Towards long-term tracking and detection of early dementia : a computerized cognitive test battery with gamification

Zeng, Zhiwei; Fauvel, Simon; Hsiang, Benny Tan Toh; Wang, Di; Qiu, Yang; Khuan, Pamela Chew Oi; Leung, Cyril; Shen, Zhiqi; Chin, Jing Jih

2018

Zeng, Z., Fauvel, S., Hsiang, B. T. T., Wang, D., Qiu, Y., Khuan, P. C. O., . . . Chin, J. J. (2018). Towards long-term tracking and detection of early dementia : a computerized cognitive test battery with gamification. Proceedings of the 3rd International Conference on Crowd Science and Engineering (ICCSE'18). doi:10.1145/3265689.3265719

<https://hdl.handle.net/10356/139062>

<https://doi.org/10.1145/3265689.3265719>

© 2018 Association for Computing Machinery. All rights reserved. This paper was published in International Conference on Crowd Science and Engineering and is made available with permission of Association for Computing Machinery.

Downloaded on 26 Apr 2023 18:02:16 SGT

Towards Long-term Tracking and Detection of Early Dementia: A Computerized Cognitive Test Battery with Gamification

Zhiwei Zeng*
LILY Research Centre, IGS
Nanyang Technological University
i160001@e.ntu.edu.sg

Di Wang
LILY Research Centre
Nanyang Technological University
WangDi@ntu.edu.sg

Cyril Leung‡
LILY Research Centre, ECE
University of British Columbia
cleung@ece.ubc.ca

Simon Fauvel
LILY Research Centre
Nanyang Technological University
sfauvel@ntu.edu.sg

Yang Qiu
LILY Research Centre
Nanyang Technological University
qiuyang@ntu.edu.sg

Zhiqi Shen§
LILY Research Centre, SCSE
Nanyang Technological University
ZQShen@ntu.edu.sg

Benny Tan Toh Hsiang
LILY Research Centre
Nanyang Technological University
BennyTanTH@ntu.edu.sg

Pamela Chew Oi Khuan†
IGA, Tan Tock Seng Hospital
pamela_ok_chew@ttsh.com.sg

Jing Jih Chin¶
IGA, Tan Tock Seng Hospital
LKC School of Medicine, NTU
jing_jih_chin@ttsh.com.sg

ABSTRACT

Effective diagnosis of mild cognitive impairment (MCI), in many cases preceding dementia, is important in determining the efficacy of dementia treatments. Inherent in the transition from normal ageing to MCI to and then to dementia is cognitive decline, which can be detected using multiple assessments over an extended period of time. Computerized cognitive tests arise as a promising way of long-term cognitive monitoring in home environment and supplementing clinical evaluation. Compared to conventional paper-and-pencil tests, they are cheaper, more repeatable, and easier to distribute and administer. Over the years, research efforts have been devoted to improve the validity, reliability and comprehensiveness of the computerized cognitive tests. However, it has long been omitted that the usability and entertainment aspects are also crucial to their overall effectiveness and user experience.

To reduce drop-out rates and improve effectiveness of long-term cognitive monitoring, we present a first-of-its-kind gamified computerized cognitive test battery, called Virtual ADL+ House. It consists of a series of mini-games, each embedded a cognitive test and featured one of the daily living activities in the Lawton IADL. Virtual ADL+ House can be used to monitor cognitive functions in long-term, alert signs of cognitive decline and provide longitudinal data to facilitate clinical diagnosis. In focus group studies conducted with doctors and older adults, we received positive feedback on the usability and entertainment value of Virtual ADL+ House. We plan to evaluate the validity and reliability of it in subsequent studies.

KEYWORDS

Dementia; early detection; computerized tests; gamification; cognitive monitoring

1 INTRODUCTION

Dementia is the leading chronic disease causing disability and dependency among the elderly population worldwide. It already places an important burden on the healthcare system, and significantly reduces the possibility of the elderly to age in place. Worldwide, a new dementia patient is diagnosed every 4.1 seconds. Currently, more than 47 million people are living with dementia, and this number is expected to triple by 2050¹. Given this trend, caring for dementia patients in the traditional way will quickly become unsustainable.

At present, most investigators believe that by the time the clinical diagnosis of dementia has been made, the neuronal damage to the brain is unlikely reversible [31] and the effectiveness of the treatments are limited. Recent advances in neuroprotective treatments [10, 20, 36] have increased the confidence that the disease progression can be effectively managed if dementia is detected early. Hence, a more appealing strategy is to identify high risk individuals at the earliest stage of impairment. The concept of mild cognitive impairment (MCI) has been applied to represent a transitional stage between normal ageing and very early stage of dementia [30]. Generally, there is a high probability that MCI individuals would gradually progress to dementia [11, 32]. In other words, MCI individuals have higher risk of developing dementia than healthy individuals. Hence, effective diagnosis of MCI provides a promising way for early detection of dementia.

Currently, due to the clinical nature of existing dementia assessment approaches, diagnosis of dementia tends to happen during advanced stage of the disease. Once the symptoms of the disease become obvious enough to cause the attention of the patient or caregivers, the best timing for treatment has often been missed. In order to detect possible dementia signs early, in MCI or even pre-MCI stage, long-term and frequent assessments are required.

*IGS stands for the Interdisciplinary Graduate School.

†IGA stands for the Institute of Geriatrics and Active Ageing.

‡ECE stands for Department of Electrical and Computer Engineering.

§SCSE stands for the School of Computer Science and Engineering.

¶NTU stands for Nanyang Technological University.

¹World Health Organization. (2015). Dementia. Retrieved from <http://www.who.int/mediacentre/factsheets/fs362/en/>

However, such assessments are costly, highly labor and resource intensive, and hence, impractical in a clinical setting.

