Using Game Technology to Automatize Neuropsychological Tests and Research in Active Aging

Dan C. Rotaru
Complutense University of
Madrid, UCM, Madrid 28040,
Spain
drotaru@ucm.es

Sara García-Herranz
UNED, Madrid, Spain
sargarcia@psi.uned.es

Manuel Freire
Complutense University of
Madrid, UCM, Madrid 28040,
Spain
manuel.freire@fdi.ucm.es

Iván Martínez-Ortiz
Complutense University of
Madrid, UCM, Madrid 28040,
Spain
imartinez@fdi.ucm.es

Baltasar Fernández-Manjón
Complutense University of
Madrid, UCM, Madrid 28040,
Spain
balta@fdi.ucm.es

Mª Carmen Díaz-
Mardomingo
UNED, Madrid, Spain
mcdiaz@psi.uned.es

ABSTRACT

Computerized neuropsychological tests provide a more systematic and easily administered assessment tool than traditional pen-and-paper tests. We consider that game technology can be effectively applied to decrease the cost of developing computerized versions of traditional tests and can even allow the creation of promising new environments to assess and researching in active aging. To study the feasibility of this approach, we developed a computer version of the 15 Objects Test and compared the performance of subjects when using the traditional paper-based version and our computer-based version, which captures all user interaction data in real time using game analytics techniques. Other relevant information, such as demographics and familiarity with technology, was also compiled by pre-post online forms. Our results show that computer and traditional pen-and-paper test versions provide similar results, while the additional interaction data captured by using game analytics techniques opens the door to new environments for active aging research.

KEYWORDS

active aging, game technology, computerized test, cognitive impairment, neuro-psychological test.

Introduction

The administration of cognitive tests through new technologies is an increasingly frequent practice in clinical and research environments [1]. Information and communication technologies (ICTs) have been shown to help objectively assess the presence of cognitive and behavioral problems in older people. Specifically, in the field of the neuropsychological assessment of aging, computerized tests have observed to be useful for the early detection of mild cognitive impairment (MCI) [2]. In this sense, it is necessary to verify whether the computerized administration of the cognitive tests significantly affects the results, and whether the computer application can be used interchangeably within clinical practice and evaluation, or on the other hand depends too heavily on the subject's previous familiarity with the technology [3]. COSMOS is a web platform configured to acquire cognitive-related metrics that can be used for research purposes [4]. Different investigations have shown that, at least in some cases, the results of computerized evaluations are equivalent to the results provided by the tests performed using the traditional pen and paper methods [5]–[8].

Computerized application of tests provides different benefits such as a more precise data collection (e.g. by accurately measuring reaction time) and the elimination of many data entry errors, which occur during traditional pen and paper data collection [9]. Other advantages of computerized tests over traditional tests include saving time through process automation, and a lower reliance on trained personnel during administration. Automated testing can directly assess task performance, such as the speed of cognitive processes, presenting the results in real time to domain experts [10]. However, we cannot ignore the need for the role of the psychology or neuropsychology domain expert, since, in order to carry out an

adequate evaluation process, not only quantitative aspects must be taken into account, but also qualitative aspects that help to understand the specific situation of the subject, and allow the evaluation process to be conducted in the most appropriate way possible [11]. This works is centered on the question of whether the scores obtained through the computerized format can be interpreted in a similar way to the scores obtained through the traditional paper format. In this sense, the computer version results should match in equivalence and in reliability.

We expect that game technology can be effectively applied to decrease the cost of developing new computerized versions of traditional tests; and that the use of game learning analytics allows the creation of new and more powerful environments for assessing and researching active aging. Use of game technology should greatly simplify the creation of highly interactive environments where interaction data can be captured for later interpretation by experts.

Objective

This work examines the equivalence between the 15 Objects Test (15-OT) [12] in its traditional, paper-based format versus the computer version of the test, implemented with game technology, when applied in a sample of cognitively healthy older adults (see Fig. 1). As a secondary goal, we attempt to understand the factors that may contribute to the effectiveness of the computerized test by acquiring and analyzing complementary information such as interaction data, demographic data or knowledge and use of technology by the target demographic.

