





Research Article

About time: neurocognitive correlates of stimulus-bound and other time setting errors in the Clock Drawing Test

Matan Soffer¹ , Meryl A. Butters², Nathan Herrmann^{3,4,5}, Sandra E. Black^{3,5,6}, Sanjeev Kumar^{1,3,4} , Bradley Pugh^{3,7}, Tarek K. Rajji^{1,3,4}, Maria Carmela Tartaglia^{3,6,8}, David F. Tang-Wai^{3,6,8} and Morris Freedman^{3,6,7}

¹Centre for Addiction and Mental Health, Toronto, Canada, ²Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA, USA, ³Toronto Dementia Research Alliance, Toronto, Canada, ⁴Department of Psychiatry, University of Toronto, Toronto, Canada, ⁵Sunnybrook Health Sciences Centre, Toronto, Canada, ⁶Department of Medicine (Neurology), University of Toronto, Toronto, Canada, ⁷Rotman Research Institute, Baycrest Health Sciences Centre, Toronto, ON, Canada and ⁸University Health Network Memory Clinic, Toronto Western Hospital, Toronto, ON, Canada

Abstract

Objective: Previous findings suggest that time setting errors (TSEs) in the Clock Drawing Test (CDT) may be related mainly to impairments in semantic and executive function. Recent attempts to dissociate the classic stimulus-bound error (setting the time to “10 to 11” instead of “10 past 11”) from other TSEs, did not support hypotheses regarding this error being primarily executive in nature or different from other time setting errors in terms of neurocognitive correlates. This study aimed to further investigate the cognitive correlates of stimulus-bound errors and other TSEs, in order to trace possible underlying cognitive deficits. **Methods:** We examined cognitive test performance of participants with preliminary diagnoses associated with mild cognitive impairment. Among 490 participants, we identified clocks with stimulus-bound errors ($n = 78$), other TSEs ($n = 41$), other errors not related to time settings ($n = 176$), or errorless clocks ($n = 195$). **Results:** No differences were found on any dependent measure between the stimulus-bound and the other TSEs groups. Group comparisons suggested TSEs in general, to be associated with lower performance on various cognitive measures, especially on semantic and working memory measures. Regression analysis further highlighted semantic and verbal working memory difficulties as being the most prominent deficits associated with these errors. **Conclusion:** TSEs in the CDT may indicate underlying deficits in semantic function and working memory. In addition, results support previous findings related to the diagnostic value of TSEs in detecting cognitive impairment.

Keywords: Clock Drawing Test; mild cognitive impairment; semantic function; executive function; working memory

(Received 20 July 2023; final revision 11 November 2023; accepted 13 November 2023)

Introduction

One of the challenges facing a patient when asked to “draw the face of a clock, put in all the numbers and set the hands to 10 after 11” is setting the hands to indicate the correct time. Following brain damage due to various pathologies, setting the correct time poses a challenge to such an extent that time setting errors (TSEs) have proven to be a significant feature in assessing dementia and its prodromal stages using the Clock Drawing Test (CDT) (Berger et al., 2008; Duro et al., 2018; Freedman et al., 1994; Lessig et al., 2008; Ricci et al., 2016).

Despite the proposed diagnostic value of TSEs, little is known about the cognitive processes that may lead to their occurrence. This discrepancy is conspicuous especially in light of the ample theoretical literature regarding putative cognitive mechanisms that may underlie specific error types in the CDT (e.g., Freedman et al., 1994; Kaplan, 1988; Rouleau et al., 1992; Eknoyan et al., 2012). Tracing compromised functions that may cause distinct error types could assist clinicians in using the CDT to identify specific

neurocognitive deficits. A considerable body of literature suggests that a qualitative approach to interpreting the CDT may have a diagnostic value in differentiating among various pathologies (Duro et al., 2018; Tan et al., 2015). Thus, validating existing classification systems of error types may increase the usefulness and accuracy of such qualitative interpretation.

One type of TSEs has been given special attention from a theoretical perspective. Setting the time to “10 minutes to 11” when instructed to set the time at “10 after 11” is thought to result from being “pulled” to the salient stimulus “10,” instead of the more complex process of translating the concept of “10 minutes past” to be represented by the digit “2” (Freedman et al., 1994; Kaplan, 1988). This error has been called a stimulus-bound error (Rouleau et al., 1992). This classic account suggests a deficit in abstract thinking (Cahn et al., 1996; Freedman et al., 1994), or compromised inhibitory control (Soffer et al., 2022; for the construct of inhibitory control, see Diamond, 2013) as possible cognitive mechanisms that may underlie this error type.

Corresponding author: M. Soffer; Email: matan.soffer@gmail.com

Cite this article: Soffer M., Butters M.A., Herrmann N., Black S.E., Kumar S., Pugh B., Rajji T.K., Tartaglia M.C., Tang-Wai D.F., & Freedman M. About time: neurocognitive correlates of stimulus-bound and other time setting errors in the Clock Drawing Test. *Journal of the International Neuropsychological Society*, 1–8, <https://doi.org/10.1017/S1355617723011396>

