Journal of Alzheimer's Disease 85 (2022) 1129–1142
DOI 10.3233/JAD-215015
IOS Press

Identifying Mild Cognitive Impairment by Using Human–Robot Interactions

Yu-Ling Chang[a,b,c,d,*], Di-Hua Luo[a], Tsung-Ren Huang[a,b,d], Joshua O.S. Goh[a,b,d,e], Su-Ling Yeh[a,b,d,e] and Li-Chen Fu[f,g,h]

[a]*Department of Psychology, College of Science, National Taiwan University, Taipei, Taiwan*
[b]*Neurobiology and Cognitive Science Center, National Taiwan University, Taipei, Taiwan*
[c]*Department of Neurology, National Taiwan University Hospital, College of Medicine, National Taiwan University, Taipei, Taiwan*
[d]*Center for Artificial Intelligence and Advanced Robotics, National Taiwan University, Taipei, Taiwan*
[e]*Graduate Institute of Brain and Mind Sciences, College of Medicine, National Taiwan University, Taipei, Taiwan*
[f]*Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan*
[g]*Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan*
[h]*MOST Joint Research Center for AI Technology and All Vista Healthcare, Taipei, Taiwan*

Handling Associate Editor: David Libon

Accepted 31 October 2021
Pre-press 10 December 2021

Abstract.

Background: Mild cognitive impairment (MCI), which is common in older adults, is a risk factor for dementia. Rapidly growing health care demand associated with global population aging has spurred the development of new digital tools for the assessment of cognitive performance in older adults.

Objective: To overcome methodological drawbacks of previous studies (e.g., use of potentially imprecise screening tools that fail to include patients with MCI), this study investigated the feasibility of assessing multiple cognitive functions in older adults with and without MCI by using a social robot.

Methods: This study included 33 older adults with or without MCI and 33 healthy young adults. We examined the utility of five robotic cognitive tests focused on language, episodic memory, prospective memory, and aspects of executive function to classify age-associated cognitive changes versus MCI. Standardized neuropsychological tests were collected to validate robotic test performance.

Results: The assessment was well received by all participants. Robotic tests assessing delayed episodic memory, prospective memory, and aspects of executive function were optimal for differentiating between older adults with and without MCI, whereas the global cognitive test (i.e., Mini-Mental State Examination) failed to capture such subtle cognitive differences among older adults. Furthermore, robot-administered tests demonstrated sound ability to predict the results of standardized cognitive tests, even after adjustment for demographic variables and global cognitive status.

Conclusion: Overall, our results suggest the human–robot interaction approach is feasible for MCI identification. Incorporating additional cognitive test measures might improve the stability and reliability of such robot-assisted MCI diagnoses.

Keywords: Cognitive assessment, dementia, health care, human–robot interaction, mild cognitive impairment, older adults

*Correspondence to: Yu-Ling Chang, PhD, Department of Psychology, National Taiwan University, No. 1, Sec. 4, Roosevelt Rd, Taipei 10617, Taiwan. Tel.: +886 2 33663105; Fax: +886 2 23629909; E-mail: ychang@ntu.edu.tw; ORCID: 0000-0003-2851-3652.

ISSN 1387-2877/$35.00 © 2022 – IOS Press. All rights reserved.

INTRODUCTION

Rapidly growing health care demand associated with the aging population and the corresponding increase in the prevalence of chronic conditions such as dementia has placed considerable strain on health care systems worldwide. Evidence from both neuropsychological and neuroimaging studies suggests that mild cognitive impairment (MCI) represents a risk factor for as well as a transitional phase in degenerative dementias such as Alzheimer's disease (AD) [1–3]. Early MCI detection may facilitate the identification of people eligible for medical or behavioral interventions and may delay the onset of dementia [4, 5], and it may help identify patients eligible for participation in clinical trials that promote the development of innovative AD therapies [6].

Assessment of cognitive function is a crucial component of the multifaceted diagnostic process for MCI. Traditionally, cognitive evaluations are conducted in a clinical setting through standardized paper–pencil tests. However, impediments such as insufficient access to health care personnel and resources [6], increased health care costs, and inadequate self-awareness among older adults for recognizing their cognitive impairment (particularly when the impairment is mild) often prevent older adults from receiving appropriate health care [6–8]. The new wave of innovation and the increasing accessibility of technology hold promise for the early diagnosis and management of MCI. Among these modern technologies, robotic technology, which involves the use of robots to perform tasks traditionally performed by humans, offers several advantages over other technology or devices, such as tablets or web-based platforms, for assessing or monitoring older adults' cognitive function. Specifically, robotic technology can provide standardized administration of cognitive tests and obtain objective and granular behavioral data efficiently and reproducibly, which are essential requirements for reliable cognitive assessments, and such technology can offer an embodied presence that may enhance user engagement during testing [9–11]. More critically, such assessments are more stimulating and interactive than other remote testing methods [12]. Robot-assisted cognitive assessments also automatically record participant responses and behaviors for further analysis by professionals. It also requires a lower user technological proficiency than that required for interfacing with other electronic devices (e.g., a computer or virtual electronic device) [13]. Hence, compared with other remote assessment methods, robotic assessments of cognitive function are more enjoyable and accessible for older adults who prefer at-home health care services [14].

Studies have suggested that robotic technology is a promising technique for assessing older adults' cognitive function [15–17]. However, these studies had certain drawbacks. For example, all these studies [15–17] have focused on cognitively healthy older adults because their primary goal was to demonstrate the feasibility of applying a robotic technology in cognitive evaluations of this population. Therefore, empirical data on robotic assessments of MCI remain limited. Moreover, these researchers have only embedded a screening test for assessing global cognitive function, such as the Mini-Mental State Examination (MMSE), Montreal Cognitive Assessment, or the Telephone Interview for Cognitive Status, in their robotic systems [15–17]. Although the aforementioned tests are commonly used in clinical and research settings, these brief screening measures often fail to capture the subtle but noteworthy cognitive and functional changes during the MCI stage [2, 18]. Moreover, the scores obtained in these screening tests may be more easily skewed by age, education, or socioeconomic biases [19]. Evidence has shown that incorporating more neuropsychological measures into MCI diagnosis improves the sensitivity and reliability of a diagnosis that predicts clinical decline [1, 20–22].

To address these knowledge gaps, two key questions were addressed in this study. First, we examined the feasibility of using robot-administered cognitive tests to detect differences among older adults with MCI, healthy older adults (HO), and healthy young adults (HY). In the present study, we designed a series of short cognitive tests focused on assessing language, episodic memory, prospective memory, and executive function, which are the cognitive domains that best discriminate age-associated cognitive changes from MCI [22–24]. Second, we examined the construct validity of the robot-administered cognitive tests. Specifically, we performed standardized neuropsychological tests to serve as a benchmark for the robot-administered tests. We hypothesized that a human–robot interaction (HRI) approach is feasible for the detection of MCI. Additionally, although age-associated cognitive changes in episodic memory, prospective memory, and aspects of executive function were expected to be observed in older adults with or without MCI relative to HY, older adults with MCI were hypothesized to have lower performances

Fig. 1. A participant interacts with the social robot, RoBoHoN, in a laboratory containing a one-way mirror. Permission to use the picture was granted by the participant.

in the aforementioned cognitive domains compared with HO. We also hypothesized that participants' performances on the robotic assessment tests could predict their performances of the corresponding standardized neuropsychological tests.

