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

A trial study of using DSST to evaluate cognitive impairment in older adults

Rahul Dasharath Gavas, Kartik Muralidharan, Ramesh Kumar Ramakrishnan
Ramesh Balaji, Harish Kumar, Srinivasa Raghavan Venkatachari
TCS Research, India

Abstract— **Non-invasive means of monitoring mild cognitive impairments (MCI) is recently gaining popularity. With the advent of easy to use physiological sensors, there have been an outburst of studies from the last decade which aim at detecting a target's mental health condition. However, not many studies present the experience or insights gained from carrying out such in-situ research work, particularly when working with older adults. Such insights could not only assist researchers in related areas when designing their study but also avoid potential pitfalls. Clinical trials were conducted by our organization in collaboration with the Geriatric Educational Research Institute, Singapore (GERI) and Singapore Management University (SMU) for detecting mild cognitive impairments in a geriatric population. Digitized versions of the standard pen & paper psychological tests were used along with gaze tracking technologies for MCI detection. Details of our user study and it's outcomes are discussed as well as a generic approach of digitizing any given psychological test battery is highlighted.**

I. INTRODUCTION

There is a significant rise in the number of senior citizens living independently, particularly in south east Asia. This has led to an active research focus in the wellness and care of ageing population [1]. By heavily instrumenting this specific group with technology, we could possibly perform better prediction and prevention of certain ailments. These include postural stability, health parameters like heart rate, calories burned, mild cognitive impairment (MCI) and so on.

In order to better understand the nuances in monitoring and obtaining these wellness measures TCS Research in collaboration with the Geriatric Educational Research Institute (GERI) and Singapore Management University(SMU) initiated a trial study wherein an identified population would be instrumented with smart devices and wearables to measure a set of identified wellness parameters.

In this work, we share our experiences in the terms of working with these participants and technology with a specific focus on our key objective of MCI detection, challenges that we have faced during its implementation and capture the lessons learned.

For detecting MCI, we deployed digitized variants of a well-known pen and paper-based psychological test battery, i.e. the Digit Symbol Substitution Test (DSST) [2]. This test aims at assessing cognitive functions like visual scanning, processing speed, working memory, cognitive processing and motor response [3]. Analogous to this, the inverted version of DSST i.e. the symbol digit substitution tests [4] are also been employed as a standard test for cognitive analysis.

Though there exists several pen and paper versions (pDSST) of the DSST [2], the need to have a digitized version was mainly because of the following reasons. The only two metrics of evaluation in pDSST are the total response times and the total score whereas the trial-wise granularity cannot be ascertained with pDSST. The inclusion of physiological sensors like eye tracking, electroencephalogram (EEG), etc is difficult in case of pDSST as the participants are supposed to write down the answers which results in muscle artifacts in the acquired signals. Also, since the mode of inputs in the pen paper variants is in written form from the participants, it comes with additional overload for the ageing category. In our initial pen and paper trials with the clinical participants, most of them were unable to take the pDSST as the writing down of around 100 entries in tiny boxes seemed to be a difficult task for many. This motivated us to have a test which requires minimal input but maximizes the number of output variables that would help us to get insights about the mental health. Since, in the digitized version, for every trial, we can either show a correct or an incorrect matched pair of digits and symbols; we therefore have 2 inputs. We further reduced it to one by restricting the inputs for only a correct match.

These tests deals with harnessing of fluid cognition, which therefore makes it a crucial marker of cognitive functions, age related variations and decline of cognitive performances [5]. Hence, this paper is mainly based on the explorations on the digitized versions of DSST (dDSST) [6], [7] to clinical trials. The variants of dDSST were used primarily to diagnose the effects of age and the effects were found to be dominant in gaze related features [6]. The study was further extended for clinical trials and the experience gained while handling the clinical participants are discussed in brief in this paper.