© The Author(s), 2023. Published by Cambridge University Press on behalf of International Neuropsychological Society. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Another mechanism which was proposed to underlie TSEs in the CDT is a deficit in conceptual processes related to retrieval of semantic knowledge, as setting the time requires the individual to rely on their previous understanding of how time is represented by the unique conventional code of the analog clock (e.g., Freedman *et al.*, 1994; Rouleau *et al.*, 1992, 1996). Support for this account was found in a study in a brain injured population which showed that TSEs were associated with lower performance on semantic and verbal tasks as well as left hemisphere lesions (Tranel *et al.*, 2008). In Alzheimer's disease (AD) and related disorders, there is indirect support for the involvement of semantic deficit in the commission of TSEs. First, in AD, which is characterized by prominent pathology in the temporal lobe association cortices and semantic deficits, stimulus-bound errors were consistently found to be more prevalent than in other dementia types (Duro *et al.*, 2018; Tan *et al.*, 2015). Moreover, it was suggested that in AD, conceptual deficits, rather than pure executive difficulties may underlie stimulus-bound errors. However, this hypothesis was not directly tested (see discussion sections in Rouleau *et al.*, 1996, Blair *et al.*, 2006). In line with this hypothesis, Duro *et al.* (2019) found stimulus-bound errors to correlate with pathology in temporal and left frontal brain regions, similarly to other errors thought to involve conceptual or semantic deficit.

Limited research exists on attempts to correlate different CDT error types with specific neuropsychological processes and domains in AD and its prodromal stages, or other types of dementia (Parsey & Schmitter-Edgecombe, 2011; Cosentino *et al.*, 2004; Rouleau *et al.*, 1992). One of the challenges in interpreting such results in the context of TSEs is related to Rouleau's qualitative classification system, which is the most prevalent and influential qualitative approach (Eknoyan *et al.*, 2012; Spenciere *et al.*, 2017). Rouleau's categorization (Rouleau *et al.*, 1992), which constituted the theoretical framework for most of these studies, does not include a distinct group of TSEs. Instead, it includes a separate category of stimulus-bound errors, while nonstimulus-bound TSEs such as placing the minute hand slightly after the 11 to indicate "10 after 11," would fall under a larger, more heterogeneous, and less defined group of "conceptual" errors. Previous attempts to validate the constructs of stimulus-bound errors or other related error types (e.g., "conceptual" errors, time representation errors) by correlating them with specific neuropsychological processes and domains in dementia and its prodromal stages were inconclusive (Cosentino *et al.*, 2004; Rouleau *et al.*, 1992; Parsey & Schmitter-Edgecombe, 2011; Umegaki *et al.*, 2021).

In a previous study, we found that stimulus-bound errors and other TSEs were associated with lower scores on semantic and executive measures without differences between stimulus-bound errors and other TSEs on those measures (Soffer *et al.*, 2022). In that study, the stimulus-bound error group was compared to an equivalent group of other TSEs in which the hour or minute target number were erroneous. This classification allows for a direct comparison of two error types for which, the classification criterion differs only in the essential defining feature of stimulus-bound errors – representing the concept of "10 after" by the digit "10," culling out clocks that indicate the correct target numbers. Despite the methodological advantages of this approach in direct comparison between stimulus-bound errors and other equivalent TSEs, the interpretation of the results was cautioned due to the small number of clocks containing these errors, and the heterogeneous sample of patients.

The aim of the current study was to further investigate the neuropsychological correlates of stimulus-bound errors and other

TSEs on the CDT using a large sample of participants with mild cognitive impairment (MCI)-related disorders only, such as MCI due to various etiologies and mild or unspecified neurocognitive disorder (NCD). Our first question was whether, both stimulus-bound errors and other TSEs are associated with lower executive and semantic functions. Special attention was given to the similarities subtended as an indicator of abstract thinking, a process originally hypothesized to underlie stimulus-bound errors. An additional goal was to further establish the discriminant validity of TSEs as a cluster of errors qualitatively different than other error types in the test. We sought to do so by comparing these errors both to a group of errorless clocks as well as a group of clocks with other miscellaneous errors, using various tasks related to executive control, semantic knowledge, working memory, and visuospatial constructional domains. We hypothesized that TSEs, when compared to other errors, would be associated with greater semantic and executive impairment, as opposed to visuospatial constructional impairment.

Methods

Data from 490 participants from the Toronto Dementia Research Alliance (TDRA) research database (Tang-Wai *et al.*, 2020) were used for this study. This database includes participants' demographic, clinical intake and cognitive assessment data, as well as preliminary diagnoses made by a physician in select academic memory clinics in Toronto. We restricted the analysis to the MCI population in order to base it on a relatively homogenous population in terms of severity of impairment, due to the possibility that in various populations these errors could be associated with various processes. In addition, inclusion of participants with dementia may result in difficulty to discriminate between the relative contribution of each cognitive domain, as all cognitive functions are expected to be significantly impaired in this population and the clocks are often characterized by significant distortions related both to time settings and constructional abilities. For this analysis we defined MCI as having a primary diagnosis of one of the following categories: MCI ($n = 332$), MCI due to vascular etiology ($n = 89$), Parkinson's disease-related MCI ($n = 20$), and mild/not otherwise specified neuro-cognitive disorder ($n = 49$). The criteria for exclusion from the analysis were preliminary diagnosis incompatible with MCI. Thus, participants with primary diagnosis of dementia or major NCD ($n = 451$), no identified cognitive impairment ($n = 420$), concussion ($n = 120$), multiple system atrophy ($n = 5$), or no identified primary preliminary diagnosis ($n = 363$). Four hundred ninety participants who completed the CDT and met the criteria for inclusion in the analysis were identified. The study was approved by Clinical Trials Ontario which represents all sites participating in the study and was completed in accordance with Helsinki Declaration. Written informed consent for the testing material to be included in the TDRA database for research purposes, including secondary analyses was obtained from all participants.