METHODS

Participants

This study group consisted of 33 HY (ages 18–35 years) and 33 older adults (aged older than 60 years). Among the older adults, 15 were individuals with MCI and 18 were HO. Participants were thoroughly screened through interviews, and individuals with a current or past diagnosis of a neurological or psychiatric disorder, a known head injury that involved loss of consciousness, a history of alcohol or substance abuse in any a given period of life, or any severe hearing or visual impairments that might lower their neuropsychological performance were excluded. The HY and HO participants were recruited from local communities with an MMSE score of ≥ 26. Individuals with MCI were recruited from the memory clinics of the National Taiwan University Hospital. The participants were classified as having MCI according to the following criteria: 1) normal activities of daily living, 2) absence of dementia, 3) a global rating of 0.5 on the Clinical Dementia Rating scale [25], and 4) objective cognitive impairment, which was operationally defined as performance ≥ 1.0 standard deviation (SD) lower than the age-appropriate norm on at least two measures in one or more cognitive domains [26]. Four neuropsychological domains were assessed using the following tests: 1) attention: the digit span forward length and digit symbol substitution (DSS) of the Wechsler Adult Intelligence Scale, third edition (WAIS-3) [27]; 2) language: the vocabulary subtest of WAIS-3 and the 30-item Boston Naming Test [28]; 3) learning and memory: the logical memory and visual reproduction subtests of the Wechsler Memory Scale, third edition (WMS-3) [29]; and 4) executive function: the design fluency test (switching condition) of the Delis–Kaplan Executive Function System (D-KEFS) [30] and the letter-number sequencing subtest of the WAIS-3. The present study was approved by the Ethics Committee and Institutional Review Board of both National Taiwan University Hospital and National Taiwan University. Before conducting the experimental procedures, written informed consent was obtained from all participants.

Social robot

RoBoHoN, a fourth-generation mobile communication robot developed by Sharp Corporation, was used to administer the cognitive tests in the HRI study. RoBoHoN is a humanoid-shaped robot that is 19.5-cm tall and has built-in functions of talking with users as well as performing actions such as walking, moving its hands, dancing, and turning its head (Fig. 1). Therefore, to build a rapport and render the test process more enjoyable for the participants, the robot was preprogrammed to initiate context-specific small talk with the participants at the beginning of the experimentation sessions and in between the tests. The robot also performed several dance moves at the end of the HRI experiment as reward for the participants' efforts in the study.

Experimental procedure

The present study was part of a larger HRI study that consisted of three sessions: 1) robot-administered cognitive testing, 2) robot-accompanied toy-playing, and 3) robot-assisted questionnaire data collection [13]. The present study only presents data collected in the first session, which had a duration of approximately 25 min. In the HRI session, each participant interacted with the robot alone in a room, with the robot placed on a table in front of the participant such that it was easy to interact and make eye contact with each other (Fig. 1). The HRI session was monitored by research staff seated behind the one-way mirror hidden from participants. All data in the HRI session was recorded by a hidden camera. Robot-administered cognitive testing data collected during the experimentation session were scored offline by a professional neuropsychologist who watched the recorded videos. A subset of the participants (17 HYs, 13 HOs, and 15 individuals with MCI) underwent a second cognitive testing session to collect their performance on selective standardized neuropsychological tests, which was used as a benchmark for the robot-administered cognitive tests. The two cognitive testing sessions were no more than one week apart. All the participants completed the MMSE first as part of the screening process, then the robot-administered cognitive test session, and finally the standardized neuropsychological test session. Overall, the testing protocol used in the present study was well received by all individuals in the cohort, and none demonstrated adverse effects resulting from the robotic assessment. No participant reported discomfort or difficulty in completing the tasks.

Cognitive tests

Robot-administered cognitive test

Five brief cognitive tests were administered by the robot to assess participants' cognitive functioning in verbal fluency, episodic memory, prospective memory, and aspects of executive function (e.g., working memory, inhibition, and shifting). Specifically, these tests included category fluency (animals) and a backward digit span test similar to the digit span subtest of the WAIS-III [27]; in this test, the participant repeated the digits (with a span ranging from two to eight digits) read by the robot in backward order.

Additionally, a word-list learning test, which consisted of three learning trials of ten concrete nouns, was administered. Each item of the list was read by the robot, and participants were asked to recall all words they remembered in any order following each of three exposures. Ten minutes after the study session, a delayed free recall test was administered, followed by an old (for ten learned items) or new (for ten lures) verbal recognition test. Three indexes generated from this task were used for analyses, namely total learning score, delayed free recall score, and delayed recognition d prime (*d*', the standard hit rate score minus the standard false alarm rate score).

An event-based prospective memory (PM) test was also administered. Participants were instructed to inform the robot of their birthdate as soon as they heard the verbal cue: the robot inquired about their need to go to a restroom (approximately 20 min after initial instruction) during the assessment. Participants could receive a maximal total score of four points: two points for the prospective component and two points for the retrospective component. For the prospective component, two points were awarded if the participant demonstrated any intention to respond to the cues (e.g., saying "I know I was supposed to say something, but I cannot remember what it is") within 15 s, one point if their response was late or required prompting (i.e., "Was there something you want to tell me?" prompted by the robot), and zero points if no action was performed. For the retrospective component, the participant received two points if a correct response was generated, one point if a similar response (e.g., telling the date of the experiment conducted instead of their birthdate) was performed, and zero points if the wrong action (e.g., answering the prospective cue literally) or no action was carried out.

Furthermore, to evaluate inhibition and shifting abilities in terms of executive function, a modified version of the Color–Word Interference Test (CWIT) of the Delis–Kaplan Executive Function System (D-KEFS) [30] was administered. In this task, verbal instructions were read out by the robot, and the stimuli were visually presented on a tablet. The participants were asked to name the color patches (condition 1), the color words (condition 2), the color of ink instead of the word written (condition 3), or to shift between naming the color of the ink or the word written (condition 4) as quickly as possible. The completion time for each condition was recorded by the robot separately through the verbal responses, but only the completion times of conditions 3 and 4 were used for further analyses.

Paper and pencil-based standardized cognitive test

As a benchmark for the robot-administered results, three standardized cognitive tests were administered: 1) the MMSE, 2) conditions 3 and 4 of the CWIT of the D-KEFS [30], which measure inhibition and shifting abilities, respectively, and 3) the California Verbal Learning Test-II (CVLT-II) [31]. Notably, these tests were not used for MCI diagnoses.

Statistical analyses

Analyses of variance (ANOVAs) or chi-squared tests were used to compare group demographics (age, education, and sex). Because educational levels significantly differed between groups, to evaluate group differences in cognitive variables, we conducted ANCOVAs controlled for educational level, with Bonferroni adjustments for Type I errors at an α level of 0.0056 for the robot-administered tests and an α level of 0.0125 for the set of paper–pencil tests. The α level was set at 0.05 for all *post hoc* comparisons. The effect sizes were calculated for pairwise comparisons of cognitive variables that were significant according to Cohen's *d* [32].

To evaluate the ability of the robot-administered cognitive tests to differentiate between individuals with and without MCI, we conducted binomial logistic regression, with an α level of 0.05. Specifically, the predictors included in the model were robot-administered cognitive test variables that were significantly different between the HO and MCI groups. The overall accuracy, sensitivity, specificity, positive predictive value, and negative predictive value of the robot-administered tests to identify individuals with MCI were calculated.

