2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

# Investigating a classical neuropsychological test in a real world context.

Nathaniel Pinkes, *Student Member, IEEE*, Zachary Fagiani, Ethan Wong, Holly Jimison, *Member, IEEE*, Misha Pavel, *Senior Member, IEEE*, Mathew Yarossi, *Member, IEEE*, Eugene Tunik

*Abstract*— **This study was performed to investigate the validity of a real world version of the Trail Making Test (TMT) across age strata, compared to the current standard TMT which is delivered using a pen-paper protocol. We developed a real world version of the TMT, the Can-TMT, that involves the retrieval of food cans, with numeric or alphanumerical labels, from a shelf in ascending order. Eye tracking data was acquired during the Can-TMT to calculate task completion time and compared to that of the Paper-TMT. Results indicated a strong significant correlation between the real world and paper tasks for both TMTA and TMTB versions of the tasks, indicative of the validity of the real world task. Moreover, the two age groups exhibited significant differences on the TMTA and TMTB versions of both task modalities (paper and can), further supporting the validity of the real world task. This work will have a significant impact on our ability to infer skill or impairment with visual search, spatial reasoning, working memory, and motor proficiency during complex real-world tasks. Thus, we hope to fill a critical need for an exam with the resolution capable of determining deficits which subjective or reductionist assessments may otherwise miss.**

## I. INTRODUCTION

The trail making test (TMT) is a common, pen and paper, neuropsychological assessment that is used worldwide to assess cognitive function including processing speed, executive function, sequencing, and mental flexibility [1]. Good performance on the TMT requires the participant to simultaneously engage their attention (search), memory, and sensory-motor control. In the standard, two-part, TMT the participant is shown an 8x11 inch paper with small circles containing numbers and/or letters printed on the page. In the first part, TMTA, the participant is asked to connect numbers in ascending order by drawing lines from one circle to the next. The second task, TMTB, requires the participant to connect numbers and letters in an alternating ascending order, i.e., 1-A-2-B- and so on. If the participant makes an error, they have to return to the last correct circle and draw the correct path. Performance on the standard TMT is assessed as total completion time, measured with a stopwatch and, in some cases, the number of errors. The TMT is used alone and as part of assessment batteries such as the Montreal Cognitive Assessment (MoCA) [2]. A number of neurological and psychiatric conditions, such as traumatic brain injury, depression, and stroke can affect performance on the TMTA and TMTB [1]. Additionally, age has been positively associated with an increase in completion time [3], [4]. There have been attempts to use the TMT, in combination with other screening tools, to predict function by measuring performance on tasks such as driving a car [5]. However, it has been found that such associations are modest, weakening the traditional TMT's potential as an effective predictor of functional ability among the healthy population [6]. Another limitation of the current version of the TMT is that it does not provide information that can be used to distinguish between different types of impairments, especially if the impairments are at the intersection of cognitive and motor domains. The scientific premise for the current investigation is that by expanding the TMT paradigm to resemble a task that emulates real-world behavior, we may be able to glean predictive validity of the test toward functional behaviors in the real world. Furthermore, by integrating more sophisticated technology to capture more fine-grained performance on the TMT, that has not been previously measured, we may be able to assess more specific characteristics of function (and dysfunction) [7], [8].

We developed, and report here as a proof of concept, a functional version of the trail making task that was capable of assessing an individual's performance at the intersection of both cognitive and motor domains while simulating the real-world task of retrieving canned goods from a kitchen cabinet. Eye-tracking technology was incorporated to extract precise information about gaze position and timing during the test. We demonstrate concurrent validity of the real-world TMT with the traditional paper TMT by correlating individual completion times and testing for the presence of well-established (on the paper test) age-dependent effects.

## II. METHODS

### A. Subjects

All protocols were conducted in conformance with the Declaration of Helsinki and were approved by the Institutional Review Board of Northeastern University. Thirty subjects, ages 18-86 years old (yo) participated after providing institutionally approved consent. All were right-handed and

* This research was funded in part by the Global Resilience Institute at Northeastern University NIH-2R01NS085122 (E.T.), NIH-2R01HD058301 (E.T.), NSF-CBET-1804550 (E.T.), and NSF-CMMI-M3X-1935337 (E.T., M.Y.), NIDDLR-90RE5028 (H.J., M.P)

NP, HJ, and MP are with the Dept. of Health Sciences, Bouvé College of Health Sciences, Northeastern University, Boston, MA 02115, USA, pinkes.n@northeastern.edu

