Decomposition of the Trail Making Test - Reliability and Validity of a Computer Assisted Method for Data Collection

Amir Poreh, Ph.D., Ashley Miller, M.A., Philipp Dines, M.D., Ph.D., Jennifer Levin, Ph.D.

Abstract

The present study describes the use of computer assisted software to decompose the Trail Making Test. The study shows that this methodology is reliable and produces data comparable to those which are produced using pencil and paper forms. Additionally, it confirms that particular sections of the Trail Making Test correlate with indexes of the Controlled Oral Word Association Test and the Five Point Test that are purportedly sensitive to executive function deficits. The present study suggests that the adaptation of computer assisted testing to clinical practice is an important evolutionary step as it provides clinicians with higher resolution for traditional measures and discerns the multiple cognitive operations within them, allowing for the identification of nonspecific error variance that impacts test performance.

Introduction

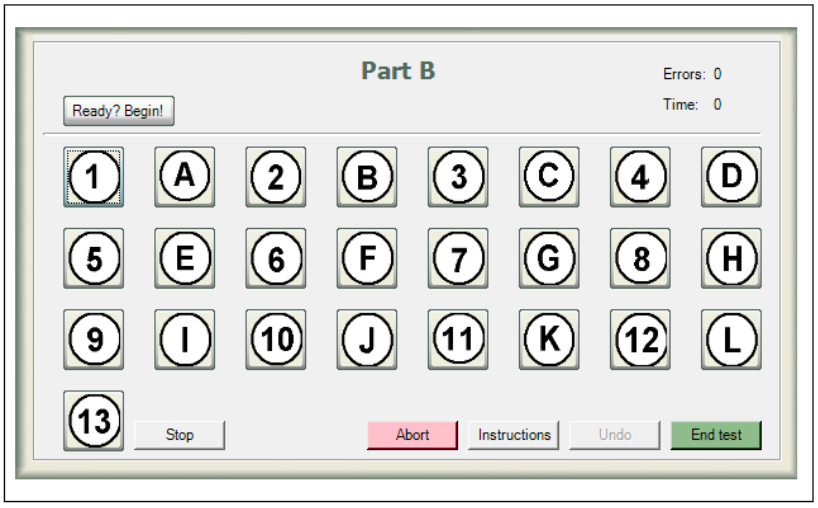
The Trail Making Test (TMT) was developed by Partington and Leiter in 1938 as a "Divided Attention Test" and was published in its current format as part of the Army Individual Test Battery (see Partington & Leiter, 1949). The test consists of two parts, A and B. The test stimuli for Part A are encircled numbers from 1 to 25, randomly spread across a sheet of paper. The subject is instructed to connect the numbers in order, beginning with 1 and ending with 25, as quickly as possible. In Part B, the subject is instructed to connect numbers and letters in an alternating pattern (1-A-2-B-3-C, etc.) as quickly as possible. Typically, the entire test can be completed in less than 5 minutes. Reitan (1958) devised the administration instruction of correcting subject errors in real time such that the subject's score, or time, includes the amount of time it takes to correct any errors. Normative data for the time it takes to complete the Trail Making Test has been summarized by Mitrushina et al. (2005) and Strauss et al. (2006). Although the use of a single score to describe the performance on the Trail Making Test has been shown to be highly reliable in the assessment of the general population, many have argued that if one is to understand the origin of a subject's poor performance on this or similar neuropsychological measures, one has to discern the cognitive substrates or processes which are required to complete this measure.

The importance of the process approach for the understanding of normal and abnormal performance on cognitive measures was first described by the renowned Czechoslovakian neurologist Arnold Pick (1931). Pick perceived normal and abnormal neurocognitive functions as an unfolding process ("microgenesis") and advocated for the use of qualitative observations to discern the underlying mechanisms of cognitive tests. Pick's ideas were later adopted by Heinz Werner (1956), Alexander Luria (1963, 1973), and Edith Kaplan's Boston Process Approach (1983, 1990). In recent years, the process approach has been criticized for its subjectivity,

