

Design and Formative Evaluation of Cognitive Assessment Apps for Wearable Technologies

Pegah Hafiz & Jakob E. Bardram

Department of Health Technology, Technical University of Denmark
Lyngby, Denmark
[pegh.jakba]@dtu.dk

ABSTRACT

Cognitive functioning is a crucial aspect of the individual's mental health and it affects human's daily activities. We have developed the Ubiquitous Cognitive Assessment Tool (UbiCAT) including three cognitive assessment apps on the Fitbit smartwatch. In this paper, we present the design and formative evaluation of the UbiCAT apps conducted with 5 participants who had a background in design and/or human-computer interaction. Moreover, we investigated the adoption of the wearable devices by our participants.

CCS CONCEPTS

• **Human-centered computing** → *Ubiquitous and mobile devices*;

KEYWORDS

cognition; cognitive function; wearable technology; application; stroop test

ACM Reference Format:

Pegah Hafiz & Jakob E. Bardram. 2019. Design and Formative Evaluation of Cognitive Assessment Apps for Wearable Technologies. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2019 International Symposium on Wearable Computers (UbiComp/ISWC '19 Adjunct)*, September 9–13, 2019, London, United Kingdom. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3341162.3347077>

1 INTRODUCTION

Cognition is a core function to the human daily activities. Cognitive measures include working memory, verbal memory, attention, psychomotor speed and executive function. Human cognition may fluctuate during the day. For example, an individual may have a better attention in the morning. The cognitive fluctuation level may vary among individuals depending on several factors including their age and mental workload. Hence, a personalized model of cognition seems practical for the individuals to reflect on their cognitive performance during the day.

Human cognition and alertness have been previously investigated via mobile apps. Dingler et al. [4] assessed alertness using three short tasks, namely Psychomotor Vigilance Task (PVT),

Go/No-Go (GNG), and Multiple Object Tracking (MOT). Abdullah et al. [1] conducted a study with 20 participants to analyze the fluctuation in individuals' alertness using PVT mobile app called PVT-Touch. Their findings showed that alertness can vary by 30% depending on the timing of the day. The standard PVT test takes 10 minutes [3] but both studies used a short version of the PVT to prevent participants' fatigue. The Cognition Kit app (<https://www.cognitionkit.com/>) is designed for the Apple Watch to assess mood and cognition in clinical context. The cognitive test used in this app is the 'N-back' test [6].

The Ubiquitous Cognitive Assessment Tool (UbiCAT) includes three cognitive assessment apps as well as mobile sensor data collection using Fitbit API. The UbiCAT cognitive assessment apps are short tests to conduct a research and assess cognition in the wild. Through UbiCAT project, we will collect two types of data: (i) cognitive performance of the individuals which are assessed via the apps and (ii) mobile sensor data including physical activity, heart rate, sleep, and mobility data. The ultimate goal of the UbiCAT project is to find correlation between individuals' cognitive performance and objective mobile data. The objectives of this paper are two-folded. First, the design and formative evaluation results of the UbiCAT apps will be presented. Second, the challenges regarding the wearable technologies for cognitive assessment will be discussed.

2 DESIGN AND STUDY

The UbiCAT apps are designed in a user-centered design process including 3 expert researchers who hold a PhD in computer science, experimental psychology and cognitive science. The design of the UbiCAT apps was revised after several meetings with the experts. In this section, we will first introduce the UbiCAT apps. Then, the formative evaluation of the apps will be explained in detail.