MY and ET are with the Department of Physical Therapy, Movement and Rehabilitation Science, Bouvé College of Health Sciences, Northeastern University, Boston, MA 02115, USA, {m.yarossi, e.tunik}@northeastern.edu

ZF is with the College of Engineering, Northeastern University, Boston, MA 02115, USA, fagiani.z@northeastern.edu

HJ and MP are with the Khoury College of Computer Sciences, Northeastern University, Boston, MA 02115, USA, {h.jimison, m.pavel}@northeastern.edu

EW is with the College of Science, Northeastern University, Boston, MA 02115, USA wong.et@northeastern.edu

free of neurological and orthopedic conditions that could interfere with the task. For analysis, the cohort was stratified into an under 50 yo (U50; 14 subjects, 6 female; mean age: 25±4 yo) and over 50 yo (O50, 16 subjects, 9 female; 66±9 yo) group.

### B. Traditional Paper Based TMT

Subjects were seated at a table for the pen-paper TMT task and asked to complete the TMT per established protocol [1]. Briefly, subjects were instructed to connect numbers (for TMTA; e.g., 1,2,3,4...) or alphanumeric sequences (for TMTB; 1,A,2,B,3,C...) in ascending order as quickly and as accurately as possible, without lifting the pencil from the paper from beginning to end. Subject first performed the TMTA, then the TMTB. The alphanumerics were shown on an 8x11 inch sheet of paper, with each number or letter shown inside a small circle with a standardized position on the page such that a line can be drawn between any two consecutive numbers without crossing another line. Prior to each task, subjects were familiarized with the task using a reduced set of numbers.

Completion time, recorded using a stopwatch, on the TMTA and the TMTB was measured as the time interval between when the pen left the circle of the first number to when the pen entered the last circle. If the subject made an error, it was immediately verbally pointed out, and the subject was instructed to move to the previous number or letter and proceed from there per the guidelines in table 1 of Bowie and Harvey [1]. The time spent correcting the errors is included in the total task completion time.

### C. Real World TMT Task

The novel real world version of the TMT (Can-TMT) was designed as a can retrieval task from an open cabinet with two shelves. Ten cans were placed on the shelves that were either labeled with the numbers 1–10 (Can-TMTA) or 1-5 and A-E (Can-TMTB). Each can was labeled with its respective number or letter and four unique visual fiducial markers called AprilTags. The markers were utilized for placement of a computer vision-derived bounding box that outlined each can's surface [9] **(Figs. 1 and 2)**


Fig. 1: The can set up for TMTA

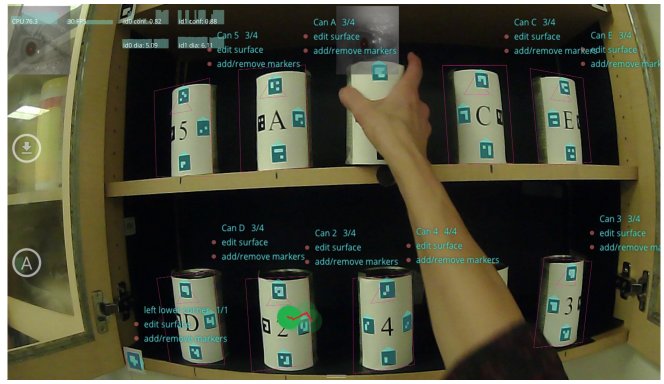


Fig. 2: The set-up for Can-TMTB as seen through the worldview camera on the eye tracker. The green dot indicates the subject's gaze recorded with the eye tracker. The fiducial markers can be seen on each can

Each subject completed the Can-TMTA followed by the Can-TMTB while wearing an head mounted eye tracker, as described below . Subjects began by looking at a black dot that was positioned just below the cabinet. *Can-TMTA*. Each subject was instructed to collect labeled cans from the shelf in ascending order from 1-10 and place them on the counter below, removing can 1, then can 2, and so on until no cans were left. Instructions were to complete the task as quickly and as accurately as possible and to only use their dominant hand. *Can-TMTB*. The Can-TMTB was performed the same way, except that subjects were required to retrieve the cans in ascending alphanumeric order (1,B,2,C...), ending the sequence at can "E". Any errors on the can TMT task were handled the same way as on the pen-paper task; the error was pointed out verbally and the subject had to return the incorrect can to it's position, move the hand back to the last can placed on the counter (the last correct can), and resume the task.