Computerized cognitive tests provide new opportunities for remotely monitoring cognitive changes and detecting early dementia. Without the need of visiting a trained assessor, computerized cognitive tests can be distributed over the Internet and self-administered. They are objective, standardized, and most importantly, more repeatable. Computerized tests provide ideal formats for generating alternative tests, improving their test-retest reliability over repeated administration in long-term monitoring. As serious disease screening and monitoring tools, the emphasis of developing computerized cognitive tests has been solely on the comprehensiveness, validity and reliability of the tests. However, it has long been omitted that computerized cognitive tests are also pieces of software and their usability and entertainment aspects are also crucial to their overall effectiveness and user experience. To reduce the drop-out rate and ensure the effectiveness of long-term cognitive monitoring, computerized cognitive tests should be designed with incentives to induce intrinsic motivation and active participation.

In this paper, we propose a gamified computerized cognitive test battery, named Virtual ADL+ House. It includes a broad spectrum of tests of memory and other cognitive functions, providing a holistic cognitive profile of the test subject. In Virtual ADL+ House, each test is designed as a mini-game related to one of the daily activities in IADL, e.g. cooking, finance, shopping and transportation. Game-like tests can help to create more enjoyable and relaxed testing experience. Moreover, the performance in the games may be used to complement the IADL self-reports to improve accuracy of IADL ratings. As the first step in validating and evaluating Virtual ADL+ House, we conducted focus groups to evaluate its usability and entertainment aspects. We received positive feedback from both doctors and older adults participated in the studies.

The rest of the paper is organized as follows. Section 2 reviews the relevant background on dementia and its early detection, cognitive domains affected by MCI and existing computerized batteries. Section 3 presents the overall design concepts and each mini-game of Virtual ADL+ House. Then, section 4 describes the focus group studies and summarizes the study results. Finally, section 5 concludes the paper.

2 BACKGROUND

2.1 Dementia and Its Early Detection

As a normal component of ageing, many people will experience a decline in their cognitive functions. When the decline becomes more significant, pathological processes might be going on. As the cognitive degeneration progresses, the cognitive and functional impairment reaches a threshold and the person is clinically diagnosed with probable dementia (Fig.1). At this point of time, the neuronal damage to the brain is already unlikely to be reversible [31] and most available treatments merely help to alleviate the symptoms rather than cure the disease. Hence, a more appealing strategy is to detect dementia in its early stage and start intervention early.

In recent years, the concept of mild cognitive impairment (MCI) has been applied to represent a transitional stage between normal ageing and very early dementia [30]. It refers to a “mild” condition in which objective memory impairment presents in the absence

Figure 1: Theoretical change in cognitive function as a person progresses to dementia (adapted from Fig.1 in [31]).

Figure 2: An illustration of different cognitive domains.

of functional disability [18, 25] —“memory complaint, preferably corroborated by an informant, objective memory impairment for age and education, normal general cognition, preserved activities of daily living, and not demented [31]”. People with MCI generally progress to dementia. Approximately 12-15% of MCI subjects per year convert to clinical dementia with functional disability [5, 11, 32]. For this reason, early detection of dementia depends upon the effective diagnosis of MCI [3].

Currently, diagnosis of dementia tends to happen during more advanced stages of the disease, and the opportunities for effective early intervention are often missed. There are several obstacles to the early detection of dementia.

Firstly, existing dementia assessment approaches are mostly clinical in nature and generally performed in a primary care setting [24]. Usually, a doctor suspecting a dementia case would first gather historical information from the patient’s next-of-kin and caregiver(s) through standard questionnaires, such as questionnaires for Activities of Daily Living (ADL) and Instrumental ADL (IADL). The doctor would then conduct clinical investigation through blood tests, brain scans, and neuropsychological tests. Based on the combined results from all of these tests, a diagnosis can be established. As these tests are often taken only as a result of clinicians’ suspicions based on patient symptoms or caregivers’ concerns, the diagnosis often comes after the onset of dementia. Moreover, most of the neuropsychological tests need to be administered by well-trained assessors,

Table 1: A brief review of self-administered computerized cognitive batteries

Test Name	Domains Covered*						Validated for Dementia	Validated for MCI	Test-retest Reliability#
	ATT	PS	LANG	VS	MEM	EF			
TDAS			✓	✓	✓	✓	[17]		
CNSVS	✓	✓			✓	✓	[15, 16]	[15, 16]: acceptable sensitivity and specificity on 3 out of 5 domains	[16]: acceptable overall reliability (0.643 - 0.79), low reliability on Stroop and Continuous Performance subtest
CANS-MCI			✓		✓	✓		[40]: significant difference between normal and MCI group in all domains	[40]: 1 month (0.61 - 0.85), 6 month (0.62 - 0.89)
CAMCI	✓	✓			✓	✓		[35]: high sensitivity and specificity	[35]: 2 - 4 weeks (0.30 - 0.74)
MicroCog	✓			✓	✓		[8]		[8]: no period (0.59 - 0.98)
CANTAB					✓		[7]	[7]: significant difference observed between normal and MCI in some domains	[22]: 4 weeks (0.09 - 0.86)
CFT		✓			✓	✓		[41]: high concurrent validity, discriminant validity not reported	
CST			✓		✓	✓	[4]	[4]: high overall concurrent and discriminant validity, no separate statistics reported for MCI	[4]: 6 weeks (Cronbach's alpha 0.76)

* The domains: ATT - attention; PS - processing speed; LANG - language; VS - visuospatial; MEM - memory; EF - executive functioning.