Method

Eighteen volunteers (12 women and 6 men), all over 60 years old (M = 65.67, SD = 5.31) were recruited from the UNED Senior University Program (Madrid). Subjects were initially randomly divided into the four experimental conditions balanced by age and sex (I, II, III and IV). Each condition was defined separately, and its participants assigned to a specific experimental condition (see **Fig. 2**). All participants provided written consent to participate in the study (see **Table 1**).

Table 1. Socio-demographic and cognitive characteristics of the participants in the four experimental conditions

|	Cond I M (SD)	Cond II M (SD)	Cond III M (SD)	Cond IV M (SD)
Participants	N=5	N=4	N=5	N=4
Female/Male	3/2	3/1	4/1	2/2
Age	67.25 (2.28)	65.75 (5.97)	63.75 (3.7)	64.25 (4.44)
Education level (years)	14.8 (1.8)	14 (1.41)	14 (2)	15.5 (0.87)
MMSE	30 (0.00)	29.50 (0.57)	29.60 (0.55)	29.25 (0.50)

(M: mean, SD: standard deviation).

Subjects with suspected MCI, neurodegenerative disease, depression/anxiety or other psychiatric disorder, or severe sensory deficit were not included in the study. The Mini-Mental State Examination (MMSE) [13] was used to exclude (*exclusion criteria*) potential subjects with scores below 24 points (out of 30). All participants had normal cognitive status and corrected-to-normal vision.

Procedure and design

Before the experiment, researchers contacted the Las Tablas UNED center in Madrid requesting permission to give an informative talk about the study. During the talk, they presented both the study and the experimental design, and allowed potential participants to sign up for the study. Pseudo anonymization was applied, so that subjects were only identified by a randomized code which was handed to subjects as a physical token, and then used to identify them throughout the study. After applying MMSE to verify the inclusion criteria, subjects were randomly assigned to the groups that were balanced by sex and age.

The study used a mixed experimental design: we compare the paper and computerized scores for each participant, using the test administration format and version as independent variables. Since there were two alternative formats, paper and computer, and two versions of the 15-OT, A and B (see **Fig. 1**), available in each format, this results in four experimental conditions, as illustrated in **Fig. 2**. The dependent variables were the number of identified objects (correct answers), the number of errors, and the time spent identifying the objects.

Subjects were tested under standardized conditions by assistants trained to present an identical set of preliminary instructions. Specific instructions about the tests were given either through oral instruction (for the paper test) or through instruction screens (for the computer-administered test). When a participant had finished the first assigned test, the researchers directed the participant to begin the second test of the pair (see **Fig. 2**). Participants could complete both of their tests in less than 25 minutes.
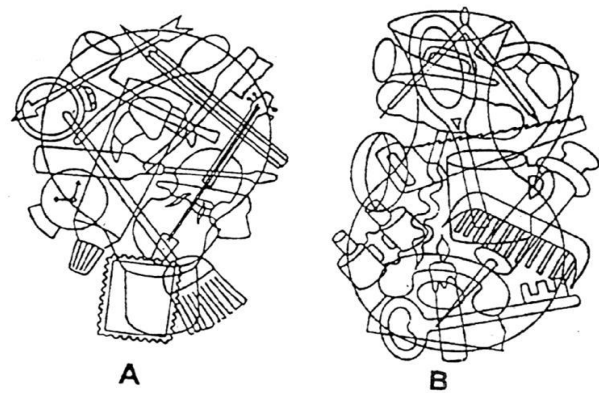
Fig. 1. The two versions, A and B, of the 15-OT [12].

Experimental conditions

Four experimental conditions were created and each participant assigned to one (see **Fig. 2**): Condition I (N = 5): pre-test on the computer, computerized test 15-OT A, post-test on the computer and test B on paper. Condition II (N = 4): test A on paper, pre-test on the computer, computerized test B and post-test on the computer. Condition III (N =5): pre-test on the computer, computerized test B, post-test on the computer and test A on paper. Condition IV (N = 4): test B on paper, pre-test on the computer, computerized test A and post-test on the computer.

