Comparison Score; O-SS, Omission Errors Standard Score; C-SS, Commission Errors Standard Score; RTV-SS, Response Time Variability Standard Score; RT-SS, Response Time Standard Score

[a]Mean with (standard deviation in parentheses)

Appendix 2

Analysis of the actual original data that are shown in Tables 5 and 8.

Appendix 3

Analysis of the actual original data that are shown in Tables 6 and 9.

Table 8 The changes in the various TOVA indices from baseline to endpoint in the two studies

Measure	2014 study ($n=24$)			2016 study ($n=67$)		
	Mean differences[a]	95% CI*	p value[b]	Mean differences[a]	95% CI*	p value[b]
ACS	−0.04 (3.8)	[−1.7, 1.6]	.957	0.01 (5.4)	[−1.3, 1.3]	.994
O-SS	0.71 (27.2)	[−11.7, 13.1]	.905	−75.70 (595.0)	[−220.8, 69.4]	.301
C-SS	−4.18 (15.0)	[−10.8, 2.5]	.206	−3.33 (27.5)	[−10.0, 3.4]	.326
RTV-SS	−1.05 (17.0)	[−8.8, 6.7]	.781	3.86 (38.7)	[−5.6, 13.3]	.418
RT-SS	0.74 (13.1)	[−4.9, 6.4]	.789	4.45 (23.7)	[−1.3, 10.2]	.130
D prime-SS	−3.69 (16.0)	[−10.6, 3.2]	.280	0.27 (32.2)	[−7.6, 8.1]	.945

ACS, Attention Comparison Score; O-SS, Omission Errors Standard Score; C-SS, Commission Errors Standard Score; RTV-SS, Response Time Variability Standard Score; RT-SS, Response Time Standard Score

*95% confidence interval

[a]Mean with (standard deviation in parentheses)

[b]Paired sample two-tailed t test of the difference between baseline and end point performance for each measure

Table 9 The Changes in the various TOVA indices from baseline to endpoint in a combined sample of both studies ($n = 91$)

TOVA Index	Mean differences[a]	95% CI*	t[b]	DF	Power ($1-\beta$)	Cohen's d	p
ACS	−0.01(5.0)	[−1.0, 1.0]	−0.01	90	0.98	−0.001	.988
O-SS	−57.5 (519.4)	[−167.5, 52.6]	−1.04	87	0.52	−0.111	.302
C-SS	−3.6 (24.9)	[−8.8, 1.7]	−1.34	88	0.51	−0.142	.183
RTV-SS	2.7 (34.8)	[−4.7, 10.1]	0.72	87	0.58	0.077	.471
RT-SS	3.5 (21.5)	[−1.0, 8.0]	1.54	89	0.50	0.163	.126
D prime-SS	−0.7 (28.9)	[−6.8, 5.3]	−0.24	89	0.81	−0.026	.807

ACS, Attention Comparison Score; O-SS, Omission Errors Standard Score; C-SS, Commission Errors Standard Score; RTV-SS, Response Time Variability Standard Score; RT-SS, Response Time Standard Score

*95% confidence interval

[a]Mean with (standard deviation in parentheses)

[b]Paired sample two-tailed t test of the difference between baseline and end point performance for each measure

References

Ben-Sheetrit J, Zurawel M, Weizman A, Manor I (2016) Symptoms versus impairment in adults with ADHD: intercorrelations of the BRIEF-A, CAARS, and TOVA. J Atten Disord. https://doi.org/10.1177/1087054716682334

Ben-Sheetrit J, Peskin M, Newcorn JH, Daniely Y, Shbiro L, Rotem A, Manor I (2018) Characterizing the placebo response in adults with ADHD. J Atten Disord. https://doi.org/10.1177/1087054718780328

Bird CM, Papadopoulou K, Ricciardelli P, Rossor MN, Cipolotti L (2003) Test-retest reliability, practice effects and reliable change indices for the recognition memory test. Br J Clin Psychol 42(4):407–425

Calamia M, Markon K, Tranel D (2012) Scoring higher the second time around: meta-analyses of practice effects in neuropsychological assessment. The Clin Neuropsychol 26(4):543–570

Chen HC, Koh CL, Hsieh CL, Hsueh IP (2009) Test–retest reliability of two sustained attention tests in persons with chronic stroke. Brain Inj 23(9):715–722

Conners CK, Erhardt D, Sparrow E (1999) Conners' ADULT ADHD Rating Scales (CAARS). Multi-Health Systems, New York

Forbes GB (1998) Clinical utility of the test of variables of attention (TOVA) in the diagnosis of attention-deficit/hyperactivity disorder. J Clin Psychol 54(4):461–476

Goldberg TE, Harvey PD, Wesnes KA, Snyder PJ, Schneider LS (2015) Practice effects due to serial cognitive assessment: implications for preclinical Alzheimer's disease randomized controlled trials. Alzheimer's Dement Diagn Assess Dis Monit 1(1):103–111

Greenberg LM, Waldman ID (1993) Developmental normative data on the test of variables of attention (TOVA™). J Child Psychol Psychiatry 34(6):1019–1030

Greenberg LM, Kindschi CL, Dupuy TR, Hughes SJ (1994) Test of variables of attention continuous performance test. The TOVA Company, Alamitos

Heilbronner RL, Sweet JJ, Attix DK, Krull KR, Henry GK, Hart RP (2010) Official position of the American Academy of Clinical Neuropsychology on serial neuropsychological assessments: the utility and challenges of repeat test administrations in clinical and forensic contexts. Clin Neuropsychol 24(8):1267–1278

Loring DW, Meador KJ (2004) Cognitive side effects of antiepileptic drugs in children. Neurology 62(6):872–877

Manor I, Ben-Hayun R, Aharon-Peretz J, Salomy D, Weizman A, Daniely Y, Adler LA (2012) A randomized, double-blind, placebo-controlled, multicenter study evaluating the efficacy, safety, and tolerability of extended-release metadoxine-XR in adults with attention-deficit/hyperactivity disorder. J Clin Psychiatry 73(12):1517–1523

McCaffrey RJ, Duff K, Westervelt HJ (eds) (2000) Practitioner's guide to evaluating change with neuropsychological assessment instruments. Springer, Berlin

Shahaf G, Nitzan U, Erez G, Mendelovic S, Bloch Y (2018) Monitoring attention in ADHD with an easy-to-use electrophysiological index. Front Hum Neurosci 12:32

Weisler R, Adler L, Rubin J, Daniely Y, Manor I (2014) A phase 3, randomized, double-blind, placebo-controlled study of metadoxine-XR extended release 1400 mg compared with placebo once daily in 300 adults with attention-deficit/hyperactivity disorder. In: Poster session presented at the annual convention of the american academy of child and adolescent psychiatry, San Diego, CA

Wilens TE, Biederman J, Faraone SV, Martelon M, Westerberg D, Spencer TJ (2009) Presenting ADHD symptoms, subtypes, and comorbid disorders in clinically referred adults with ADHD. J Clin Psychiatry 70(11):1557

Zabel TA, von Thomsen C, Cole C, Martin R, Mahone EM (2009) Reliability concerns in the repeated computerized assessment of attention in children. Clin Neuropsychol 23(7):1213–1231