Separate hierarchal multiple regression analyses were conducted on data for the full cohort to determine the predictive value of robot-administered tests for performance of the corresponding standardized neuropsychological tests. In the regression models, age, education, sex (dummy coded as 1 = men, 0 = women), and MMSE scores were entered in the first step as predictors. The robot-administered test results that exhibited significant differences between the HO and MCI groups, namely the word-list total learning scores (model 1), delayed recognition *d*' scores (model 2), PM prospective scores (model 3), modified CWIT condition 3 scores (model 4), and modified CWIT condition 4 scores (model 5), were entered in the second step individually as a predictor for each model. The corresponding dependent variables were CVLT-II total learning (model 1), long-delayed free recall (model 2), D-KEFS CWIT condition 4 (shifting condition, model 3), and D-KEFS CWIT conditions 3 (model 4) and 4 (model 5). The mappings between the predictors and the dependent variables were based on the similarity of the test constructs and underlying cognitive process between the predictors and dependent variables (e.g., word-list total learning to the CVLT-II total learning). Additionally, in model 2, CVLT-II delayed free recall was selected as the dependent variable because evidence suggests that this variable, rather than the delayed recognition variable, is the optimal neuropsychological predictor of progression from MCI to Alzheimer's disease [33, 34]. Notably, PM has been reported to be strongly correlated with shifting ability in executive function [35, 36], and therefore, D-KEFS CWIT condition 4, which assesses shifting ability, was used as the dependent variable in our regression analysis for the robotic PM test. Results were reported as significant at the threshold of $p < 0.01$ (Bonferroni correction) for each overall model, and significant predictors were selected using the forward selection method.

RESULTS

Demographic characteristics

The three groups did not differ in their frequency distribution based on sex. Significant group differences were observed for age, but the HO group and the MCI group were comparable in age. Significant group differences were observed for educational level, with the MCI group having a lower level of educational attainment compared with the HY and HO groups ($p = 0.006$ and 0.005, respectively), but no difference between the HY and HO groups was discovered (Table 1).

Cognitive performance

Group comparison results of robot-administered tests

Because educational levels significantly differed among groups, the analyses for the robot-administered cognitive tests were adjusted for educational attainment. The ANCOVA results (Table 1) revealed

Table 1
Demographic and cognitive characteristics of the healthy young adults (HY), healthy old adults (HO), and older adults with mild cognitive impairment (MCI)

	HY (mean SD)	HO (mean SD)	MCI (mean SD)	*Statistics*
Demographics variables				
Age	21.52 (2.24)	71.11 (3.55)	73.53 (6.39)	$F_{(2,63)} = 1414.1, p < 0.001^{\dagger}$
Education	15.06 (1.97)	15.44 (3.76)	12.67 (2.74)	$F_{(2,63)} = 5.09, p = 0.009^{\ddagger\S}$
Gender (women/men)	17/16	8/10	9/6	$\chi^2_{(2,N=66)} = 0.79, p > 0.05$
Robot-administered cognitive test				
Category fluency	24.06 (5.91)	17.94 (5.56)	14.40 (4.47)	$F_{(2,62)} = 15.24, p < 0.001^{\dagger}$, partial $\eta^2 = 0.33$
Digit span backward length (max = 8)	5.44 (1.48)	2.83 (1.43)	2.71 (1.20)	$F_{(2,62)} = 26.75, p < 0.001^{\dagger}$, partial $\eta^2 = 0.47$
WLT total learning (max = 30)	21.36 (3.14)	19.83 (2.77)	17.27 (4.59)	$F_{(2,62)} = 7.15, p = 0.002^{\ddagger\S}$, partial $\eta^2 = 0.19$
WLT delayed free recall (max = 10)	7.58 (1.32)	6.69 (2.06)	5.60 (2.80)	$F_{(2,62)} = 5.70, p = 0.005^{\ddagger}$, partial $\eta^2 = 0.16$
WLT delayed recognition *d*'	0.34 (0.01)	0.33 (0.03)	0.28 (0.08)	$F_{(2,62)} = 7.98, p = 0.001^{\ddagger\S}$, partial $\eta^2 = 0.21$
PM prospective component (max = 2)	1.72 (0.58)	1.11 (0.90)	0.40 (0.73)	$F_{(2,62)} = 19.56, p < 0.001^{\dagger\S}$, partial $\eta^2 = 0.39$
PM retrospective component (max = 2)	1.79 (0.55)	1.17 (0.99)	0.53 (0.92)	$F_{(2,62)} = 13.40, p < 0.001^{\dagger\S}$, partial $\eta^2 = 0.30$
Modified CWIT condition 3 (s)	39.73 (9.19)	60.56 (11.62)	84.33 (26.49)	$F_{(2,62)} = 37.98, p < 0.001^{\dagger\S}$, partial $\eta^2 = 0.55$
Modified CWIT condition 4 (s)	51.24 (13.08)	76.61 (23.07)	121.67 (59.10)	$F_{(2,62)} = 20.70, p < 0.001^{\dagger\S}$, partial $\eta^2 = 0.40$
Standardized cognitive tests				
MMSE (max = 30)	29.41 (0.80)	27.98 (1.46)	26.40 (2.23)	$F_{(2,62)} = 10.89, p < 0.001^{\dagger}$, partial $\eta^2 = 0.32$
CVLT-II total learning (max = 80)	55.47 (5.43)	42.77 (7.97)	33.40 (11.10)	$F_{(2,41)} = 22.99, p < 0.001^{\dagger\S}$, partial $\eta^2 = 0.53$
CVLT-II long-delayed free recall (max = 16)	13.00 (2.26)	9.62 (2.22)	6.13 (3.72)	$F_{(2,41)} = 19.62, p < 0.001^{\dagger\S}$, partial $\eta^2 = 0.49$
D-KEFS CWIT condition 3 (s)	33.59 (7.69)	59.69 (12.72)	88.73 (27.97)	$F_{(2,41)} = 29.80, p < 0.001^{\dagger\S}$, partial $\eta^2 = 0.59$
D-KEFS CWIT condition 4 (s)	44.71 (11.01)	70.15 (14.22)	128.67 (50.90)	$F_{(2,41)} = 22.22, p < 0.001^{\dagger\S}$, partial $\eta^2 = 0.52$

All scores are raw scores. CVLT-II, California Verbal Learning test, second version; D-KEFS CWIT, the Color–Word Interference Test (CWIT) of the Delis–Kaplan Executive Function System; MMSE, Mini-Mental State Examination; PM, prospective memory; WLT, word-list test. *p* indicates the results of the overall intergroup comparisons. All results of cognitive tests, including the MMSE, were based on analyses with educational level as a covariate; † indicates significant differences between the HY group and both the HO and MCI groups; ‡ indicates a significant difference between the HY group and the MCI group; § indicates a significant difference between the HO group and the MCI group.

that the three groups significantly differed across all tests administered. The *post hoc* analyses demonstrated an effect of aging (i.e., HY > HO = MCI) on the category fluency test and digit span backward test. Specifically, the HY group outperformed both the HO ($p < 0.001$, $d = 1.07$ versus $p < 0.001$, $d = 1.80$) and MCI ($p < 0.001$, $d = 1.84$ versus $p < 0.001$, $d = 2.03$) groups, whereas the two older groups exhibited similar performance.

An effect of risk factor status (i.e., HY = HO > MCI) was observed on the results of the word-list total learning and delayed recognition *d*' measures. Specifically, the MCI group had poorer performance on the two measures compared with both the HY ($p < 0.001$, $d = 1.04$ versus $p < 0.001$, $d = 0.99$) and HO groups ($p = 0.034$, $d = 0.52$ versus $p = 0.004$, $d = 0.82$), but the HY and HO groups did not differ between one another (Fig. 2). By contrast, on the word-list delayed free recall test, the MCI group had significantly lower performance ($p = 0.002$, $d = 0.91$) than the HY group and marginally lower performance ($p = 0.089$, $d = 0.44$) compared with the HO group. However, no difference between the HO and HY groups was observed on the same measure.