Unless otherwise stated, the metrics used for test-retest reliability is Pearson's correlations.

making the process lengthy, costly, resource-intensive, and prone to subjective biases.

Secondly, long-term tracking, i.e. multiple assessments over an extended period of time, is required to monitor the change in cognitive functions timely and effectively [2]. However, most neuropsychological tests themselves have not been designed for repeated administrations, especially at very short re-test intervals [9]. Improvements in test performance due to repeated exercise, a.k.a. practice effects, can reduce the sensitivity of the test to true changes in cognitive functions. With limited health-care resources, it is unrealistic to carry out long-term tracking in clinical environment. Moreover, as long-term tracking programs can be quite repetitive and time-consuming, it is also a non-trivial task to design incentives to ensure the compliance to the programs.

2.2 Cognitive Domains Impacted by MCI

Since MCI is a transitional stage between normal ageing and very early dementia, effective diagnosis of MCI can help to identify individuals with high risk of developing dementia. As illustrated in Fig.2, although memory impairment is the most representative symptom of MCI, multiple cognitive domains other than memory are compromised in most MCI individuals [2, 5, 29, 46].

Memory is a high-level cognitive function. Once dysfunctions are reported in memory, the symptoms go beyond forgetfulness to affect other cognitive domains such as attention, processing speed, language, visuospatial and executive functioning [1, 5, 15].

In order to catch cognitive decline in different cognitive domains, assessments for MCI should include an array of brief tests to assess the overall cognitive status, enabling the clinician to quickly acquire a broad cognitive profile of the subject [3].

2.3 Existing Computerized Batteries

Traditional paper-and-pencil neuropsychological tests have been proven to distinguish among cognitively healthy elderly, individuals with MCI and with dementia [2, 14, 46]. However, paper-and-pencil tests have many limitations. Firstly, they are generally impractical for widespread use due to high cost and lengthy administration time. Secondly, they are not suitable for long-term tracking due to the lack of alternative forms. Thirdly, they generally need to be administered by trained assessors, and thus do not support self-administration.

Computerized cognitive tests have the potential to address the limitations of paper-and-pencil tests and be an effective way to monitor cognitive change in long-term. Computerized tests can be easily distributed via the Internet and administered with or without the supervision of an assessor. Computerized test sessions can be shorter and less labor-intensive than their paper-and-pencil counterparts. They also maintain standardized and consistent test procedures across subjects, and thus are not susceptible to the variations in assessors [45]. Standardized computerized batteries provide ideal formats for generating alternative tests, enabling serial assessments for monitoring in long-term [5]. Another benefit brought by computerized tests is the ability to capture more high-fidelity data which increases the test sensitivity and accuracy of performance measurements.

Recently, there are many attempts in developing and validating computerized batteries to assist psychologists and researchers for screening and follow-up of MCI [3, 5, 24, 38, 43]. Systematic reviews of computerized cognitive tests can be found in [39, 44, 47]. For our purpose, we are more interested in the ones that are:

- self-administered,
- covering broad cognitive domains,
- suitable for long-term monitoring, and

- sensitive enough for effectively discriminating individuals with MCI.

In Table 1, we briefly review existing computerized tests according to above criteria. Among the batteries reviewed in [47], we selected 8 cognitive assessment batteries that can be self-administered. 6 out of the 8 of them, except TDAS and MicroCog, have been tested with MCI subjects. 4 of them, CANS-MCI, CAMCI, CANTAB, and CST, have demonstrated acceptable sensitivity and specificity as well as high discriminative validity. Significant differences can be observed between the test performance of the control group and the MCI group, showing that these batteries can effectively distinguish the two groups. The test-retest reliability of the batteries have been studied over test-retest intervals ranging from 1 day to 6 weeks. For most of the batteries, the test-retest reliability differs greatly depending on the subtest.

3 A NOVEL COMPUTERIZED TEST BATTERY

3.1 Virtual ADL+ House - The Gamified Test Battery

Virtual ADL+ House is an online virtual reality home environment to promote long-term cognitive assessment. It is a game consists of a series of mini-games with different themes. Gaming activities performed in Virtual ADL+ House are abstractions of many of the daily living routines, modelled after the Lawton Instrumental Activities of Daily Living Scale [12], including using telephone, food preparation, housekeeping, laundry and handling finances. A cognitive assessment battery covering multiple cognitive domains is integrated into the game design. When an individual is playing a mini-game in Virtual ADL+ House, he/ she is actually performing a “disguised” neuropsychological test. Game playing data, such as the time spent to finish each session, the number of attempts made on each tasks, the accuracy, and types of errors made, will be collected and used to determine the test performance.

Compared with other computerized cognitive assessment batteries, Virtual ADL+ House has two major advantages:

- (1) Gamification elements can induce intrinsic motivation to complete multiple assessments in long-term monitoring.
- (2) Game activities designed in the light of IADL can provide useful proxy source to complement self-reports, improving the accuracy of IADL scoring.

Gamification. In long-term monitoring programs, multiple assessments are often required to track the cognitive change over time. Although variations of the test instruments are used in different assessments to avoid practice effects, the test processes can still be quite repetitive and time-consuming. The drop-out rate was about 11%² in a 2-year Finnish trial for preventing cognitive decline [28] and about 38% in a 6-year Dutch trial for preventing dementia [42]. Among those who dropped out, at least 15% were due to lack of motivation and time [28]. Thus, to reduce drop-outs and ensure the effectiveness of long-term monitoring programs, it is important to design incentives to promote active participation and compliance.