Clock Drawing Test

Participants were provided with plain 8 ½" × 11" paper, in landscape orientation, and were instructed to "Draw the face of clock, put in all the numbers, and set the hands at 10 after 11." Instructions were repeated as needed. All clock drawings were collected as part of standard cognitive assessment conducted in the memory clinics affiliated with the TDRA. This assessment includes the Toronto Cognitive Assessment (TorCA), an assessment tool

comprised of various short subtests in various domains aimed to detect mild cognitive changes (Freedman et al., 2018). The clock drawings were scored by trained assessors such as health care professionals or graduate students, certified psychological associates, or psychologists. TorCA administration and scoring training was provided to all sites by Baycrest team members to promote consistency across sites.

The TorCA CDT scoring system is a modification of the Freedman scoring system (Freedman et al., 1994; see appendix 1 in supplementary material for scoring criteria), which has previously showed high levels of reliability relatively to other scoring systems (e.g., Souder, O'Sullivan & Pechenik, 1999; Suhr et al., 1998). In order to identify stimulus-bound errors or other TSEs, all clocks were reviewed by the first author (MS). Stimulus-bound errors were identified according to Rouleau's criteria: (A) "The hands are set for 10 to 11 instead of 10 after 11" ($n = 51$). (B) "The time is written (in letters and/or numbers) besides the '11' or between '10 and 11'" ($n = 3$). In addition, we included other clocks that approximated Rouleau's criteria, clearly representing the minute by the number "10" instead of "2" in any way, and representing the hour by the number "11" ($n = 7$). Finally, similarly to Cahn et al. (1996) and Cahn & Kaplan (1997), clocks with a hand pointing to the digit "10" (such as clocks showing "10 past 10") were also included in this category ($n = 17$). The remaining TSEs, in which the clock indicated wrong time (including no indication of time at all), were classified as other TSEs, for example, 11:05, or a single hand pointing slightly after the 11 (e.g., see Cahn et al., 1996 Fig. 2: "conceptual deficit"; Rouleau et al., 1996, Fig. 1: "misrepresentation of the time on the clock"). Errors in hand length were not considered TSEs, as these are most likely minor errors not related to the deciphering of the dual code needed to represent both minutes and hours by the target numbers (see Berger et al., 2008 for discussion).

Grouping procedure

Out of the 490 clocks, 78 clocks with stimulus-bound errors and 41 with other TSEs were identified. Assignment to the stimulus-bound and other TSE groups took precedence, and these clocks were grouped to these categories regardless of the existence of other error types. These 2 error groups correspond to the "hour/minute target number indicated in some way" based on the Freedman et al. (1994) scoring system (see items 10–11 in Appendix 1 in supplementary material). The first comparison group included all clocks that were error free (no-error group, $n = 176$) according to the TDRA database. The second comparison group, the non-TSE group ($n = 195$), included all clocks that contained only errors which were not defined as TSEs according to the criteria above (see consort diagram in appendix 2 in supplementary material).

Neuropsychological tests of interest

All 490 participants completed the TorCA as part of their routine clinical assessment in the period between July 2017 and May 2020. The tasks were chosen to represent four neuropsychological domains that may be associated with CDT performance: Executive Control, Semantic Function, Working Memory, Visuo-Spatial Construction (complex figure copy). In addition, due to the high relevance of this subtask to the construct of stimulus-bound errors as a marker impaired abstraction, this test was also examined separately in addition to its inclusion in the composite score. A possible interfering factor – English proficiency, was also represented by TorCA variables (see description below). The

intertwined nature of executive functions as well as the relatively executive nature of the Similarities subtest posed a challenge in terms of grouping the TorCA variables executive control, semantic, and working memory composites. Thus, we further confirmed the grouping of variables in these three domains using a factor analysis with Direct Oblimin Method (nonorthogonal) rotation. Of note, although the TorCA battery includes the Trail Making Test, this test was not included in the analysis for 2 reasons: (1) high percentage of data Missing Not/Not-Completely At Random (MNAR/MNCAR; 8.8% did not complete Trail-B due to various reasons), and (2) low loading of the index of interest – Trail Making B-A, on the relevant factors in the factor analysis. The factor analysis was conducted twice, with and without the Trail Making Test, in order to examine the possible bias the missing data in the test could have on the results. In both models the analysis yielded the same 3 factor solution. Results of the full factor analysis model in support of the division to these three cognitive composites can be found in appendix 3 in supplementary material.

TorCA measures

The subtests composite scores and indexes used for the purpose of this analysis are presented in Table 1. Further detail regarding the TorCA battery can be found in Freedman et al. (2018).

Statistical analyses

The statistical analysis was performed using IBM SPSS Statistics software. First, we compared the 4 error groups on the executive control and semantic composites, as well as the working memory index and complex figure copy score. In addition, the Similarities subtest was tested both as a part of the semantic composite and independently. To minimize multiplicity of comparisons we used 3 planned post hoc contrast to answer 3 different questions: Direct comparison between stimulus-bound errors and other TSEs, in order to assess whether those two groups are similar (i.e., lack difference on the observed measures) in terms of underlying cognitive deficits as suggested by Soffer et al. (2022). The evaluation of similarity between the 2 error groups was qualitative, and the groups were compared with significance level of $p \geq 0.2$, since statistical tests are designed to detect differences between groups rather than similarity, and a lower p value (e.g., $p < .05$) consists a stronger indication for a possible difference. Comparison between TSEs (both stimulus-bound errors and other TSEs as one group) and no-error, in order to assess which cognitive deficits are associated with existence of these errors. Lastly, in order to further establish the construct validity of TSEs we compared them against the non-TSEs group which contained other errors as appeared in the TDRA data platform. The criterion for significance in the two pairwise comparison between the TSE and the two reference groups was compatible with the Bonferroni correction for 2 post hoc contrasts, $p \leq .025$. Age, English command and education, were added as covariates in the group comparisons. Since Levene's test was significant in most variables, we used a linear model with Huber-White (Huber, 1967; White, 1980) Heteroskedastic-consistent standard errors in all models.