On the PM and modified CWIT tests, a combined effect of aging and risk factor status (i.e., HY > HO > MCI) was observed. Specifically, the HY group outperformed both the HO and MCI groups on PM prospective score ($p = 0.007$, $d = 0.81$ versus $p < 0.001$, $d = 1.99$), PM retrospective score ($p = 0.009$, $d = 0.78$ versus $p < 0.001$, $d = 1.67$), and modified CWIT condition 3 ($p < 0.001$, $d = 1.99$ versus $p < 0.001$, $d = 2.25$) and condition 4 ($p = 0.007$, $d = 1.35$ versus $p < 0.001$, $d = 1.65$) completion times. The HO group outperformed the MCI group on PM prospective score ($p = 0.002$, $d = 0.86$), PM retrospective score ($p = 0.023$, $d = 0.67$), and modified CWIT condition 3 ($p = 0.001$, $d = 1.16$) and condition 4 ($p = 0.001$, $d = 1.00$) completion times.

Distinguishing MCI by using robot-administered tests

A binomial logistic regression was performed to determine whether the robot-administered cognitive tests were useful in distinguishing between older adults with MCI and those without MCI (young adults were not included in this analysis). Six cognitive test variables (i.e., word-list total learning and delayed

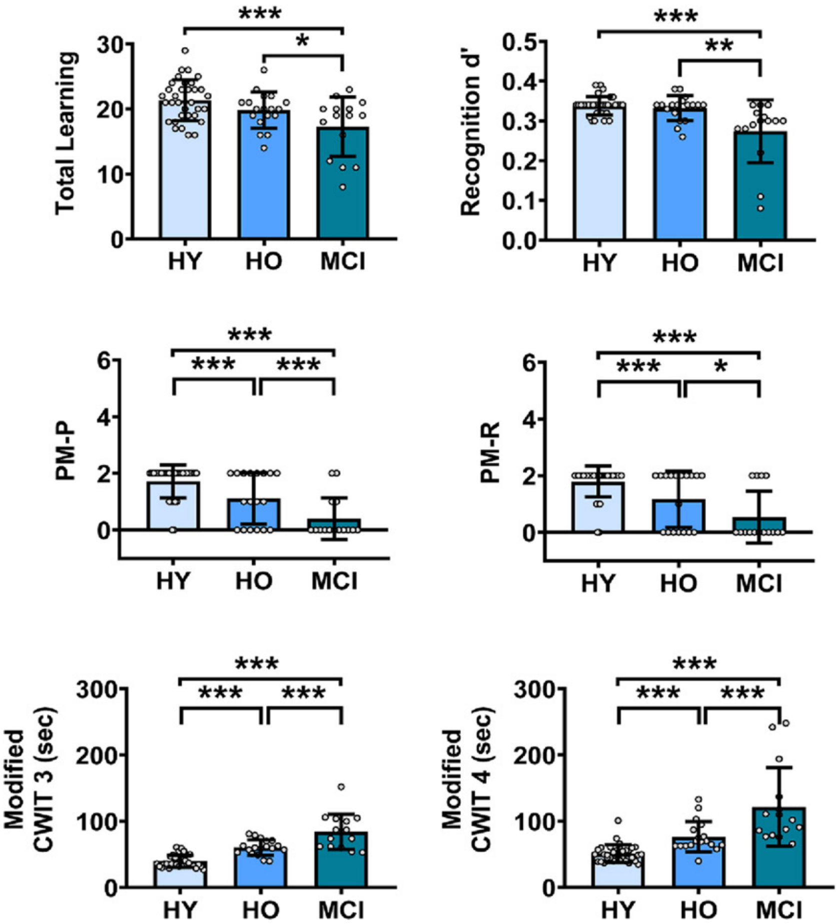

Fig. 2. Performance on the cognitive tests administered by the robot for young healthy adults (HY), older healthy adults (HO), and individuals with mild cognitive impairment (MCI). Error bars denote the standard error of the mean. PM-P, prospective memory prospective score; PM-R, prospective memory retrospective memory; Modified CWIT 3 and 4, modified Color-Word Interference Test condition 3 and condition 4. All analyses were conducted with education level as a covariate. $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.005$.

recognition *d*' measures, PM tests, and modified CWIT tests) were included in the model based on the between-group comparisons showing significant group differences between the HO and MCI groups. The results suggest that the logistic regression model was statistically significant, [χ^2 (6, $N=33$) = 17.39, $p=0.009$], explaining 56.4% (Nagelkerke R^2) of the variance in MCI classification and correctly classifying 84.8% of cases (OH group: 16/18; MCI group: 12/15), with a sensitivity of 80%, specificity of 88.9%, positive predictive value of 85.7%, and negative predictive value of 84.2%.

Performance across robot-administered tests

We calculated *z* scores for each older adult participant by using the mean and SD from the HO group in each of the six robotic measures (Fig. 2) that differentiated the HO and MCI. Individuals with scores 1 SD lower than the HO group average were identified for the cohort with MCI and older controls. In the MCI group, the percentages of participants with low performance (i.e.,<1 SD lower than the HO group average) on the robotic word-list total learning, delayed recognition *d*', PM prospective and retrospective components, and modified CWIT conditions 3 and 4 indexes were 33%, 67%, 73%, 73%, 67%, and 53%, respectively. In HO, the percentages for the corresponding measures were 16%, 22%, 33%, 39%, 17%, and 17%, respectively (Fig. 3A). We further examined differences between the HO and MCI groups through a frequency distribution of people who had low performance on one or more of the tests. The results revealed that the MCI group contained more people who were low performers compared with the HO group on the measures of word-list delayed recognition *d*' [χ^2

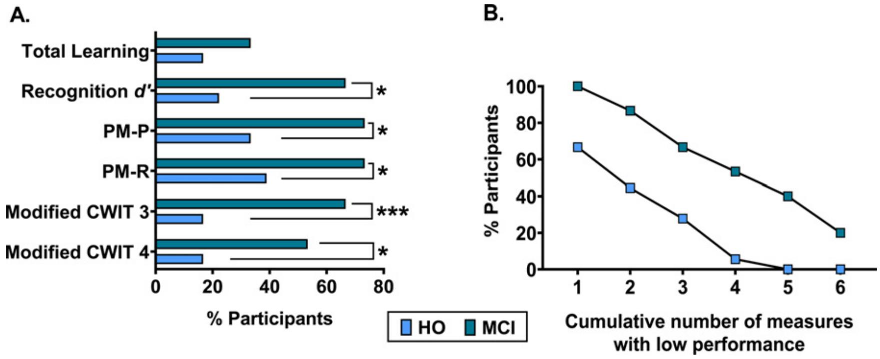

Fig. 3. A) Group summary of proportion (as a percentage) of healthy older adults (HO) and individuals with mild cognitive impairment (MCI) who had a performance score more than 1 SD below the HO average in the robotic tests. $^{*}p<0.05$; $^{***}p<0.005$. B) Cumulative proportion (as a percentage, y-axis) of participants who recorded low performance on one to six of the robot-administered cognitive tests (x-axis).

(1, $N=33)=6.62$, $p=0.01$], PM prospective (χ^2 (1, $N=33)=5.24$, $p=0.022$), PM retrospective component (χ^2 (1, $N=33)=3.92$, $p=0.048$), and modified CWIT condition 3 (χ^2 (1, $N=33)=8.57$, $p=0.003$) and condition 4 (χ^2 (1, $N=33)=4.95$, $p=0.026$).

We further calculated cumulative percentages of people with low performance across the six measures. In the MCI group, 100% (15), 87% (13), 67% (10), 53% (8), 40% (6), and 20% (3) of people had low performance on at least one test, at least two tests, at least three tests, at least four tests, at least five tests, and on all six tests, respectively. By contrast, in the HO group, 67% (12), 44% (8), 28% (5), 6% (1), 0% (0), and 0% (0) of people had low performance on at least one test, at least two tests, at least three tests, at least four tests, at least five tests, and on all six tests, respectively (Fig. 3B).

