Virtual Reality (2025) 29:109
https://doi.org/10.1007/s10055-025-01187-0

ORIGINAL ARTICLE

Validity and reliability of a virtual reality system as an assessment tool for cognitive impairment based on the six cognitive domains

Jeremy Wui Ming Pang[1] · Jie En Lim[2] · Eileen Yi Ling Koh[2] · Ngiap Chuan Tan[2,3]

Received: 10 September 2024 / Accepted: 20 June 2025 / Published online: 16 July 2025
© The Author(s) 2025

Abstract
The prevalence of neurocognitive disorders, including dementia is increasing in ageing populations globally. Conventional pen-and-paper neuropsychological assessments like the Montreal Cognitive Assessment (MoCA) are limited by their inability to correlate clinical cognitive scores with real-world functional performance. Efficacious, more ecologically valid and less operator-dependent assessment tools are needed to identify at-risk persons for early intervention. Technology-based tools like virtual reality (VR) are increasingly applied in healthcare, such as a novel "Cognitive Assessment using VIrtual REality" (CAVIRE-2) software which has been developed to assess the six domains of cognition automatically in 10 min. The study aimed to validate the CAVIRE-2 as a tool based on a matrix of scores and time to complete the 13 VR scenarios to discriminate persons who are cognitively healthy from those with MCI. Multi-ethnic Asian adults aged 55–84 years were recruited at a public primary care clinic in Singapore. Both CAVIRE-2 and MoCA were administered to each participant independently. 280 participants completed the study, of which 244 were found to be cognitively normal and 36 were cognitively impaired by MoCA. CAVIRE-2 showed moderate concurrent and convergent validity with MoCA and MMSE respectively. CAVIRE-2 demonstrated good test–retest reliability with Intraclass Correlation Coefficient of 0.89 (95% CI = 0.85–0.92, $p < 0.001$), and good internal consistency with Cronbach's alpha = 0.87. CAVIRE-2 displayed good discriminative ability with area under curve (AUC) of 0.88 (95% CI = 0.81–0.95, $p < 0.001$), and an optimal cut-off score of < 1850 (88.9% sensitivity, 70.5% specificity, Youden's = 0.59). CAVIRE-2 is potentially a valid and reliable assessment tool comparable to MoCA, which can distinguish cognitive status.

Keywords Virtual reality · Validity · Reliability · Screening · Cognitive impairment · Cognition

1 Introduction

Dementia is a major neurocognitive disorder marked by a pronounced cognitive decline in one or more of the six cognitive domains: perceptual motor, executive function, complex attention, social cognition, learning and memory, and language; the cognitive decline significantly interferes with the individual's daily functional abilities (Black and Grant 2014). Dementia lies on a cognitive continuum, beginning with its prodromal stage known as mild cognitive impairment (MCI). MCI also involves decline in one or more of the six cognitive domains, but unlike dementia, the individual's daily functional abilities are preserved; symptoms often progress insidiously with individuals having poor insight into their deficits. (Black and Grant 2014; Scharre 2019; Kumar et al. 2022). Approximately 10–18% of patients who fulfil the criteria of amnestic MCI eventually progress to Alzheimer's-type dementia within a year (Amieva et al. 2004; Bruscoli and Lovestone 2004; Thaipisuttikul et al. 2022). However, only 8% of individuals with expected MCI are diagnosed during primary care assessments, suggesting that cognitive decline remains grossly underdiagnosed (Liu 2024).

Jeremy Wui Ming Pang and Jie En Lim are joint first authors with equal contributions.

✉ Ngiap Chuan Tan
tan.ngiap.chuan@singhealth.com.sg

[1] Duke-NUS Medical School, Singapore, Singapore

[2] SingHealth Polyclinics, 167, Jalan Bukit Merah Connection One, Tower 5, #15-10, Singapore 150167, Singapore

[3] SingHealth Duke-NUS Family Medicine Academic Clinical Programme, Duke-NUS Medical School, Singapore, Singapore

More countries worldwide are experiencing burgeoning ageing populations. The proportion of individuals aged 65 years and above in Singapore has increased from 11.7% in 2013 to 19.1% in 2023. By the year 2030, approximately 1 in 4 individuals in Singapore will be aged 65 years and above (National Population and Talent Division Singapore 2023). The trend is significant as increasing age is a major risk factor for dementia (Alzheimer's Disease International 2023). In Singapore, the prevalence of MCI is around 12.5%, which translates to 1 in 8 individuals (Liu et al. 2021). Effective screening and early identification of individuals with MCI is thus pivotal, providing an opportunity for timely intervention to slow down the progression of MCI, and preventing its further deterioration to dementia (Ng et al. 2021; Yu et al. 2021; Granland et al. 2022).

Currently, objective testing of MCI relies predominantly on paper-and-pencil neuropsychological assessments. The limitations include the inability to correlate cognitive scores obtained in a controlled experimental environment with real-world functional behaviour, a term coined as ecological validity (Sbordone and Long 1996; Chaytor and Schmitter-Edgecombe 2003; Spooner and Pachana 2006; Tarnanas et al. 2013). The construct of ecological validity can be conceptually dissected into two approaches: veridicality and verisimilitude. Veridicality pertains to the extent to which neuropsychological assessments predict real-world outcomes. In contrast, verisimilitude is defined as the degree to which cognitive demands presented by these tests mirror those encountered in the naturalistic environment (Sbordone and Long 1996).

Traditional neuropsychological assessments like the Montreal Cognitive Assessment (MoCA) adopt a veridicality-based methodology to cognitive function testing, which can be susceptible to relatively weaker correlations with real world outcome measures (Sbordone and Long 1996). This approach is also constrained by the ambiguity of measurement constructs, limited perspectives of a person's daily behaviour and oversight of compensatory mechanisms, leading to underestimation of an individual's real-world performance (Chaytor and Schmitter-Edgecombe 2003). Conversely, neuropsychological assessments grounded in the verisimilitude paradigm have demonstrated superiority compared to a battery of other assessments, with results more consistently aligned with real-world functional outcome measures (Parsons 2015).

Virtual reality (VR) based technological applications are gaining traction in the field of cognitive screening (Neguț et al. 2016; Kourtesis 2020; Lim et al. 2020; Yan et al. 2021; Christova et al. 2022). It harbours the ability to faithfully reproduce the naturalistic environment and consequently improving the ecological validity (Bohil et al. 2011; Climent-Martínez et al. 2014). The virtual environment (VE) re-creations facilitate a verisimilitude approach in cognitive testing, whereby participants can carry out tasks mimicking both basic and instrumental activities of daily living (BADL and IADL) to accurately assess their cognitive capability in real time (Atkins 2015; Boz et al. 2020). Such interactive functions within the VE allow tailored experiences for individuals, circumvent testing fatigue, and reduce anxiety resulting from the traditional face-to-face paper-based assessments (Valladares-Rodríguez et al. 2016; Jang et al. 2022). Automation of the screening process standardizes the evaluation procedures, reduces administration variability, and optimizes time and manpower resource utilisation.