In order to examine the independent relative contribution of each variable of interest, namely, the four cognitive indices, taking into account the interfering variables English command, age, and education, we conducted multinomial logistic regression with TSEs as the reference category, comparing it to both the no-error group and the non-TSEs group. Absence of multicollinearity was examined both by correlation coefficients between the covariates,

Table 1. TorCA subtests used for creation of composites and other outcome measure

Variable	Subtask	Description	Scale	Composite/index calculation
Executive control composite	Serial Subtraction 3s ¹	subtract by 3s or 7s, starting from 90 or 100 (for 3s and 7s, respectively)	0–13	Averaging the z transformed values of each variable
	Serial Subtraction 7s ¹		0–13	
Semantic composite	Luria's Alternating Sequences	draw an alternating graphic sequence	0–2	Averaging the z transformed values of each variable
	Multilingual Naming Test ^{2,†}	Name drawings of common objects	0–15	
	Semantic Fluency ³ Similarities ⁴	Generate as many animal names as possible in 60 seconds Define in how two objects are similar. Each item is scored on a scale of 0–2 with partial score given to answers which are partially correct but too concrete	0–10*	
Working memory	Semantic Knowledge	determine which one of 4 response items (4 pictures) is most related to the target picture	0–10*	Sum of scores
	Digit Span Forward	Repeat a string of numbers. Main outcome measure is the longest string recalled correctly	0–9	
	Digit Span Backward	Repeat a string of numbers in reverse order. Main outcome measure is longest string recalled correctly	0–8	
Visuospatial constructional English command composite	Benson Complex Figure ⁵	Complex figure copy	0–17	Single subtask score
English command composite	Word Reading	read the word aloud correctly	0–12*	Averaging the z transformed values of each variable
	Sentence Comprehension	read a sentence and point to the picture that matches the sentence	0–8*	

*Each item is scored on a scale of 0–2. Partial score is given in some circumstances such as more than 5 s to read a word, self-correction, or request for repeated presentation of the stimulus. †15-item split form was used.

¹Bristow et al. (2016).

²Gollan et al. (2012).

³Strauss et al. (2006).

⁴Modified from Darvesh et al. (2005).

⁵Possin et al. (2011).

and by examination of variance inflation factor (VIF) for the predictors, which were all well within the accepted ranges. The linearity assumption was examined through inspection of the interaction terms between the predictors and their log transformation, as operationalized in Field, 2014, and was met at a satisfactory level. Inspection of outliers (High Cook's distance and leverage value) was performed to assure that the results were not affected by a few extreme cases.

Data handling: Missing education and age data (no variable exceeded 1.02% of missing data) were replaced by the mean. For the composite scores, in case of missing subtests scores (no cognitive variable exceeded 1% of missing data), the composite score was the average of the z transformed existing scores for the individual within the same composite. Extreme values (1.8% at most) in the cognitive indexes and the English command composite were winsorized at the -3.29 SDs threshold (e.g., Field, 2014).

Results

Demographic and sample characteristics of the error groups are presented in Table 2.

Means, SDs, and results of omnibus group comparison tests for the dependent variables of interest are presented in Table 3.

Planned contrasts

First contrast: are participants with stimulus-bound errors different than participants with other TSEs

The evaluation of similarity between the two error groups was qualitative, and the groups were compared with significance level of $p \geq 0.2$. No significant differences were found on any of the cognitive measures between the two TSEs groups namely, the stimulus-bound group and other TSEs group. Executive control

composite ($Wald \chi^2 = .96, p = .328, d = .19$), semantic composite: ($Wald \chi^2 = .03, p = .862, d = .03$), working memory index ($Wald \chi^2 = .88, p = .349, d = .17$), complex figure copy ($Wald \chi^2 = .72, p = .396, d = .22$), and the Similarities subtest ($Wald \chi^2 < .01, p = .948, d = .01$).

Second contrast: are participants with TSEs different from participants with other errors?

Participants with clocks with TSE performed significantly worse than those with non-TSEs on the working memory index ($Wald \chi^2 = 8.47, p = .004, d = .32$) and the semantic composite ($Wald \chi^2 = 5.68, p = .017, d = .25$). The groups did not differ on the Similarities subtest ($Wald \chi^2 = 3.74, p = .053, d = .24$), the executive control composite ($Wald \chi^2 = .27, p = .609, d = .06$), or the complex figure copy ($Wald \chi^2 = .20, p = .656, d = .06$).

Third contrast: are participants with TSEs different from participants with no error clocks?