Performance on the standardized paper–pencil tests

The ANCOVA results adjusted for educational level on paper–pencil standardized cognitive tests revealed that the three groups significantly differed across the various measures used (Table 1). Specifically, the *post hoc* analyses showed that an effect of aging was observed in the MMSE score, with the HY group outperforming both the HO ($p=0.005$, $d=1.06$) and MCI ($p<0.001$, $d=1.80$) groups, whereas the two older groups had similar performance.

On the results of CVLT-II and D-KEFS CWIT measures, a combined effect of aging and risk factor status (i.e., HY > HO > MCI) was observed. Specifically, the HY group outperformed both the HO and MCI groups on measures of CVLT-II total learning ($p<0.001$, $d=1.86$ versus $p<0.001$, $d=2.53$), CVLT-II long-delayed free recall ($p=0.003$, $d=1.51$ versus $p<0.001$, $d=2.23$), standardized CWIT condition 3 ($p<0.001$, $d=2.48$ versus $p<0.001$, $d=2.69$), and standardized CWIT condition 4 ($p=0.03$, $d=2.00$ versus $p<0.001$, $d=2.28$). The HO group outperformed the MCI group on CVLT-II total learning ($p=0.015$, $d=0.97$), CVLT-II long-delayed free recall ($p=0.005$, $d=1.14$), standardized CWIT condition 3 ($p=0.001$, $d=1.34$), and standardized CWIT condition 4 ($p<0.001$, $d=1.57$).

When referencing clinically age-appropriate norms, in the HO group, only 1 (8%) participant registered a score of more than 1 SD below the age-appropriate norm on the standardized CWIT condition 3, and no participants recorded such a score on CVLT-II total learning, long-delayed free recall, and CWIT condition 3 variables. In the MCI group, 7 (47%), 8 (53%), 8(53%), 9 (60%) individuals recorded scores more than 1 SD below the age-appropriate norm for CVLT-II total learning, long-delayed free recall, and standardized CWIT conditions 3 and 4, respectively.

Predictive abilities of the performance of robotic tests to standardized cognitive tests

The results of separate hierarchical regression analyses (Table 2) revealed that the performance of the five robot-administered tests that showed differences between the HO and MCI groups were predictors of performance on the corresponding standardized neuropsychological tests, above and beyond the contribution of demographic variables and MMSE performance (Fig. 4).

Table 2
Hierarchical regression analyses of demographic, clinical, and cognitive predictors of the standardized cognitive test scores

	Model 1		Model 2		Model 3		Model 4		Model 5	
	CVLT-II total learning		CVLT long-delayed free recall		D-KEFS CWIT condition 4		D-KEFS CWIT condition 3		D-KEFS CWIT condition 4	
Predictors	β	p	β	p	β	p	β	p	β	P
Step1:										
Age	–0.59	<0.001***	–0.51	<0.001***	0.43	0.005***	0.69	<0.001***	0.43	0.005***
Education	0.13	0.36	0.12	0.41	–0.28	0.10	–0.20	0.19	–0.28	0.10
Sex (0, women; 1, men)	–0.21	0.05	–0.30	0.009**	0.14	0.28	0.08	0.50	0.14	0.28
MMSE	0.20	0.21	0.24	0.13	–0.18	–0.14	–0.14	0.93	–0.18	0.33
R^2	0.64		0.62		0.49		0.59		0.49	
Adjusted R^2	0.60		0.58		0.44		0.54		0.44	
F	17.49	<0.001***	16.05	<0.001***	9.61	<0.001***	14.12	<0.001***	9.61	<0.001***
Step2: **Robotic test**										
WLT total learning (Model 1)	0.41	0.001***	–	–	–	–	–	–	–	–
WLT delayed recognition *d*' (Model 2)	–	–	0.32	0.005**	–	–	–	–	–	–
PM prospective score (Model 3)	–	–	–	–	–0.41	0.007**	–	–	–	–
Modified CWIT condition 3 (Model 4)	–	–	–	–	–	–	0.85	<0.001***	–	–
Modified CWIT condition 4 (Model 5)	–	–	–	–	–	–	–	–	0.69	<0.001***
R^2	0.73		0.69		0.58		0.90		0.80	
Adjusted R^2	0.70		0.65		0.52		0.88		0.77	
ΔR^2	0.10		0.07		0.09		0.31		0.31	
ΔF	13.94	0.001***	8.82	0.005***	8.19	0.007**	114.39	<0.001***	57.87	<0.001***

CVLT-II, California Verbal Learning test, second version; CWIT, the Color–Word Interference Test, D-KEFS, the Delis–Kaplan Executive Function System; MMSE, Mini-Mental State Examination; PM, prospective memory; WLT, word-list test. *$p<0.05$; **$p<0.01$; ***$p<0.005$.

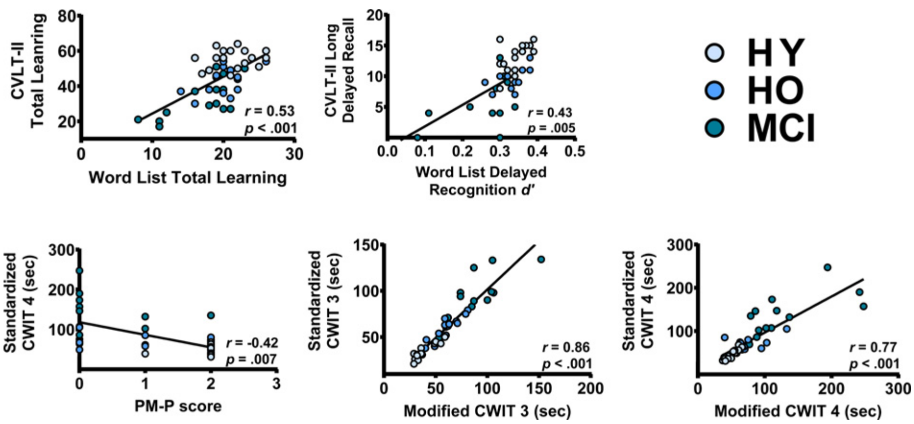

Fig. 4. Graphic presentation of statistically significant associations between robot-administered tests and standardized neuropsychological measures. Correlational coefficients indicated partial correlation after adjustments for age, sex, education, and Mini-Mental State Examination scores.

DISCUSSION

The present study demonstrated the feasibility of evaluating older adults on a range of cognitive tests by using the HRI approach. All participants were able to complete the robot-assisted tasks, and no reports of adverse reactions or discomfort were received. Our primary finding was that robot-administered cognitive tests could distinguish between older adults undergoing normal cognitive decline associated with aging from those with MCI with adequate sensitivity and specificity. Moreover, robot-administered tests demonstrated sound ability to predict the results of standardized cognitive tests conducted by a licensed psychologist, even after adjustments were made for demographic variables and MMSE performance, indicating good construct validity.

This paper presents the first evidence that through a relatively brief cognitive test battery conducted by a social robot, older adults with a risk for developing dementia could be identified. The robotic tests that revealed robust discriminability between MCI and HO were those that assessed delayed episodic memory, PM, and aspects of executive function (i.e., inhibition and shifting). These cognitive domains, especially episodic memory tasks in conjunction with tasks that rely on executive control, serve as excellent indicators for discriminating MCI from cognitive function associated with normal aging as well as predictors for future progression from MCI to AD, as identified in studies using standardized neuropsychological tests [22, 37–39]. Notably, we found that the MMSE, a common clinical screening tool employed in other HRI studies, was unable to differentiate the MCI group from the HO group. Similarly, our robotic evaluation revealed that both HO and MCI groups had a high percentage of people (67% versus 100%) with low performance on at least one of the selected tests, indicating a high possibility of falsely identifying HOs as individuals MCI when the classification was based on a single test result. The group discriminability was high when low performance was observed on multiple tests, which is consistent with the results of other studies [40, 41].