The pre-test and post-test were intended to gain additional subject data. The pre-test was carried out just before starting the computerized version of the 15-OT and was composed of a set of six questions designed assess the participant's previous knowledge of technology, as well as demographic information such as age and education level. The main questions of the pre-test inquired what type of devices, such as smartphones, PCs or tablets, were used by participants; how often they were used; and for what general purpose, such as staying in touch, playing, or working. The post-test was carried out after the completion of the computerized version of 15-OT, and asked participants about their satisfaction with the game (technology acceptance) and its perceived level of difficulty. Also, participants were requested to fill in additional information regarding their knowledge and use of different specific social networks (WhatsApp, Facebook, etc.), to complement the pre-test results regarding their familiarity with technology.

Materials

There are different tests to evaluate the visuo-perceptive and visuo-spatial functions. In this study, we have focused on the assessment of visual recognition, cognitive processing speed and object naming, through the 15-OT (see **Fig. 1**) [12]. This test, based on Poppelreuter (1914-1917), evaluates the identification and discrimination of overlapping objects. The task consists of 2 illustrations composed of superimposed outlines of 15 common objects (versions A and B). The subject is asked to name and point to, as fast as possible, each of the objects in the illustration (be it on paper or on a computer screen). There is no set time for the completion of the test, but the time it takes for the participants to perform the task reflects the processing speed as well as other cognitive mechanisms involved. It is obvious that in the computer version the subject spends more time on the task than in the traditional version since the writing process is slower than oral production. Not only were the correct answers computed, but also the errors, the execution time and the difficulties shown in the execution. A response was considered correct if the participant pointed to and named the corresponding object from the composite picture. All responses which discerned only a detail of the object or which did not permit identification of the corresponding object were considered incorrect. The subjects were encouraged to continue searching for new objects until they considered that they had recognized all objects; but otherwise, no indication of the number of identified objects or whether there were any remaining objects to be identified was provided. This tool was selected, among other aspects, because it is comparatively short, and because it is an evaluation tool that does not exhibit the influence of schooling and is free from practice effects, as it is expected to be entirely new for participants. The design of the 15-OT was easily transferable to a computer application using game technology that can be also produced for touch screen devices such as tablets with minimal cost.

Computer application

The computer version was designed to perform versions A and B of the 15-OT. The application presents the user with a pre-test, followed by a short tutorial, the 15-OT A or B version (based on the participant's experimental group), and ends with the post-test (see **Fig. 2**). The tutorial, which is displayed just before the 15-OT itself, uses a shorter and simpler version of the test activity, displaying a different image with only five objects, where the subject is guided through sequential informative steps on how to select objects using the mouse, and how to type their names using the keyboard, until all objects have been identified.

The computer version tracks all subject interactions, sending them to a server for analysis. If the cloud-based server cannot be reached, the interactions are stored locally, and will be sent later once the connection is restored. Interactions are sent and stored in the Experience API (xAPI) format, which is designed to represent the sequence of events that occur during a subject's interaction with an interactive application such as a learning activity or serious game [14]. Storing the data in a standard format simplifies latter reuse and sharing of the data if the experimental design allows such reuse.