In recent years, several iterations of VE have been implemented in the screening of mild cognitive impairment. VE modelled after supermarkets (Cipresso et al. 2014; Yan et al. 2021), food kiosks (Tarnanas et al. 2013), daily activities (Tarnanas et al. 2013; Atkins 2015), and mazes (da Costa et al. 2022) have shown promising screening potential of neurocognitive decline. However, these VR tools have yet to assess cognition comprehensively, varying between two to a maximum of five out of total of six cognitive domains (Liu et al. 2023). Furthermore, the studies were not conducted in a primary care setting (Liu et al. 2023) where the target at-risk population resides. This reduces the generalizability of the results, and the scalability potential of these VR tools remains to be determined.

A fully immersive and automated VR assessment system, Cognitive Assessment by VIrtual REality (CAVIRE), has been developed as a tool to assess the six domains of cognition (Wong et al. 2021; Tan et al. 2022). An earlier study assessed the performance of participants using the first version of the VR software (CAVIRE-1). Their performance is based on a matrix of scores and completion time as the participants completed the tasks across 13 scenarios in its VE. They were separated a-priori into two groups, "cognitively healthy" and "cognitively impaired" based on their MoCA cut-off scores of 26 before using the CAVIRE-1. The results revealed the differential performances between the "cognitively healthy" and "cognitively impaired" groups of participants across six cognition domains upon the completion of the CAVIRE-1 assessment (Tan et al. 2024).

An enhanced second iteration of the VR software (CAVIRE-2) has been developed and is postulated to be capable of assessing all six cognitive domains of participants based on its performance matrix. CAVIRE-2 was developed based on the feedback of the participants who used CAVIRE-1 in the previous study. Enhancements were made to improve the user experience of CAVIRE, for example optimisations of the voice and text instructions, visualisation and interaction with virtual objects, and scene transitions. Hence, the primary aim of this study is to evaluate the validity (concurrent and convergent) and reliability

of CAVIRE-2 as an independent assessment tool to differentiate the cognition of older adults.

In addition, the previous study involving cognitively healthy participants demonstrated that individuals in younger age groups achieved higher CAVIRE-1 scores compared to those in older age groups (Tan et al. 2022). Extant literature also elucidates a negative correlation between neuropsychological assessment scores and age, along with a positive correlation between neuropsychological assessment scores and education attainment exceeding 12 years (Jefferson et al. 2011; Freitas et al. 2012; Malek-Ahmadi et al. 2015). Therefore, the secondary aim of this study is to explore the effects of age and education on CAVIRE-2's ability to distinguish the cognitive status of older individuals.

2 Materials and methods

2.1 Instrumentation

The CAVIRE-2 software is a fully immersive VR system, allowing users to perform virtual tasks in three dimensional (3D) virtual environments (VE). The content is jointly designed by an interdisciplinary team consisting of experienced family physicians cum researchers with postgraduate qualifications, and special interests in the field of geriatrics (Lim et al. 2020). They partnered an external multimedia solutions company in the system design and development. The latter is responsible for the technical development, covering user interface design, game mechanics, and virtual reality game interface prototyping.

CAVIRE-2 comprises 14 discrete scenes, inclusive of one starting tutorial session and 13 virtual scenes simulating both basic and instrumental activities of daily living (BADL and IADL) in local residential and community settings (Fig. 1A, B). The virtual residential blocks and shophouses were modelled with a high degree of realism. Such design aims to bridge the gap between an unfamiliar virtual game environment and participants' lived experiences, facilitating a more immersive and realistic emulation. Automated visual and audio instructions are embedded to help guide participants through individual segments, which they are required to complete through specific hand gestures and speech. An automated scoring system is embedded in CAVIRE (Lim et al. 2020). The scores are automatically calculated for each of the 13 scenes and for the entire assessment. The matrix of scores is structured by: (a) the proportion of correctly completed tasks for each scene; (b) the number of attempts taken to complete the task for each scene; and (c) the completion of tasks within the stipulated time for each scene.

A summary of the 13 virtual scenes is as follows:

1. Brush and rinse teeth.
2. Prepare peanut butter bread for breakfast.
3. Identify photos of important persons.
4. Watch television and listen to the weather forecast on the radio simultaneously.
5. Name and remember fruits in a shopping list.
6. Choose appropriate clothes to wear for grocery shopping.
7. Leave the house and take the umbrella.
8. Take the lift to level 1.
9. Cross the road safely.
10. Choose the correct stall, based on the shopping list.
11. Select the correct fruits from memory.
12. Calculate and pay the correct sum of money.
13. Select the appropriate emotion, based on the given scenarios.

The performances of participants using the CAVIRE-2 system collectively assess their six cognitive domains, as stipulated in the 5th edition of the Diagnostic and Statistical

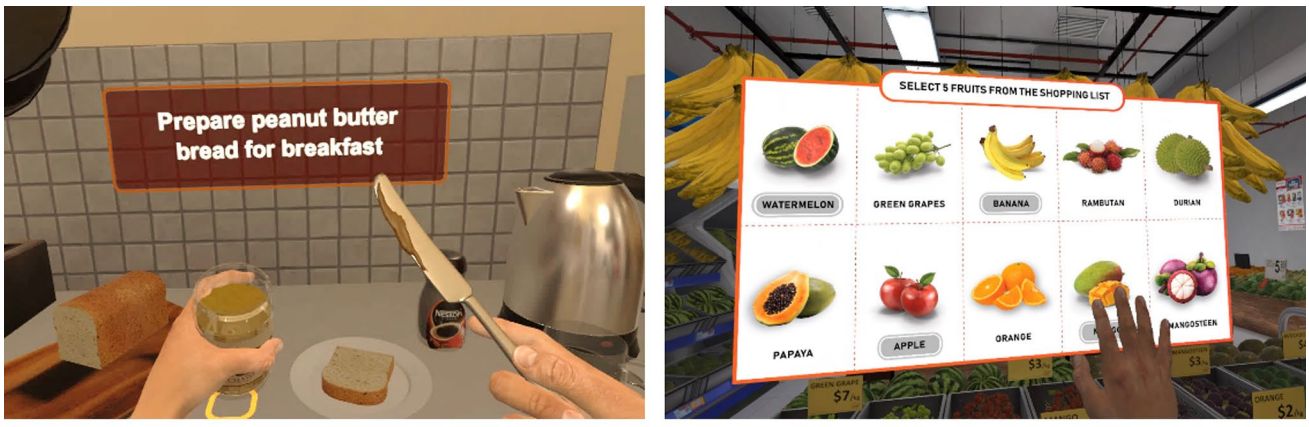


Fig. 1 **A** Scene 2—Prepare peanut butter bread for breakfast. **B** Scene 11—Select the correct fruits from memory

Manual of Mental disorders (DSM-5) (Black and Grant 2014). The respective cognitive domain evaluation in each virtual scenario is detailed in Fig. 2.

To maintain a balanced framework evaluation, individual cognitive domains are assessed separately over 4 different scenes. For example, complex attention is assessed in Scenes 4, 9, 10, and 12. For Scene 4, divided attention is assessed, where the user needs to watch television and listen to the radio at the same time. For Scene 9, sustained attention is assessed, where the user needs to focus on the pedestrian crossing traffic light and press the button to start walking when the light turns green. For Scenes 10 and 12, selective attention is assessed, where the user needs to choose the correct store and pay the correct sum of money in the presence of distractors.