Notably, the PM measures appear to be sensitive to the effects of aging, and a high proportion of HOs performed poorly on the PM measures compared with on other robot-administered tests. Determining whether factors related to the psychometric property (e.g., skewed score distribution and limited score range) of the PM measures or the lack of test preparedness among the participants (the PM test was introduced within 5 minutes of commencing the HRI session) accounted for this result was difficult because we did not include a standardized PM test to verify the robot-assisted test results. However, the finding that HOs exhibited age-associated decline in PM is consistent with the results of several studies [42–44]. Nevertheless, these findings together support the notion that the incorporation of more comprehensive neuropsychological testing could improve the reliability of an MCI diagnosis that predicts future cognitive decline [40, 41].

Consistent with our hypothesis, the high predictive capacity of the robotic tests on episodic memory, PM, and executive function to ascertain the standardized paper–pencil test results was observed. Despite the promising results, differences between our robotic assessment session compared with a clinical evaluation setting should be highlighted. First, a robotic assessment session demands greater test participant's independence and will to perform the cognitive tests compared with that required in a clinical setting given that the clinicians are more flexible and adaptive to the participant's needs (e.g., talk slower as appropriate).

Second, a standardized neuropsychological evaluation typically takes place in a quiet and distraction-free environment. By contrast, in the robotic session, participants may have had greater difficulty ignoring the distraction of the robot's movements or ambient noise while performing the tests, which may resemble real-life environmental demands and therefore improve the ecological validity of the cognitive tests [45]. We observed that a higher proportion of HO group members recorded results of ≥ 1 SD below the age-appropriate norm in at least one of the robot-administered cognitive tests compared with that in the standardized cognitive tests. The performance gap between the two types of tests may be due to the greater number of environmental distractions in the robotic cognitive test setting than in the standardized neuropsychological evaluation setting. Alternatively, because all participants received the robot-administered tests before the paper-and-pencil tests, a practice-related improvement may have affected the standardized cognitive test results, particularly for the CWIT measures [46]; this might account for the performance gap between the two testing sessions among HO group participants.

Third, a formal neuropsychological evaluation always adapts a multimethod–multisource evaluation approach, in which testing data is incorporated with other types of data, such as information obtained from patient or informant interviews as well as behavioral observations (e.g., participant's effort, psychological status, and any language, cultural, or physical barriers that might affect the interpretation of the test results), to reach the final interpretation of a patient's status of cognitive function. By contrast, in the present study, we focused purely on objective testing results conducted by the robot, which may have biased our interpretation of the results. Despite these differences from a standardized neuropsychological evaluation, our finding indicated that introducing robotic systems as a support tool for the health care system offers new prospects for using social robots as a user interface to improve access to cognitive screening for community-dwelling older adults.

Self-administered cognitive tests using technologies such as the tablets, smartphones, or web-based platforms have become prevalent in human society, allowing greater flexibility in terms of how cognitive abilities are assessed. Moreover, in this study, given several key features of the robot, the HRI sessions could provide a superior cognitive testing experience compared with other self-administered technologies. For example, the robot was capable of engage users more reliably [15, 47] due to its physical presence as well as its ability to verbally interact and communicate with users in real time. This differed markedly from other types of human–machine interactions not involving a physically embodied device. Evidence suggests that older adults with cognitive impairment consistently preferred a robot over a computer [11] or virtual agent [48] because of the robot's physical presence and social interaction. Other studies have also reported that the physical presence of robots elicited more likeability ratings and positive mood changes among both adults and children compared with tablet-only approaches [49, 50]. In a recent study undertaken by our team [51], we also observed that compared with a tablet, participants faced with a robot were more likely to engage in prosocial behaviors (i.e., sorting the waste for recycling). Similar to some existing technologies (e.g., smartphones), robots can continuously record multiple types of behavioral data (e.g., verbal responses, gestures, and facial expressions) over a given time period, with such data available for later analysis by professionals. Moreover, this multidimensional data recording process can be automatically initiated without requiring activation by the user. The physical presence and data collection capabilities of a robot facilitate microlongitudinal evaluations (i.e., evaluations of real-time fluctuations in the behaviors and cognitive function of individuals over a given time period) that can enhance the ecological validity and reliability of cognitive assessment. Furthermore, robots can extend current technologies (e.g., computerized adaptive testing) by providing personalized test administration or cognitive stimulation programs, the type, difficulty, and variety of which can be dynamically adjusted according to user performance and test usage history in real time; this is a topic that warrants further exploration (particularly when it is combined with machine learning–based algorithms). Although our robot in the present study was primitive compared with forthcoming prototypes, our study and findings provide critical initial platforms for future development of more enhanced social robots with technologies such as advanced speech and image recognition and machine learning algorithms to serve in capacities not feasible for other human–machine interfaces.

The impact of artificial intelligence (AI)–based solutions or technologies in health care, as well as how best to incorporate those systems in health care remain open questions [6, 52]. However, a study indicated that although a lone pathologist outperformed a deep-learning system in identifying metastatic breast cancer, a combined approach with both deep learning and pathologists resulted in a significant reduction in the error rate, suggesting that cooperation between AI and human was the most powerful means of improving clinical service [53]. A study reported that some older Japanese adults are more willing to make first contact for cognitive screening with a robot than with a human neuropsychologist because of having fewer psychological barriers to completing the tests (e.g., avoiding nervousness or embarrassment) [17]. Although further investigations are warranted in the domain of cognitive assessments among older adults, we echo the findings of previous studies [6, 9, 15, 17] that with appropriate test selection and interpretation supervised by professionals (i.e., neuropsychologists or health care providers), robot-assisted cognitive evaluations can help deliver cost-effective health services to communities (e.g., working in parallel with psychometrists or technicians to administer the screening tests), improve health care resource allocation, and quickly identify individuals requiring a more comprehensive diagnostic evaluation.

To our knowledge, this is the first study that went beyond a single screening measure and employed a series of social robot–assisted tests targeting multiple cognitive domains to identify older adults with MCI. Despite its innovations, the current study had some limitations. First, our robot–initiated conversations and the testing session without placing any operational demands on participants. However, evidence has demonstrated that many older adults are uneasy about interactions with new technologies due to concerns of overwhelming challenges or their perceived inability to use the device [54]. In the present study, we did not consider variables (e.g., technological skills, attitudes toward robots) [55, 56] that may have generated confounding factors (e.g., inducing anxiety) in participants' cognitive performance during the HRI study.

Second, the present study enrolled a small sample, especially for the MCI group, in which some outliers were noted. The large intragroup performance variation in the MCI group is consistent with the results of our prior work [1, 22, 23]. Nevertheless, we conducted *post hoc* analyses on several robot-administered cognitive test variables, including the episodic memory and CWIT parameters, after removing aforementioned outliers from the MCI group data. The results revealed the same pattern as that without removing these outliers, indicating the robustness of our findings. Furthermore, in the sensitivity analysis conducted, a statistical power of 0.87, exceeding the standard requirement of 0.8, was obtained according to the estimated power curve for the final sample size of 66 and a mean Cohen's *f* effect size of 0.65. Despite that, the relatively small sample size may have limited our ability to examine the heterogeneity among people with MCI because our MCI sample comprised a mixture of patients with amnestic and nonamnestic MCI. Future studies with larger sample sizes are essential to replicate our findings and identify patterns among different subtypes of MCI.