In our case, the interaction data is sent to preexisting cloud Learning Analytics infrastructure, where it is stored, analyzed, and made available for visualization by authorized users. In this case the results, displayed in real-time, include the identified correct answers and the candidate wrong answers. Wrong answers require an additional interpretation by experts because some subject answers might have additional synonyms that the application has
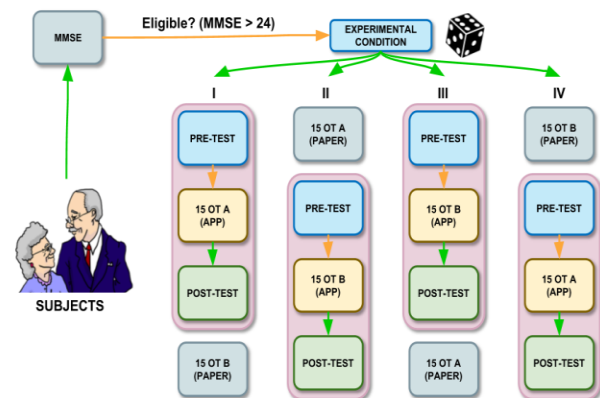
Fig. 2. The flow of the experiment. Subjects were assessed with the MMSE, before being randomly assigned to one of the four experimental conditions (I, II, III, IV).
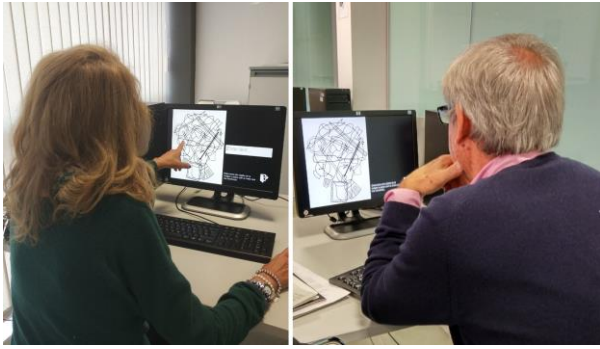
Fig. 3. Subjects identifying objects in the 15-OT version A.

not considered; or include a typo that made a correct answer seem incorrect. Tracking provides additional data on how participants spent their time within the computer version, the objects that they attempted to identify multiple times, and other information that can provide insights into how they interacted with the game or even the mental processes involved.

Unity3D was used to develop the 15-OT computer application (see **Fig. 3**), due to its popularity for developing video games. Unity reduces the cost of creating highly interactive environments such as the ones needed for psychological assessment, and provides support for multiple platforms (e.g. PC, Android, iOS) with the same codebase. For the initial study, the computer application has been exported for Windows operating systems on PCs, matching the existing infrastructure of UNED computer labs. However, we are also planning a version for tablets.

Statistical analysis

In addition to a descriptive analysis of the data, the nonparametric Kruskal-Wallis test was applied to analyze the differences between the experimental groups in the execution of the 15-OT (considering score and time) in their paper or computerized versions. In addition, a correlation analysis using Spearman´s Rho was carried out to analyze the equivalence between both versions. All statistical analyses were performed using SPSS (v. 24.0), and a significance level of 0.05 was adopted.

Results

In terms of performance, the average score for correct answers in the paper version was 13.94 (SD = 1.11), compared to 12.89 (SD = 1.45) for the computer version. Regarding execution time (measured in seconds), participants using the paper version required 71.83 (SD = 20.05), compared to 111.00 (SD = 35.47) in the computer version.

No significant differences were found between the groups, either in the scores or in test application time, regardless of the form of application (paper or computer version). The two presentation versions of the 15-OT showed similar results. A comparison of the mean scores per experimental condition can be found in **Fig. 4**. The subjects of conditions II, III and IV have similar scores on paper and in the computer version. The subjects in condition I scored
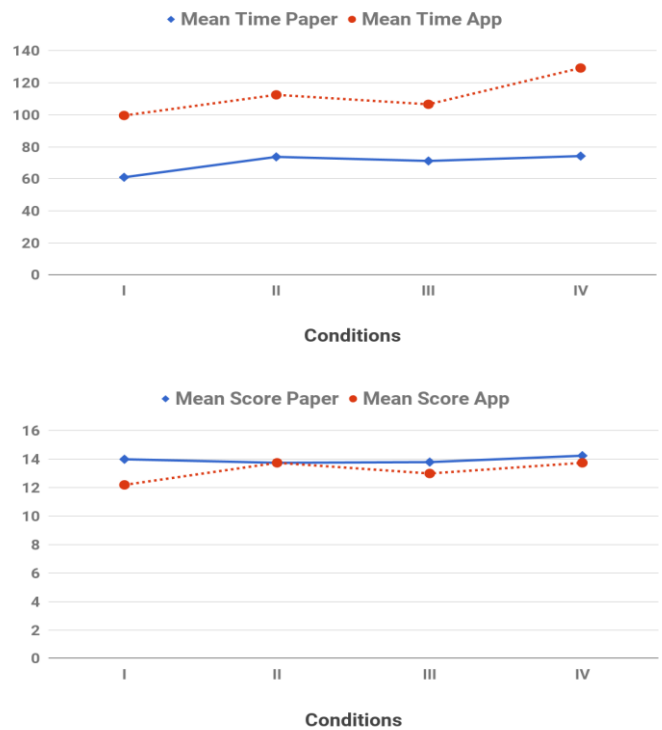

Fig. 4. Mean time to complete the activity, in seconds (top); and mean count of correct answers (bottom) per condition of the 15-OT paper vs computer application.

slightly lower on the computer version than the rest of the conditions, but the difference is not significative.