Each scene may evaluate a single cognitive domain or a combination of cognitive domains simultaneously. For example, in Scene 2 (Fig. 1A), the participant is given the task of preparing peanut butter bread for breakfast. The completion of this task requires the participant to plan out the steps in a logical sequence and pick up the correct items using the appropriate hand. Hence, executive function and perceptual motor function are assessed in this task. In Scene 11 (Fig. 1B), the participant is required to select the 5 correct fruits, based on a shopping list that was displayed in several scenes before that, i.e., Scene 5. Therefore, the cognitive domain of learning and memory is assessed during this task.

CAVIRE-2 is built upon the Unity cross-platform game engine. Participants were inducted to use the HTC VIVE Pro Head Mounted Display (HMD), a high-performance virtual reality headset capable of projecting a 110-degree field of view at a 90 Hz refresh rate. It enables presentation of the VE to the participants, concurrently computing their gaze direction and head movements. Several additional peripherals were included such as a Leap Motion Controller, which allows for 3D spatial recognition of hand positions and intricate fine motor dexterity. Furthermore, the utilisation of Lighthouse, an infra-red based inside-out positional tracking system helps track the orientation of the HMD. Voice recognition was enabled via the integrated Application Programming Interface (API) and audio input was captured using a Rode VideoMic Pro.

2.2 Study site and population

This study was conducted at Eunos Polyclinic, a public primary care clinic located in the eastern region of Singapore. It serves the primary healthcare needs of 118,565 residents in the estate (Department of Statistics Singapore 2021). The study population included patients seeking medical consultation; accompanying family members; volunteers; or external visitors of the clinic.

The inclusion criteria were community-dwelling multiethnic Asian adults (1) aged between 55 and 84 years old; (2) able to speak English or Mandarin; and (3) willing to complete the study procedures. Those excluded from the study were adults with (1) pre-existing diagnosis of dementia, or on current dementia medication as documented in the electronic medical records; (2) neurological deficits affecting movement, speech, or vision; (3) history of medical conditions (motion sickness, epilepsy) that could be triggered through immersion in virtual reality; and (4) any disability rendering the individual incapable of providing informed consent.

2.3 Study design

A 2-visit study was conducted between August 2023 and January 2024. The study was designed such that the study team was blinded to the participants' cognitive status, i.e., whether MCI or cognitively healthy, in order to reduce potential bias while administering the CAVIRE-2 and MoCA assessments. Written informed consent was obtained from each participant after screening for their fulfilment of their eligibility criteria. A research assistant inducted the participants to the use of the headset, checked for their comfort and potential adverse reaction before they proceeded with the VR assessment using CAVIRE-2 system. Their performance indices were electronically tabulated automatically by the system.

Fig. 2 The six cognitive domains assessed in each scene of CAVIRE-2

Cognitive Domains	Scene												
	1	2	3	4	5	6	7	8	9	10	11	12	13
Perceptual motor	■	■					■	■					
Executive function		■					■		■			■	
Complex attention				■					■	■		■	
Social cognition						■			■	■			■
Learning and memory			■	■	■						■		
Language	■		■		■							■	

Next, a second research assistant, blinded to the CAVIRE-2 assessment, administered the MoCA assessment to the participant to reduce potential bias. Participants with ≤ 6 years of education received a 1-point adjustment to their final MoCA scores (Khaw et al. 2021). A MoCA score of 26 and above is regarded as "cognitively healthy" based on current local clinical practice guidelines. Any incidental findings, including (1) MoCA score of < 26 and/or (2) low CAVIRE-2 scores were shared with participants according to their informed consent. They were advised to seek medical consultation to further evaluate their cognitive status.

A subset of the study population comprising 100 participants were invited to schedule a second appointment in the next 2 to 4 weeks to repeat the identical CAVIRE-2 assessment. The performance scores from both CAVIRE-2 assessments were computed to determine its reliability.

2.4 Sample size calculation

Participants were recruited from the general population where the study team was unaware of their cognitive status, whether they were cognitively healthy or cognitively impaired.

For the study aim on assessing reliability, a minimum of 100 participants was selected, in accordance with the COSMIN Study Design Checklist (Terwee et al. 2012), which recommends this threshold for adequate estimation of internal consistency (e.g., Cronbach's alpha) and test–retest reliability, ensuring acceptable levels of precision and standard error in the reliability coefficients.

Based on current literature, the prevalence of MCI in Singapore is estimated at 12.5% (Liu et al. 2021). While CAVIRE-2 is designed as a screening rather than diagnostic tool, screening instruments must still demonstrate the ability to discriminate between individuals with and without mild cognitive impairment (MCI). The required sample size to assess the ability of CAVIRE-2 to distinguish between individuals with and without mild cognitive impairment (MCI) was estimated using the area under the receiver operating characteristic curve (AUROC). Assumptions based on existing literature were used, including an anticipated AUROC of ≥ 0.80, a significance level (α) of 0.05, a statistical power (1–β) of 80%, and a ratio of 1:7 (35 MCI cases, 245 cognitively healthy controls) based on 12.5% prevalence of MCI in Singapore. These parameters yielded a required total sample size of 280 participants, sufficient to detect an AUROC significantly greater than 0.5 (Obuchowski and McClish 1997).

2.5 Data management and monitoring

Data from the study were transcribed into the Research Electronic Data Capture (REDCap) secure research database, and subsequently audited by a data management officer in the institution research department for potential errors. Participants' data were anonymized and linked to unique study identification numbers, void of personal identifiers. Participants who were unable to complete CAVIRE-2 due to procedural side effects were similarly encoded in the final database but categorised as study drop-outs. The de-identified data set was then handed over to the data analytics team to compute the results.

2.6 Statistical analysis

All data analysis was conducted using IBM SPSS 29.0. Descriptive statistics were presented in frequencies and percentages for categorical variables, while mean and standard deviation were utilized for continuous variables. Table 1 assessed the difference in baseline characteristics between cognitively healthy and cognitively impaired individuals using Chi-square test and independent t-tests. Normality of the Total CAVIRE-2 scores, individual domain scores, and time spent between groups were assessed using the Shipiro-Wilk's test and median and interquartile ranges were presented for skewed scores. For difference in Total CAVIRE-2 scores, individual domain scores, and time spent to complete the VR assessment between cognitively impaired individuals and their cognitively healthy counterparts, the non-parametric Mann–Whitney U test was performed. Statistical significance was predetermined at $p < 0.05$ and 95% confidence intervals (CI) quantified the degree of agreement.

The presence of floor and ceiling effects was determined by > 15% of participants achieving the lowest, or highest possible score, respectively. Floor and ceiling effects were used to assess the distribution of responses at extreme ends of the scoring. These effects can indicate whether the test is too difficult or too easy, which may limit the ability to detect differences among participants.