Third, the HO participants were recruited based on the criterion of an MMSE score of ≥ 26 and without extensive neuropsychological evaluation, and therefore, we cannot exclude the possibility of misclassification (i.e., identifying individuals with MCI as HOs), particularly for those in the early stage of MCI. Fourth, the modified CWIT test was not independently administered by the robot due to the lack of a display screen in the robot. Future studies involving robots with visual display capability are warranted. Finally, although a robot was capable of recording various data, we focused on analyzing coarse-grained quantitative data (e.g., total scores, total time used) but not qualitative data (e.g., test-taking strategies, response styles) or fine-grained quantitative data (e.g., types of error responses, reaction time for memory tests) due to our relatively small sample size. Combining these data may increase clinical utility by providing more comprehensive information for differentiating various etiologies (e.g., AD, vascular-related) related to the subtle cognitive impairments observed in people with MCI [45].

In conclusion, through a social robot, we administered a short yet effective series of cognitive tests to older adults with and without MCI as well as healthy young adults. Our results demonstrated the feasibility of using this HRI approach to differentiate older adults with cognitive faculties associated with the normal aging process from those who have risk factors for later dementia development. Although more methodologically rigorous evaluations of using robots to assess cognitive function in older adults outside of the traditional clinical setting are warranted, the present study highlights the possibility of incorporating a social robot–based approach into the health care system when such approaches are undertaken with professional supervision.

ACKNOWLEDGMENTS

The authors would like to thank Yu-Wei Liu, Shin-Min Hsu, Yi-Jhen Chen, Hsin-Yi Hung, Chih-Wei Ning, Yu-Lan Cheng, Yu-Wei Lu, Yun-Shiuan Chuang, and Uan-Luen Hsieh for assistance in data collection, and Professor Kang-Ling Lay for space support.

This work was supported by the Ministry of Science and Technology, Taiwan (grant numbers 109-2629-H-002-001-MY3 and 108-2410-H-002-106-MY2 to Y.L.C.; MOST 110-2634-F-002-042 under the MOST Joint Research Center for AI Technology and All Vista Healthcare).

Authors' disclosures available online (https://www.j-alz.com/manuscript-disclosures/21-5015r2).

REFERENCES

[1] Chang YL, Bondi MW, Fennema-Notestine C, McEvoy LK, Hagler DJ, Jr., Jacobson MW, Dale AM, Alzheimer's Disease Neuroimaging Initiative (2010) Brain substrates of learning and retention in mild cognitive impairment diagnosis and progression to Alzheimer's disease. *Neuropsychologia* **48**, 1237-1247.

[2] Chang YL, Bondi MW, McEvoy LK, Fennema-Notestine C, Salmon DP, Galasko D, Hagler DJ, Jr., Dale AM, Alzheimer's Disease Neuroimaging Initiative (2011) Global clinical dementia rating of 0.5 in MCI masks variability related to level of function. *Neurology* **76**, 652-659.

[3] Petersen RC, Morris JC (2005) Mild cognitive impairment as a clinical entity and treatment target. *Arch Neurol* **62**, 1160-1163; discussion 1167.

[4] Berres M, Monsch AU, Spiegel R (2021) Using historical data to facilitate clinical prevention trials in Alzheimer disease? An analysis of longitudinal MCI (mild cognitive impairment) data sets. *Alzheimers Res Ther* **13**, 97.

[5] Yang C, Moore A, Mpofu E, Dorstyn D, Li Q, Yin C (2020) Effectiveness of combined cognitive and physical interventions to enhance functioning in older adults with mild cognitive impairment: A systematic review of randomized controlled trials. *Gerontologist* **60**, 633-642.

[6] Sabbagh MN, Boada M, Borson S, Chilukuri M, Doraiswamy PM, Dubois B, Ingram J, Iwata A, Porsteinsson AP, Possin KL, Rabinovici GD, Vellas B, Chao S, Vergallo A, Hampel H (2020) Rationale for early diagnosis of mild cognitive impairment (MCI) supported by emerging digital technologies. *J Prev Alzheimers Dis* **7**, 158-164.

[7] Chao RY, Chen TF, Chang YL (2021) Executive Function predicts the validity of subjective memory complaints in older adults beyond demographic, emotional, and clinical factors. *J Prev Alzheimers Dis* **8**, 161-168.

[8] Chang YL, Chao RY, Hsu YC, Chen TF, Tseng WI (2021) White matter network disruption and cognitive correlates underlying impaired memory awareness in mild cognitive impairment. *Neuroimage Clin* **30**, 102626.

[9] Van Patten R, Keller AV, Maye JE, Jeste DV, Depp C, Riek LD, Twamley EW (2020) Home-based cognitively assistive robots: Maximizing cognitive functioning and maintaining independence in older adults without dementia. *Clin Interv Aging* **15**, 1129-1139.

[10] Darragh M, Ahn HS, MacDonald B, Liang A, Peri K, Kerse N, Broadbent E (2017) Homecare robots to improve health and well-being in mild cognitive impairment and early stage dementia: Results from a scoping study. *J Am Med Dir Assoc* **18**, 1099.e1-1099.e4.

[11] Tapus A, Tapus C, Mataric M (2009) The role of physical embodiment of a therapist robot for individuals with cognitive impairments. In *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 103-107.

[12] Law M, Sutherland C, Ahn HS, MacDonald BA, Peri K, Johanson DL, Vajsakovic DS, Kerse N, Broadbent E (2019) Developing assistive robots for people with mild cognitive impairment and mild dementia: A qualitative study with older adults and experts in aged care. *BMJ Open* **9**, e031937.

[13] Huang TR, Liu YW, Hsu SM, Goh JOS, Chang YL, Yeh SL, Fu LC (2021) Asynchronously embedding psychological test questions into human-robot conversations for user profiling. *Int J Soc Robot* **13**, 1359-1368.

[14] Robinson H, MacDonald B, Broadbent E (2014) The role of healthcare robots for older people at home: A review. *Int J Soc Robot* **6**, 575-591.

[15] Sorrentino A, Mancioppi G, Coviello L, Cavallo F, Fiorini L (2021) Feasibility study on the role of personality, emotion, and engagement in socially assistive robotics: A cognitive assessment scenario. *Informatics* **8**, 23.

[16] Varrasi S, Di Nuovo S, Conti D, Di Nuovo A (2018) A social robot for cognitive assessment. In *HRI 2018 Companion* ACM/IEEE International Conference on Human-Robot Interaction, Chicago, IL, USA, pp. 269-270.

[17] Takaeda K, Kamimura T, Inoue T, Nishiura Y (2019) Reliability and acceptability of using a social robot to carry out cognitive tests for community-dwelling older adults. *Geriatr Gerontol Int* **19**, 552-556.

[18] Tang-Wai DF, Knopman DS, Geda YE, Edland SD, Smith GE, Ivnik RJ, Tangalos EG, Boeve BF, Petersen RC (2003) Comparison of the short test of mental status and the minimental state examination in mild cognitive impairment. *Arch Neurol* **60**, 1777-1781.

[19] Lancu I, Olmer A (2006) [The minimental state examination–an up-to-date review]. *Harefuah* **145**, 687-690, 701.

[20] Saxton J, Snitz BE, Lopez OL, Ives DG, Dunn LO, Fitzpatrick A, Carlson MC, Dekosky ST, Investigators GEMS (2009) Functional and cognitive criteria produce different rates of mild cognitive impairment and conversion to dementia. *J Neurol Neurosurg Psychiatry* **80**, 737-743.

[21] Chehrehnegar N, Nejati V, Shati M, Rashedi V, Lotfi M, Adelirad F, Foroughan M (2020) Early detection of cognitive disturbances in mild cognitive impairment: A systematic review of observational studies. *Psychogeriatrics* **20**, 212-228.

[22] Chang YL, Jacobson MW, Fennema-Notestine C, Hagler DJ, Jr., Jennings RG, Dale AM, McEvoy LK, Alzheimer's Disease Neuroimaging Initiative (2010) Level of executive function influences verbal memory in amnestic mild cognitive impairment and predicts prefrontal and posterior cingulate thickness. *Cereb Cortex* **20**, 1305-1313.