Participants with clocks with TSEs differed from those with errorless clocks in all dependent variables. The most prominent effect size was evident on the semantic composite ($Wald \chi^2 = 23.65, p < .001, d = .50$), followed by working memory ($Wald \chi^2 = 16.20, p < .001, d = .47$), the Similarities subtest ($Wald \chi^2 = 12.16, p < .001, d = 0.42$), the executive control composite ($Wald \chi^2 = 7.69, p = .006, d = .32$), and finally, the complex figure copy ($Wald \chi^2 = 5.84, p = .016, d = .28$).

Relative contribution of neuropsychological domains

The overall multinomial model was significant ($n = 490, \chi^2 = 109.35, Cox \& Snell R^2 = .200$). In differentiating between clocks with TSEs and clocks with non-TSEs, while taking into account all components of variance, only the semantic composite

Table 2. Demographic and sample characteristics

	S-B (<i>n</i> = 78)	Other TSE (<i>n</i> = 41)	Non-TSE (<i>n</i> = 195)	No-error (<i>n</i> = 176)	<i>p</i> [†]	All significant pairwise comparisons [†]
Gender <i>n</i> (% female)	43 (55.1%)	18 (43.9%)	101 (51.8%)	92 (52.3%)	NS	
MCI	49	30	145	120	NS	
NCD mild/NOS	8	4	14	23		
VCI	15	4	29	29		
PD-MCI	6	2	7	4		
Education	13.63 ¹ (4.16)	14.25 ² (3.24)	14.95 ³ (3.60)	15.15 (3.19)	<i>p</i> = .011	S-B < non-TSE S-B < no-error
Age	74.3 ⁴ (10.24)	75.32 (10.02)	72.12 (11.07)	68.63 (11.63)	<i>p</i> < .001	S-B > no-error Other TSE > no-error Non-TSE > no-error
CDT	11.68 (2.11)	10.61 (2.91)	13.04 (1.71)	15.00 (0)	<i>p</i> < .001 ⁵	S-B < non-TSE S-B < no-error Other TSE < non-TSE Other TSE < no-error
English command composite	-.48 (1.22)	-.28 (1.23)	.04 (0.93)	0.24 (0.81)	<i>p</i> < .001	Non-TSE < no-error S-B < non-TSE S-B < no-error Other TSE < no-error

CDT = clock drawing; MCI = mild cognitive impairment; NCD = neurocognitive disorder; NOS = not otherwise specified; PD = Parkinson's disease; S-B = stimulus-bound error; TSE = time setting error; VCI = vascular cognitive impairment.

Test; ¹*n* = 75; ²*n* = 40; ³*n* = 194; ⁴*n* = 77; ⁵The assumption of homogeneity of variance was not met, however, a robust test of equality of means could not be performed because the no-error group had no variance.

[†]Unadjusted.

**p* < .01.

Table 3. Means, standard deviations, and omnibus results for the effect of error group on the variables of interest, with age, education, and English proficiency index as covariates

	Stimulus-bound errors	Other TSEs	Non-TSEs (other errors)	No-error	<i>df</i>	Wald χ^2	<i>p</i>
Executive control	-.185 (.81)	-.264 (.81)	-.05 (.83)	.182 (.64)	3	12.96	.005
Semantic	-.424 (.80)	-.351 (.89)	-.016 (.67)	.283 (.54)	3	26.826	<.001
Working memory	9.56 (1.87)	10.05 (2.02)	10.67 (1.94)	11.07 (2.06)	3	17.67	<.001
Complex figure copy	15.05 (2.02)	14.71 (2.99)	15.08 (2.15)	15.88 (1.08)	3	19.67	<.001
Similarities	8.17 (2.18)	8.29 (2.33)	8.99 (1.59)	9.45 (1.14)	3	13.74	.003

TSE = time setting error.

Table 4. Full multinomial model with TSEs as reference group

		Wald statistic	<i>p</i>	Odds ratio	Odds ratio 95% CI		
					Low	High	
Predicting time setting errors or non-time setting error group (other errors)	Intercept	1.49	.223				
	Executive control composite	.32	.575	.93	.72	1.20	
	Semantic composite	4.99	.025	1.38	1.04	1.83	
	Complex figure	1.08	.298	.89	.70	1.11	
	Working memory	8.85	.003	1.53	1.16	2.02	
	English composite	1.65	.199	1.20	.91	1.57	
	Age	2.00	.157	.98	.96	1.01	
	Education	1.41	.235	1.04	.97	1.12	
	Predicting time setting error or no-error group	Intercept	5.39	.020			
		Executive control composite	.42	.52	1.11	.81	1.51
Semantic composite		15.69	<.000	2.03	1.43	2.87	
Complex figure		4.11	.043	1.46	1.01	2.10	
Working memory		9.63	.002	1.62	1.19	2.19	
English composite		1.98	.160	1.26	.91	1.73	
Age		6.33	.012	.97	.94	.99	
Education		.037	.85	1.01	.93	1.09	

CI = confidence interval.

Cognitive variables (executive control, semantic, complex figure, working memory, and English command) are standardized to ease interpretation.

and the working memory index had significant independent contribution. In differentiating between TSE and no-error clocks, the semantic composite, working memory index, age and the complex figure copying had significant contribution. For the full model see Table 4.

Discussion

This study examined the relationship between stimulus-bound errors and other TSEs on the CDT and performance on tasks related to various neurocognitive domains in MCI. First, we aimed to further examine the possibility that both stimulus-bound errors

and other TSEs are similar in terms of executive and semantic correlates. We included a specific measure of abstraction abilities, since abstraction has been suggested to underlie stimulus-bound errors in particular (Cahn *et al.*, 1996; Freedman *et al.*, 1994) and has never been empirically tested. No differences were found between the two TSE groups on the Similarities test or any other measure. Nonetheless, the significant differences in the Similarities subtest between both TSE groups and the reference groups may suggest a deficit in abstract thinking as a putative additional underlying mechanism in the formation of both stimulus-bound errors and other TSEs.