Gamification can help in this regard. In Virtual ADL+ House, each neuropsychological test is designed as a mini-game related to

one of the daily activities in IADL. While resembling the core idea of the test, gamification elements are added to provide meanings to the test actions. For example, the conventional Digit Span test is designed to be a mini-game which requires the player to remember and dial telephone numbers. The enjoyment of playing the games can induce intrinsic motivation to perform the assessments repeatedly. Moreover, gamification helps to create a more relaxed testing environment. Some individuals may be nervous and pressurized in a clinical testing environment. More accurate test performance may be collected in a relaxed testing environment when the tested individual is at ease.

Complementing IADL Self-reports. ADL and IADL are widely used to refer to a person’s daily self-care activities. The ability or inability to perform ADLs indicates a person’s functional status. ADL and IADL can be scored based on self-reports of activity, a proxy source or performance-based measurements through direct observation of behavior. For elderly living alone without any caregiver providing proxy source, self-reports may be used to determine the ADL or IADL ratings. However, individuals with cognitive impairment have a general tendency to over-rate their abilities [6]. There may be significant differences between self-reports of the patients and performance-based measurements [34]. Hence, when it is impractical to carry out direct observation, ADL or IADL scores obtained solely based on self-reports can be very inaccurate.

In Virtual ADL+ House, the game activities are designed in accordance with the daily activities in IADL. Although the mini-games do not involve many physical elements and have much simplified processes compared to the real-life activities, performance in the games can still be indicative of the ability or inability to perform the corresponding real-life activity, especially for activities like finance, medication and using telephone. Thus, the performance in the games may be used to complement the IADL self-reports to improve the accuracy of IADL scoring.

3.2 Test Description of Virtual ADL+ House

The Virtual ADL+ House consists of 11 tests (or 11 mini-games), covering 6 cognitive domains. Table 2 summarizes the tests according to the cognitive domains and shows the related IADL task of each test. The computerized battery is delivered as an iPad application. The tests mainly require tapping and speech inputs. No keyboard text inputs are required.

Rapid Visual Processing (Cleaning)

Rapid Visual Processing is a form of tapping test that is used for attention testing. It mainly assesses visual sustained attention, but also requires selective attention and working memory. It is widely used in both paper-and-pencil and computerized batteries, e.g. MoCA [27], CAMCI [35], and CANTAB [7]. Our version of the Rapid Visual Processing test follows a novel design that uses graphical stimuli instead of numbers. In a cleaning/ housekeeping activity setting, the test subject is instructed to pack a sequence of items together into a paper box, e.g. sock-scarf-sweater. Items are displayed in the center of the screen at the rate of 100 items per minute. An opened paper box is displayed below the items. When the test subject observes the target sequence, her/she needs to tap on the paper box to pack the items in the sequence into the box.

Simple Reaction Task (Cooking)

²The drop-out rate is calculated using the number of participants discontinued intervention divided by the total number of participants enrolled

Table 2: Components of the Virtual ADL+ House Cognitive Assessment Battery

Cognitive Domain	Virtual ADL+ House Test	Related IADL Task
Attention	Rapid Visual Processing	Cleaning
Processing Speed	Simple Reaction Task	Cooking
	Motor Screening Task	Cooking
	Digit Symbol Substitution Test	Finance
Language	Semantic Word Fluency	Cleaning
Visuospatial	Visual Spatial Imagery	Transportation
Memory	Digit Span	Using Telephone
	Paired Associate Learning	Shopping
	Benton Visual Retention Test	Medication
Executive Functioning	Go-No Go Test	Cooking
	Stroop Test Multitasking Version	Laundry

Adapted from the conventional Finger Tapping Test, the Simple Reaction Task is meant to assess motor speed as well as kinesthetic and visual-motor ability [26]. As shown in Fig. 3b, a cooking pot is placed in the center of the screen. Red tomatoes are presented in the center of the cooking pot, one at a time with different time intervals in between. The test subject must tap on the tomato on the screen as quickly as possible to prevent it from being overcooked.

Motor Screening Test (Cooking)

Adapted from the Motor Screening Test in the CANTAB battery [7], our version of the test is also designed to assess mainly visual-motor ability. Similar to the Simple Reaction Task, the game design is also getting tomatoes out of the cooking pot to prevent overcooking. The difference is that the tomatoes appear in different locations.

Digit Symbol Substitution Test (Finance)

The Digit Symbol Substitution Test is a variant of the Coding Test in WAIS-IV [21] which is mainly for assessing the processing speed, visual acuity and associative memory. Similar to NCGG-FAT [24], we reduce the number of stimuli and the requirements over motor functions to adapt the test for self-administration on tablets. The game scenario is understanding and paying bills. The test subject receives bills and writes checks in “foreign languages” which consist of unfamiliar symbols. Given the correspondence between symbols and numbers, the game is designed with two different modes - decoding the symbols into numbers and encoding the numbers into symbols. In order to figure out the amount to be paid, the test subject needs to decode the symbols by choosing the corresponding number for each symbol in sequence. In order to write a check to pay a bill, the test subject needs to perform the reversed process - encode the numbers digit by digit by choosing the corresponding symbol.

Semantic Word Fluency (Cleaning)

The Semantic Word Fluency Test is adapted from the conventional Category Fluency Test. It is designed to not only assess language and semantic memory, but also working memory and cognitive flexibility [19]. The test subject has to generate a list of

items, as many as possible, in a given category within a given period of time. The speech response of the test subject is recorded and analyzed. To fit into the cleaning theme, the categories used include “food” and “clothing” (to tidy up the fridge and the closet). Both the size of the clusters and the strategies used to generate the words will be considered for determining the performance.

Visual Spatial Imagery (Transportation)

The Visual Spatial Imagery is adapted from the Visual-Spatial Test in Mindstreams battery [5]. The game scenario is way-finding and map comprehension. The test subject is first given a computer-generated 3D map and tasked to move from a starting point to an end point via a stipulated route. The test subject is required to imagine the view from certain points along the route. At each viewpoint, four possible views are listed. The test subject needs to select the correct one among the four choices.