leading to the emergence of the Quantified Process Approach (QPA; Poreh, 2000). This approach calls for the quantitative documentation and analysis of the process data and subsequent validation in controlled studies. Poreh (2006) reviewed the neuropsychological literature and identified three methodologies that are often employed to quantify the Boston Process Approach: the utilization of "satellite tests," such as in the development of the Rey Complex Figure Test Recognition Trial (Meyers & Meyers, 1995); the composition of new indexes, such as in the introduction of the Clustering and Switching Indexes for the analysis of the Controlled Oral Word Association Test (Troyer, Moskovitch & Winocur, 1997); and the decomposition of the test scores, such as in the development of the Global and Local Indexes for the Rey Complex Figure Test (Poreh, 2006) and the decomposition of the Rey Auditory Verbal Learning Gained and Lost Access Indexes (Woodard, Dunlosky, & Salthouse, 1999). Each of these methodologies has been employed in dozens of research studies yet has been too cumbersome to be carried out in day to day clinical practice because of the use of pencil and paper forms as well as a traditional stop watch to collect and record the data. In the case of the Trail Making Test, some researchers have advocated the use of difference and ratio scores that compare the performance on Part A and B. The use of such indexes, however, is controversial since it increases the error variance, produces skewed distributions, and results in high false positives (Nunnally, 1978; Cohen & Cohen, 1983). Moreover, studies suggest that these TMT indexes fail to demonstrate sensitivity to degree of head injury or to reliably identify simulators (Martin, Hoffman, & Donders, 2003

The present study describes a method for decomposing the Trail Making Test using software that presents to the examiner the elements of the TMT on a computer screen. The subject is asked to perform the test using a pencil and the original test form, while the examiner clicks on the mouse each time the subject moves the pencil from one element to the next. Figure 1 provides a screen shot of the computer software.

Figure 1
Screenshot of the Trail Making Test (TMT) Software



This software provides several benefits beyond the traditional pencil and paper data collection format. First, the examiner is guided through the administration process by pop-up instructions which can be read or played via a loudspeaker, ensuring high quality control and the

standardized administration of the test. Second, the test scores can be decomposed into five sections with five elements in each section (TMT A: 1-5, 6-10, 11-15, 16-20, 21-25; TMT B: 1A2B3, C4D5E, 6F7G8, H9I10J, 11K12L13), allowing for the analysis of within task performance. Finally, as these data are collected and stored on the computer, it allows for "on the fly" scoring of existing subtraction (A-B) and ratio (A/B) indexes as well as the decomposition of the test into subsections

The present study set out to examine the reliability and equivalence of this computerized method for collecting data in the general population. Additionally, the study examined the validity of decomposition of the Trail Making Part B in a sub-sample of elderly subjects by administering to this sample two additional measures that purport to assess executive functions. The elderly sub-sample was chosen since according to the literature, they are more likely to exhibit deficits on the Trail Making Test (Ashendorf et al., 2008).

Methods

Participants. The database used for this study consisted of 271 subjects. A sample of 138 subjects was collected in a series of unpublished experiments investigating the effects of age on planning. Another 53 older adult subjects were collected to investigate strategies employed by older adults on commonly used neuropsychological measures (Yocum, 2008). Finally, a sample of 80 subjects was collected for assessing the correlation between the TMT and the Stroop Test (Miller, 2010). All of the subjects were Caucasian community dwelling volunteers. In the first study, much like Tombaugh's (2004), subjects were recruited from social organizations, places of employment, psychology classes, and word of mouth. In the second study, subjects were residents of independent-living facilities for older adults who maintained autonomy in their own apartments and were responsible for their own shopping and bill paying. These independent living sites were located in a suburb of a Midwestern city. The older adult subjects were contacted via a flier sent to each apartment from a research coordinator advertising participation for psychological research. Those interested residents contacted their community coordinator who screened for neurological and psychiatric illness. Once the coordinator was satisfied with the appropriateness of the subject, the contact information was released to the examiner. The examiner used this contact information to call subjects and schedule an appointment in their apartment for the testing. During the initial portion of the exam the examiner documented the subject's demographic background and verified that they indeed did not have a history of psychiatric or neurological illness. The average age of the subjects was 38.2 (SD = 21.29), ranging from 18 to 92. The education level ranged from 6 to 23 (M = 14.5 , SD = 2.5). The majority of subjects were female (58.6%) and right handed (94.7%).

Procedure. Before administering the tests, the examiners were trained as to how to administer the test using the guidelines presented by Strauss, Sherman, and Spreen (2006), and the in the use of the software. After they demonstrated proficiency in administering the test and operating the software, including the ability to promptly correct subjects when they made errors, the examiners collected a small sample of subjects. The research coordinator for the study examined the data for administration errors and gave her final approval. Each subject completed a consent form prior to the test(s) administration. A copy of the informed consent was given to each subject. Next, the examiner opened the computer laptop and set it up so that the subject could not see the screen. The computer screen provided the examiner with the instructions for

administering the tests. In this fashion the standardization and quality control of the test administration were maintained. All subjects were administered Trail Making Tests A and B. A sub-sample extracted from Yocum's (2008) database of elderly subjects (N = 53) was also administered the Five Point Test (Regard, 1982) and the Controlled Oral Word Association Test (COWAT; Benton & Hamsher, 1976).