II. USER STUDY SETUP

DSST is a well-known evaluation tool for assessing cognitive functioning. The test involves a lookup table with numbers from 1 through 9, each of which is paired with a unique, easy-to-draw symbol such as '$>$' ,'V', or '$+$'. Below the lookup table are a series of numbers from 1 to 9 in random order, along with an empty box associated with it. The participant is allowed to fill in those gaps with corresponding symbols for every number. The time duration is 90-120 seconds depending on the test version. The task requires the participant to scan the corresponding number-symbol entry provided in the lookup table and then write

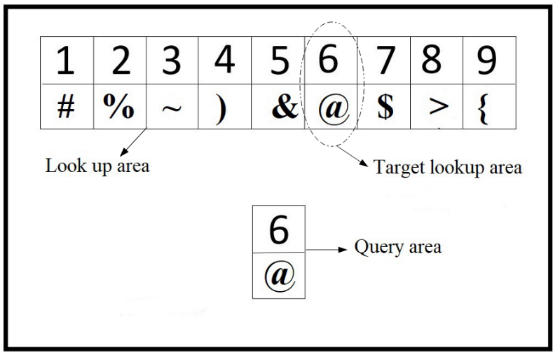

Fig. 1: Digitized DSST (dDSST) [6]. Three versions of this test are: (i) *Version 1*: The lookup area (LUA) entries are fixed for all trials while the query area (QA) entries changes randomly with trials. (ii) *Version 2*: The entries in LUA and QA changes randomly with trials. (iii) *Version 3*: The LU entries are fixed, the entries and the position of QA changes randomly with trials.

the correct symbol in the empty box provided against each number.

DSST assesses an assortment of cognitive functions. Good performance in DSST requires intact attention, motor speed and visio-perceptual functions, including scanning and ability to write/draw (basic manual dexterity). Associative learning can also affect the performance. For instance, if number-symbol associations are rapidly learned following the first few trials, then the performance speed improves as the participant does not require to check the association for every test entry. Consciously engaging oneself in this learning strategy to improve the performance speed calls for the executive functions of planning and strategising. Working memory, another executive function, is likely required to hold in mind the task rules and for the continual updating of required digit-symbol pairs.

A schematic layout of our proposed digitized DSST test is shown in Fig. 1. The layout includes the following: (i) A lookup area (LUA) which is fixed at the top containing the digit-symbol pairs. (ii) query digit-symbol pair appearing at a specific position on the screen, termed as query area (QA). (iii) Target LUA (TLUA) which is the region of LUA having same digit as QA. The participant is supposed to press the space bar button when the QA entries match with that of the TLUA. If there is no match, the participant is supposed to wait 3 seconds for the next trial. For every trial, the participant checks the digit-symbol pair in the QA, search for the same digit in the lookup region, match the digit-symbol pairs of TLUA and QA and respond accordingly.

To get additional insights about the cognitive functions of the participant, we used an infrared eye tracker with 60 Hz sampling rate from Gazepoint [8] to study their gaze patterns. The eye tracker is placed below the computer screen. As the matching step is at the core of the DSST, gaze analysis is a good means of studying the behavior of the participants. The usage of gaze tracking is beneficial in understanding the implications of paired associations made during recall. Completing a trial, i.e. matching a digit to its symbol, without using the LUA is indicative of the use of learned/memorized paired associations. In order to study these effects, we selected different versions of the DSST. In one of the versions, the lookup table entries change with trials. This provides vital information which elucidates the importance of paired associations. A typical experiment setup using eye tracker and a chin rest is as shown in Fig. 2.

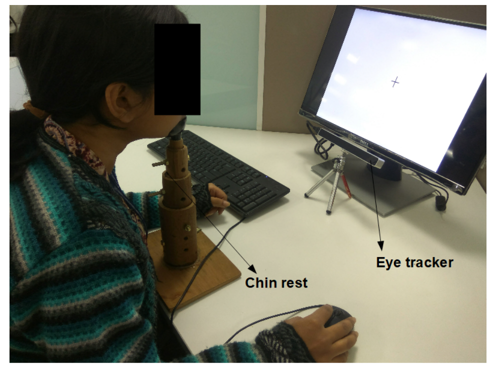

Fig. 2: Experimental setup

Each participant was required to perform the DSST-90 test (a pen & paper version) followed by three variants of our computerized version of the DSST. The computerized version of the test required participants to perform an initial calibration step necessary for the eye tracker. Participants were briefed on the study while a staff member demonstrated how to press the spacebar when the symbols matched. The participants where then administered the test. Below is a description of the dDSST variants.