In relation to the study of the equivalence between the two 15-OT versions, statistically significant and direct correlations were found between the scores of both versions (r = 0.481 p < 0.043). Likewise, direct and significant correlations were found between execution times in the computer and traditional versions (r = 0.771, p < 0.001).

In addition, in terms of the individual performance of each subject in the two versions, the results reflect similarities in the
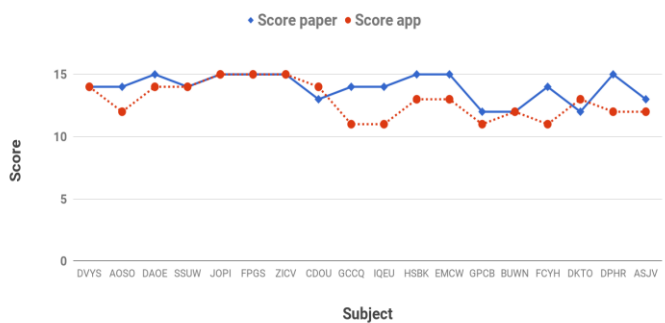
Fig. 5. Correct answers per subject on the 15-OT, on paper vs. computer versions.
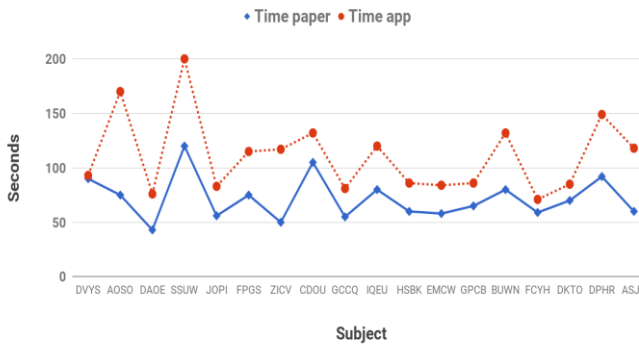
Fig. 6. Time in seconds per subject of the 15 OT paper vs. computer application.

performance of both versions of the 15-OT (see **Fig. 5**). The number of correct answers in the different versions shows that scores in the computer application are generally lower than those from in the paper version. Subjects that score higher on the traditional 15-OT also score higher on the computer version, and vice versa.

The times required for each subject to identify and denominate each of the items that make up the 15-OT in each of the versions (paper vs computer) can be found in **Fig. 6**. As the data shows, the time spent on the computer version is always longer than the time spent on the paper version. This can be generally explained because participants using the computer version had to type the names of identified objects, while those using the paper version only had to speak the names out loud.

The technological knowledge of the subjects has been determined from the pre-test answers (see **Fig. 7**), and it has been computed taking into account how many days per week the subject uses technological devices, how many different technological devices subjects claim to know and have used at least once and, finally, subjects' goals for using technological devices, including, for example, communication, management, or education. Subjects with higher technological knowledge scored higher and finished tests, in both paper and computer versions, faster.
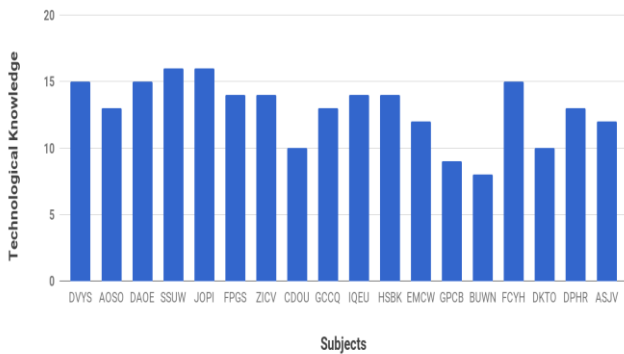
Fig. 7. Technology use per subject.