The current analysis was characterized by a relatively large sample size, and a wider array of tests, taking into account confounding variables such as age, English command and education. Thus, this study substantially strengthens our preliminary findings (Soffer *et al.*, 2022), suggesting that in at risk populations for dementia, stimulus-bound errors do not have a different cognitive mechanism compared to other TSEs and do not appear to represent a specific executive deficit related to abstraction or executive control.

Second, we aimed to further examine differences between TSEs and errorless clocks. The association between presence of TSEs and semantic deficits was found to be exceptionally robust. Differences were present in all domains tested, as well as the interfering factors of English command, age, and education. The regression model revealed that semantic function and working memory were the only cognitive factors with significant independent contribution to differentiation between the groups. Thus, our results further confirm the association between TSEs and semantic deficits (Soffer *et al.*, 2022; Tranel *et al.*, 2008), as well as executive deficits (Soffer *et al.*, 2022). Nonetheless, the regression analysis underscores working memory and not executive control as the main executive contributor to the occurrence of these errors.

By comparing participants with TSEs to a group with non-TSEs, we attempted to further establish the classification of TSEs as a qualitatively distinct cluster of errors. Group comparisons revealed that TSEs were associated with worse performance relatively to non-TSEs on the semantic and working memory measures, but not on the complex figure copy and the executive control composite. Thus, it is possible that the role executive control and visuospatial processing deficits play in the commission of other common errors in the task is larger relative to the lower severity of those errors, compared to TSEs. The regression model again revealed significant independent variance related only to semantic function and working memory. This comparison contributes to the distinction of TSEs as representing a more severe semantic and working memory deficit. Moreover, while the TSEs group was not free of other CDT errors, the only differentiating criterion between the two groups was existence of a TSE in the clock. This is consistent with findings related to TSEs as being one of the most “telling” error categories in the test in regards to detection of dementia and its prodromal stages (Berger *et al.*, 2008; Duro *et al.*, 2018; Lessig *et al.*, 2008; Ricci *et al.*, 2016).

Our grouping criterion may also explain the discrepancies between our findings and other analyses which did not find strong support for discriminant validity between different error types in dementia and MCI (Umegaki *et al.*, 2021; Parsey & Schmitter-Edgecombe, 2011). These studies used Rouleau’s division which distinguishes between stimulus-bound errors and nonstimulus-bound TSEs, which are included in the variegated category of

“conceptual” errors. Thus, it is possible that the invalid division between stimulus-bound errors and other TSEs and the heterogeneity of errors within the “conceptual” cluster or other categories would render previous group comparisons or correlation matrices hard to interpret.

This study has several limitations. First, although the TorCA variables chosen to represent the relevant domains include a relatively wide variety of tests, several have restricted ranges. The grouping into composite scores, enabled us to generate index variables with greater variance. However, it is possible that those variables differed in terms of psychometric properties. In particular, it is possible that the restricted range of the visuospatial-constructional and the executive control measures resulted in under representation compared to the semantic composite in the regression model. Nonetheless, large effect sizes were detected for the Similarities subtest, which also had limited range, suggesting that the effects obtained for the semantic composite, might indeed represent the pivotal role of semantic processing in the occurrence of TSEs. Moreover, these findings also align with previous findings regarding the relationship between semantic deficits in CDT performance in MCI (Ahmed *et al.*, 2016).

Our findings do not suggest the usage of CDT and TSEs as a stand-alone measure of domain specific impairment. As any attempt to examine underlying deficient processes in test performance, one should always take into account the possibility that different cognitive impairments may lead to similar results in tests. This is relevant especially for TSEs which are binary and categorical in nature. Thus, as suggested earlier, it is possible that some individuals commit stimulus-bound errors due to strong “pull” to the stimulus “10,” lapse in attention, or other reasons (e.g. Rouleau *et al.*, 1992). This reasoning also applies to the interpretation of the relationship between working memory and TSEs. It is possible that for some individuals, the multiple simultaneous demands of the task create cognitive load, which in turn, may compromise the ability to retrieve essential knowledge related to time representation in the analog clock system. This would explain the association of TSEs with decreased working memory capacity, as working memory and the ability to cope with cognitive load are two closely related constructs. Alternatively, it is also plausible that other individuals simply could not retain the time setting instructions in memory. Thus, clinicians should always integrate a single test’s performance in the context of a greater picture that includes other domain specific measures, behavioral observation and clinical manifestation.

Future research should include investigation of neurocognitive correlates of other error types in the CDT. This would enable better validation of qualitative interpretation of the test. It would also be informative to determine whether stimulus-bound errors and other TSEs may be associated with other deficient functions such as executive control or abstraction in other populations, such as individuals with brain injury. This will enable better interpretation of clock drawings as a possible assisting tool for tracking specific cognitive deficits. Moreover, better understanding of how simultaneous activation of processes such as working memory, grapho-motor activity and retrieval of semantic knowledge, affects CDT performance is warranted. This may shed more light on mechanisms underlying neurocognitive constructs such as “cognitive load,” “interference,” and more.

Supplementary material. For supplementary material accompanying this paper visit <https://doi.org/10.1017/S1355617723011396>

Acknowledgements. The authors thank Marcos Sanches for his statistical support.