Measures. The Trail Making Test was administered to each subject using the exact instructions provided by Reitan (1958) and Lezak, Howieson, and Loring (2004). Each subject was provided with the test form and a pencil. As the subject listened to the instructions from the computer, the examiner pointed to the circles on the page. Once the subject started the test, the examiner carefully observed the subject and clicked on the mouse each time the subject connected the circles with his or her pencil. The clicking on the mouse was done when the line crossed the circle's circumference or when the subject was almost touching, by a reasonable distance, the target circle's circumference and made a graphomotor movement toward the next target circle. This method was applied to the test sections of the Trails A and B. Since the software is designed such that the arrow of the mouse jumps from one circle to the next, the examiner is able to fully attend to the subject and readily observe his or her movements across the page. Whenever the subject makes an error, the examiner clicks on the erroneous circle, stops the subject, and instructs him or her to go back to the circle where the error was made. The computer keeps track of the time it took for the subject to correct the error. In addition to total time for Trails A and B, the software records the latency for each movement from one circle to the next and then collapses the data into 5 separate indexes, each index composed of 5 circles. The decomposition into five sections was chosen because the number of circles (i=25) can only be divided evenly by five, the nature of the test instructions, and the need to smooth out the examiners' reaction time. The resulting decomposition allows the examiner to evaluate the speed during the demonstration portion of the test (the first 4 items on Trails A and the first 5 items on Trails B) and the process by which the task was performed, such as slowing down or stopping in a particular section.

The Controlled Oral Word Association Test (COWAT; Benton & Hamsher, 1976) was administered using the exact instructions described in the literature (Lezak, Howieson, & Loring, 2004; Strauss, Sherman, & Spreen, 2006). Subjects were asked to say as many animals as they could during a 60 second period and to say as many words that start with the letters F, A, and S, each during a 60 second period. The responses of the subjects were recorded directly into the examiner's laptop using software developed by the first author. To enhance the typing speed, an automated word completion as well as a digital tape recording was embedded within the software. Following the administration of the test, an automated database driven software scored the composition indexes, Clustering and Switching, using the guidelines that are described by Troyer, Moscovitch, and Winocur (1997). Clustering during the phonemic fluency task is defined as successively generating words that begin with the same first two letters or phonemes. In the Animal fluency portion, clusters are defined as successively generated words belonging to the same semantic subcategories, such as pets, zoo animals, birds, etc. The size of the cluster is counted beginning with the second word in each cluster. Switches are defined as the number of transitions between clusters, including single words. The raw score is used here because it has been found that adjusted score and percentages do not capture significant group differences in fluency performance (Epker, Lacritz, & Cullum, 1999; Troster et al., 1998). Numerous studies have studied the Clustering and Switching Indexes. These studies show that switching indexes are more sensitive to executive function deficits while clustering is more sensitive to word

finding deficits (Troyer, Moscovich, Winocur, Alexander, & Stuss, 1998; Rich, Troyer, Bylsma, & Brandt, 1999).

Regard's Five Point Test (Regard, 1982) was administered using the exact same instructions described in the literature. In this test, the subject is asked to create as many unique designs as possible using a series of 5 points, much like those which appear on a dice, which are printed across a page. The subject is instructed to create designs across the page, during a pre-determined time (3 minutes), from left to right. The instructions were read to the subject via the computer so as to ensure the highest level of quality control. Since it is difficult to follow a subject as they complete this task, the data entry was carried out after the testing session was over.

Data Analysis. Data was analyzed using SPSS Release 11.5. Whenever multiple comparisons were made, a Bonferroni correction was utilized to account for Type-I error. Pearson correlation coefficients were used to assess the relationship between the present data and Trail Making normative data that have been published in the literature. Whenever correlations in the high 80s or 90s were noted, the authors did not include probability values.

Results. To assess the equivalence of our norms to those of previous studies, Pearson correlations were performed between the presented data and the tabulated norms of both Tombaugh (2004) and Drane, Yuspeh, Huthwaite, and Klingler (2002). These norms were chosen because they are both thorough and fairly recent. Since the age grouping of the norms were not identical, the closest categories for each age group were chosen. The correlation between Tombaugh's norms and the current study was $r = 0.978$ and $r = 0.957$ for Trails A and B, respectively. The correlation between Tombaugh's norms (Strauss et al., 2006, p. 663) for the two derived indexes of division and subtraction of the Trail Making Test's Part A and B was also high ($r = 0.897$ and $r = 0.919$, respectively). The correlation between Drane et al.'s (2002) published norms and the current study for Trails A sum score was $r = 0.979$ and $r = 0.986$ for Trails B. Table 1 shows the mean scores for each of the traditional indexes as well as the results of the repeated independent samples t-tests that were conducted between the sum total Trail Making Test Part A and B across Tombaugh's (2004) seven age groups. One sees that among the 45 to 54 and 55 to 59 age groups, significant differences emerged with the present method yielding a significant faster latency. It should be noted that these differences would have washed out if we were to apply a Bonferroni correction to address the multiple comparisons.