Concurrent validity is measuring the degree to which the new tool correlates and performs with the gold-standard tool when administered simultaneously. CAVIRE is meant to be a screening tool that differentiates individuals with MCI from those with normal cognition, just like MoCA. Hence, in this study, MoCA was used as a gold-standard tool to correlate with CAVIRE. Convergent validity measures the correlation with other measures assessing the same construct that theoretically should be related. MMSE evaluates cognitive domains on memory, attention, language, and perceptual motor function, which are similar to the domains evaluated in CAVIRE. Hence, we have used MMSE to correlate with

Table 1 Demographic characteristics of the study population

	Total	NC	MCI	*p*-value[#]
	n = 280*	n = 244	n = 36	
Gender				0.634
Male, n (%)	127 (45.4)	112 (45.9)	15 (41.7)	
Female, n (%)	153 (54.6)	132 (54.1)	21 (58.3)	
Age, mean (s.d.)	67.3 (6.6)	66.5 (0.4)	72.9 (1.3)	< 0.001
Ethnicity				0.377
Chinese, n (%)	249 (88.9)	218 (89.3)	31 (86.1)	
Malay, n (%)	13 (4.6)	12 (4.9)	1 (2.8)	
Indian, n (%)	15 (5.4)	12 (4.9)	3 (8.3)	
Others, n (%)	3 (1.1)	2 (0.8)	1 (2.8)	
Marital status				0.785
Married, n (%)	211 (75.4)	182 (74.6)	29 (80.6)	
Single, n (%)	32 (11.4)	28 (11.5)	4 (11.1)	
Divorced/Separated, n (%)	21 (7.5)	20 (8.2)	1 (2.8)	
Widowed, n (%)	16 (5.7)	14 (5.7)	2 (5.6)	
Occupancy			0.237	
Living alone, n (%)	37 (13.2)	30 (12.3)	7 (19.4)	
Living with family, n (%)	243 (86.8)	214 (87.7)	29 (80.6)	
Housing				0.029
Public Housing, n (%)	192 (68.6)	173 (70.9)	19 (52.8)	
Private housing, n (%)	88 (31.4)	71 (29.1)	17 (47.2)	
Number of years of education				< 0.001
≤ 10 years, n (%)	139 (49.6)	110 (45.1)	29 (80.6)	
> 10 years, n (%)	141 (50.4)	134 (54.9)	7 (19.4)	
Employment status				0.143
Working, n (%)	117 (41.8)	106 (43.4)	11 (30.6)	
Not working/Retired, n (%)	163 (58.2)	138 (56.6)	25 (69.4)	
Barthel's Index (BADL), mean (s.d.) (max = 100)	98.9 (3.8)	99.2 (0.2)	96.8 (1.1)	< 0.001
Lawton's Scale (IADL), mean (s.d.) (max = 23)	22.6 (1.9)	22.9 (0.1)	20.7 (0.7)	< 0.001
MMSE, mean (s.d.) (max = 30)	28.1 (2.1)	28.6 (0.1)	25.0 (0.5)	< 0.001
MoCA, mean (s.d.) (max = 30)	27.4 (2.6)	28.1 (0.1)	22.1 (0.6)	< 0.001

NC, Normal (Cognitively Healthy) Control; MCI, Mild Cognitive Impairment; n, number of individuals; BADL, Basic Activities of Daily Living; IADL, Instrumental Activities of Daily Living; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; CAVIRE, Cognitive Assessment using VIrtual Reality; s.d., Standard deviation

*Total number of participants excluding one drop-out

[#]*p*-values are for comparisons between the two groups, NC and MCI

CAVIRE. Spearman's correlation was performed between participants' CAVIRE-2 scores and the respective instrument scores (MoCA and MMSE). A correlation coefficient of < 0.25 indicates small concurrence, 0.25 to 0.50 indicates moderate concurrence, 0.5 to 0.75 indicates good concurrence and over 0.75 indicates excellent concurrence (Portney and Watkins 2007).

Internal consistency and test–retest reliability are the two measures used to determine the reliability of CAVIRE-2. Internal consistency was evaluated using Cronbach's Alpha, which describes the extent to which all items in CAVIRE-2 measure the construct of cognitive function and is linked to the inter-relatedness of the items within the test (Tavakol and Dennick 2011). A Cronbach's alpha value of ≥ 0.7 is generally considered acceptable, with a higher alpha suggesting increased inter-relatedness between items (Tavakol and Dennick 2011). The temporal stability, degree of correlation and agreement between CAVIRE-2 scores were assessed using Intraclass Correlation Coefficient (ICC) to ascertain test–retest reliability. ICC scores within the range of 0–0.5, > 0.5–0.75, > 0.75–0.9 and > 0.9 were rated as poor, moderate, good, and excellent, respectively (Koo and Li 2016).

Receiver operating characteristic (ROC) curve analysis was selected as an indicator of responsiveness of the CAVIRE-2 system. The area under the receiver operating characteristic curve (AUROC) indicates the degree of accuracy of CAVIRE-2 in discriminating between cognitively healthy and cognitively impaired individuals. For CAVIRE-2 to be clinically relevant, an AUROC of > 0.8, and sensitivity of > 80% would be required to demonstrate that CAVIRE-2 possesses sufficient discriminative power in distinguishing between normal (cognitively healthy) adults and those with mild cognitive impairment (Fan et al. 2006; Nahm 2022). The optimal cut-off of for the CAVIRE-2 score was determined using Youden's index (*J*), which identifies a suitable test cut-off where the difference between true positives (sensitivity) and false positives (1-specificity) rates are maximised (Youden 1950).

To evaluate the secondary aim of this study, ROC subanalyses were performed to explore the effects of age and education on the discriminative ability of CAVIRE-2 to differentiate between cognitively healthy and cognitively impaired individuals.

3 Results

3.1 Descriptive results

A total of 281 community-dwelling participants were recruited for this study. One participant was excluded from

Table 2 Concurrent and convergent validity of CAVIRE-2

	Spearman's correlation, r_s
Concurrent validity	
MOCA	0.482
Convergent validity	
MMSE	0.427

MMSE: Mini-Mental State Examination; MoCA: Montreal Cognitive Assessment

the final analysis as she could not complete the VR testing, citing nausea during VE immersion. The drop-out rate was 1 out of 281 participants, i.e., 0.36%. The remaining 280 participants completed the CAVIRE-2 testing and study procedures.

The mean age of the 280 participants was 67.3 years (Table 1). Those in the MCI group were older, mean age of 72.9 years versus 66.5 years in the cognitively healthy counterparts ($p<0.001$) and were more likely to reside in private housing (47.2%) compared to 29.1% among healthy participants ($p=0.029$).

More participants without MCI had > 10 years of education (A-levels/diploma, tertiary): 54.9% compared to 19.4% of those with MCI. The proportion of MCI amongst all participants with ≤ 10 years of education (no formal education, primary school, secondary school) is also similarly higher at 20.9%, and decreases to just 5.0% in those with > 10 years of education.

For neuropsychological assessment, the mean MMSE (28.6 vs. 25.0, $p<0.001$) and MoCA (28.1 vs. 22.1, $p<0.001$) scores significantly differed between the cognitively healthy and impaired groups (Table 1). The lower Barthel's Index scores (measuring functionality in basic activities of daily living), along with the lower Lawton's Scale scores (measuring subjective ability to perform instrumental activities of daily living), reflected poorer functional ability amongst participants with MCI ($p<0.001$).

3.2 Floor and ceiling effects

No floor or ceiling effect was found for both first and second CAVIRE-2 assessment. None of the participants scored the lowest (0) or highest possible score (2400).