Funding statement. This work was made possible by a grant from the Canada Brain Research Fund of Brain Canada, with the financial support of Health Canada and the Toronto Dementia Research Alliance.

Competing interests. MF is listed on a patent related to methods and kits for differential diagnosis of AD vs frontotemporal dementia using blood biomarkers. TKR has received research support from Brain Canada, Brain and Behavior Research Foundation, BrightFocus Foundation, Canada Foundation for Innovation, Canada Research Chair, Canadian Institutes of Health Research, Centre for Aging and Brain Health Innovation, National Institutes of Health, Ontario Ministry of Health and Long-Term Care, Ontario Ministry of Research and Innovation, and the Weston Brain Institute. TKR also received for an investigator-initiated study in-kind equipment support from Newronika, and in-kind research online accounts from Scientific Brain Training Pro, and participated in 2021 and 2022 in an advisory activity for Biogen Canada Inc. TKR is also an inventor on the United States Provisional Patent No. 17/396,030 that describes cell-based assays and kits for assessing serum cholinergic receptor activity. MAB is supported by the US National Institutes of Health. MF is listed on a patent related to methods and kits for differential diagnosis of AD vs frontotemporal dementia using blood biomarkers. All other Authors report no competing interest.

References

- Ahmed, S., Brennan, L., Eppig, J., Price, C. C., Lamar, M., . . . & Delano-Wood, L. (2016). Visuoconstructional impairment in subtypes of mild cognitive impairment. *Applied Neuropsychology. Adult*, 23(1), 43–52. <https://doi.org/10.1080/23279095.2014.1003067>
- Berger, G., Frölich, L., Weber, B., & Pantel, J. (2008). Diagnostic accuracy of the clock drawing test: The relevance of “time setting” in screening for dementia. *Journal of Geriatric Psychiatry and Neurology*, 21(4), 250–260. <https://doi.org/10.1177/0891988708324939>
- Blair, M., Kertesz, A., McMonagle, P., Davidson, W., & Bodi, N. (2006). Quantitative and qualitative analyses of clock drawing in frontotemporal dementia and Alzheimer's disease. *Journal of the International Neuropsychological Society*, 12(2), 159–165. <https://doi.org/10.1017/S1355617706060255>
- Bristow, T., Jih, C. S., Slabich, A., & Gunn, J. (2016). Standardization and adult norms for the sequential subtracting tasks of serial 3s and 7s. *Applied Neuropsychology. Adult*, 23(5), 372–378. <https://doi.org/10.1080/23279095.2016.1179504>
- Cahn, D. A., & Kaplan, E. (1997). Clock drawing in the oldest old. *Clinical Neuropsychologist*, 11(1), 96–100. <https://doi.org/10.1080/13854049708407036>
- Cahn, D. A., Salmon, D. P., Monsch, A. U., Butters, N., Wiederholt, W. C., Corey-Bloom, J., & Barrett-Connor, E. (1996). Screening for dementia of the alzheimer type in the community: The utility of the clock drawing test. *Archives of Clinical Neuropsychology*, 11(6), 529–539. [https://doi.org/10.1016/0887-6177\(95\)00041-0](https://doi.org/10.1016/0887-6177(95)00041-0)
- Cosentino, S., Jefferson, A., Chute, D. L., Kaplan, E., & Libon, D. J. (2004). Clock drawing errors in dementia: Neuropsychological and neuroanatomical considerations. *Cognitive and Behavioral Neurology*, 17(2), 74–84. <https://doi.org/10.1097/01.wnn.0000119564.08162.46>
- Darvesh, S., Leach, L., Black, S. E., Kaplan, E., & Freedman, M. (2005). The behavioural neurology assessment. *Canadian Journal of Neurological Sciences*, 32(2), 167–177.
- Duro, D., Cerveira, P., Santiago, B., Cunha, M. J., Pedrosa De Lima, J. M., Botelho, M. A., & Santana, I. (2019). Clock drawing test in mild cognitive impairment: Correlation with cerebral perfusion in single-photon emission computed tomography. *Neuropsychology*, 33(5), 617–632. <https://doi.org/10.1037/neu0000549>
- Duro, D., Tábuas-Pereira, M., Freitas, S., Santiago, B., Botelho, M. A., & Santana, I. (2018). Validity and clinical utility of different clock drawing test scoring systems in multiple forms of dementia. *Journal of Geriatric Psychiatry and Neurology*, 31(3), 114–122. <https://doi.org/10.1177/0891988718774432>
- Eknoyan, D., Hurley, R. A., & Taber, K. H. (2012). The clock drawing task: Common errors and functional neuroanatomy. *Journal of Neuropsychiatry and Clinical Neurosciences*, 24(3), 260–265. <https://doi.org/10.1176/appi.neuropsych.12070180>
- Field, A. (2014). *Discovering statistics using IBM SPSS statistics* (4th ed). London: SAGE Ltd.
- Freedman, M., Leach, L., Carmela Tartaglia, M., Stokes, K. A., Goldberg, Y., Spring, R., . . . & Tang-Wai, D. F. (2018). The Toronto cognitive assessment (TorCA): Normative data and validation to detect amnesic mild cognitive impairment. *Alzheimer's Research and Therapy*, 10, 65. <http://doi.org/10.1186/s13195-018-0382-y>.
- Freedman, M., Leach, L., Kaplan, E., Winocur, G., Shulman, K. I., & Delis, D. C. (1994). *Clock drawing: A neuropsychological analysis*. Oxford: Oxford University Press.
- Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2012). Self-ratings of spoken language dominance: A multi-lingual naming test (MINT) and preliminary norms for young and aging Spanish-English bilinguals. *Bilingualism: Language and Cognition*, 15(3), 594–615. <https://doi.org/10.1017/S1366728911000332>
- Huber, P. J. Under nonstandard conditions. In: *Proceedings of the fifth berkeley symposium on mathematical statistics and probability: Weather modification*, Berkeley, CA: University of California Press, 1967, 221.
- Kaplan, E. (1988). A process approach to neuropsychological assessment. In T. Boll, & B. K. Bryant (Eds.), *Clinical neuropsychology and brain function: Research, measurement, and practice*. Washington, DC: American Psychological Association
- Lessig, M. C., Scanlan, J. M., Nazemi, H., & Borson, S. (2008). Time that tells: Critical clock-drawing errors for dementia screening. *International Psychogeriatrics*, 20(03), 459–470. <https://doi.org/10.1017/S1041610207006035>
- Parsey, C. M., & Schmitter-Edgecombe, M. (2011). Quantitative and qualitative analyses of the clock drawing test in mild cognitive impairment and Alzheimer disease: Evaluation of a modified scoring system. *Journal of Geriatric Psychiatry and Neurology*, 24(2), 108–118. <https://doi.org/10.1177/0891988711402349>
- Possin, K. L., Laluz, V. R., Alcantar, O. Z., Miller, B. L., & Kramer, J. H. (2011). Distinct neuroanatomical substrates and cognitive mechanisms of figure copy performance in Alzheimer's disease and behavioral variant frontotemporal dementia. *Neuropsychologia*, 49(1), 43–48. <https://doi.org/10.1016/j.neuropsychologia.2010.10.026>
- Ricci, M., Pigliautile, M., D'Ambrosio, V., Ercolani, S., Bianchini, C., Ruggiero, C., . . . & Mecocci, P. (2016). The clock drawing test as a screening tool in mild cognitive impairment and very mild dementia: A new brief method of scoring and normative data in the elderly. *Neurological Sciences*, 37(6), 867–873. <https://doi.org/10.1007/s10072-016-2480-6>
- Rouleau, I., Salmon, D. P., & Butters, N. (1996). Longitudinal analysis of clock drawing in Alzheimer's disease patients. *Brain and Cognition*, 31(1), 17–34. <https://doi.org/10.1006/brcg.1996.0022>
- Rouleau, I., Salmon, D. P., Butters, N., Kennedy, C., & McGuire, K. (1992). Quantitative and qualitative analyses of clock drawings in Alzheimer's and Huntington's disease. *Brain and Cognition*, 18(1), 70–87. [https://doi.org/10.1016/0278-2626\(92\)90112-Y](https://doi.org/10.1016/0278-2626(92)90112-Y)
- Soffer, M., Melichercik, A., Herrmann, N., Bowie, C. R., Fischer, C. E., Flint, A. J., . . . & Butters, M. A. (2022). Time setting errors in the Clock Drawing Test are associated with both semantic and executive deficits. *Applied Neuropsychology: Adult*. <https://doi.org/10.1080/23279095.2021.2023154>.
- Souder, E., O'Sullivan, P., & Pechenik, G. (1999). Comparison of scoring criteria for clock drawing test. *Journal of Clinical Geropsychology*, 5(2), 139–145.
- Spenciere, B., Alves, H., & Charchat-Fichman, H. (2017). Scoring systems for the clock drawing test: A historical review. *Dementia and Neuropsychologia*, 11(1), 6–14. <https://doi.org/10.1590/1980-57642016dn11-010003>
- Strauss, E., Sherman, E. M., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary*. Washington, DC: American Chemical Society.