A question is raised as to whether the use of a computer to collect the data instead of a traditional stopwatch could have altered the reliability of the results. To test this hypothesis, the inter-rater reliability of the recording software was examined by having two pairs of examiners record the performance of 18 subjects on the test. The inter-rater reliability for the Trail Making Test A and B sum scores was 0.996 and 0.998, respectively.

Table 1
Means and Standard Deviations of the Trail Making Test Indexes and Comparisons of the Sum Scores with Tombaugh's (2004) Normative Data

		TMT A Mean Latency in Seconds		TMT B Mean Latency in Seconds		A/B		B-A		TMT A norms †	TMT B norms †
Age	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*t*	*t*
18-24	85	22.53	7.28	46.41	15.3	2.27	.72	24.02	13.3	*t*= 0.25	*t*=0.86
25-34	53	24.49	9.35	47.24	16.5	2.15	.77	23.07	13.7	*t*=0.04	*t*=1.11
35-44	34	28.47	10.97	56.45	18.9	2.33	.87	29.54	16.6	*t*=0.02	*t*=0.42
45-54	31	26.39	9.33	52.58	17.0	2.13	.61	14.84	12.6	*t*=2.59	*t*=3.42 ***
55-59	19	28.47	6.02	55.79	17.6	2.46	.72	31.39	15.7	*t*=1.64	*t*=1.85
60-64	20	33.95	7.82	70.15	28.4	2.15	73	25.59	5.7	*t*=0.19	*t*=0.09
65-69	7	34.57	7.50	76.28	30.5	2.36	1.06	31.34	11.8	*t*=0.79	*t*=0.35
70-90	22	58.31	26.81	125.45	68.4	2.31	.87	56.32	25.5	*t*=0.11	*t*=0.95

***p<0.001
† Tombaugh's (2004)

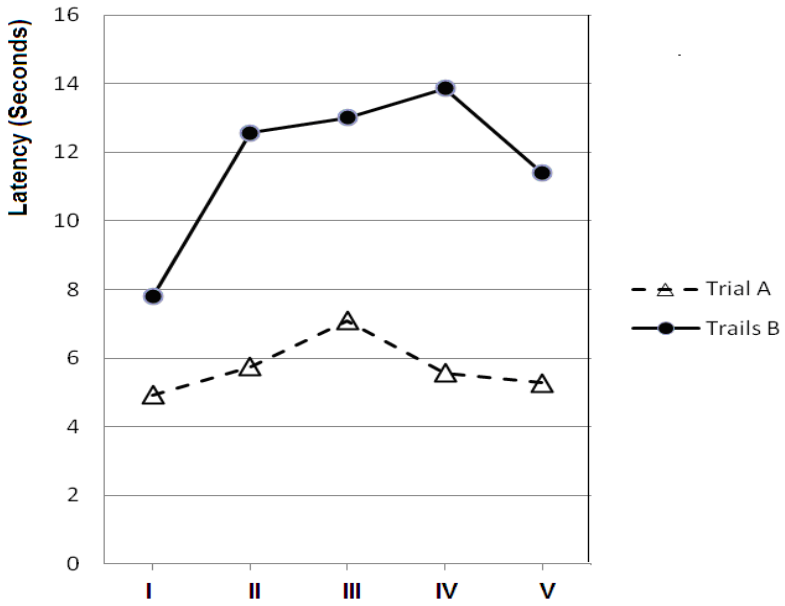
Table 2 shows the inter-rater reliability correlation coefficients for the Trail Making Test A and B subsections. One sees that the inter-rater reliability coefficients for the Trail Making Part A five subsections ranged from 0.933 to 0.994 and the Part B five subsections ranged from 0.881 and 0.997.

Table 2
Inter-rater Correlation Coefficients for the Trails A and B Subsections

Trails A subsections	*r*	Trails B subsections	*r*
1-5	0.954	1 A 2 B 3	0.997
6-10	0.994	C 4 D 5 E	0.962
11-15	0.980	6 F 7 G 8	0.891
16-20	0.933	H 9 I 10 J	0.972
21-25	0.944	11 K 12 L 13	0.881

Figure 2 shows the mean latency of the subjects on the Trails A and B subsections. Repeated Measures ANOVA of the whole sample showed that both the linear and quadratic models were statistically significant (F = 7.56, p<.006 and F = 35.1, p<.0001). Simple contrast analysis of Trails A subsections with the mean score serving as the reference category shows that the first section, made up of circles 1 to 5, was completed significantly faster (F = 37, p<0.001) than the other sections, and that section 3 took the longest to complete (F = 39, p<0.0001). Namely, subjects start quickly, slow down, and then recover and quickly complete the rest of the test.