3.3 Concurrent and convergent validity

Based on Table 2, concurrent validity was supported by a moderate correlation between CAVIRE-2 score and MoCA score ($r_s=0.482$). Comparing with MMSE, CAVIRE-2 has moderate convergent validity with a correlation of 0.427.

Table 3 Test–retest reliability and internal consistency of CAVIRE-2 scores (n = 100)

	ICC (95% CI)	Cronbach's alpha	
		First assessment	Second assessment
CAVIRE-2 total score	0.892 (0.846–0.924)	0.869	0.878
Perceptual motor	0.675 (0.535–0.772)		
Executive function	0.749 (0.642–0.825)		
Complex attention	0.807 (0.725–0.865)		
Social cognition	0.752 (0.646–0.826)		
Learning and memory	0.716 (0.595–0.801)		
Language	0.796 (0.709–0.857)		

ICC: Intraclass Correlation Coefficient

3.4 Reliability

Test–retest reliability and internal consistency were evaluated with the Intraclass Correlation coefficient (ICC), and Cronbach's Alpha respectively (Table 3). CAVIRE-2 assessment shows good test–retest reliability with a total score ICC estimate of 0.89 (95% CI = 0.85–0.92, $p<0.001$). ICC estimates of the six cognitive domains ranged from moderate to good test–retest reliability. The first and second assessments using CAVIRE-2 reported a Cronbach's Alpha of 0.869 and 0.878, respectively, highlighting a close interrelatedness amongst CAVIRE-2 testing segments in assessing cognitive function.

3.5 CAVIRE-2 scores and completion time

A lower CAVIRE-2 total score was noted amongst participants with MCI during their first assessment, compared to those in the cognitively healthy (NC) group (MCI: Median 1538, IQR 1331–1763 vs. NC: Median 1925, IQR 1775–2050, $p<0.001$) (Table 4 and Fig. 3A). The same trend is evident in each of the six cognitive domains: perceptual motor, executive function, complex attention, social cognition, learning and memory, language (all $p<0.001$). The participants with MCI also took correspondingly longer time to complete the CAVIRE-2 assessment overall (MCI: Median 9.7 min, IQR 8.7–10.8 min vs. NC: Median 8.1 min, IQR 7.5–8.8 min, $p<0.001$) (Fig. 3B). Participants with MCI also took significantly longer time to complete the MoCA assessment (MoCA: Median 16.3 min, IQR 13.0–18.7 min vs CAVIRE-2: Median 9.7 min, IQR 8.7–10.8 min).

3.6 Responsiveness

Figure 4A shows an area under the receiver operating characteristic curve (AUROC) of 0.88 (95% CI = 0.81–0.95, $p<0.001$) (Table 5), indicating a high level of accuracy in

Table 4 CAVIRE-2 six cognitive domain scores and completion time for the first assessment

	Total	NC	MCI	*p*-value
	n = 280	n = 244	n = 36	
CAVIRE-2 total score	1888 (1725–2025)	1925 (1775–2050)	1538 (1331–1763)	< 0.001
Perceptual motor	375 (350–400)	375 (350–400)	350 (281–375)	< 0.001
Executive function	350 (300–375)	350 (325–375)	275 (225–300)	< 0.001
Complex attention	213 (175–250)	225 (200–275)	125 (100–175)	< 0.001
Social cognition	275 (231–300)	275 (250–300)	250 (200–250)	< 0.001
Learning and memory	325 (300–350)	325 (300–350)	250 (206–300)	< 0.001
Language	375 (350–400)	375 (350–400)	300 (275–344)	< 0.001
Completion time (s)	491 (450–542)	484 (447–525)	580 (521–645)	< 0.001

Values are summarized as median (interquartile range (IQR)). *p*-values are based on the Mann–Whitney U test

NC, Normal (Cognitively Healthy) Control; MCI, Mild Cognitive Impairment

classifying between cognitively healthy controls and participants with mild cognitive impairment. A cut-off that maximises Youden's index and sensitivity reveals an optimal cut-off score at 1850 (Fig. 4D). Using this cut-off score, CAVIRE-2 testing shows sensitivity of 88.9%, specificity of 70.5%, and a Youden's index of 0.59.

ROC sub-analysis (Fig. 4B) of participants with ≤ 10 years and > 10 years of education shows an AUROC of 0.86 (95% CI = 0.77–0.95, $p < 0.001$) and 0.87 (95% CI = 0.76–0.99, $p = 0.003$) respectively (Table 5). The optimal cut-off score for participants with ≤ 10 years of education was < 1825 (sensitivity = 89.7%, specificity = 58.2%), while those with > 10 years of education had a cut-off of < 1900 (sensitivity = 85.7%, specificity = 71.6%).

Another ROC sub-analysis by age shows a difference in discriminative potential (Fig. 4C). For participants aged 55–69 years old, the AUROC was 0.77 (95% CI 0.61–0.93, $p = 0.024$), at the optimal cut-off score of < 1900, sensitivity = 81.8% and specificity = 60.8%. Conversely, for participants aged 70–84 years old, the AUROC was higher at 0.91 (95% CI 0.84–0.98, $p < 0.001$), and at a lower cut-off of < 1700, sensitivity = 88.0% and specificity = 80.9% (Table 5).

4 Discussion

CAVIRE-2 is the second iteration of a fully immersive, verisimilitude-based, automated VR platform developed to assess six domains of cognition. The study results show that CAVIRE-2 is comparable to the conventional MoCA assessment in its discriminative ability for cognitive assessment, with similar sensitivity (88.6%) and greater specificity (70.4%).

The advocacy for immersive VR cognitive assessments hinges on its proposed ecological validity– the ability to