Digit Span (Using Telephone)

The Digit Span Test is adapted from the conventional Digit Span Forward, mainly assessing short-term memory [19]. The game is designed to be a dialing task. The first screen displays a phonebook with a series of numbers to call. After pressing the “Make Call” button, the test subject is directed to the second screen where a telephone is displayed. The test subject needs to recall and dial the numbers by tapping the corresponding keys on the telephone.

Paired Associate Learning (Shopping)

The Paired Associate Learning Test is adapted from the corresponding test in the CANTAB battery [7]. It is meant to assess visual memory. The game scenario is item finding in the garage. Closed paper boxes are displayed at different locations on the screen and are opened one by one in randomized order (see Fig. 3e). One or more of them will contain a pattern. The appeared patterns are then displayed one by one in the center of the screening. When a pattern is displayed, the test subject needs to tap on the box in which the pattern was contained. The patterns used are complex geometric shapes which are hard to be given any verbal meanings.

Benton Visual Retention Test (Medication)

The Benton Visual Retention Test is widely used for assessing visual memory, visual perception, and visual construction. Our

Figure 3: Screen shots of Virtual ADL+ House

adapted version of the test is designed with a game scenario of taking medication. The first screen displays the medications and the doses to be taken. Each medication is represented by a pill bottle labeled with a tag that has a geometric shape on it. Different geometric shapes represent different medications. Dose is indicated by a number, representing the number of pills to be taken. When the time for encoding is up, the test subject needs to select the correct pill bottles from a list of bottles by recognizing the patterns on the tags and choose the correct dose.

Go-No Go Test (Cooking)

The Go-No Go Test is a variant of the Continuous Performance Task (CPT) which is mainly for assessing executive functioning and attention [33]. Similar to TOVA [13] and Mindstreams [5], we use graphical stimuli instead of English letters. The game scenario is cooking vegetables. A series of vegetables are presented in the center of a cooking pot, one at a time with different time intervals in between. The test subject needs to tap only a specified vegetable while ignoring others, e.g. tap only onions while ignoring other vegetables.

Stroop Test Multitasking Version (Laundry)

The Stroop test is a well-established test [23] for executive functioning. It assesses the ability to suppress a usual response in favor of an unusual one and to switch perceptual set according to changing demands [37]. Our Stroop Test Multitasking Version is adapted from the Stroop Inference Test of the Mindstreams battery [5] and designed with a laundry theme. The task of the test subject is to separate laundries into correct baskets according to their colors

to prevent color transfer. Each piece of clothes is labeled by a tag with a colored word to indicate its color, while the clothes itself is colorless (refer to Fig. 3f). However, the letters of the word may or may not be in the color named by the word. For example, the word “red” may be in red or green. The test has three different modes. In the first mode, the color of the letters matches the color named by the word. The player needs to separate laundries according to the color of the letters. In the second mode, the color of the letters does not match the color named by the word. The player needs to separate laundries according to the color named by the word. In the third mode, the stimuli of the previous two modes are mixed, i.e. the color of the laundries may be named by the words or color of the letters. The test subject needs to switch between the two perceptual sets in the third mode.

4 PILOT STUDY

4.1 Study Design

The goal of our project is to develop a computerized cognitive test battery for long-term tracking and detection of early dementia. The test and evaluation of Virtual ADL+ House is divided into three phases:

- **Phase I:** focus groups to evaluate the usability and entertainment aspects of Virtual ADL+ House.
- **Phase II:** small scale pilot study to evaluate the validity (concurrent and discriminant), test-retest reliability of the battery and adjust cut-off scores.

Table 3: Score breakdown for each of the 12 statements

No.	Statement	Doctor Mean	Seniors Mean	Overall Mean
1	I think that elderly users would like to use this platform frequently.	4.00	4.00	4.00
2	I think that the platform is unnecessarily complex.	4.33	4.20	4.28
3	I think that the platform is easy to use.	4.50	4.40	4.46
4	I think that elderly users would need the support of a technical person to be able to use this platform.	3.29	4.20	3.64
5	I think that the various functions in this platform are well integrated.	4.14	4.00	4.09
6	I think that there is too much inconsistency in this platform.	4.14	4.00	4.09
7	I think that most people would learn to use this platform very quickly.	4.00	4.00	4.00
8	I think that this platform is very cumbersome to use.	4.29	3.80	4.10
9	I think elderly users would feel very confident using the platform.	4.00	4.20	4.08
10	I think elderly users would need to learn a lot of things before they could get going with this platform.	3.86	4.20	3.99
11	Overall, I find the tasks included in the platform to be fun/interesting/entertaining for elderly users.	4.38	4.00	4.23
12	The rewards/incentives and motivation mechanisms are sufficient to keep elderly users engaged and entertained in a sustained manner.	4.25	4.40	4.31

* The rows shaded in gray are negative statements. For positive statements, high scores indicate strongly agree. For negative statements, high scores indicate strongly disagree.

Table 4: Phase I evaluation results - usability and entertainment

Contents	Mean	SD	95% CI
Doctors	4.10	0.79	3.45, 4.74
Seniors	4.12	0.65	3.38, 4.86
Overall	4.10	0.73	3.67, 4.54

- **Phase III:** large scale study to collect trajectory data in the long-term, analyze disease progression patterns and identify high-risk individuals.

Since Virtual ADL+ House is a gamified and self-administered cognitive battery, the usability and entertainment aspects become prominent and need to be evaluated. We have already carried out Phase I of the evaluation and will present the results in the remaining part of this section.