1) Version 1: This version is like the conventional pen and paper DSST. Here, the entries in LUA are fixed for each trial and the QA appears at the centre of the screen. Test-specific parameters like response time per trial, total time, score along with the metrics corresponding to gaze analysis are recorded. This version aimed at studying the working memory.

2) Version 2: This version is like version 1 except that the entries in LUA changes pseudo-randomly with each trial. The randomness in presenting the entries of LUA helped in overcoming the possibilities of participants memorizing the LUA entries.

3) Version 3: This version aims at studying the performance characteristics when the QA location changes with trials and the participant is expected to have better spatial visuomotor coordination for accomplishing the task. It also intends to assess the effects associated with positional changes of the QA that might indicate one's visual neglect.

The study involved a total of 35 participants (*Female=20, Male=15*) with an average age of 71 years. Based on diagnosis for MCI by a medical professional, the participants were placed into two double blind groups (*Group A=14, Group B=21*). However, out of these 35 participants only 28 (*Group A=10, Group B=18*) performed the computerized DSST and 15 (*Group A=6, Group B=9*) completed the initial eye-tracker calibration step correctly. Institutional ethics review board approvals were obtained (reference number: 2015/01076).

TABLE I: User Study Participation Statistics

Total Number of Participants	With Computerized DSST Scores	With MCI Grouping Information	With DSST-90 (pen & paper) Scores	With Eye Tracking Data (calibrated)	With Eye Tracking Data (uncalibrated)
35	29	28	35	17	18

	V1 Score	V2 Score	V3 Score	V1 Time	V2 Time	V3 Time	DSST-90
V1 Score	1						
V2 Score	0.70	1					
V3 Score	0.62	0.88	1				
V1 Time	-0.39	-0.22	-0.16	1			
V2 Time	-0.52	-0.13	-0.13	0.75	1		
V3 Time	-0.55	-0.36	-0.16	0.66	0.73	1	
DSST-90	0.63	0.37	0.26	-0.43	-0.54	-0.62	1

Fig. 3: Correlations in the test metrics

III. Results and Discussions

We initially formulated the following hypotheses. *Null Hypothesis* H_0: Participants diagnosed with MCI will not exhibit significant performance degradation (test time, test score) when subjected to the dDSST. *Null Hypothesis* H_1: Participants diagnosed with MCI will not exhibit any significant patterns in eye tracking features when subjected to the dDSST.

Fig. 3 shows the inter-correlations of the test metrics between the pDSST and the 3 variants of dDSST. It is to be noted that the correlation between the pDSST score and the version 1 of dDSST is quite high. This is obvious as the base version of dDSST, i.e. version 1 is similar to pDSST when compared to the other two versions.

However, this score metric of dDSST is not statistically significant in distinguishing the two participant categories under test. This can be attributed to the several unforeseen variables corresponding to MCI which cannot be easily captured using the conventional features. We therefore explored other features like the key hold time, gaze related features like gaze durations in TLUA, LUA and QA mentioned in [6]. In addition to this, we further explored the blink related features like average blink counts, blink rate variability related time domain parameters [9] obtained by considering the series formed from inter blink gaps. Table II shows the results by considering the top performing features. Note that as the main objective of this paper is to summarize the experiences of the clinical trial, we don't discuss the results in finer details.

TABLE II: Results of the User Study. The only non-significant difference (using student's t-test with $p = 0.05$) is between the DSST-90 Group A and Group B results. All other differences are significant at $p = 0.05$.

	Group A	Group B	p Value
DSST-90 (mean score)	34.5	31.7	0.2
dDSST (mean time (s))	130.6	134.1	0.04
Blink Rate Variability (mean time between blinks)	1.3	0.6	0.02

As mentioned earlier in Section II the participants were diagnosed by a medical professional for MCI and placed into two double blind groups. We performed a student's t-test across the multiple variants between the two groups. While no difference was observed between the groups for the pen and paper version and it's equivalent computerized version, we did observe a significant difference in the time taken to perform the dDSST test as well a difference in the skewedness of the blink rate variability.