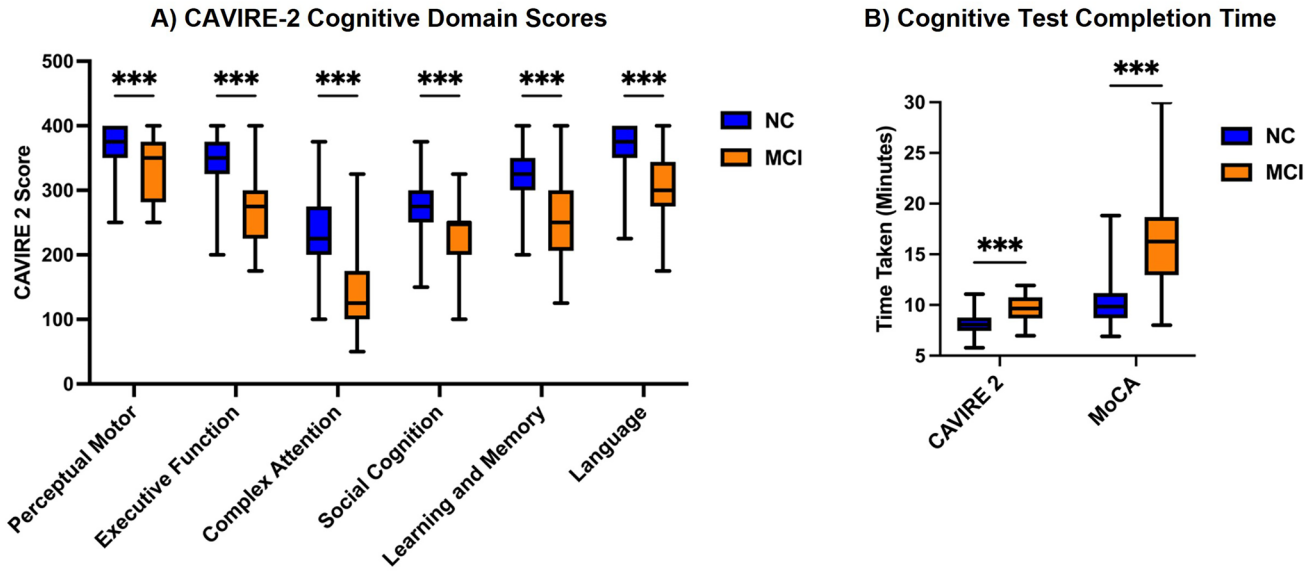


Fig. 3 **A** Box plot of CAVIRE-2 six cognitive domain scores, with comparison between NC and MCI participants. **B** Box plot of the completion time for the two cognitive tests: CAVIRE-2 assessment and MoCA assessment, with comparison between NC and MCI participants. Abbreviations: NC, Normal (Cognitively Healthy) Control; MCI, Mild Cognitive Impairment. ***Represents statistical significance of < 0.001

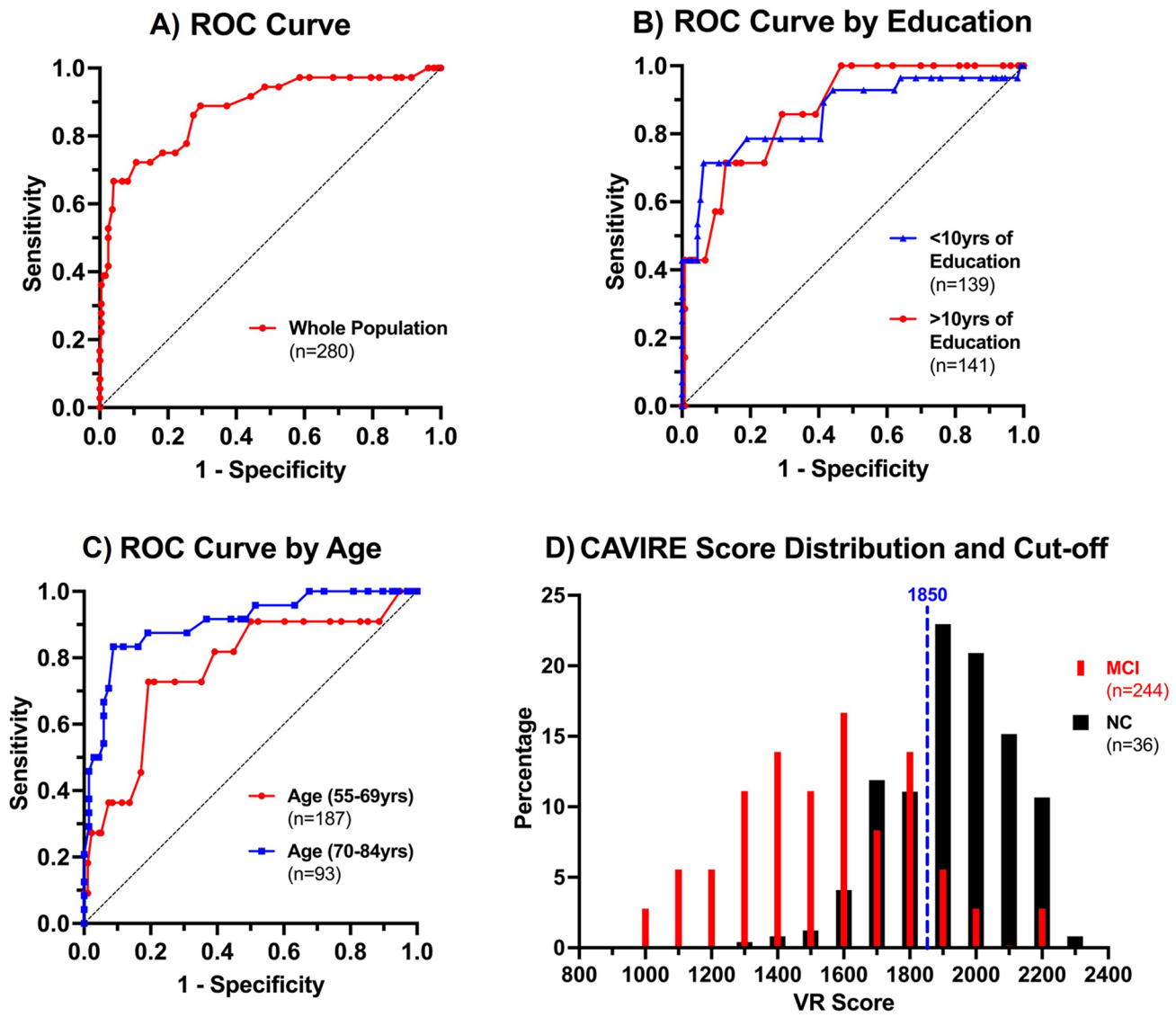


Fig. 4 **A** CAVIRE-2 receiver operating characteristic (ROC) curve for mild cognitive impairment (MCI) assessed against the Montreal Cognitive Assessment (MoCA), whole population (n = 280). **B** CAVIRE-2 ROC curve for MCI assessed against MoCA, by education. **C** CAVIRE-2 ROC curve for MCI assessed against MoCA, by age. **D** CAVIRE-2 score distribution between normal controls vs. mild cognitive impairment with selected optimal cut-off, whole population (n = 280)

Table 5 ROC performance of CAVIRE-2 in the study population stratified by education levels and age (n = 280)

	NC	MC	AUC (95% CI)	*p*-value	Optimal cut-off	Sensitivity	Specificity	Youden's Index
	n							
All	244	36	0.879 (0.811–0.947)	<0.001[#]	<1850	0.889	0.705	0.59
Education ≤ 10 years	110	29	0.861 (0.772–0.950)	<0.001[#]	<1825	0.897	0.582	0.48
Education > 10 years	134	7	0.872 (0.758–0.985)	0.003[#]	<1900	0.857	0.716	0.57
Age 55–69 years	176	11	0.768 (0.611–0.925)	0.024[#]	<1900	0.818	0.608	0.43
Age 70–84 years	68	25	0.909 (0.838–0.979)	<0.001[#]	<1700	0.880	0.809	0.69

NC, Normal (Cognitively Healthy) Control; MCI, Mild Cognitive Impairment; AUC, Area Under Curve; 95% CI, 95% Confidence Interval

[#]Null Hypothesis: True area = 0.5

correlate quantitative scores with real-world functional capacity. The participants with MCI achieved significantly lower scores ($p<0.001$) for both the Barthel's Index (BADL) and Lawton's Scale (IADL), as compared to those who were cognitively healthy (NC). Early identification of cognitive impairment of adults using tools such as CAVIRE-2 allows healthcare professionals to carry out further evaluation of their challenges coping with their activities of daily living, recommend measures to mitigate their cognition decline and to seek family and other community resources to maintain their quality of life.