Discussion

Our main question is whether game technology can be effectively applied to the creation of new environments for assessing and researching active aging. The current study evaluates the equivalence of the 15-OT presented in the traditional pen-and-paper version versus the computer version developed using game technology. The results demonstrate an equivalence between both versions of the 15-OT test. The traditional 15-OT consists of 15 overlapping objects displayed on a paper that must be identified by the subjects, either by saying their name, or, in the computer version, using the mouse to click on it, and the keyboard to type the object's name.

The computer application automatically collects and sends anonymous data of the subjects' interactions to a Learning Analytics (LA) infrastructure in the cloud. Received data is analyzed and displayed to domain experts. LA visualizations display the performance of the subject in real-time and the performance of the whole group, the duration of the entire experiment and every interaction of the subject with the objects in the screen. Feedback provided by the LA is meant to aid and support the decision of the domain experts in their diagnostic of the subjects' cognitive state.

Some subjects, even if technologically savvy, may still prefer the traditional pen-and-paper test, which avoids the use of a pointing device (mouse) and having to type object names on a keyboard. However, each traditional test requires a fully dedicated researcher per individual subject, to oversee the test and mark correct object identifications; while the computer version can be administered to large groups with minimal intervention, avoids potential sources of bias, and gathers, analyzes and provides access to potentially valuable complementary data which would be difficult to capture with the traditional test. In this sense, additional data can be very useful in determining the cognitive profile of the individual, as they provide accurate information on the type and way the subject being assessed generates responses. Aspects such as the sequencing of responses, the type of errors – perseverance, intrusions – and omissions can be considered as signs of alteration compatible with possible cognitive impairment. Evidence of these signs may contribute to the early diagnosis of cognitive pathologies in older people. Therefore, the computer versions of the test are another alternative in clinical practice, as they can offer the expert additional and objective information that can help to determine and improve the clinical diagnosis [15].

Generally, there is an increasing interest in creating motivating tools and serious games for older people previous to any appearances of cognitive impairment for preventing measures [16]. Moreover, serious games require additional cognitive functions, having an advantage over more traditional computerized assessment in training cognitive skills and preventing cognitive decline.

Conclusions and Future Work

Validated cognitive assessment computer applications for older people can provide additional benefits in clinical practice, such as automatic and more precise collection of interaction data, real-time feedback that aids in the decision-making process, and automatization of the process of performing tests and integration of pre-post testing without further intervention. Subjects have shown interest in participating in the study and have given positive feedback after the activity. These positive attitudes support the idea of designing and elaborating more computer applications in prevention and diagnosis for older people at a larger scale.

Another aspect to be taken into account when carrying out this type of work is not only the difficulty in selecting healthy older people, but also the fact that the application of the tests has been carried out individually and not collectively, which entails more time spent in administering the tests (even if in this first round of this experiment the sample is still small). Despite this limitation, the current study provides results about the efficacy and feasibility in using game technology to automatize neuropsychological tests and research in active aging. The results of the current study show that the computer application of the 15-OT can be used as an additional validation tool that helps domain experts spotting cognitive age-related cognitive issues. We plan to continue the study and to increase the number of subjects implied. We are also planning to extend the study with a version for tablets. This new version will be more ecological as we will use a voice recognition system as main input and users will not need to write their answers with the keyboard.

Instances of games or applications for non-entertainment purposes using game technology have been successfully applied in education, medicine and other fields [17], and are collectively called "serious games". We are developing multiple serious games for testing cognitive impairment that focus on different cognitive skills. The future work of this study is to be validated with further game-like applications. The subjects will be presented with different situations that require quick reaction and focus on the details of the situation.

ACKNOWLEDGMENTS

REFERENCES

[1] A. Shah, K. R. Kraemer, C. R. Won, S. Black, and W. Hasenbein, "Developing Digital Intervention Games for Mental Disorders: A Review," *Games Health J.*, vol. 7, no. 4, pp. 213–224, Aug. 2018.