While the small sample size prevents us from drawing any conclusion, the observations do encourage us to explore this domain further. The study also provided us insights to improve our protocol which we talk about in the next section.

IV. Guidelines/Future Roadmap

In particular to MCI detection using dDSST, this exercise helped us gain the following insights based on participant feedback and our own data observation:

(i) **Eye-tracker Calibration**: Systematic error, essentially a static offset from the gazed target location, is a common error in most eye-trackers. Hence, the calibration is essential to eliminate this error. In our work, several participants failed to perform this calibration step correctly rendering the eye tracking data unusable. We therefore need to investigate methods for reduction/removal of this eye calibration step.

Next Steps: The current dDSST layout is inspired from the traditional pen and paper test. This layout is concentrated at the centre of the screen and hence requires more precise tracking. We are therefore investigating a layout that utilizes the entire screen without impacting the test outcome. Fig. 4 shows the new design with a sample simulated gaze map for a given trial in the DSST. For an ideal search task, the scanpath follows a trail from the QA to the target LUA and back to the QA. Quick results have shown that even with a systematic error above 2 degrees, owing to the sparsely placed entities, the gazed entity can be predicted based on the nearest neighbor approach [10]. This design also seems suitable for use with a webcamera for eye-tracking - which we are exploring in order to avoid use of the chin-rest.

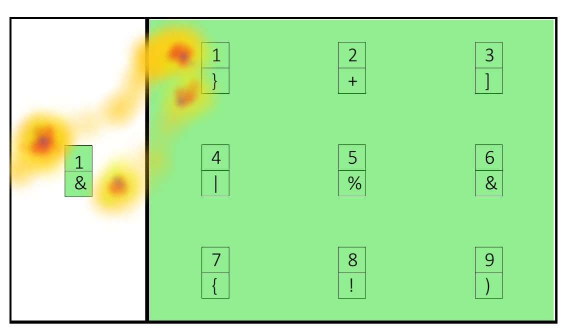

Fig. 4: Sparse designs for DSST to solve the issues with eye tracking

(ii) **Elimination of the chin rest**: Eye tracking studies typically require chin rests in order to reduce any errors due to head movements during the data collection exercise (Fig. 2). In our study, feedback received clearly indicated discomfort in prolonged usage of the chin rest and hence opens up a research thread to explore methods of eliminating it.

Next Steps: We are currently exploring techniques to perform pose independent (and distance independent) eye tracking using a webcamera. This will hopefully not only eliminate the need for a chin-rest but will also provide a scalable solution as a dedicated eye-tracking device will no longer be needed. As a first step we have implemented a gaze estimation algorithm using deep learning techniques utilizing the RT-GENE dataset [11]. The gaze information is converted to screen coordinates using a regression based model giving us close to 100% accuracy for a 3×3 grid with free head movement within a head box size, Fig. 5.

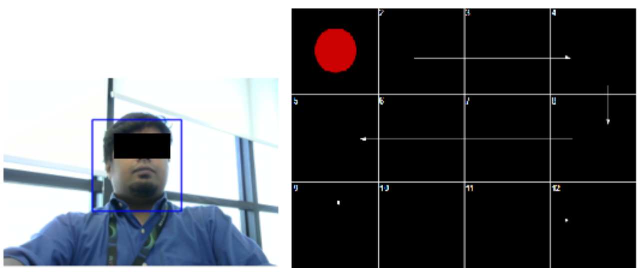

Fig. 5: Illustration of the in-house developed 3×3 grid-based eyetracker

(iii) **Duration of the test**: Participants also provided feedback with regards to the duration of the study. While this study required participants to perform the dDSST over multiple variants (which will be reduced to one in a normal scenario) we are investigating options of decreasing the test duration without compromising on the outcome.

Next Steps: In our current dDSST version, participants are evaluated over 50 trials. This is in contrast to the pen and paper version where the test is time bound. We are exploring how to optimize these parameters, number of trials and time, to find the sweet spot between the two.