Figure 2
Mean Latency for the Trails A and B Subsections



Tables 3 and 4 provide the mean and standard deviations of the latency scores on the Trails A and B subsections. Repeated Measures ANOVA of the Trails B subsections showed that both the linear and quadratic models were statistically significant (F=46.9, p<.0001 and F=100.77, p<.00001). Contrast analysis of mean difference shows that the first section (including circles 1, A, 2, B, and 3) was completed significantly faster than any other section (mean differences ranged from -3.5 to -6.2, p<0.0001 for all sections). Section 5 (11, K, 12, L, and 13) was completed faster than sections 3 and 4, but slower than section 1 (mean differences for the comparison between section 5 and sections 3 and 4 was -1.87 and -2.7, p<0.001).

Table 3
Means and Standard Deviations of the Trail Making Test Part A Latency by Subsections

		Part A Subsections									
Age Group	*n*	I		II		III		IV		V	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
18-20	28	4.51	4.3	5.50	3.9	6.60	5.1	5.27	3.4	4.90	3.2
21-25	65	2.89	1.4	4.21	1.9	4.79	2.0	4.34	2.1	3.93	1.4
26-40	30	3.06	1.9	4.44	2.3	4.46	1.7	4.06	2.1	3.92	1.8
31-50	13	3.90	2.3	4.50	1.9	6.90	5.3	4.36	1.8	3.70	1.7
51-60	19	4.00	2.3	3.84	1.9	5.00	5.3	4.36	1.8	4.69	1.7
61-70	33	3.72	2.6	5.94	3.0	5.50	2.9	4.72	1.8	4.50	1.9
71-92	20	6.13	2.8	6.40	2.3	8.40	3.1	6.20	2.2	6.13	1.1

Table 4
Means and Standard Deviations of the Trail Making Test Part B Latency by Subsections

Age Group	*n*	Part B Subsections									
		I		II		III		IV		V	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
18-20	28	5.21	3.5	9.07	3.3	10.71	3.6	12.35	6.5	8.5	4.1
21-25	65	5.49	3.4	8.51	3.6	9.46	3.1	9.51	6.6	9.1	4.2
26-40	30	6.66	4.0	10.4	4.9	11.56	4.8	12.16	6.5	10.47	4.2
31-50	13	8.38	5.0	9.77	4.2	9.29	4.7	9.00	3.4	6.61	2.5
51-60	19	5.11	2.8	8.44	3.2	10.8	4.1	9.66	2.6	9.27	3.1
61-70	33	10.25	8.7	17.06	7.3	20.0	8.8	16.40	5.6	14.26	5.8
71-92	20	16.28	2.3	31.52	4.1	25.21	3.2	36.22	7.2	27.61	3.3

Another question is raised as to whether the use of recording the errors using the software might slow down the subjects. If this was the case, then we would expect the introduction of systematic error into both the Trails A and B scores among the elderly, a subgroup that produces a high rate of errors. Initial analysis shows that the error rate of the elderly sample was comparable to the error rate that has been reported in the literature (Ashendorf, 2008). Namely, 56.2% and 42.1% of the elderly subjects ages 61-70 and 71-92 produced no errors, and 12.4% and 31.7% of the subjects produced two or more errors. An examination of the Mayo Older Adults normative study (Steinberg, Bieliauskas, Smith, & Ivnik, 2005) shows those subjects between the ages of 62 and 72 who score between 75 and 84 seconds on the Trail Making B are placed in the 50^{th} percentile rank. Similarly, subjects between the age of 77 and 99 who score between 115-141 seconds also fall in this range. Tombaugh (2004) reported a mean score of 74.5 seconds (*SD* = 19.5) for ages 60 to 64 and 130.61 seconds (*SD* = 45.74) for ages 75 to 79. As seen in Table 1, our sample obtained similar results to those that were produced by Tombaugh (2004). Therefore, the concern that the computer assisted scoring of errors might inflate the subjects' scores was not supported.

To examine the incremental validity of the decomposition of the Trail Making Test into subsections, the sub-sample of older adults completed the COWAT (Animals and FAS, Benton

& Hamsher, 1976)) and the Five Point Test (Regard, 1982). Table 5 shows that the COWAT scores were consistent with previously published data (see Troyer, Moscovitch, & Winocur, 1997, p. 374). The table also shows that the Five Point Test unique designs scores produced by this sub-sample was comparable to those reported by Marianne Regard (1991) for the two age groups (Ages 61 to 70, Mean score = 27.3 *SD* = 7.4, *N* = 20) of older adults (Age 71-92; Mean score = 22.5, *SD* = 10.2, *N* = 20) .