A total of 13 participants attended our focus group studies. Two focus groups were organized to gather the feedback on Virtual ADL+ House. One focus group consisted of 8 doctors from the Tan Tock Seng hospital (the doctors' group), including psychologists, occupational therapists, and doctors and nurses from the memory clinic. The other group consisted of 5 healthy older adults aged above 60 (the seniors' group). During each focus group, the participants were first given brief demos of each test. They were given about 10 minutes to explore and experience the tests themselves after the demos. Then, guided by a moderator, the participants described their experience qualitatively. The moderator followed a list of 12 statements and noted down the degree to which each participant agreed with each statement on a scale from 1 to 5. For positive statements, 1 represents totally disagree while 5 represents totally agree. For negative statements, it is the reverse — 1 represents totally agree while 5 represents totally disagree. In general, a higher score indicate better usability or entertainment value.

4.2 Results and Discussion

Generally, we received positive feedback on Virtual ADL+ House. As shown in Table 4, the average score for all 12 statements is 4.1, with a standard deviation of 0.73 and a 95% confidence interval [3.67, 4.54]. The average score of the seniors' group is 4.12 (SD = 0.65, 95% CI = [3.38, 4.86]), slightly higher than that of the doctor's group which is 4.1 (SD = 0.79, 95% CI = [3.45, 4.74]).

According to the score breakdown of the 12 statements in Table 3, statement 3 received the highest overall mean score (4.46), indicating that the participants strongly agreed that Virtual ADL+ House is easy to use. Following statement 3 is statement 12, which received an overall mean score of 4.31. The participants highly agreed that Virtual ADL+ House can provide enough incentives to keep elderly users engaged and entertained in the long-term.

The results also reveal some areas for improvements. Two statements received overall mean scores below 4. Statement 4 received the lowest overall mean score (3.64). Since it is a negative statement, a low score indicates that the participants somewhat agree with the statement. Hence, some participants thought that the elderly would need the support of a technical person in order to use Virtual ADL+ House. Similar for statement 10 which received a mean score of 3.99, some participants agreed that the elderly users would need to learn a lot of things before they can use the battery smoothly. Generally, the doctors' group and the seniors' group had similar attitudes towards Virtual ADL+ House, which is manifested by very close group means for most of the statements. However, they had divergent views on statement 4 and 10. The seniors' group tended to be more positive. They gave much higher scores to these two statements. The mean score of the seniors' group for statement 4 is 4.2, nearly 1 point higher than that of the doctors' group. The mean score of the seniors' group for statement 10 is 4.2, also much higher than the 3.86 rated by the doctors' group. On one hand, the seniors demonstrated their confidence that they can deal with Virtual ADL+ House by themselves without much problems. While on the other hand, the doctors showed more worries towards leaving a comprehensive computerized battery completely self-administered.

5 CONCLUSIONS

Dementia is the leading chronic disease that already places an important burden on the healthcare system. With no effective treatment in its advanced stage, a promising strategy of treating dementia is to identify individuals that have high risk of developing dementia in its early stage, i.e. in MCI or even pre-MCI stage. The detection of early dementia requires long-term and frequent assessments of cognitive functions.

We presented Virtual ADL+ House which is, to the best of our knowledge, the first-ever gamified computerized cognitive test battery. We embedded neuropsychological tests covering a range of cognitive functions into a series of mini-games. Each game is designed in the context of one of the daily activities of IADL, yet preserving the test mechanism of the original neuropsychological test. In the focus groups conducted to evaluate Virtual ADL+ House, we received mostly positive feedback from doctors and older adults regarding its usability and entertainment value. However, the doctors' group expressed concerns over the steep learning curve required to get familiarized with Virtual ADL+ House. To address these concerns, we will consider to improve the game descriptions, add short game tutorials before the first play, and conduct training sessions to the participants of later studies.

Formal clinical investigations, including blood tests, brain scans, and conventional neuropsychological test, are still required to reach a final diagnostic decision. Virtual ADL+ House can be used to monitor cognitive functions in home environment, alert early dementia signs and provide longitudinal assessment data to supplement clinical evaluation.

ACKNOWLEDGMENTS

This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its IDM Futures Funding Initiative, the Singapore Ministry of Health under its National Innovation Challenge on Active and Confident Ageing (NIC Project No. MOH/NIC/COG04/2017), and the Interdisciplinary Graduate School, Nanyang Technological University under its Graduate Research Scholarship.