2.1 Overview of the Apps

The UbiCAT includes three apps namely Color Test (CT), Letter Test (LT) and Arrow Test (AT). Each app provides a set of short instructions as well as the test itself. All tests are timed and users should respond as fast as possible. The apps are implemented on the FitBit Ionic smartwatch. The CT is adapted from the Stroop color-word test [5], which presents a set of color names one by one, each with either the same (congruent) or different (incongruent) ink color. Figure 1 shows a screenshot of the CT with an incongruent stimuli. Users are presented with four colors to select the correct ink color of the stimuli. The Stroop test examines sustained attention and the performance measures are the Stroop effect and average response time.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

UbiComp/ISWC '19 Adjunct, September 9–13, 2019, London, United Kingdom

© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-6869-8/19/09...\$15.00

<https://doi.org/10.1145/3341162.3347077>



Figure 1: A screenshot of the Color test adapted from the Stroop color-word test

The LT is adapted from the ‘N-back’ test, where the stimulus are a set of letters shown one by one and the user is asked to memorize *N* letter back in the sequence. ‘N-back’ is used by psychologists and psychiatrists to assess working memory. Figure 2 shows a screenshot of the LT where *N* is 1 and users should tap on either ‘Yes’ if the current letter (‘T’) is the same as one letter showed back in the sequence or ‘No’ otherwise. This test becomes more difficult as *N* increases.



Figure 2: A screenshot of the Letter test adapted from the N-back test

Choice reaction test has variations such as the GNG and stop signal test. A computer-based four-choice reaction time test is developed and tested on adults by Deary et al. [2]. The AT is a two-choice

reaction time test. In this test, each stimulus is an arrow pointing either to the left or right side. A screenshot of a sample AT test is shown in Figure 3. The users are required to tap on the right/left app button in case the arrow points to the right/left.



Figure 3: A screenshot of the Arrow test adapted from the two-choice reaction time test

2.2 Formative Evaluation

The aim of this study was to (i) evaluate app design and improve the user interface of the apps and (ii) explore the adoption of wearable technologies. The participants and the procedure of this study are presented below.

2.2.1 Participants. The participants of this study were selected from the researchers who had a background in app design and/or human-computer interaction at Technical University of Denmark.

2.2.2 Procedure. A consent form was handed to the participants before beginning the study. Then, we asked the participants about their experience of using a wearable device, the duration, and reason to stop using it (if any). The procedure to work with each app involved three steps. First, the participant was asked to wear the FitBit smartwatch and launch the app, read the test instructions, take a test, and view his/her score. We asked the participants to verbalize their thought as in ‘think-aloud’ method. Then, participants were asked to fill in a questionnaire form including 7 questions. We selected relevant questions of a reliable and objective tool called Mobile App Rating Scale (MARS) [7] to evaluate UbiCAT apps. The chosen questions were taken from three sections of the MARS namely aesthetics (3 questions), functionality (2 questions) and information (2 questions). The 5-point rating scale was used to give a score to the participants’ response to each question. Finally, a semi-structured interview was conducted with each participant. The participants’ interaction with the apps and the interviews were recorded.

Table 1: Usability results of the UbiCAT apps

	Aesthetics	Functionality	Information
Arrow Test	3.93 ± .61	4.6	4.4
Letter Test	4 ± .2	3.3 ± 1.84	2.6
Color Test	4 ± .2	4.2 ± .28	3.9 ± .14

3 RESULTS

Five people participated in this study (1 female, 4 male; age=28 ± 4.35), 1 with a PhD degree, 3 PhD students and 1 Master's degree student. The experiment took approximately 1 hour per participant. The interviews with the participants were transcribed after the experiment. The video recordings were also checked multiple times to identify where participants struggled with the apps.

3.1 Usability

Table 1 shows the usability results for the UbiCAT apps reported separately in terms of aesthetics, functionality and information. The range of scores in Table 1 is from 1 (lowest) to 5 (highest).

3.2 Wearable Technology Adoption

We investigated whether our participants use a wearable device or not. Of the 5 participants, 3 of them did not need to use a wearable activity tracker:

"I feel a bit overwhelmed when I have too much data about myself. I prefer not to have numbers to define myself. I do not even use a normal watch" (P1).

"[I have used] only normal watches. I have no need for it. No need to track the data" (P2).