(iv) **Incentives**: Incentives plays a very important role in acquiring good quality data from the test participants [12]. Non clinical data collection to be specific, requires proper incentives to be given to the participants in order to keep them well motivated during the experiment, lack of which leads them to lose their interest during the task. In case of clinical trials, the participants, however, are still motivated as they consider it to be a part of their medication process. However, such clarifications need to be made well in advance and proper written consents have to be taken along with IRB approvals.

V. Conclusions

For any diagnostic type of studies which basically involves 2 class classification (healthy and unhealthy), there exists two sorts of test batteries, viz., physio and mental oriented tests. In case of MCI, the effects are manifested in both the above test batteries. There are standard conventional pen and paper tests which have been used so far. However, as seen in the results, an additional sensor aids in providing multiple variables to test the effect, which seems to be beneficial over the conventional test metrics. Test batteries usually come with a norm of values created by considering a large sample of participants from various age groups, demographies, educational backgrounds, gender and so on. Hence, any study which target the digitization of such batteries, should undergo the same set of exercise and create suitable norms. If any physiological sensing is involved, the norms should also contain the metrics corresponding to the dominant features obtained from the sensors. This comes with additional overhead of sensor calibration, sensor data preprocessing and feature engineering.

VI. Acknowledgements

We would like to thank the Geriatric Education and Research Institute (GERI), Singapore and the School of Information Systems, Singapore Management University, Singapore for facilitating and assisting us in various stages of this work.

References

[1] Hoimonty Mazumder, Md Mahbub Hossain, and Anupom Das, “Geriatric care during public health emergencies: lessons learned from novel corona virus disease (covid-19) pandemic,” *Journal of Gerontological Social Work*, vol. 63, no. 4, pp. 257–258, 2020.

[2] Shobini L Rao, DK Subbakrishna, and K Gopukumar, *NIMHANS neuropsychology battery-2004, manual*, National Institute of Mental Health and Neurosciences, 2004.

[3] Stephen Joy, Edith Kaplan, and Deborah Fein, “Speed and memory in the wais-iii digit symbol—coding subtest across the adult lifespan,” *Archives of Clinical Neuropsychology*, vol. 19, no. 6, pp. 759–767, 2004.

[4] Claudia Cornelis, Livia Joanna De Picker, Wouter Hulstijn, Glenn Dumont, Maarten Timmers, Luc Janssens, Bernard GC Sabbe, and Manuel Morrens, “Preserved learning during the symbol digit substitution test in patients with schizophrenia, age-matched controls and elderly,” *Frontiers in psychiatry*, vol. 5, pp. 189, 2015.

[5] Timothy A Salthouse, “The processing-speed theory of adult age differences in cognition.,” *Psychological review*, vol. 103, no. 3, pp. 403, 1996.

[6] Chatterjee et al, “Evaluating age-related variations of gaze behavior for a novel digitized-digit symbol substitution test,” *Journal of Eye Movement Research*, vol. 12, no. 1, 2019.

[7] Chatterjee et al, “Eye movements-an early marker of cognitive dysfunctions,” in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 4012–4016.

[8] “Gazepoint,” `https://www.gazept.com`, [Online; accessed 27-April-2021].

[9] Gavas et al, “Blink rate variability: a marker of sustained attention during a visual task,” in *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*, 2020, pp. 450–455.

[10] Yunfeng Zhang and Anthony J Hornof, “Mode-of-disparities error correction of eye-tracking data,” *Behavior research methods*, vol. 43, no. 3, pp. 834–842, 2011.

[11] Tobias Fischer, Hyung Jin Chang, and Yiannis Demiris, “RT-GENE: Real-Time Eye Gaze Estimation in Natural Environments,” Oct. 2018, This work was supported in part by the Samsung Global Research Outreach program, and in part by the EU Horizon 2020 Project PAL (643783-RIA).

[12] Georg-Christoph Haas, Frauke Kreuter, Florian Keusch, Mark Trappmann, and Sebastian Bähr, “Effects of incentives in smartphone data collection,” *Big Data Meets Survey Science: A Collection of Innovative Methods*, pp. 387–414, 2020.