[2] D. Darby, P. Maruff, A. Collie, and M. McStephen, "Mild cognitive impairment can be detected by multiple assessments in a single day," *Neurology*, vol. 59, no. 7, pp. 1042–1046, 2002.

[3] B. B. Brodey, N. L. Gonzalez, K. A. Elkin, W. J. Sasiela, and I. S. Brodey, "Assessing the Equivalence of Paper, Mobile Phone, and Tablet Survey Responses at a Community Mental Health Center Using Equivalent Halves of a 'Gold-Standard' Depression Item Bank," *JMIR Ment. Heal.*, vol. 4, no. 3, p. e36, Sep. 2017.

[4] A. Aeberhard *et al.*, "Introducing COSMOS: a Web Platform for Multimodal Game-Based Psychological Assessment Geared Towards Open Science Practice," *J. Technol. Behav. Sci.*, Sep. 2018.

[5] N. E. Bush, N. Skopp, D. Smolenski, R. Crumpton, and J. Fairall, "Behavioral screening measures delivered with a smartphone app: Psychometric properties and user preference," *J. Nerv. Ment. Dis.*, vol. 201, no. 11, pp. 991–995, 2013.

[6] M. Clionsky and E. Clionsky, "Psychometric equivalence of a paper-based and computerized (iPad) version of the memory orientation screening test (most ®)," *Clin. Neuropsychol.*, vol. 28, no. 5, pp. 747–755, 2014.

[7] L. A. Goldstein *et al.*, "Outcome Assessment via Handheld Computer in Community Mental Health: Consumer Satisfaction and Reliability," *J. Behav. Health Serv. Res.*, vol. 38, no. 3, pp. 414–423, 2011.

[8] C. J. Gwaltney, A. L. Shields, and S. Shiffman, "Equivalence of Electronic and Paper-and-Pencil Administration of Patient-Reported Outcome Measures: A Meta-Analytic Review," *Value Heal.*, vol. 11, no. 2, pp. 322–333, Mar. 2008.

[9] J. Morganstein, "Mobile applications for mental health providers.," *Psychiatry Interpers. Biol. Process.*, vol. 79, no. 4, pp. 358–363, 2016.

[10] K. A. Wesnes, "Moving beyond the pros and cons of automating cognitive testing in pathological aging and dementia: The case for equal opportunity," *Alzheimer's Res. Ther.*, vol. 6, no. 8, pp. 1–8, 2014.

[11] M. Alegret *et al.*, "Detection of visuoperceptual deficits in preclinical and mild Alzheimer's disease," *J. Clin. Exp. Neuropsychol.*, vol. 31, no. 7, pp. 860–867, 2009.

[12] B. Pillon *et al.*, "Cognitive slowing in Parkinson's disease fails to respond to levodopa treatment: the 15-objects test.," *Neurology*, vol. 39, no. 6, pp. 762–8, Jun. 1989.

[13] M. F. Folstein, S. E. Folstein, and P. R. McHugh, "'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician," *J. Psychiatr. Res.*, vol. 12, no. 3, pp. 189–198, 1975.

[14] Á. Serrano-Laguna, I. Martínez-Ortiz, J. Haag, D. Regan, A. Johnson, and B. Fernández-Manjón, "Applying standards to systematize learning analytics in serious games," *Comput. Stand. Interfaces*, vol. 50, pp. 116–123, 2017.

[15] D. Martínez-Pernía *et al.*, "Using game authoring platforms to develop screen-based simulated functional assessments in persons with executive dysfunction following traumatic brain injury," *J. Biomed. Inform.*, vol. 74, pp. 71–84, Oct. 2017.

[16] V. Vallejo *et al.*, "Evaluation of a new serious game based multitasking assessment tool for cognition and activities of daily living: Comparison with a real cooking task," *Comput. Human Behav.*, vol. 70, pp. 500–506, May 2017.

[17] E. J. Lyons, "Review of *Games for Health: Proceedings of the 3 rd European Conference on Gaming and Playful Interaction in Health Care*", *Games Health J.*, vol. 3, no. 1, pp. 49–52, 2014.