### D. Eye Tracking

A Pupil Labs Core tracker (Pupil Labs, Berlin, Germany) was used for binocular eye tracking [10] with data collected using the open source software suite provided by Pupil Labs (https://github.com/pupil-labs). Two infrared cameras (200 Hz, 640X480 pixels) captured eye images and one RGB camera (60 Hz, 1920X1080 pixels) mounted to the eye glass frames captured the subject's view of the scene. Each eye camera was adjusted to best capture its respective pupil (confidence of 0.95 or better for each eye model). The scene camera was adjusted to capture the study environment in its entirety. A manual single marker calibration utilizing the vestibulo–ocular reflex was used for calibrating the eye tracker. Task completion time was recorded using the eye tracking data as the duration between the time at which the eye gaze entered the bounding box of the first object and the time at which the object entirely left the scene from the world view camera. As in the pen-paper task, any task errors were included in the measurement of task completion time.

Preprocessing of eye tracking data was performed in Pupil Player (Pupil Labs, Berlin, Germany). Each trial was individually reviewed for overt errors that could affect the data quality. Surface tracking in Pupil Player utilized visual fiducial markers, AprilTags, from the APRIL Robotics Laboratory at the University of Michigan [9]. As described, each can was labeled with four markers to allow for robust surface tracking. The Pupil Player program allowed for predetermined bounding boxes to be placed around each can so that gaze data on each can could be analyzed. The bounding boxes used to define the can surface were the same for all cans and for all subjects. Following surface tracker activation, each trial was watched again to ensure stability of the surface tracking. If tracking was unstable, the video parameters were adjusted and the trials re-processed. When tracking was satisfactory, the processed data was exported to Matlab for analysis. Data that were available for analysis included, for each can: 3D can position, x-y eye position for each eye, 2D and 3D gaze estimation position, the duration of time spent looking at the can, and time stamps. Only the gaze position data and time stamps were analyzed for this publication.

### *E. Data Analysis*

Two methods were used for comparison of Can-TMT and Paper-TMT performance. The Pearson's correlation coefficient describing the relationship between the two test modalities (Can-TMT, Paper-TMT) was calculated separately for the TMTA and the TMTB versions of the tasks. To normalize for the different number of items in Can-TMT (10) and Paper-TMT (25) each completion time was divided by the number of items in order to obtain an average time per item. Separate 2x2 mixed factorial ANOVAs ($\alpha$ = .05) with a within-subjects factor of TEST_MODALITY (Can-TMT, Paper-TMT), and a between-subjects factor of AGE (U50,O50) were performed for the Time-Per-Item on the TMTA and TMTB. Significant main effects and interactions were followed with post-hoc t-test with Bonferonni correction for multiple comparisons.

## III. RESULTS

All subjects completed all components of the experiment, and there were no reported adverse events. The completion times for the Paper-TMTA were $21.42 \pm 5.87$ seconds for the U50 group and $32.45 \pm 12.809$ sec. for the O50 group (Fig 3). Completion times for the Paper-TMTB were $34.97 \pm 9.15$ sec. for the U50 group and $59.45 \pm 26.99$ sec. for the O50 group (Fig 4). Completion times for the Can-TMTA were $15.45 \pm 2.38$ sec. for the U50 group and $20.53 \pm 4.34$ sec. for the O50 group. Completion times for the Can-TMTB were $14.59 \pm 2.33$ sec. for the U50 group and $18.50 \pm 4.34$ sec. for the O50 group.

There was a significant correlation between Can-TMTA and Paper-TMTA completion times (r=0.704, p<0.001) **(Fig. 3 Top.)**. For Time-Per-Item on the TMTA there was a significant main effect of TEST_MODALITY [F(1,28)=128.07, p<0.001] a significant main effect of AGE [F(1,28)=14.43, p=0.001], but no significant TEST_MODALITY x AGE interaction [F(1,28)=0.269, p=0.608]. Post-hoc t-tests with