REFERENCES

- [1] Heiko Braak and Eva Braak. 1991. Neuropathological staging of Alzheimer-related changes. *Acta neuropathologica* 82, 4 (1991), 239–259.
- [2] David Darby, P Maruff, A Collie, and M McStephen. 2002. Mild cognitive impairment can be detected by multiple assessments in a single day. *Neurology* 59, 7 (2002), 1042–1046.
- [3] Glen M Doniger, David M Zucker, Avraham Schweiger, Tzvi Dwolatzky, Howard Chertkow, Howard Crystal, and Ely S Simon. 2005. Towards practical cognitive assessment for detection of early dementia: a 30-minute computerized battery discriminates as well as longer testing. *Current Alzheimer Research* 2, 2 (2005), 117–124.
- [4] John H Dougherty Jr, Rex L Cannon, Christopher R Nicholas, Lorin Hall, Felicia Hare, Erika Carr, Andrew Dougherty, Jennifer Janowitz, and Justin Arunthamkun. 2010. The computerized self test (CST): an interactive, internet accessible cognitive screening test for dementia. *Journal of Alzheimer's Disease* 20, 1 (2010), 185–195.
- [5] Tzvi Dwolatzky, Victor Whitehead, Glen M Doniger, Ely S Simon, Avraham Schweiger, Dena Jaffe, and Howard Chertkow. 2003. Validity of a novel computerized cognitive battery for mild cognitive impairment. *BMC geriatrics* 3, 4 (2003).
- [6] Mary Margaret Edwards. 1990. The reliability and validity of self-report activities of daily living scales. *Canadian Journal of Occupational Therapy* 57, 5 (1990), 273–278.
- [7] Anikó Egerházi, Roland Berecz, Enikő Bartók, and István Degrell. 2007. Automated Neuropsychological Test Battery (CANTAB) in mild cognitive impairment and in Alzheimer's disease. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 31, 3 (2007), 746–751.
- [8] Richard W Elwood. 2001. MicroCog: assessment of cognitive functioning. *Neuropsychology review* 11, 2 (2001), 89–100.
- [9] Marina G Falletti, Paul Maruff, Alexander Collie, and David G Darby. 2006. Practice effects associated with the repeated assessment of cognitive function using the CogState battery at 10-minute, one week and one month test-retest intervals. *Journal of clinical and experimental neuropsychology* 28, 7 (2006), 1095–1112.
- [10] Martin Farlow, Steven Potkin, Barbara Koumaras, Jeffrey Veach, and Dario Mirski. 2003. Analysis of outcome in retrieved dropout patients in a rivastigmine vs placebo, 26-week, Alzheimer disease trial. *Archives of neurology* 60, 6 (2003), 843–848.
- [11] Charles Flicker, Steven H Ferris, and Barry Reisberg. 1991. Mild cognitive impairment in the elderly predictors of dementia. *Neurology* 41, 7 (1991), 1006–1006.
- [12] Carla Graf. 2009. The Lawton instrumental activities of daily living (IADL) scale. *The gerontologist* 9, 3 (2009), 179–186.
- [13] Lawrence M Greenberg and Irwin D Waldman. 1993. Developmental normative data on the Test of variables of attention (TOVA&Dc). *Journal of Child Psychology and Psychiatry* 34, 6 (1993), 1019–1030.
- [14] Michael Grundman, Ronald C Petersen, Steven H Ferris, Ronald G Thomas, Paul S Aisen, David A Bennett, Norman L Foster, Clifford R Jack Jr, Douglas R Galasko, Rachelle Doody, et al. 2004. Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials. *Archives of neurology* 61, 1 (2004), 59–66.
- [15] C Thomas Gualtieri and Lynda G Johnson. 2005. Neurocognitive testing supports a broader concept of mild cognitive impairment. *American Journal of Alzheimer's Disease & Other Dementias* 20, 6 (2005), 359–366.
- [16] C Thomas Gualtieri and Lynda G Johnson. 2006. Reliability and validity of a computerized neurocognitive test battery, CNS Vital Signs. *Archives of Clinical Neuropsychology* 21, 7 (2006), 623–643.
- [17] Masashi Inoue, Daiki Jimbo, Miyako Taniguchi, and Katsuya Urakami. 2011. Touch Panel-type Dementia Assessment Scale: a new computer-based rating scale for Alzheimer's disease. *Psychogeriatrics* 11, 1 (2011), 28–33.
- [18] Noor Jehan Kabani, John G Sled, Anita Shuper, and Howard Chertkow. 2002. Regional magnetization transfer ratio changes in mild cognitive impairment. *Magnetic resonance in medicine* 47, 1 (2002), 143–148.
- [19] Elke Kalbe, Josef Kessler, Pasquale Calabrese, R Smith, AP Passmore, Met al Brand, and R Bullock. 2004. DemTect: a new, sensitive cognitive screening test to support the diagnosis of mild cognitive impairment and early dementia. *International journal of geriatric psychiatry* 19, 2 (2004), 136–143.
- [20] Francesco Landi, Matteo Cesari, Graziano Onder, Andrea Russo, Sergio Torre, and Roberto Bernabei. 2003. Non-steroidal anti-inflammatory drug (NSAID) use and Alzheimer disease in community-dwelling elderly patients. *The American journal of geriatric psychiatry* 11, 2 (2003), 179–185.
- [21] Elizabeth O Lichtenberger and Alan S Kaufman. 2009. *Essentials of WAIS-IV assessment*. Vol. 50. John Wiley & Sons.
- [22] Christine Lowe and Patrick Rabbitt. 1998. Test-retest reliability of the CANTAB and ISPOCD neuropsychological batteries: theoretical and practical issues. *Neuropsychologia* 36, 9 (1998), 915–923.
- [23] Colin M MacLeod. 1991. Half a century of research on the Stroop effect: an integrative review. *Psychological bulletin* 109, 2 (1991), 163.
- [24] Hyuma Makizako, Hiroyuki Shimada, Hyuntae Park, Takehiko Doi, Daisuke Yoshida, Kazuki Uemura, Kota Tsutsumimoto, and Takao Suzuki. 2013. Evaluation of multidimensional neurocognitive function using a tablet personal computer: Test-retest reliability and validity in community-dwelling older adults. *Geriatrics & gerontology international* 13, 4 (2013), 860–866.
- [25] Fadi Massoud, Howard Chertkow, Victor Whitehead, Olga Overbury, and Howard Bergman. 2002. Word-reading thresholds in Alzheimer disease and mild memory loss: a pilot study. *Alzheimer Disease & Associated Disorders* 16, 1 (2002), 31–39.
- [26] Maura Mitrushina, Kyle B Boone, Jill Razani, and Louis F D'Elia. 2005. *Handbook of normative data for neuropsychological assessment*. Oxford University Press.
- [27] Ziad S Nasreddine, Natalie A Phillips, Valérie Bédirian, Simon Charbonneau, Victor Whitehead, Isabelle Collin, Jeffrey L Cummings, and Howard Chertkow. 2005. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society* 53, 4 (2005), 695–699.
- [28] Tiia Ngandu, Jenni Lehtisalo, Alina Solomon, Esko Levälähti, Satu Ahtiluoto, Riitta Antikainen, Lars Bäckman, Tuomo Hänninen, Antti Jula, Tiina Laatikainen, et al. 2015. A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. *The Lancet* 385, 9984 (2015), 2255–2263.
- [29] JM Olichney, SK Morris, C Ochoa, DP Salmon, LJ Thal, M Kutas, and VJ Iragui. 2002. Abnormal verbal event related potentials in mild cognitive impairment and incipient Alzheimer's disease. *Journal of Neurology, Neurosurgery & Psychiatry* 73, 4 (2002), 377–384.
- [30] Ronald C Petersen. 2000. Aging, mild cognitive impairment, and Alzheimer's disease. *Neurologic Clinics* 18, 4 (2000), 789–805.