[23] Chang YL, Chen TF, Shih YC, Chiu MJ, Yan SH, Tseng WY (2015) Regional cingulum disruption, not gray matter atrophy, detects cognitive changes in amnestic mild cognitive impairment subtypes. *J Alzheimers Dis* **44**, 125-138.

[24] Chang YL, Chen TF, Tseng WY (2020) White matter pathways underlying Chinese semantic and phonological fluency in mild cognitive impairment. *Neuropsychologia* **149**, 107671.

[25] Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL (1982) A new clinical scale for the staging of dementia. *Br J Psychiatry* **140**, 566-572.

[26] Jak AJ, Bondi MW, Delano-Wood L, Wierenga C, Corey-Bloom J, Salmon DP, Delis DC (2009) Quantification of five neuropsychological approaches to defining mild cognitive impairment. *Am J Geriatr Psychiatry* **17**, 368-375.

[27] Chen HY, Chen RH (2002) *Wechsler Adult Intelligence Scale-Third Edition (WAIS-III) Manual for Taiwan*, The Chinese Behavioral Science Corporation, Taipei, Taiwan.

[28] Kaplan E, Goodglass H, Weintraub S (1983) *Boston Naming Test*, Lea & Febiger, Philadelphia.

[29] Hua MS, Chang BS, Lin KN, Yang JM, Lu SR, Chen HY (2005) *Wechsler Memory Scale (WMS), Third Edition.* The Chinese Behavioral Science Corporation, Taipei, Taiwan.

[30] Delis DC, Kaplan E, Kramer JH (2001) *Delis-Kaplan Executive Function System: D-KEFS*, Pearson, San Antonio, TX.

[31] Delis DC, Kramer JH, Kaplan E, Ober BA (2000) *California Verbal Learning Test – second edition.* Psychological Corporation, San Antonio, TX.

[32] Cohen J (1977) *Statistical power analysis for the behavioral sciences*, Academic Press, New York.

[33] Rabin LA, Pare N, Saykin AJ, Brown MJ, Wishart HA, Flashman LA, Santulli RB (2009) Differential memory test sensitivity for diagnosing amnestic mild cognitive impairment and predicting conversion to Alzheimer's disease. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn* **16**, 357-376.

[34] Gainotti G, Quaranta D, Vita MG, Marra C (2014) Neuropsychological predictors of conversion from mild cognitive impairment to Alzheimer's disease. *J Alzheimers Dis* **38**, 481-495.

[35] Ji L, Zhao Q, Gu H, Chen Y, Zhao J, Jiang X, Wu L (2021) Effect of executive function on event-based prospective memory for different forms of learning disabilities. *Front Psychol* **12**, 528883.

[36] Kinsella GJ, Pike KE, Wright BJ (2020) Who benefits from cognitive intervention in older age? The role of executive function. *Clin Neuropsychol* **34**, 826-844.

[37] Battista P, Salvatore C, Berlingeri M, Cerasa A, Castiglioni I (2020) Artificial intelligence and neuropsychological measures: The case of Alzheimer's disease. *Neurosci Biobehav Rev* **114**, 211-228.

[38] Belleville S, Fouquet C, Duchesne S, Collins DL, Hudon C (2014) Detecting early preclinical Alzheimer's disease via cognition, neuropsychiatry, and neuroimaging: Qualitative review and recommendations for testing. *J Alzheimers Dis* **42 Suppl 4**, S375-382.

[39] Delprado J, Kinsella G, Ong B, Pike K, Ames D, Storey E, Saling M, Clare L, Mullaly E, Rand E (2012) Clinical measures of prospective memory in amnestic mild cognitive impairment. *J Int Neuropsychol Soc* **18**, 295-304.

[40] Bondi MW, Smith GE (2014) Mild cognitive impairment: A concept and diagnostic entity in need of input from neuropsychology. *J Int Neuropsychol Soc* **20**, 129-134.

[41] Edmonds EC, Delano-Wood L, Clark LR, Jak AJ, Nation DA, McDonald CR, Libon DJ, Au R, Galasko D, Salmon DP, Bondi MW, Alzheimer's Disease Neuroimaging Initiative (2015) Susceptibility of the conventional criteria for mild cognitive impairment to false-positive diagnostic errors. *Alzheimers Dement* **11**, 415-424.

[42] Kliegel M, Ballhausen N, Hering A, Ihle A, Schnitzspahn KM, Zuber S (2016) Prospective memory in older adults: Where we are now and what is next. *Gerontology* **62**, 459-466.

[43] Maylor EA (1996) Age-related impairment in an event-based prospective-memory task. *Psychol Aging* **11**, 74-78.
[44] Park DC, Hertzog C, Kidder DP, Morrell RW, Mayhorn CB (1997) Effect of age on event-based and time-based prospective memory. *Psychol Aging* **12**, 314-327.
[45] Diaz-Orueta U, Blanco-Campal A, Lamar M, Libon DJ, Burke T (2020) Marrying past and present neuropsychology: Is the future of the process-based approach technology-based? *Front Psychol* **11**, 361.
[46] Davidson DJ, Zacks RT, Williams CC (2003) Stroop interference, practice, and aging. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn* **10**, 85-98.
[47] Mancioppi G, Fiorini L, Timpano Sportiello M, Cavallo F (2019) Novel technological solutions for assessment, treatment, and assistance in mild cognitive impairment. *Front Neuroinform* **13**, 58.
[48] Wu YH, Wrobel J, Cristancho-Lacroix V, Kamali L, Chetouani M, Duhaut D, Le Pevedic B, Jost C, Dupourque V, Ghrissi M, Rigaud AS (2013) Designing an assistive robot for older adults: The ROBADOM project. *IRBM* **34**, 119-123.
[49] Rosenberg-Kima R, Koren Y, Yachini M, Gordon G (2019) Human-robot-collaboration (HRC): Social robots as teaching assistants for training activities in small groups. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 522-523.
[50] Zhexenova Z, Amirova A, Abdikarimova M, Kudaibergenov K, Baimakhan N, Tleubayev B, Asselborn T, Johal W, Dillenbourg P, CohenMiller A, Sandygulova A (2020) A comparison of social robot to tablet and teacher in a new script learning context. *Front Robot AI* **7**, 99.
[51] Lo SY, Lai YY, Liu JC, Yeh SL (2021) Robots and sustainability: Robots as persuaders to promote recycling. *Int J Soc Robot*, in press.
[52] van der Cammen TJ, Albayrak A, Voute E, Molenbroek JF (2017) New horizons in design for autonomous ageing. *Age Ageing* **46**, 11-17.
[53] Wang D, Khosla A, Gargeya R, Irshad H, Beck A (2016) Deep learning for identifying metastatic breast cancer. *ArXiv*, abs/1606.05718.
[54] Turnbull ML, Jin Y, Yau AHY, Lai MSY, Cheung MYC, Kwan WYW, Watson BM (2021) mHealth in hyperconnected Hong Kong: Examining attitudes and access to mobile devices and health information among older Chinese residents. *Mhealth* **7**, 43.
[55] Chien S-e, Chu L, Lee H-H, Yang C-C, Lin F-H, Yang P, Wang T-M, Yeh S (2019) Age difference in perceived ease of use, curiosity, and implicit negative attitude toward robots. *ACM Trans Hum Robot Interact* **8**, 1-19.
[56] Chu L, Chen HW, Cheng PY, Ho P, Weng IT, Yang PL, Chien SE, Tu YC, Yang CC, Wang TM, Fung HH, Yeh SL (2019) Identifying features that enhance older adults' acceptance of robots: A mixed methods study. *Gerontology* **65**, 441-450.