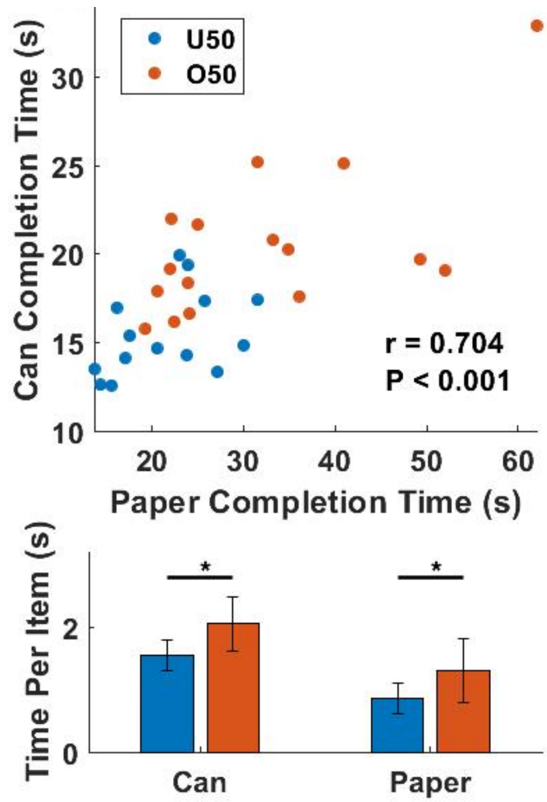


Fig. 3: **TMTA.** *Top.* Correlation between Can-TMT and Paper-TMT. *Bottom.* Per-Item-Time for two different age groups. * indicate significant differences p<0.025.

Bonferroni correction for two comparisons ($\alpha$=.025) revealed a significant difference between U50 and O50 participants for the Can-TMTA [t(28)=3.89, p=0.001], and the Paper-TMTA [t(21.63)=3.09, p=0.005] **(Fig. 3 Bottom.)**.

There was a significant correlation between Can-TMTB and Paper-TMTB (r=0.737, p<0.001) **(Fig. 4 Top.)**. For Time-Per-Item on the TMTB there was a significant main effect of AGE [F(1,28)=11.73, p=0.002] and a significant TEST_MODALITY x AGE interaction [F(1,28)=5.972, p=0.21, but no significant main effect of TEST_MODALITY [F(1,28)=3.79, p=0.062]. Post-hoc t-tests with Bonferroni correction for two comparisons ($\alpha$=0.025) revealed a significant difference between U50 and O50 participants for the Can-TMT [t(28)=3.01, p=0.006], and the Paper-TMT [t(18.82)=3.41, p=0.003] **(Fig. 4 Bottom.)**.

## IV. DISCUSSION

In this study, we set out to determine if a more functionally relevant task, emulated the TMT, could capture similar performance features. Our data highlight a strong and significant correlation between the real world and paper versions of the TMT (TMTA and TMTB versions) establishing criterion validity of the real world task. This indicates that inter-subject variation in completion times is likely to be conserved in the real world task. Moreover, the real world TMT task demonstrated statistically significant capacity to distinguish between the two age strata tested in our study, providing further confirmation of the test's validity. This work sets a

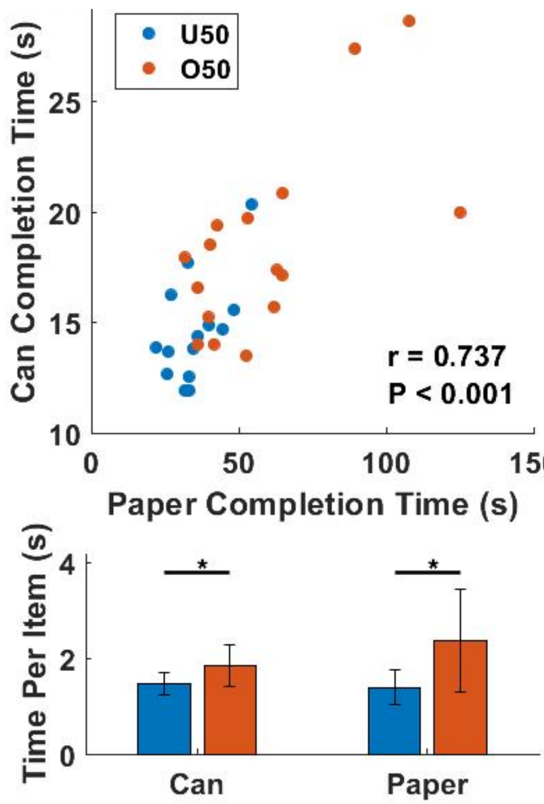


Fig. 4: **TMTB.** *Top.* Correlation between Can-TMT and Paper-TMT. *Bottom.* Per-Item-Time for two different age groups. * indicate significant differences p<0.025.