The study has its strengths. The study has validated CAVIRE-2 as an alternative, reliable and ecological valid tool to assess and identify people with cognition impairment in primary care. The VR system is portable, operator independent and can be implemented in various settings within the local community, extending its outreach to people in isolation or in non-clinical facilities. Most participants reported no prior experience with immersive VR technology, yet their completion rates were high, reflecting usability and acceptability of newer technology by the older adults.

Currently, CAVIRE-2 is known to be the only VR cognitive assessment that evaluates all six cognitive domains through different scenes involving activities of daily living. Other VR technologies are limited to assessing between two to five out of the six cognitive domains (Liu et al. 2023). Moreover, in contrast to other existing VR cognitive assessments, CAVIRE-2 omits the use of external devices to interact with the VR environment, for example: mouse, touch screen, joystick. In CAVIRE-2, the individual can perform the VR tasks using hand movements and speech. This provides a more naturalistic environment, thus having better ecological validity for the assessment of cognitive function.

The results in this study also affirm that age and extent of education are significant variables affecting the neuropsychological assessments using VR. The different CAVIRE-2 optimal cut-off scores should be considered when the VR system is used for people of different length of their education. Similarly, the performance matrix of the CAVIRE-2 system needs to be adjusted to assess the cognition of older adults, with a lower optimal cut-off score of 1700 for those aged 70–84 years compared with 1900 in younger adults aged 55–69 years. This observation is in keeping with current studies that advocate a lower MoCA threshold in older participants when classifying cognitive impairment (Rossetti et al. 2011; Freitas et al. 2012), suggesting that a tiered approach may apply when using the CAVIRE-2 system.

Its limitations should be recognised as the VR scenes are created to simulate local environments in a developed and highly urbanised community, in which the almost 80% of the population resides in apartments. The validation of the VR software is based on the MoCA value of 26 (Ng 2013), but opinions vary on its cut-off score. A recent meta-analysis reports an optimal cut-off score of 23 produced the highest classification accuracy and reduced the false positive rate (Carson et al. 2018). Indeed, adopting a lower MoCA cut-off may impact the performance of CAVIRE-2 by reducing the number of false positives and thus increasing its accuracy.

5 Conclusion

The study demonstrated the utility of CAVIRE-2 as a valid and reliable cognitive assessment tool comparable to traditional neuropsychological assessments like MoCA. Further proof of value study will be conducted when it is implemented in a larger scale across more diverse primary care and community settings such as nursing homes and daycare centres for senior adults.

Acknowledgements The study team would like to acknowledge FXMedia Internet Pte Ltd, the software company, for their partnership in developing the virtual reality system. The execution of this study was possible through the funding provided by the Healthy Longevity Catalyst Awards HLCA 2022 (HLCA22Feb-0026). The study team would also like to acknowledge Prof Chirk Jenn Ng from the Centre for Population Health Research and Implementation, SingHealth Regional Health System, Singapore for his contributions in the study design, and obtaining the study funding and ethics board approval.

Author contributions JEL and NCT contributed to the development of the virtual reality assessment and the study design. JEL and NCT secured the study funding. JWMP and JEL facilitated the recruitment of participants, administration of study procedures, and collection of data. JWMP and EYLK performed the statistical analysis. JWMP, JEL, EYLK and NCT interpreted and discussed the results. JWMP and JEL contributed to the first draft of the manuscript. All authors commented on the manuscript revisions and approved the final manuscript.

Data availability The data from this study are not publicly available due to privacy and ethical reasons.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethics approval The study was approved by the SingHealth Centralised Institutional Review Board (CIRB Ref. No.: 2022/2221).

Consent to participate All participants joined this study voluntarily, provided written informed consent, and had the right to withdraw their participation from the study at any point of time.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share

adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

References

Alzheimer's Disease International (2023) Risk Factors and Risk Reduction. London. Available at: https://www.alzint.org/about/risk-factors-risk-reduction/. Accessed 10 Feb 2024.

Amieva H et al (2004) Annual rate and predictors of conversion to dementia in subjects presenting mild cognitive impairment criteria defined according to a population-based study. Dementia Geriatr Cogn Disord 18(1):87–93. https://doi.org/10.1159/000077815

Atkins AS et al (2015) Assessment of age-related differences in functional capacity using the virtual reality functional capacity assessment tool (VRFCAT). J Prev Alzheimer's Dis 2(2):121–127. https://doi.org/10.14283/jpad.2015.61

Black DW, Grant JE (2014) DSM-5 guidebook: the essential companion to the diagnostic and statistical manual of mental disorders, 5th edn. American Psychiatric Publication

Bohil CJ, Alicea B, Biocca FA (2011) Virtual reality in neuroscience research and therapy. Nat Rev Neurosci 12(12):752–762. https://doi.org/10.1038/nrn3122

Boz HE et al (2020) A new tool to assess amnestic mild cognitive impairment in Turkish older adults: virtual supermarket (VSM). Neuropsychol Dev Cogn Sect B Aging Neuropsychol Cogn 27(5):639–653. https://doi.org/10.1080/13825585.2019.1663146

Bruscoli M, Lovestone S (2004) Is MCI really just early dementia? A systematic review of conversion studies. Int Psychogeriatr IPA 16(2):129–140. https://doi.org/10.1017/s1041610204000092

Carson N, Leach L, Murphy KJ (2018) A re-examination of Montreal Cognitive Assessment (MoCA) cutoff scores. Int J Geriatr Psychiatry 33(2):379–388. https://doi.org/10.1002/gps.4756

Department of Statistics Singapore (2021) Census of population 2020 statistical release 2: households, geographic distribution, transport and difficulty in basic activities. https://www.singstat.gov.sg/-/media/files/publications/cop2020/sr2/cop2020sr2.ashx. Accessed 18 Dec 2023

Chaytor N, Schmitter-Edgecombe M (2003) The ecological validity of neuropsychological tests: a review of the literature on everyday cognitive skills. Neuropsychol Rev 13(4):181–197. https://doi.org/10.1023/b:nerv.0000009483.91468.fb

Christova M et al (2022) Mixed reality prototype of multimodal screening for early detection of cognitive impairments in older adults: protocol development and usability study. JMIR Res Protoc 11(10):e39513. https://doi.org/10.2196/39513

Cipresso P et al (2014) Virtual multiple errands test (VMET): a virtual reality-based tool to detect early executive functions deficit in Parkinson's disease. Front Behav Neurosci 8:405. https://doi.org/10.3389/fnbeh.2014.00405

Climent-Martínez G et al (2014) Neuropsychological evaluation of the executive functions by means of virtual reality. Rev Neurol 58(10):465–475

da Costa RQM et al (2022) Two immersive virtual reality tasks for the assessment of spatial orientation in older adults with and without cognitive impairment: concurrent validity, group comparison, and accuracy results. J Int Neuropsychol Soc JINS 28(5):460–472. https://doi.org/10.1017/S1355617721000655

Fan J, Upadhye S, Worster A (2006) Understanding receiver operating characteristic (ROC) curves. Can J Emerg Med 8(1):19–20. https://doi.org/10.1017/S1481803500013336

Freitas S et al (2012) Montreal Cognitive Assessment: influence of sociodemographic and health variables. Arch Clin Neuropsychol Off J Natl Acad Neuropsychol 27(2):165–175. https://doi.org/10.1093/arclin/acr116