Table 5
Means and Standard Deviations for Verbal Fluency and the Five Point Test among Older Adults

	Age Groups			
	61-70		71-92	
	M	*SD*	*M*	*SD*
Semantic Cluster	1.139	.65	0.993	.30
Semantic Switching	10.27	.65	6.54	.31
Semantic Sum Score	22.27	4.5	15.25	5.07
Phonemic Cluster	.286	1.96	.254	1.62
Phonemic Switching	29.41	8.9	28.08	11.7
Phonemic Sum	39.33	14.6	36.25	15.3
Five Point Test Unique Designs	33.64	7.3	24.6	11.8

Two exploratory stepwise regression analyses were conducted with the independent variables being the five sections of the Trail Making Test B and the dependent variable being the COWAT's Phonemic and Semantic Fluency tests sum scores for each analysis. The results indicated that only the last section of Trails B (11, K, 12, L, and 13) entered the equations (r = 0.42, p = 0.02 and r = 0.46, p = 0.011 for Phonemic and Semantic fluency, respectively). When the Semantic Switching Index served as the dependent variable, the last section was once more the only one that entered into the equation (r = 0.48, p = 0.007). When the Semantic Clustering Index served as the dependent variable, none of the Trail Making B subsections entered into the equation (r = 0.26, p = 0.879). Similar findings were discovered on the Phonemic Fluency task. When the Phonemic Switching Index served as the dependent variable, only the first section entered into the equation (r = 0.40, p<0.26). However, when the Phonemic Clustering Index served as the dependent variable, the fourth section entered into the equation, although the resulting model was relatively weak (r = 0.36, p = 0.049). Additional support for the above findings was obtained with an exploratory stepwise regression analysis with the Trail Making Test B subsections serving as independent variables and the Five Point Test serving as the

dependent variable (see Table 6). This analysis produced an equation in which only the last section of the Trail Making Test Part B entered ($r = 0.59$, $p<0.000$).

Discussion

The present study supports the fact that the use of computer assisted one-on-one testing, which uses software to collect the data of traditional neuropsychological measures, produces comparable data to those obtained using the traditional manner of data collection. Namely, that the tabulated sum scores of this measure highly correlate with previously published tabulated norms. The COWAT, which was also administered using a computer assisted method, likewise produced equivalent scores for the sum and Clustering and Switching Indexes to those that have been reported in the literature. The present method of data collection was specifically designed to circumvent the perils that are associated in the literature with the adoption of conventional psychological and neuropsychological measures into computerized formats (Schulenberg & Yutrzenka, 2004 for a review of this topic). Such formats change the modalities of the stimuli presentation and create subtle changes in the testing paradigm which could potentially affect the ability of the computerized test to tap into the same neurocognitive constructs that were evaluated by the original tests. Salo, Henik, and Robertson (2001) show that a change in the form of presentation of the test stimuli may have an effect on the test outcome. They show that a clinical version of the Stroop Color and Word Test and the computerized single-trial version of the task do not produce the same magnitude of the interference effect among psychiatric or brain damaged subjects. The authors attribute the results to the fact that the clinical version of the test, which presents more than one stimuli at a time, produces a higher degree of interference effects and lower performance among schizophrenics and frontal lobe injured subjects. Since the software in the present study used the exact versions of the original clinical tasks and did not change the original test paradigm, no such effects were produced.

The first goal of the study was to assess the reliability and validity of the computer generated Trail Making Test decomposition indexes. The high inter-rater reliability coefficients of the section by section subscores suggest that an examiner who is versed with the software can reliably record the latency of subjects' pencil movements between the circles using the computer assisted one-on-one software. The high inter-rater reliability confirms the reliability of the data collected using the new methodology. It also suggests that the use of other more sophisticated and costly technologies to collect such data, such as touch sensitive screens or digital pens, for collecting the section by section data, is not necessary.

As for the differences in latency within the Trail Making Test, this study clearly shows that the first portions of both Trails A or B are completed faster, most likely because subjects are trained, following Reitan's (1958) instructions of pointing out to the subjects the location of the first five items (circles) as well as the training during the pretest portion of the test. The last section of the Trail Making Test, particularly Test Part B, section 5 (11, K, 12, L, and 13), also emerged across all age groups as being completed faster than all but the first section. In addition, this is the only section among older adults that differentially correlates with other measures and indexes of executive functions used in this study.