- [31] Ronald C Petersen. 2003. Mild cognitive impairment clinical trials. *Nature Reviews Drug Discovery* 2, 8 (2003), 646.
- [32] Ronald C Petersen, James C Stevens, Mary Ganguli, Eric G Tangalos, JL Cummings, and ST DeKosky. 2001. Practice parameter: Early detection of dementia: Mild cognitive impairment (an evidence-based review) Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology* 56, 9 (2001), 1133–1142.
- [33] Cynthia A Riccio and Cecil R Reynolds. 2001. Continuous performance tests are sensitive to ADHD in adults but lack specificity. *Annals of the New York Academy of Sciences* 931, 1 (2001), 113–139.
- [34] Mark A Sager, Nancy Cross Dunham, Amy Schwantes, Laurie Mecum, Kaia Halverson, and Diane Harlowe. 1992. Measurement of Activities of Daily Living in Hospitalized Elderly: A Comparison of Self-Report and Performance-Based Methods. *Journal of the American Geriatrics Society* 40, 5 (1992), 457–462.
- [35] Judith Saxton, Lisa Morrow, Amy Eschman, Gretchen Archer, James Luther, and Anthony Zuccolotto. 2009. Computer assessment of mild cognitive impairment. *Postgraduate medicine* 121, 2 (2009), 177–185.
- [36] V Solfrizzi, F Panza, and A Capurso. 2003. The role of diet in cognitive decline. *Journal of neural transmission* 110, 1 (2003), 95–110.
- [37] Otfried Spreen and Esther Strauss. 1998. *A compendium of neuropsychological tests: Administration, norms, and commentary*. Oxford University Press.
- [38] Rachel Swainson, JR Hodges, CJ Galton, J Semple, A Michael, BD Dunn, JL Iddon, TW Robbins, and BJ Sahakian. 2001. Early detection and differential diagnosis of Alzheimer's disease and depression with neuropsychological tasks. *Dementia and geriatric cognitive disorders* 12, 4 (2001), 265–280.
- [39] Mary C Tierney and Miriam A Lerner. 2010. Computerized cognitive assessment in primary care to identify patients with suspected cognitive impairment. *Journal of Alzheimer's Disease* 20, 3 (2010), 823–832.
- [40] Jane B Tornatore, Emory Hill, Jo Anne Laboff, and Mary E McGann. 2005. Self-administered screening for mild cognitive impairment: initial validation of a computerized test battery. *The Journal of neuropsychiatry and clinical neurosciences* 17, 1 (2005), 98–105.
- [41] Catharine Trustring Eve and Celeste A Jager. 2014. Piloting and validation of a novel self-administered online cognitive screening tool in normal older persons: the Cognitive Function Test. *International journal of geriatric psychiatry* 29, 2 (2014), 198–206.
- [42] Eric P Moll van Charante, Edo Richard, Lisa S Eurelings, Jan-Willem van Dalen, Suzanne A Ligthart, Emma F Van Bussel, Marieke P Hoevenaar-Blom, Marinus Vermeulen, and Willem A van Gool. 2016. Effectiveness of a 6-year multidomain vascular care intervention to prevent dementia (preDIVA): a cluster-randomised controlled trial. *The Lancet* 388, 10046 (2016), 797–805.
- [43] Roberta F White, Kenneth E James, Jennifer J Vasterling, Richard Letz, Karen Marans, Richard Delaney, Maxine Krengel, Fredric Rose, and Helena C Kraemer. 2003. Neuropsychological screening for cognitive impairment using computer-assisted tasks. *Assessment* 10, 1 (2003), 86–101.
- [44] Katherine Wild, Diane Howieson, Frank Webbe, Adriana Seelye, and Jeffrey Kaye. 2008. Status of computerized cognitive testing in aging: a systematic review. *Alzheimer's & dementia: the journal of the Alzheimer's Association* 4, 6 (2008), 428–437.
- [45] David L Woods, John M Wyma, Timothy J Herron, and E William Yund. 2015. The effects of aging, malingering, and traumatic brain injury on computerized trail-making test performance. *PloS one* 10, 6 (2015), e0124345.
- [46] Gelin Xu, John S Meyer, John Thornby, Munir Chowdhury, and Minh Quach. 2002. Screening for mild cognitive impairment (MCI) utilizing combined minimal-cognitive capacity examinations for identifying dementia prodromes. *International journal of geriatric psychiatry* 17, 11 (2002), 1027–1033.
- [47] Stelios Zygoris and Magda Tsolaki. 2015. Computerized cognitive testing for older adults: a review. *American Journal of Alzheimer's Disease & Other Dementias* 30, 1 (2015), 13–28.