- Suhr, J., Grace, J., Allen, J., Nadler, J., & McKenna, M. (1998). Quantitative and qualitative performance of stroke versus normal elderly on six clock drawing systems. *Archives of Clinical Neuropsychology*, 13(6), 495–502.
- Tan, L. P., Herrmann, N., Mainland, B. J., & Shulman, K. (2015). Can clock drawing differentiate Alzheimer's disease from other dementias? *International Psychogeriatrics*, 27(10), 1649–1660. <https://doi.org/10.1017/S1041610215000939>
- Tang-Wai, D. F., Strother, S. C., Pugh, B., Spring, R., Vica, C., Nourhaghighi, N., ... & Freedman, M. (2020). Toronto dementia research alliance (TDRA) dementia clinical-research platform: An example of research embedded into clinical care. *Alzheimer's and Dementia*, 16(S11), e041756. <https://doi.org/10.1002/alz.041756>
- Tranel, D., Rudrauf, D., Vianna, E. P. M., & Damasio, H. (2008). Does the clock drawing test have focal neuroanatomical correlates? *Neuropsychology*, 22(5), 553–562. <https://doi.org/10.1037/0894-4105.22.5.553>
- Umegaki, H., Suzuki, Y., Komiya, H., Watanabe, K., Yamada, Y., Nagae, M., & Kuzuya, M. (2021). Frequencies and neuropsychological characteristics of errors in the clock drawing test. *Journal of Alzheimer's Disease [Preprint]*, 82(3), 1291–1300. <https://doi.org/10.3233/JAD-210456>
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4), 817–838. <https://doi.org/10.2307/1912934>