How can one explain these latter seemingly incongruent findings? One common assumption among clinicians is that the Trail Making B is composed of two factors. The first, the scanning factor, requires the subject to scan the page while searching for the sequential circle. The second, the mental flexibility factor, relates to the subject's ability to maintain the alternating sequence of numbers and letters in his or her mind. As the subject approaches the end of the task, the impact of the scanning factor is reduced since the number of move

permutations decreases. At the same time, the letter and number sequencing becomes more complex and the ability to shift attention from the previous target and move his or her attention to the next target becomes more important. Thus, the last section of the test, which involves greater mental flexibility but less scanning could potentially be a more pure measure of executive functions than the rest of the test. To study this hypothesis, one would need to replicate the study with young and older adults as their eye movements are being monitored. This methodology would allow the researcher to document with precision the forward and backward eye movements of the subject, and evaluate whether such scanning behavior is reduced as the subject reaches the end of the task.

The second goal of the study was to examine the validity of the decomposition of the Trail Making Part B by administering two additional measures that purport to assess executive functions to a sample of elderly subjects. It was reasoned that since this population often shows a decline in executive functions, certain sections of the Trail Making Test Part B would correlate with such measures while others would not. The results of the present study supported our hypothesis. Using multiple regression analyses the study shows that only the last section of Trail Making Test Part B correlated with COWAT sum scores and Switching Indexes, indexes that have been previously shown to correlate with frontal lobe insult (Troyer et al., 1998). In contrast, none of the Trail Making Part B sections significantly correlated with the Phonemic and Semantic Clustering Indexes (Troyer et al., 1998). Similar findings were obtained when these subsections were correlated with the Five Point Test, a measure of nonverbal fluency. Hence, these findings lend preliminary support for the incremental validity of the Trail Making Test decomposition in its potential ability to identify more subtle executive function deficits.

Although the data in the present study are very promising, there are several limitations to the current work. First, our validation data are limited to elderly subjects and were not collected on all of the subjects in the study. Second, we do not provide detailed analyses of the error location made by our subjects, despite the fact that such error scores could potentially have clinical relevance. We did not do so since the base rate of errors in our sample was relatively small. Finally, we did not study any clinical populations using our methodology. Thus, we can only hypothesize which sections of the Trail Making Test would provide incremental validity over the traditional sum scores. Specifically, following Salo, Henik, and Robertson's (2001) theory regarding the variables that underlie the Stroop Color and Word Test performance, it could be that clinical populations with executive function deficits, such as schizophrenic patients, will exhibit more errors on the 3rd and 4th sections of the Trail Making Test Part B, the sections that require the highest mental load, assessing both visual scanning and mental flexibility.

In sum, the results of the present study show that computer assisted testing that employs noninvasive technologies to administer traditional neuropsychological tests produces comparable results as the traditional stopwatch method of administration, increases the resolution of these measures, and has potential to discern the processes that underlie both cognitive deficits and normal brain functioning. Additionally, such technologies show promise in decreasing the administration errors and increasing the standardization of the test administration by introducing quality control indexes into the testing process (such as time stamping), and reducing and/or eliminating the need for the cumbersome manual scoring of data.

About the Authors

Amir M Poreh, Ph.D. is an Associate Professor at Cleveland State University Department of Psychology Associate Clinical Professor of Psychiatry Case Western Reserve University School of Medicine.
Email: *aporeh@gmail.com*

Ashley Miller, M.A. is a third year doctoral student at the University of Tulsa studying clinical psychology.
Email: *ashley-miller@utulsa.edu*

Philipp Dines, Ph.D. is an Assistant Professor of Psychiatry Case Western Reserve University School of Medicine Medical Director Inpatient Neurogeropsychiatry Service Geropsychiatry Fellowship Program University Hospitals Case Medical Center Distinguished Fellow American Psychiatric Association.
Email: *philipp.dines@uhhospitals.org*

Jennifer Levin, Ph.D. is an Assistant Professor, Department of Psychiatry Case Western Reserve University School of Medicine University Hospitals Case Medical Center.
Email: *jennifer.levin@uhhospitals.org*

References

Ashendorf, L., Jefferson, A. L., O'Connor, M. K., Chaisson, C., Green, R. C., & Stern, R. A. (2008). Trail Making Test errors in normal aging, mild cognitive impairment, and dementia. *Archives of Clinical Neuropsychology, 23,* 129-137.

Benton, A., & de Hamsher, K. S. (1976). *Multilingual aphasia examination*. Iowa City: University of Iowa.

Drane, D. L., Yuspeh, R. L., Huthwaite, J. S., & Klingler, L. K. (2002). Demographic characteristics and normative observations for derived-trail making test indices. *Neuropsychiatry, Neuropsychology and Behavioral Neurology, 15(1),* 39-43.

Epker, M. O., Lacritz, L. H., & Cullum, C. M. (1999). Comparative analysis of qualitative verbal fluency performance in normal elderly and demented populations. *Journal of Clinical and Experimental Neuropsychology, 21,* 425-34.