foundation to exploit the power of an instrumented real world version of the TMT that can capture cognitive-motor domains in a functional behavioral paradigm. Although we did not analyze all of the available eye tracking data collected in this study, future investigations will be able to explore more granular features of the visual search, such as the time spent searching for each item, the time spent fixating on each item, the number of items searched, and the number of saccades used, thereby allowing us to develop computational models of visual search efficiency, working memory, and cognitive capacity for set switching [7], [8]. We are currently exploring computer vision-based motion tracking and pose estimation via software such as DeepLabCut [11] to track motion of the hand/arm during such experiments to further facilitate adaptability of this task to the real world. This work will have a significant impact on our ability to infer skill or impairment with visual search, spatial reasoning, working memory, and motor proficiency during complex real-world tasks. Thus, we hope to fill a critical need for an exam with the resolution capable of determining deficits which subjective or reductionist assessments may otherwise miss. Such information could be used to develop person-specific profiles to plan rehabilitation programs or interventions and generalize other laboratory-based assessments to the real world. This work has important implications for detecting so called "hidden impairments" in aging populations, individuals with stroke [12], and mild cognitive impairment that may only be revealed during complex cognitive-motor interactions such as locating and retrieving items from a store shelf.

## V. ACKNOWLEDGMENT

We would like to thank Leila Kiernan DPT, Meghan MacDonald DPT, Alison Natter DPT, Katrina Pollard DPT, Daniel Tsai DPT, Bailey Uitz DPT for help with data collection and project development.

## REFERENCES

[1] C. R. Bowie and P. D. Harvey, "Administration and interpretation of the trail making test," *Nature Protocols*, vol. 1, no. 5, pp. 2277–2281, 2006. [Online]. Available: https://doi.org/10.1038/nprot.2006.390

[2] Z. S. Nasreddine, N. A. Phillips, V. Bédirian, S. Charbonneau, V. Whitehead, I. Collin, J. L. Cummings, and H. Chertkow, "The montreal cognitive assessment, moca: a brief screening tool for mild cognitive impairment," *Journal of the American Geriatrics Society*, vol. 53, no. 4, pp. 695–699, 2005.

[3] T.-B. R. Wahlin, L. Bäckman, Å. Wahlin, and B. Winblad, "Trail making test performance in a community-based sample of healthy very old adults: Effects of age on completion time, but not on accuracy," *Archives of Gerontology and Geriatrics*, vol. 22, no. 1, pp. 87–102, 1996.

[4] P. Amodio, H. Wenin, F. Del Piccolo, D. Mapelli, S. Montagnese, A. Pellegrini, C. Musto, A. Gatta, and C. Umiltà, "Variability of trail making test, symbol digit test and line trait test in normal people. a normative study taking into account age-dependent decline and sociobiological variables," *Aging clinical and experimental research*, vol. 14, no. 2, pp. 117–131, 2002.

[5] T. A. Novack, J. H. Baños, A. L. Alderson, J. J. Schneider, W. Weed, J. Blankenship, and D. Salisbury, "Ufov performance and driving ability following traumatic brain injury," *Brain injury*, vol. 20, no. 5, pp. 455–461, 2006.

[6] P. Vaucher, D. Herzig, I. Cardoso, M. H. Herzog, P. Mangin, and B. Favrat, "The trail making test as a screening instrument for driving performance in older drivers; a translational research," *BMC geriatrics*, vol. 14, no. 1, pp. 123–123, 2014.

[7] S. Hagler, H. B. Jimison, and M. Pavel, "Assessing executive function using a computer game: Computational modeling of cognitive processes," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 4, pp. 1442–1452, 2014.

[8] T. Singh, J. Fridriksson, C. M. Perry, S. C. Tryon, A. Ross, S. Fritz, and T. M. Herter, "A novel computational model to probe visual search deficits during motor performance," *Journal of neurophysiology*, vol. 117, no. 1, pp. 79–92, 2017.

[9] E. Olson, "AprilTag: A robust and flexible visual fiducial system," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 5 2011, pp. 3400–3407.

[10] M. Kassner, W. Patera, and A. Bulling, "Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction," in *Adjunct Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '14 Adjunct. New York, NY, USA: ACM, 2014, pp. 1151–1160. [Online]. Available: http://doi.acm.org/10.1145/2638728.2641695

[11] A. Mathis, P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis, and M. Bethge, "Deeplabcut: markerless pose estimation of user-defined body parts with deep learning," *Nature neuroscience*, vol. 21, no. 9, pp. 1281–1289, 2018.

[12] H. R. Williams, L. Y. Leung, C. M. Gordon-Davis, E. Tunik, M. Pavel, M. Yarossi, and H. B. Jimison, ""a huge, life-changing thing:" a qualitative survey of the priorities and goals of young stroke survivors," *Archives of Physical Medicine and Rehabilitation*, vol. 100, no. 7, pp. e5–e6, 2019.