"I never felt I have to. I already tracked the things that I wanted on my smartphone on the street or park and I do not think I need it" (P3).

Two of the participants had already used at least one wearable device for various time periods: *"It (Apple watch) did not really give me anything. It was too obtrusive and pervasive. You are always reminded of something. A classic mechanical watch does not disturb me since it does not collect data and it is passive" (P4 used Basic Pick and Apple Watch for several months).*

"I used Nokia smartwatch for six months but I felt I was motivated without it" (P5 used Nokia steel HR).

3.3 Feedback

UbiCAT apps give feedback to the participants when a user responds to a stimuli. A sample screenshot of the feedback to a correct response in the CT test is shown in Figure 4. We received the following comment regarding the feedback to the user's response:

"I am not sure how feedback affects my performance. How could it affect? If it indicates that I entered a response...for couple of times it distracted me to think if I was right or wrong. It is not about the amount of the feedback. Cognitively, I might be doing something else. Maybe I want to reduce my error rate instead of responding as fast as possible" (P4).

The rest of the participants did not report any issue in regard to the app feedbacks during the tests.



Figure 4: A screenshot of a feedback to a correct response in the Color test app

3.4 Test Score

All UbiCAT apps use the same user interface to display a score after the test. Upon finishing a test, the number of correct responses is displayed on the smartwatch screen. During the user-centered design meetings, we decided to make it as simple as possible due to the negative impact of the low score on the individual's mood. The following question was asked during the interviews: *"Did your score at the end of the test help you to understand your performance? If yes, how?"*

All of the participants mentioned that a maximum score is essential for them to understand their performance, for instance, 24 (number of correct responses) out of 30 (maximum score). We perceived that quantifying user's performance is essential.

4 CONCLUSION AND FUTURE WORK

The usability study showed promising outcomes in terms of aesthetics, functionality and information of the UbiAT apps. The score of the information section for the LT gave average score which indicates that the instruction set of this test should be improved. Through this study, we assessed the usability of the UbiCAT apps and investigated the participants' interaction with the wearable devices. The participants' comments regarding wearable devices will help us in improving the usability of the wearable devices in particular smartwatches. We have planned to conduct a study to evaluate the UbiCAT apps with more participants.

ACKNOWLEDGMENTS

This project is developed as a part of PH's PhD. PH is an early stage researcher of Technology Enabled Mental Health (TEAM) for Young People. TEAM has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 722561. We wish to thank our participants.

REFERENCES

- [1] Saeed Abdullah, Elizabeth L Murnane, Mark Matthews, Matthew Kay, Julie A Kientz, Geri Gay, and Tanzeem Choudhury. 2016. Cognitive rhythms: Unobtrusive and continuous sensing of alertness using a mobile phone. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 178–189.
- [2] Ian J Deary, David Liewald, and Jack Nissan. 2011. A free, easy-to-use, computer-based simple and four-choice reaction time programme: the Deary-Liewald reaction time task. *Behavior research methods* 43, 1 (2011), 258–268.
- [3] David F Dinges and John W Powell. 1985. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior research methods, instruments, & computers* 17, 6 (1985), 652–655.
- [4] Tilman Dingler, Albrecht Schmidt, and Tonja Machulla. 2017. Building Cognition-Aware Systems: A Mobile Toolkit for Extracting Time-of-Day Fluctuations of Cognitive Performance. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 47.
- [5] Charles J Golden and Shawna M Freshwater. 1978. Stroop color and word test. (1978).
- [6] Wayne K Kirchner. 1958. Age differences in short-term retention of rapidly changing information. *Journal of experimental psychology* 55, 4 (1958), 352.
- [7] Stoyan R Stoyanov, Leanne Hides, David J Kavanagh, Oksana Zelenko, Dian Tjondronegoro, and Madhavan Mani. 2015. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. *JMIR mHealth and uHealth* 3, 1 (2015), e27.