Kaplan, E. (1983). A process approach to neuropsychological assessment. In T. Boll & B. K. Bryant (Eds.) *Clinical neuropsychology and brain function: Research, measurement, and practice* (pp. 125-167). Washington D.C.: American Psychological Press.

Kaplan, E. (1990). The process approach to neuropsychological assessment of psychiatric patients. *Journal of Neuropsychiatry and Brain Functions: Research Measurement, and Practice, 2,* 72-87.

Lezak, M., Howieson D.B., & Loring D.W. (2004). *Neuropsychological assessment 4th Edition.* New York: Oxford University Press.

Luria, A. R. (1963). *Restoration of function after brain injury.* Oxford, England: Pergamon Press.

Luria, A. R. (1973). *The man with a shattered world: The history of a brain wound.* Cambridge, MA: Harvard University Press.

Martin, T. A., Hoffman, N. M., & Donders, J. (2003). Clinical utility of the Trail Making Test ratio score. *Applied Neuropsychology, 10 (3), 163* – 169.

Meyers, J. E., & Meyers, K. R. (1995). *Rey Complex Figure and recognition trial*. Odessa, FL: Psychological Assessment Resources.

Miller, A. (2010). *Examining the Errors and Self-Corrections on the Stroop Test.* (Master's Thesis). Cleveland State University, Cleveland, OH.

Mitrushina, M., Boone, K. B., Rzani, J., & D'Elia, L. F. (2005). *Handbook of normative data for neuropsychological assessment (2nd Edition).* City?: Oxford University Press.

Pick, A. & Thiele, R. (1931). Aphasie. In A. Bethe (ed.), *Handb. d. Norm. u. Pathol. Physiol. Vol. XV, 2.* Berlin: Springer.

Poreh, A. (2000). The quantified process approach: An emerging methodology to neuropsychological assessment. *The Clinical Neuropsychologist, 14 (2),* 212 – 222.

Poreh, A. (2006). Introduction to the Quantified Process Approach. In A. Poreh (Ed.),*The Quantified Process Approach to neuropsychological assessment* (pp. 3-15). New York: Taylor & Francis Group, Psychological Press.

Regard, M. (1982). Children's production of verbal and nonverbal fluency tasks. *Perceptual and Motor Skills, 55,* 839-844.

Regard, M. (1991). *The perception and control of emotions: Hemispheric differences and the role of the frontal lobes* (Habilitationsschrift (Post Doctoral Thesis)). University Hospital, Department of Neurology, Zurich, Switzerland.

Reitan, R. M. (1958). Validity of the Trail Making Test as an indication of organic brain damage. *Perceptual and Motor Skills, 8,* 271-276.

Rich, J. B., Troyer, A. K., Bylsma, F. W., & Brandt, J. (1999). Longitudinal analysis of phonemic clustering and switching during word-list generation in Huntington's disease. *Neuropsychology, 13,* 525-531.

Salo, R., Henik, A., & Robertson, L. C. (2001). Interpreting Stroop interference: An analysis of differences between task versions. *Neuropsychology, 15,* 462-71.

Schulenberg, S. E., & Yutrzenkab, B. A. (2004). Ethical issues in the use of computerized assessment. *Computers in Human Behavior, 20,* 477–490.

Steinberg, B. A. Bieliauskas, L. A., Smith, G. E., & Ivnik, R. J. (2005). Mayo's Older Americans Normative Studies: Age- and IQ-Adjusted Norms for the Trail-Making Test, the Stroop Test, and MAE Controlled Oral Word Association Test. *Clinical Neuropsychologist, 19,* 329-377.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests – Administration, norms and commentary (3rd Edition).* Oxford: Oxford University Press.

Tombaugh, T. N. (2004). Trail Making Test A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology, 19(2),* 203-14.

Troster, A. I., Fields, J. A., Testa, J. A., Paul, R. H., Blanco, C. R., Hames, K. A., et al. (1998). Cortical and subcortical influences on clustering and switching in the performance of verbal fluency tasks. *Neuropsychologia, 36,* 295-304.

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology, 11,* 138-146.

Troyer, A. K., Moscovitch, M., Winocur, G., Alexander, M. P., & Stuss, D. (1998). Clustering and switching on verbal fluency: the effects of focal frontal- and temporal-lobe lesions. *Neuropsychologia, 36,* 499-504.

Yocum, A. (2008). *Employing strategy in measures of executive functioning: Young vs. old adults* (Master's Thesis). Cleveland State University, Cleveland, OH.

Werner, H. (1956). Microgenesis and aphasia. *The Journal of Abnormal and Social Psychology, 52,* 347-353.

Woodard, J. L., Dunlosky, J. A., & Salthouse, T. A. (1999). Task decomposition analysis of the inter trial free recall performance on the Rey Auditory Verbal Learning Test in normal aging and Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology, 21,* 666-676.