Granland KA, Thompson CL, Dong Y (2022) "Train your brain" cognitive intervention group program for singaporean older adult patients with mild cognitive impairment: a pilot feasibility study. J Geriatr Psychiatry Neurol 35(3):442–449. https://doi.org/10.1177/08919887211002661

Jang S et al (2022) Virtual reality-based monitoring test for MCI: a multicenter feasibility study. Front Psychiatry Front Res Found 13:1057513. https://doi.org/10.3389/fpsyt.2022.1057513

Jefferson AL et al (2011) A life course model of cognitive activities, socioeconomic status, education, reading ability, and cognition. J Am Geriatr Soc 59(8):1403–1411. https://doi.org/10.1111/j.1532-5415.2011.03499.x

Khaw J et al (2021) Current update on the clinical utility of MMSE and MoCA for stroke patients in Asia: a systematic review. Int J Environ Res Public Health. https://doi.org/10.3390/ijerph18178962

Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 15(2):155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kourtesis P (2020) Immersive virtual reality methods in cognitive neuroscience and neuropsychology: the virtual reality everyday assessment lab (VR-EAL), an immersive neuropsychological test battery of everyday cognitive functions. Dissertation, University of Edinburgh. https://doi.org/10.7488/era/1227

Kumar A et al (2022) Alzheimer disease. StatPearls. StatPearls Publishing, Treasure Island. https://www.ncbi.nlm.nih.gov/pubmed/29763097.

Lim JE et al (2020) A fully-immersive and automated virtual reality system to assess the six domains of cognition: protocol for a feasibility study. Front Aging Neurosci 12:604670. https://doi.org/10.3389/fnagi.2020.604670

Liu LY et al (2021) Prevalence, risk and protective factors for mild cognitive impairment in a population-based study of Singaporean elderly. J Psychiatr Res 145:111–117. https://doi.org/10.1016/j.jpsychires.2021.11.041

Liu Q et al (2023) Virtual reality technology in the detection of mild cognitive impairment: a systematic review and meta-analysis. Ageing Res Rev 87:101889. https://doi.org/10.1016/j.arr.2023.101889

Liu Y et al (2024) Detection rates of mild cognitive impairment in primary care for the United States Medicare Population. J Prev Alzheimer's Dis 11(1):7–12. https://doi.org/10.14283/jpad.2023.131

Malek-Ahmadi M et al (2015) Age- and education-adjusted normative data for the Montreal Cognitive Assessment (MoCA) in older adults age 70–99. Neuropsychol Dev Cogn Sect B Aging Neuropsychol Cogn 22(6):755–761. https://doi.org/10.1080/13825585.2015.1041449

Nahm FS (2022) Receiver operating characteristic curve: overview and practical use for clinicians. Korean J Anesthesiol 75(1):25–36. https://doi.org/10.4097/kja.21209

National Population and Talent Division Singapore (2023) Population in brief 2023. Singapore. https://www.population.gov.sg/media-centre/articles/population-in-brief-2023-key-trends/. Accessed 18 Dec 2023

Neguț A et al (2016) Virtual reality measures in neuropsychological assessment: a meta-analytic review. Clin Neuropsychol 30(2):165–184. https://doi.org/10.1080/13854046.2016.1144793

Ng A et al (2013) Effectiveness of Montreal Cognitive Assessment for the diagnosis of mild cognitive impairment and mild Alzheimer's

disease in Singapore. Singap Med J 54(11):616–619. https://doi.org/10.11622/smedj.2013220

Ng PEM et al (2021) Implementation and effectiveness of a multidomain program for older adults at risk of cognitive impairment at neighborhood senior centres. Sci Rep 11(1):3787. https://doi.org/10.1038/s41598-021-83408-5

Obuchowski NA, McClish DK (1997) Sample size determination for diagnostic accuracy studies involving binormal ROC curve indices. Stat Med 16:1529–1542

Parsons TD (2015) Ecological validity in virtual reality-based neuropsychological assessment. In: Encyclopedia of information science and technology, 3rd ed. IGI Global, pp 1006–1015. https://doi.org/10.4018/978-1-4666-5888-2.ch095

Portney LG, Watkins MP (2007) Foundations of clinical research: applications to practice, 3rd edn. Prentice Hall, New Jersey

Rossetti HC et al (2011) Normative data for the Montreal Cognitive Assessment (MoCA) in a population-based sample. Neurology 77(13):1272–1275. https://doi.org/10.1212/WNL.0b013e318230208a

Sbordone RJ, Long C (1996) Ecological validity of neuropsychological testing. CRC Press, United States

Scharre DW (2019) Preclinical, prodromal, and dementia stages of Alzheimer's disease. Pract Neurol 15:36–47

Spooner DM, Pachana NA (2006) Ecological validity in neuropsychological assessment: a case for greater consideration in research with neurologically intact populations. Arch Clin Neuropsychol Off J Natl Acad Neuropsychol 21(4):327–337. https://doi.org/10.1016/j.acn.2006.04.004

Tan NC et al (2022) Age-related performance in using a fully immersive and automated virtual reality system to assess cognitive function. Front Psychol 13:847590. https://doi.org/10.3389/fpsyg.2022.847590

Tan NC et al (2024) A virtual reality cognitive screening tool based on the six cognitive domains. Alzheimer's & Dementia (Amsterdam, Netherlands) 16(4):e70030. https://doi.org/10.1002/dad2.70030

Tarnanas I et al (2013) Ecological validity of virtual reality daily living activities screening for early dementia: longitudinal study. JMIR Serious Games 1(1):e1. https://doi.org/10.2196/games.2778

Tavakol M, Dennick R (2011) Making sense of Cronbach's alpha. J Int Assoc Med Sci Educ JIAMSE 2:53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Terwee CB et al (2012) Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Research Int J Quality Life Aspects Treat Care Rehabil 21(4):651–657. https://doi.org/10.1007/s11136-011-9960-1

Thaipisuttikul P et al (2022) Rate of conversion from mild cognitive impairment to dementia in a Thai hospital-based population: a retrospective cohort. Alzheimer's Dementia J Alzheimer's Assoc 8(1):e12272. https://doi.org/10.1002/trc2.12272

Valladares-Rodríguez S et al (2016) Trends on the application of serious games to neuropsychological evaluation: a scoping review. J Biomed Inform 64:296–319. https://doi.org/10.1016/j.jbi.2016.10.019

Wong WT et al (2021) Comparison of Time taken to assess cognitive function using a fully immersive and automated virtual reality system vs. the montreal cognitive assessment'. Front Aging Neurosci 13:756891. https://doi.org/10.3389/fnagi.2021.756891

Yan M et al (2021) A virtual supermarket program for the screening of mild cognitive impairment in older adults: diagnostic accuracy study. JMIR Serious Games 9(4):e30919. https://doi.org/10.2196/30919

Youden WJ (1950) Index for rating diagnostic tests. Cancer 3(1):32–35. https://doi.org/10.1002/1097-0142(1950)3:1%3c32::aid-cncr2820030106%3e3.0.co;2-3

Yu J et al (2021) Mindfulness intervention for mild cognitive impairment led to attention-related improvements and neuroplastic changes: Results from a 9-month randomized control trial. J Psychiatr Res 135:203–211. https://doi.org/10.1016/j.jpsychires.2021.01.032

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.